Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

No evidence for discrete biomarker-based subtypes of Parkinson's Disease

Cosimo Guerra
University of Salerno
Cristiano Sorrentino
University of Salerno
Paolo Barone
University of Salerno
Roberto Erro
rerro@unisa.it
University of Salerno

Article

Keywords: seed amplification assay, cluster analysis, disease subtypes, null models, cerebrospinal fluid biomarkers

Posted Date: September 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10866905/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Biomarker-driven subtyping of Parkinson's disease represents a translational gap in the field. Here, we evaluated whether combining two disease-defining biomarkers, namely α-synuclein seed amplification assay (SAA) kinetics and dopaminergic denervation, with age reveals discrete biological subtypes. Across a primary cohort ($n = 595$) and an internal replication cohort ($n = 333$), statistical evaluations of clustering tendency, partition reproducibility, and stability against calibrated null distributions revealed no evidence of genuine subgroup structure. Instead, the data reflected a continuous spectrum. Conventional subtyping metrics frequently crossed significance thresholds even under calibrated reference distributions generated without latent mixture membership. Additionally, SAA kinetics did not consistently predict longitudinal motor or non-motor progression, and observed correlations between SAA parameters and cerebrospinal fluid analytes were primarily driven by a non-specific shared factor, suggesting a possible matrix effect rather than distinct pathological processes. Consequently, PD-defining biomarkers align along a continuum rather than displaying discrete clustering structures characteristic of distinct biological subtypes.

Introduction

The clinical heterogeneity of Parkinson’s disease (PD) has sustained a long tradition of subtype research [1]. Its premise rests in the observed variability in clinical presentation and disease course, which would reflect distinguishable entities, possibly sustained by different pathophysiological mechanisms. Hence, their identification would allow progression to be forecast, patients to be stratified for clinical trials, and treatments to be directed [2]. Moving beyond earlier hypothesis-driven approaches, the dominant framework over the past decades exploited data-driven cluster analysis, which simultaneously considers a broad set of variables without an *a-priori* hypothesis to identify subtypes [3]. Accordingly, several subtyping solutions, mostly relying on clinical features only, have been generated, but subsequently identified to have significant methodologic shortcomings, questionable clinical applicability and unknown biological relevance [3]. Replication studies have further yielded mixed results, with partial confirmations that depended substantially on the clustering variable set and the cohort employed [4]. In one systematic attempt, eight published data-driven classification systems were re-derived in a single well-characterised external cohort and a Delphi panel judged none of them being reproducible [5]. An alternative proposal that has been put forward, but never empirically applied, posits that only biomarker-defined subtypes of PD would reflect homogenous underlying biological and molecular abnormalities, which might more likely respond differentially to a given therapeutic approach, particularly for strategies aimed at neuroprotection [6].

Along this theoretical shift, the recent development of sensitive and specific tools to detect α-synuclein (α-syn) pathology *in vivo*, especially of seed amplification assay (SAA) [7], has instigated two proposals for a biological definition of PD [8, 9]. In this context, the α-syn SAA does not only return a dichotomous outcome (e.g., positive/negative), but also quantitative kinetic aggregation parameters, that are under active investigation based on the hypothesis that faster/higher-amplitude kinetic parameters might

reflect a more aggressive pathological process. However, recent evidence relating these parameters to disease progression or subtyping solutions is genuinely mixed. While SAA kinetic have not been correlated to longitudinal motor outcome in one study [10], shorter reaction times at baseline have been reported to predict membership of a diffuse-malignant subtype at ten years [11] and faster kinetics predicted cognitive decline independently of Alzheimer-type co-pathology [12]. Conversely, another study found no consistent improvement when incorporating SAA kinetics to predict five-year motor, cognitive, or medication outcomes, with any association confined to a single REM-sleep item [13]. Furthermore, a PPMI analysis using the 24-hour protocol reported that higher amplitude parameters accompanied slower cognitive decline, the reverse of what one would expect, and cautioned that this finding may reflect assay methodology rather than biology [14].

Against this background, SAA kinetic parameters might still offer a distinct biological dimension that, alongside dopaminergic denervation, could establish a more robust, biomarker-informed subtyping scheme. However, the question that follows is whether combining these two disease-defining biomarkers with age, which likely influences both independently of clinical status [15, 16], delineates discrete biological subtypes rather than variations of a continuum.

In fact, a recurring and rarely confronted methodological problem that has transversally characterized previous subtyping studies rests in the evidence that partitioning algorithms return *k* groups whenever *k* is requested, including when applied to data with absent group structure. Assessment methods of clustering tendency, evaluation of partition stability under resampling, and systematic comparison against explicitly specified reference distributions have been available in the statistical literature for decades [17–21], yet have been never applied in PD subtyping research. In their absence, the between-group clinical differences reported as supportive of a taxonomy are expected by construction whenever the partition cuts across a clinically relevant gradient, but cannot discriminate between a discrete-subtype output and a non-partitionable continuum.

Therefore, the primary objective of this work was to determine whether biomarker-driven subtypes of PD can be identified. This was addressed using a battery of statistical procedures designed to establish, *a priori*, whether the data exhibit genuine cluster structure, a necessary precondition for any subsequent determination of subtype number. A negative clustering result is informative only insofar as the method employed is demonstrably capable of detecting clusters when they are present. This was assessed by injecting synthetic clusters of known but varying separation, size, and orientation into the real dataset, and measuring how often each clustering criterion signaled significant structure in response, compared against how often each criterion signaled significant structure when no synthetic cluster had been added. This procedure establishes whether each criterion crosses its significance threshold under known conditions, not whether it correctly recovers which specific data points belong to the injected cluster. Given this distinction, the analysis constitutes a descriptive check of the method sensitivity rather than a formal statistical test establishing the absence of clusters.

Three further sets of secondary analyses, none of which was preregistered, are presented: description of the continuous biomarker dimensions and their clinical correlates; longitudinal models relating baseline SAA kinetics to the rate of change in motor, dopaminergic denervation, and non-motor outcomes; and a set of CSF analyses prompted by an unexpected pattern in the correlation matrix of CSF analytes.

Results

2.1 Participants

The primary cohort comprised 595 genetically negative, α-syn SAA + PD patients (24h SAA protocol); the replication cohort comprised 333 α-syn SAA + PD patients (150h SAA protocol), of whom 45 (13.5%) overlapped with the primary cohort. Full flow and baseline characteristics, including the excluded-subject comparison, are in Supplemental Material (§ 1.1, Tables S1 and S2, Fig. S1).

2.2 Correlation structure of the variables

No correlation between SAA kinetics and dopamine transporter (DAT) imaging outcomes exceeded $|\rho| = 0.119$ (Fig. 1), with the highest correlation between kinetic parameters and age being observed for maximum slope [SLOPEmax (ρ =-0.111; Fig. 1)]. Within the kinetic block, maximum fluorescence (Fmax), area under the fluorescence curve (AUC) and SLOPEmax formed a single amplitude factor (ρ = 0.72–0.88), distinct from timing parameters (Fig. 1). Principal-component loadings are in § 1.2 of Supplemental Material.

2.3 No cluster structure is supported: tendency, number, and stability statistics

The Hopkins statistic showed a small but significant departure from the fitted Gaussian-copula null (observed 0.794 vs. null median 0.765, 95% null interval 0.753–0.777; Monte Carlo p = 0.0020; Fig. 2). When read against the conventional threshold of 0.5 this value would be misleadingly read as strong clustering. Hartigan's dip test rejected unimodality for none of the seven variables or seven principal components after Bonferroni correction (lowest uncorrected: log TTT, p = 0.026, corrected p = 0.360; Fig. 3; supplemental material § 1.3); of 1,000 random one-dimensional projections, 2 were nominally multimodal, and in the matched 200-projection comparison the observed count (1) did not exceed the null (Monte Carlo p = 0.058). The gap statistic attained its maximum at k = 1 and selected k = 1 by the one-standard-error rule in 100% of 500 null realisations (Fig. 4; supplemental material § 1.4). Prediction strength never reached the conventional 0.80 threshold (maximum 0.757 at k = 2) and was indistinguishable from its null (median 0.736, 95% interval 0.622–0.837; p = 0.383); the maximum silhouette index (0.264) was likewise not significant (null median 0.221; p = 0.052). Gaussian-mixture BIC selected a different number of components depending on the null used for comparison alone, illustrating that the criterion tracks density shape, not group existence. Read against its own null, the bootstrap Jaccard index at k = 2 (0.918) was not informative (null median 0.901, 95% interval 0.816–0.949; p =

0.279). Between-algorithm agreement was low and below its null (mean adjusted Rand index 0.141 vs. null median 0.246 at k = 2; p = 0.894). HDBSCAN assigned 87.9% of observations to noise at minimum cluster size 10–20, and all observations to noise from size 30 upward. Of seven battery statistics evaluated at k = 2 in both cohorts, only the Hopkins statistic exceeded the copula null in each (Fig. 5; Table S5). When treated as one family, only the primary-cohort Hopkins result survived Bonferroni correction. The departure was confined to the statistic measuring local density and absent from every statistic measuring partition reproducibility. Full details across all four analyses, including per-variable dip-test results, the BIC comparison under alternative nulls, and the HDBSCAN parameter sweep, are in Supplemental Material (§§ 1.3–1.5, 1.9; Tables S3-S5).

2.4 Robustness of the null clustering result

The null result was unchanged across three alternative variable weightings (gap statistic selected k = 1 in every case; prediction strength 0.731–0.771). A positive-control analysis using SAA status itself (n = 678, including 66 assay-negative and 5-inconclusive subjects) recovered a two-group structure tracking assay outcome (adjusted Rand index 0.820; Fig. 6), confirming the pipeline is not inert, though the gap statistic still selected k = 4 rather than k = 2, and the benchmark labels are not independent of the clustering inputs. The battery reproduced the null result in the 150h SAA cohort (Fig. S3; Table S5): prediction strength 0.633 vs. null 0.606 (p = 0.332), Jaccard 0.816 vs. 0.771 (p = 0.282), gap statistic k = 1 in 100% of null realisations. Excluding the 45 overlapping subjects yielded consistent results: in the remaining sample (n = 288), the Hopkins statistic remained the only metric to exceed the copula null distribution (p = 0.010). Conversely, prediction strength, Jaccard stability, silhouette width, and inter-algorithm agreement all fell well within their expected null reference intervals (all p ≥ 0.16). Removing one collinear parameter (AUC, correlated − 0.970 with T50) raised prediction strength to 0.804 and Jaccard to 0.937 with no accompanying hallmark of real structure — a direct demonstration that these thresholds are not valid decision rules. Across the complete diagnostic-curve family (76 values), 8 nominally exceeded the reference but none survived correction (smallest adjusted p = 0.166). Assay run explained two-thirds of the variance of log TTT in the primary cohort (η^2 = 0.667, ICC = 0.656) and centre effects were comparatively small (η^2 0.043–0.106). Repeating the battery on run-residualised kinetics pushed prediction strength above 0.80 in the primary cohort (0.827) without the value being distinguishable from its own null (p = 0.129), reinforcing that threshold-crossing was not evidence of structure. With nothing injected, the bootstrap Jaccard index already crossed its threshold in 100% of replicates in the primary cohort, and prediction strength in 27% (Fig. 7), suggesting these baseline rates show that the thresholds are uninformative alone. Configurations reaching ≥ 80% crossing probability in all three directions required large separations: in the primary cohort, this threshold was met at Δ ≥ 2.5 SD only with 50% prevalence, and at Δ ≥ 3.0 SD with prevalence ≥ 30%; in the replication cohort, only Δ = 3.0 at 50% prevalence satisfied this criterion. Full detail across all four robustness analyses, including the overlap-free and collinear-parameter checks and the entire injection grid, is provided in the supplemental material (§§ 1.6, 1.8, 1.10, 1.11, Tables S5-S7).

2.5 Imposed partitions and continuous axes

An imposed two-group (amplitude-based) partition differentiated only 2 of 38 external variables after correction (sex; BMI $\varepsilon^2 = 0.019$, $p = 0.0011$). An imposed three-group (age–DAT) partition differentiated 23 of 38, but the largest effects were mathematical derivatives of the construction variables (e.g., mean striatal binding $\varepsilon^2 = 0.345$) and genuinely external correlates were modest at best (neurofilament light $\varepsilon^2 = 0.171$, UPSIT 0.074, SCOPA-AUT 0.047). Full detail on both imposed partitions and the continuous-axis correlates is in supplemental material (§ 1.7, Fig. S2).

2.6 Longitudinal models: motor, dopaminergic, and non-motor outcomes

In the primary cohort, 3 of 28 kinetics-by-time interactions survived correction, all involving amplitude parameters on DAT binding: Fmax–caudate ($\beta = -0.0082$/SD/year, BH $p = 0.03$), Fmax–putamen ($\beta = -0.0082$, BH $p = 0.024$), SLOPEmax–caudate ($\beta = -0.0083$, BH $p = 0.03$). Under progressive adjustment for run and centre, both caudate associations remained nominally significant (Fmax, $p = 0.010$; SLOPEmax, $p = 0.014$). For putaminal binding, the Fmax association was no longer nominally significant ($p = 0.072$), whereas the SLOPEmax association remained significant at this stage ($p = 0.036$) but lost significance after run residualisation ($p = 0.088$). These differing attenuation patterns would be compatible with region-specific technical confounding. None replicated in the 150h SAA cohort; no motor-outcome interaction was significant at any stage. None of twelve pre-specified non-motor domains survived correction (two nominal: RBDSQ $p = 0.016$, MDS-UPDRS I $p = 0.045$; Fig. S4). The MoCA coefficient ($\beta = 0.050$ points/year/SD of T50, BH $p = 0.170$) was stable in direction and magnitude (0.025–0.065) across thirteen reported specifications of trajectory shape, attrition handling, and censoring-weight adjustment, after excluding the ≥ 10-year follow-up stratum from the pattern-mixture analysis, the estimate of which was too imprecise to be informative ($\beta = 0.0969$, SE = 0.1179). The extended seven-outcome neuropsychological battery, corrected separately, likewise yielded nothing (0 of 7). Reliability was excellent in the 24h SAA protocol (ICC(1,3) 0.88–0.92 across parameters) but only poor-to-moderate in the 150h SAA protocol (T50 ICC(1,3) = 0.52; maximum-slope coefficient of variation median 11.2%), which is expected to attenuate the 150h longitudinal associations toward the null. Across the 54 estimable specifications of the MoCA analysis (kinetic parameter, aggregation, estimation strategy, random-slope inclusion, dose adjustment), all 54 produced a coefficient in the same direction (median $p = 0.0064$), with 37 of 54 nominally significant — direction was robust; nominal significance was not. Full detail across all four analyses, including the technical-adjustment ladder, the thirteen MoCA sensitivity specifications, and the specification curve, is in supplemental material (§§ 1.12–1.15, Tables S8-S11, Fig. S5-S6).

2.7 CSF biomarker correlates of SAA kinetics

In the replication cohort, TTT and T50 correlated positively with all seven available CSF analytes across distinct biochemical families and platforms, including total tau and Aβ42 moving in the same direction, which would be incompatible with a genuine pathological signature (Fig. 8). These analytes shared one

dominant component accounting for 67.7% of variance, with uniformly positive loadings. After adjusting for this component, no kinetics–biomarker associations survived multiplicity correction, dropping from 4 of 12 and 22 of 35 associations to zero in the primary and replication cohorts, respectively.. A leave-one-analyte-out control, on the 150h cohort, excluded simple circularity. The component correlated only weakly with CSF albumin (ρ = −0.021 and − 0.129), excluding a total-protein explanation. Independently of the shared component, CSF albumin was directly associated with the kinetic parameters in both cohorts: in the replication cohort (n = 266), with T50 (partial ρ = 0.240), TTT (0.234), SLOPEmax (0.243) and AUC (− 0.234, all BH p ≤ 0.0002); in the primary cohort (n = 55), with Fmax (− 0.514), SLOPEmax (− 0.500) and AUC (− 0.447). The association survived adjustment for a haemoglobin-contamination indicator and replicated when CSF-to-serum albumin quotient was substituted for albumin (all p ≤ 0.0018), consistent with a blood–brain barrier permeability effect; replication across protocols was limited to AUC, the only parameter significant in both. Full detail across both analyses, including the platform and haemoglobin-contamination checks, is in the supplemental material (§§ 1.16–1.17, Tables S12, S13).

Discussion

No reproducible partition was detected within the variable space defined by α-syn SAA kinetics, DAT imaging, and age. This null result was replicated internally across both SAA protocols (e.g., on 595 and 333 subjects) and remained consistent when shared participants were also excluded. This consistency across replication attempts, exceeding the weight of any single internal statistic, constitutes the strongest evidence of the current study. However, these findings should not be framed as uniformly negative. While every statistic evaluating partition reproducibility failed to deviate from the null, a local-density statistic departed modestly from the specified null distribution in both cohorts, though no evidence demonstrates that the data separate into stable, recoverable groups. Because the reference distribution, while being more conservative than a naive Gaussian model cannot account for every potential dependence within the data, this departure remains indicative rather than conclusive. This null finding withstands two key methodological objections. First, it is not an artifact of an insensitive analytical pipeline, as our procedure successfully detected a known, gross contrast in a validation control. However, because those control labels were derived from the same baseline measurements, this check merely proves that the pipeline functions, without proving that it can detect subtle biological subtypes. Second, the null result is remarkably robust, surviving virtually all variations in variable weighting, aggregation, and covariate adjustments. The sole instance where a p-value crossed a conventional significance threshold, following the removal of a collinear parameter, proved instructive rather than contradictory: because no secondary indicators of genuine grouping co-occurred, it highlights that passing a statistical threshold is not equivalent to discovering true biological structure.

Our findings highlight a crucial distinction between boundary reproducibility and true cluster separation: a dividing line can be highly reproducible across datasets simply because the underlying continuous distribution has a stable, elongated shape, not because discrete biological subtypes exist. This principle clarifies why partitions that seem clinically informative can be imposed on unsupported data despite

carrying minimal evidential weight: dividing a continuous clinical gradient into discrete slices inevitably produces between-group differences that merely mirror that gradient. Interpreting such differences as proof of the existence of discrete groups represents a recognized methodological limitation across the broader subtyping literature. While the clinical heterogeneity of PD is not under discussion [22], our data do not support the concept that this heterogeneity resolves into discrete, recoverable categories within this biomarker-informed variable space [23] and at this sample size. The question that follows is whether these two disease-defining biomarkers can fundamentally serve as complementary indicators of the underlying pathophysiology for the identification of discrete, meaningful subtypes.

Evidence linking dopaminergic denervation, as measured by DAT-SPECT, to motor impairment is well established [24], yet their relationship is not tightly coupled. It weakens from the early stage, when substantial presynaptic and circuit-level compensation attenuate the clinical expression of a given degree of denervation [25], to a more advanced stage, when the involvement of non-dopaminergic pathways contributes to axial and gait disturbance [26]. Yet, as the present analysis demonstrates directly, dopaminergic denervation is itself a continuous outcome lacking discrete structure, which therefore does not resolve into separable subgroups. Conversely, SAA kinetics likewise varied continuously across the sampled population, but have not to date correlated robustly with any clinical outcome, cross-sectionally or longitudinally, [10–14, 27, 28] nor have they been shown to be partitionable in the current analyses. This supports the concept that SAA may function principally as a qualitative, categorical marker of pathological status (e.g., aggregation-competent seed present or absent) [29], rather than as a quantitative marker of pathology severity whose continuous range would correspond to a meaningful clinical gradient. This pattern, whereby a biomarker of pathological status proves diagnostically informative but prognostically uninformative, is not unique to synucleinopathies. In the AD field, amyloid-PET functions as the canonical marker of pathology, defining "A+" status in the ATN framework [30] and serving essentially as a binary gate for trial eligibility and diagnosis, yet longitudinal change in amyloid burden does not track cognitive decline [31, 32].

Under this framing, the continuous variation captured by the kinetic parameters may reflect measurement noise and biochemical variability around a fundamentally binary state, rather than a graded biological signal. The secondary CSF findings would support this interpretation. SAA kinetic parameters correlated with several unrelated analytes across opposing pathological processes, but these correlations vanished after controlling for a shared dominant component. While the nature of this common factor remains undetermined, caution is warranted when interpreting unadjusted literature correlations. The finding that total protein content (via a proxy marker) independently correlated with kinetic parameter might suggest a technical matrix effect, pointing toward the need for assay protocol standardization rather than a novel biological insight. Separately, CSF albumin was associated with kinetics independently of this factor. This would not point to a total-protein effect, and its persistence with the albumin quotient suggests a link to blood–brain barrier permeability.

An additional trend observed in our secondary analyses would associate slower kinetics with better cognition as measured with MOCA. While this trend did not reach statistical significance in adjusted

analyses and was absent for any other neuropsychological outcomes, its direction was invariant across all 54 estimable specifications, despite nominal significance holding in only about two thirds of them. Accordingly, this signal might be possibly compatible with a small effect that the current study was underpowered to detect. Predictably, past findings have been mixed: two of five studies detected a similar relationship [11, 12], while three failed to do so [13, 14, 27]. Therefore, dedicated, higher-powered studies are needed to determine whether this directional trend reflects a genuine biological effect.

In summary, we found no evidence of discrete, biomarker-driven subtypes when testing for group structure within the space defined by the two core biological anchors of PD: pathological α-syn aggregation and dopaminergic denervation. Of note, the weak rank-correlation between the two sets of variables, while not establishing absolute statistical independence, might further suggest they represent distinct phenomena. This would align with evidence that initiation of nigrostriatal neurodegeneration might occur independently of α-syn aggregation and can be tau-mediated [33].

The primary strength of this work lies in its methodological rigor at a conventionally overlooked stage: verifying whether data support discrete group structure prior to assuming its existence, calibrating criteria against matched null distributions, and explicitly mapping detection limits. Evaluation of the latter demonstrated that certain stability statistics conventionally cited as evidence for subtype existence were proven to be uninformative: one statistic reached saturation even in the complete absence of injected structure, whereas another responded only marginally to a substantial, deliberately introduced separation. These behaviours highlight a fundamental vulnerability of common clustering metrics, raising important questions about the methodological validity of the broader PD subtyping literature. However, it should be noted that

detection sensitivity heavily depends on the axis along which genuine subtypes separate, a property which is unknowable *a priori*. Consequently, the conclusions are bounded: only relatively large subgroups would be reliably detected, leaving open the possibility of smaller, structurally distinct, or noise-obscured subgroups. However, it remains to be seen if these smaller sized subtypes exist and whether they are clinically meaningful.

Notably, technical (assay-run) variance was non-trivial and unmodeled technical noise might have acted against structure detection by degrading genuine signal. The weak correlation observed between amplitude and timing parameters might be consistent with this account: nucleated-polymerization theory treats nucleation and growth as mechanistically separable steps governed by distinct rate constants [34, 35], but parameters are also known to differ in their susceptibility to technical noise, with time-to-threshold reported as the most reliable kinetic parameter across laboratories and experimental conditions [36], so the observed dissociation may reflect genuine kinetic independence, differential measurement reliability across parameters, or both, and the present data cannot distinguish between them.

Some limitations warrant consideration. First, the replication cohort provides internal rather than external validation. Second, longitudinal outcomes were modeled assuming linearity, which was formally verified

for only a single outcome, and measurement reliability differed between assay protocols, leading to unequal evidentiary weight between cohorts in longitudinal analyses. Third, secondary analyses were specified post hoc following initial data inspection. Accordingly, confidence in these secondary findings should be tempered relative to our primary structural conclusions. Finally, our analyses relied on an enriched cohort of early-stage, treatment-naïve patients and may not fully reflect the broader PD population. However, we anticipate no strong theoretical reason why this specific biomarker structure would fundamentally differ across the wider disease spectrum.

Methods

This is a secondary observational analysis of the Parkinson's Precision Medicine Initiative (PPMI), reported per STROBE guidelines (cross-sectional items applied to baseline analyses); full methodological detail and all numerical diagnostics are provided in the Supplement under the section numbers cited below (§ 2.1–2.22).

Demographic, clinical, imaging, and CSF data were drawn from the PPMI Curated Data Cut (11 May 2026); α-syn SAA kinetic parameters were obtained from the biospecimen results file (§ 2.1). Two cohorts corresponding to PPMI two SAA protocols were assembled (§ 2.2): a primary cross-sectional cohort (24-h protocol, Project 237; n = 595 analysable, type-1/Lewy-body-pattern-positive patients) and an internal replication cohort (150-h pilot protocol, Project 155; n = 333). Both were restricted to genetically negative patients by excluding carriers of known PD-risk variants or subjects lacking genetic classification (§ 2.2.1); 45 subjects (13.5% of the replication cohort) overlap between cohorts, so replication is internal rather than external, and the full battery was repeated after their removal (§ 1.8). Cohort roles were assigned by follow-up duration: the 24-h cohort (median follow-up 3.0 years) served as the primary population for cross-sectional clustering, the 150-h cohort (median follow-up 9.0 years) for longitudinal modeling, with cross-replication of both (§ 2.2.2).

SAA kinetic parameters (Fmax, time-to-threshold (TTS), area under the curve, time-to-maximum-slope [24h] (TSmax; 24h) or time-to-half-aggregation [150h] (T50), and maximum slope (SLOPEmax) were derived from triplicate wells, aggregated as the mean (24h) or median (150h, chosen after inspection of replicate outliers; § 2.3). DAT-SPECT striatal binding ratios (lowest-uptake putamen, mean caudate) were extracted from the curated pipeline (§ 2.4). Seven primary variables (four log-transformed, standardized kinetic parameters; two DAT-SPECT measures; age) were retained after removing two variables for near-collinearity with retained counterparts (§ 2.6); no imputation was performed.

Before any partitioning was attempted, clustering tendency was formally assessed (§ 2.7) using the Hopkins statistic, Hartigan's dip test (on variables, principal components, and 1,000 random projections), and comparison against a fitted Gaussian-copula null preserving observed marginals and rank dependence — explicitly not assumed unimodal. Four algorithms (k-means, Ward and average-linkage hierarchical clustering, Gaussian mixtures) were applied across k = 2–10, whereas HDBSCAN was evaluated separately using cluster sizes of 10 to 75 (§ 2.8). A partition was considered supported under

a deliberately permissive, pre-specified rule: any one of three k = 2 reproducibility criteria exceeding the null, or either partition-quality curve surviving Benjamini-Hochberg (BH) correction across k and cohorts, would suffice; a local-density result or a conventional threshold crossing alone would not. Stability was further assessed via bootstrap Jaccard index, proportion of ambiguous clustering, and between-algorithm agreement (§ 2.9), with six pre-specified sensitivity analyses testing dependence on variable weighting, population definition, replicate aggregation, cohort overlap, collinearity, and assay-run adjustment (§ 2.10). Characterisation of the partitions and of the continuous axes are detailed in (§ 2.11)

Because a negative clustering result is only informative if the method could have detected structure had it existed, synthetic clusters of known separation, prevalence, and direction were injected into both cohorts, and the threshold-crossing rate of each criterion was compared against its rate under no injection (§ 2.12, § 2.13), a descriptive sensitivity study, not an equivalence test: it establishes threshold-crossing probability, not recovery of injected class membership, and conclusions are worded accordingly.

Longitudinal associations between baseline kinetics and rate of change in motor, dopaminergic, and non-motor outcomes were modeled via linear mixed-effects models with random slopes (§ 2.14), convergence verified via an explicit Newton-decrement criterion across all fits (§ 2.15), and robustness assessed via a 54-specification curve (§ 2.16). An unexpected positive-correlation pattern among CSF analytes prompted exploratory analysis of a shared, uncharacterized component (§ 2.17), evaluated against blood contamination, substrate concentration, and CSF albumin/Qalb as candidate explanations, sample replicate reliability, and against platform/run confounding (§ 2.18, 2.19).

Multiple-testing correction was applied within eleven pre-specified hypothesis families ((§ 2.20, Table S14), with uncorrected values reported throughout. Software, seeds, and reproducibility materials are detailed in § 2.21; no analysis plan was preregistered, and post-hoc choices are flagged at the point they occur.

Declarations

Competing interests

The authors declare no competing interests.

Author Contribution

C.G. co-conceptualized the study, designed and implemented all computational models, conducted data analysis, and led the experimental workflow and the writing of the manuscript. C.S and P.B. contributed

to manuscript revision, writing, and clinical analysis and interpretation; R.E. conceptualized the study, supervised the overall research direction, and contributed to manuscript writing and revision.

Data Availability

Data used in the preparation of this article were obtained on June 16th 2026, from the Parkinson's Precision Medicine Initiative (PPMI) database (www. ppmi-info.org/access-dataspecimens/download-data). The PPMI investigators have not participated in reviewing the data analysis or content of this manuscript.

Code availability

The underlying code for this study is available from the corresponding author upon request.

References

1. Deng X, Mehta A, Xiao B, Ray Chaudhuri K, Tan EK, Tan LC. Parkinson's disease subtypes: Approaches and clinical implications. Parkinsonism Relat Disord. 2025;130:107208. doi: 10.1016/j.parkreldis.2024.107208.
2. Marras C, Fereshtehnejad SM, Berg D, Bohnen NI, Dujardin K, Erro R, Espay AJ, Halliday G, Van Hilten JJ, Hu MT, Jeon B, Klein C, Leentjens AFG, Mollenhauer B, Postuma RB, Rodríguez-Violante M, Simuni T, Weintraub D, Lawton M, Mestre TA. Transitioning from Subtyping to Precision Medicine in Parkinson's Disease: A Purpose-Driven Approach. Mov Disord. 2024;39(3):462–471. doi: 10.1002/mds.29708.
3. Mestre TA, Fereshtehnejad SM, Berg D, Bohnen NI, Dujardin K, Erro R, et al. Parkinson's disease subtypes: critical appraisal and recommendations. J Parkinsons Dis. 2021;11(2):395–404. doi:10.3233/JPD-202472
4. Cash TV, Lessov-Schlaggar CN, Foster ER, Myers PS, Jackson JJ, Maiti B, et al. Replication and reliability of Parkinson's disease clinical subtypes. Parkinsonism Relat Disord. 2024;124:107016. doi:10.1016/j.parkreldis.2024.107016
5. Mestre TA, Eberly S, Tanner C, Grimes D, Lang AE, Oakes D, Marras C. Reproducibility of data-driven Parkinson's disease subtypes for clinical research. Parkinsonism Relat Disord. 2018;56:102–106. doi:10.1016/j.parkreldis.2018.07.009
6. Espay AJ, Schwarzschild MA, Tanner CM, Fernandez HH, Simon DK, Leverenz JB, Merola A, Chen-Plotkin A, Brundin P, Kauffman MA, Erro R, Kieburtz K, Woo D, Macklin EA, Standaert DG, Lang AE. Biomarker-driven phenotyping in Parkinson's disease: A translational missing link in disease-modifying clinical trials. Mov Disord. 2017;32(3):319–324. doi: 10.1002/mds.26913.
7. Concha-Marambio L, Pritzkow S, Shahnawaz M, Farris CM, Soto C. Seed amplification assay for the detection of pathologic alpha-synuclein aggregates in cerebrospinal fluid. Nat Protoc.

2023;18(4):1179–1196. doi: 10.1038/s41596-022-00787-3.

8. Simuni T, Chahine LM, Poston K, Brumm M, Buracchio T, Campbell M, et al. A biological definition of neuronal α-synuclein disease: towards an integrated staging system for research. Lancet Neurol. 2024;23(2):178–190. doi:10.1016/S1474-4422(23)00405-2
9. Höglinger GU, Adler CH, Berg D et al (2024) A biological classification of Parkinson's disease: the SynNeurGe research diagnostic criteria. Lancet Neurol 23:191–204
10. Siderowf A, Concha-Marambio L, Lafontant DE, Farris CM, Ma Y, Urenia PA, et al. Assessment of heterogeneity among participants in the Parkinson's Progression Markers Initiative cohort using α-synuclein seed amplification: a cross-sectional study. Lancet Neurol. 2023;22(5):407–417. doi:10.1016/S1474-4422(23)00109-6
11. Grillo P, Riboldi GM, Pisani A, Kang UJ, Fereshtehnejad SM. Faster reaction times of CSF alpha-synuclein seed amplification assay predict the diffuse malignant subtype of Parkinson's disease at 10-year follow-up. J Parkinsons Dis. 2026;16(2):233–245. doi:10.1177/1877718X261417468
12. Orrú CD, Vaughan DP, Vijiaratnam N, Real R, Martinez-Carrasco A, Fumi R, et al. Diagnostic and prognostic value of α-synuclein seed amplification assay kinetic measures in Parkinson's disease: a longitudinal cohort study. Lancet Neurol. 2025;24(7):580–590. doi:10.1016/S1474-4422(25)00157-7
13. Belete D, Mattjie C, Huxford B, Bestwick JP, Noyce AJ, Simonet C. Evaluating the role of α-synuclein seed amplification as a disease progression marker: evidence and uncertainties. Mov Disord Clin Pract. 2026;13(5):1228–1235. doi:10.1002/mdc3.70434
14. Rodriguez-Antiguedad J, Puig-Davi A, Ruiz-Barrio I, Campo-Caballero D, Vacchi E, Melli G, et al. Association of α-synuclein seed-amplification kinetics with cognitive decline in idiopathic Parkinson's disease. Parkinsonism Relat Disord. 2026;150:108856. doi:10.1016/j.parkreldis.2026.108856
15. Varrone A, Dickson JC, Tossici-Bolt L, Sera T, Asenbaum S, Booij J, Kapucu OL, Kluge A, Knudsen GM, Koulibaly PM, Nobili F, Pagani M, Sabri O, Vander Borght T, Van Laere K, Tatsch K. European multicentre database of healthy controls for [123I]FP-CIT SPECT (ENC-DAT): age-related effects, gender differences and evaluation of different methods of analysis. Eur J Nucl Med Mol Imaging. 2013;40(2):213–27. doi: 10.1007/s00259-012-2276-8.
16. DelleDonne A, Klos KJ, Fujishiro H, Ahmed Z, Parisi JE, Josephs KA, Frigerio R, Burnett M, Wszolek ZK, Uitti RJ, Ahlskog JE, Dickson DW. Incidental Lewy body disease and preclinical Parkinson disease. Arch Neurol. 2008;65(8):1074–80. doi: 10.1001/archneur.65.8.1074.
17. Hartigan JA, Hartigan PM. The dip test of unimodality. Ann Stat. 1985;13(1):70–84. doi:10.1214/aos/1176346577
18. Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. J R Stat Soc Series B Stat Methodol. 2001;63(2):411–423. doi:10.1111/1467-9868.00293
19. Tibshirani R, Walther G. Cluster validation by prediction strength. J Comput Graph Stat. 2005;14(3):511–528. doi:10.1198/106186005X59243

20. Hennig C. Cluster-wise assessment of cluster stability. Comput Stat Data Anal. 2007;52(1):258–271. doi:10.1016/j.csda.2006.11.025
21. Şenbabaoğlu Y, Michailidis G, Li JZ. Critical limitations of consensus clustering in class discovery. Sci Rep. 2014;4:6207. doi:10.1038/srep06207
22. Mulroy E, Erro R, Bhatia KP, Hallett M. Refining the clinical diagnosis of Parkinson's disease. Parkinsonism Relat Disord. 2024;122:106041. doi: 10.1016/j.parkreldis.2024.106041.
23. Landolfi A, Sorrentino C, Barone P, Erro R. Biological frameworks for Parkinson's disease: the heterogeneity SAAgged. J Neurol. 2025;272(4):318. doi: 10.1007/s00415-025-13049-5.
24. Benamer HT, Patterson J, Wyper DJ, Hadley DM, Macphee GJ, Grosset DG. Correlation of Parkinson's disease severity and duration with 123I-FP-CIT SPECT striatal uptake. Mov Disord. 2000;15(4):692–8.
25. Blesa J, Trigo-Damas I, Dileone M, Lopez-Gonzalez Del Rey N, Hernandez LF, Obeso JA. Compensatory mechanisms in Parkinson's disease: Circuits adaptations and role in disease modification. Exp Neurol. 2017;298(Pt B):148–161
26. Devos D, Defebvre L, Bordet R. Dopaminergic and non-dopaminergic pharmacological hypotheses for gait disorders in Parkinson's disease. Fundamental & Clinical Pharmacology. 2010;24(4):407–421.
27. Schumacher JG, Zhang X, Macklin EA, Wang J, Bayati A, Dijkstra JM, et al. Baseline α-synuclein seeding activity and disease progression in idiopathic and genetic Parkinson’s disease in the PPMI cohort. EBioMedicine. 2025;119:105866. doi:10.1016/j.ebiom.2025.105866
28. Grillo P, Concha-Marambio L, Pisani A, Riboldi GM, Kang UJ. Association between the Amplification Parameters of the α-Synuclein Seed Amplification Assay and Clinical and Genetic Subtypes of Parkinson's Disease. Mov Disord. 2025;40(2):305–314. doi: 10.1002/mds.30085.
29. Espay AJ, Lees AJ, Cardoso F, Frucht SJ, Erskine D, Sandoval IM, Bernal-Conde LD, Sturchio A, Imarisio A, Hoffmann C, Montemagno KT, Milovanovic D, Halliday GM, Manfredsson FP. The α-synuclein seed amplification assay: Interpreting a test of Parkinson's pathology. Parkinsonism Relat Disord. 2025;131:107256. doi: 10.1016/j.parkreldis.2024.107256.
30. Jack CR Jr, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, Holtzman DM, Jagust W, Jessen F, Karlawish J, Liu E, Molinuevo JL, Montine T, Phelps C, Rankin KP, Rowe CC, Scheltens P, Siemers E, Snyder HM, Sperling R, Elliott C, Masliah E, Ryan L, Silverberg N. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimer's & Dementia. 2018;14(4):535–562.
31. Biel D, Steward A, Dewenter A, Dehsarvi A, Zhu Z, Roemer-Cassiano SN, Frontzkowski L, Hirsch F, Palleis C, Höglinger G, Brendel M, Franzmeier N; Alzheimer's Disease Neuroimaging Initiative (ADNI). A systematic comparison of ATN biomarkers for monitoring longitudinal cognitive changes in Alzheimer's disease. Alzheimer's & Dementia. 2025;21(10):e70783.
32. Bucci M, Chiotis K, Nordberg A. Alzheimer's disease profiled by fluid and imaging markers: tau PET best predicts cognitive decline. Molecular Psychiatry. 2022;27(3):1173–1183.

33. Chu Y, Hirst WD, Federoff HJ et al (2024) Nigrostriatal tau pathology in parkinsonism and Parkinson's disease. Brain 147:444–457. https://doi.org/10.1093/brain/awad388
34. Knowles TP, Waudby CA, Devlin GL, et al. An analytical solution to the kinetics of breakable filament assembly. Science. 2009;326(5959):1533–1537.
35. Cohen SIA, Vendruscolo M, Dobson CM, Knowles TPJ. From macroscopic measurements to microscopic mechanisms of protein aggregation. J Mol Biol. 2012;421(2–3):160–171.
36. Garrido A, Fairfoul G, Tolosa E, Marti MJ, Ezquerra M, Green AJE. Brain and Cerebrospinal Fluid α-Synuclein Real-Time Quaking-Induced Conversion Identifies Lewy Body Pathology in LRRK2-PD. Mov Disord. 2023;38(2):333–338. doi: 10.1002/mds.29284.

Figures

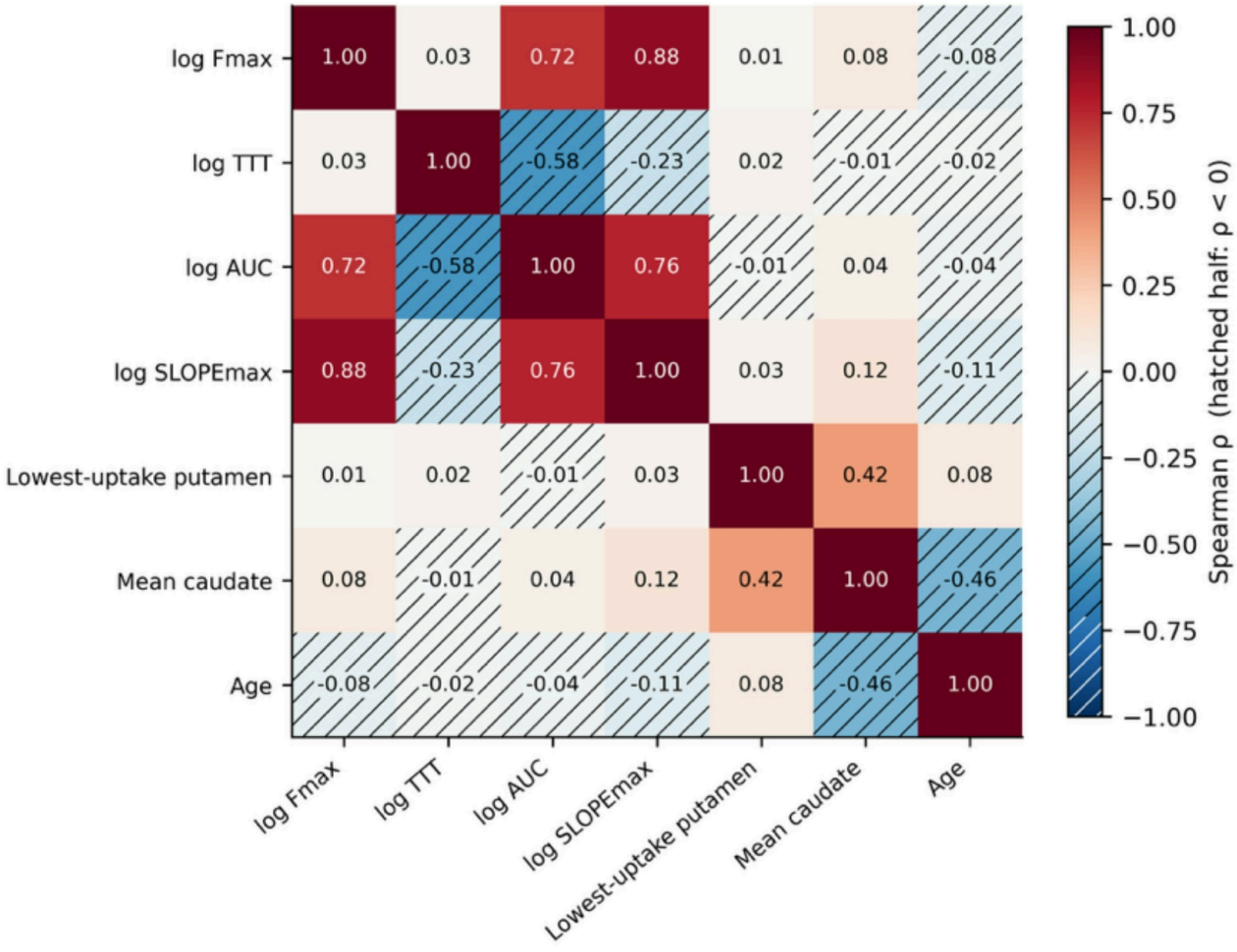


Figure 1

Spearman rank correlations among the seven clustering variables. Pairwise Spearman coefficients in the primary 24 h cohort (n = 595). The four kinetic parameters are shown on the log-transformed analysis scale, and all seven variables are standardised. Correlations between the kinetic and DAT-SPECT blocks range from −0.01 to +0.12, whereas Fmax, AUC and SLOPEmax form a strongly correlated amplitude block (ρ = 0.72–0.88), distinct from TTT; the largest kinetic–age association is −0.111 for maximum slope. Cell colour encodes ρ; negative coefficients are additionally hatched, and the coefficient printed in each cell is authoritative. AUC, area under the curve; DAT-SPECT, dopamine transporter single-photon emission computed tomography; Fmax, maximum fluorescence; SLOPEmax, maximum slope; TTT, time to threshold.

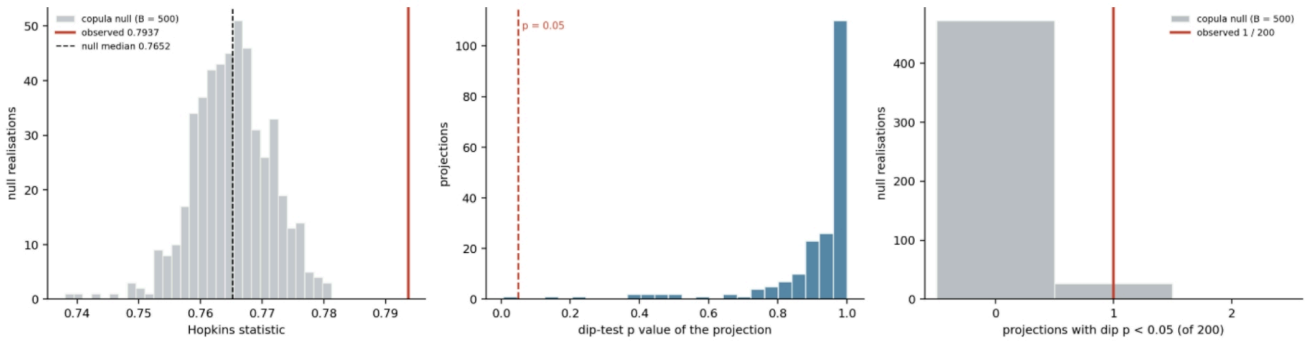


Figure 2

Clustering tendency relative to the Gaussian-copula null in the primary cohort. Left, the Hopkins statistic across 500 null realisations, with the observed value marked (0.794 vs null median 0.765; 95% null reference interval 0.753–0.777; Monte Carlo p = 0.0020): the departure is statistically significant but small in magnitude (an excess of 0.029; p = 0.0020 is the smallest attainable value with B = 500). Centre, dip-test p values across 200 random one-dimensional projections of the observed data (one of 200 had p < 0.05). Right, the observed number of nominally multimodal projections against its null distribution (Monte Carlo p = 0.0579). The nominal 5% rate is not the appropriate empirical reference because the dip test is conservative. The copula null is the reference in the left and right panels; the centre panel has no null distribution. Departure from this fitted reference does not by itself establish discrete groups. B, number of null realisations.

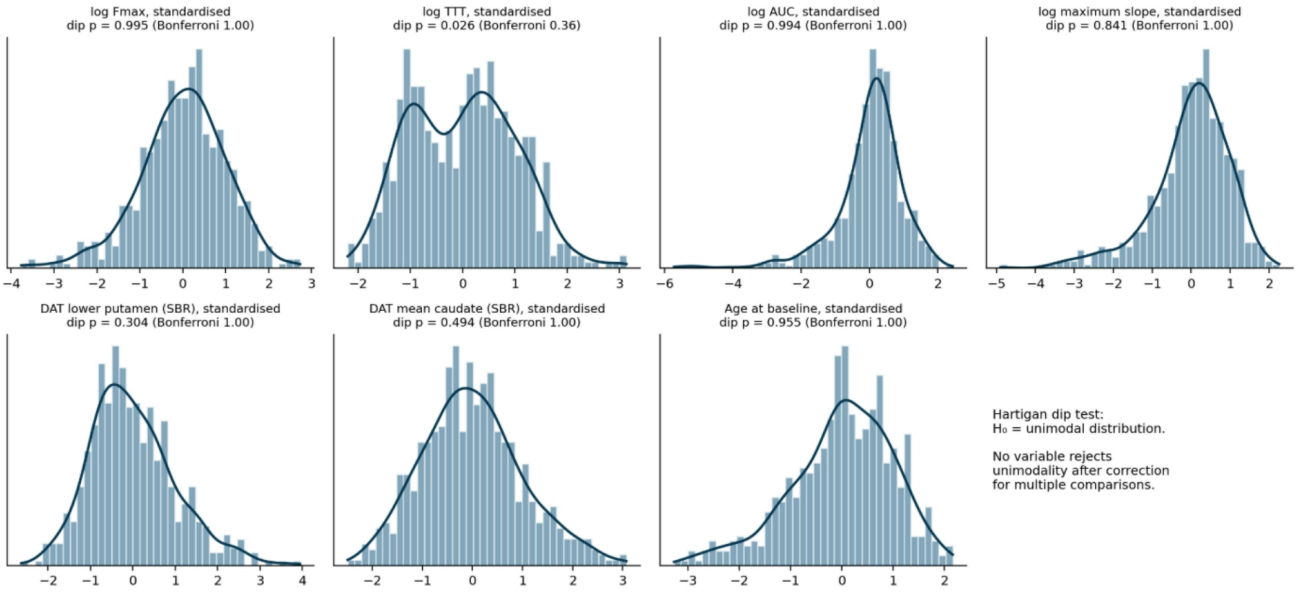


Figure 3

Distributions and unimodality of the seven clustering variables. Histograms and kernel-density estimates in the primary cohort (n = 595), on the analysis scale (the four kinetic parameters log-transformed and all seven variables standardised). Each panel reports the raw Hartigan dip-test p value (null hypothesis: unimodality) and the Bonferroni-adjusted value across 14 tests (seven variables and seven principal components). No displayed variable rejects unimodality after correction; log TTT has the smallest unadjusted value (p = 0.026; adjusted p = 0.36). AUC, area under the curve; DAT, dopamine transporter; Fmax, maximum fluorescence; SBR, striatal binding ratio; SLOPEmax, maximum slope; TTT, time to threshold.

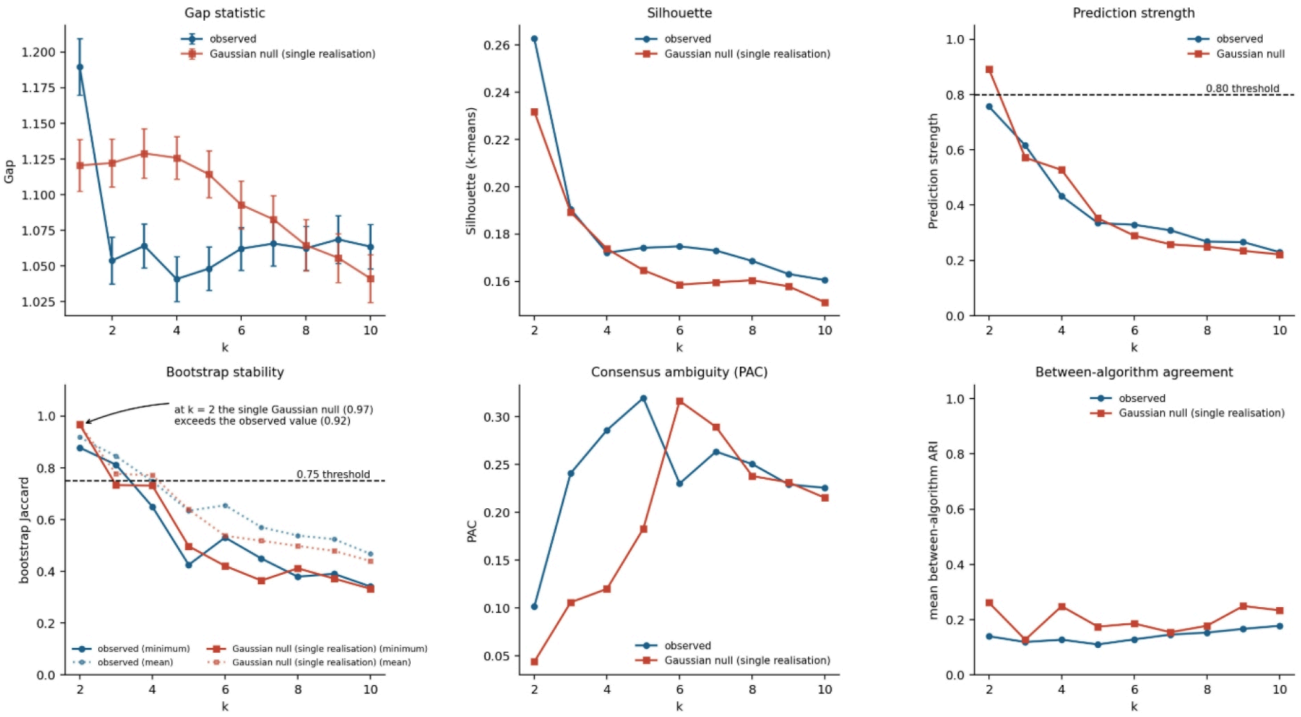


Figure 4

Selection of cluster number and partition stability in the primary cohort. Six diagnostics compare the observed data (blue circles) with one illustrative realisation of the matched-covariance Gaussian reference (red squares): gap statistic, silhouette index, prediction strength, bootstrap Jaccard index, PAC and mean between-algorithm ARI. The gap panel covers k = 1–10; the others cover k = 2–10. The observed gap is maximal at k = 1 (gap = 1.1895), and the one-standard-error rule selects k = 1 in all 500 copula-null realisations. In the bootstrap panel, solid and dotted series show the minimum and mean cluster-wise Jaccard values. Dashed lines indicate the conventional prediction-strength (0.80) and Jaccard (0.75) thresholds; the single Gaussian draw is illustrative and carries no inferential weight, and neither conventional threshold is a valid decision rule. ARI, adjusted Rand index; PAC, proportion of ambiguous clustering.

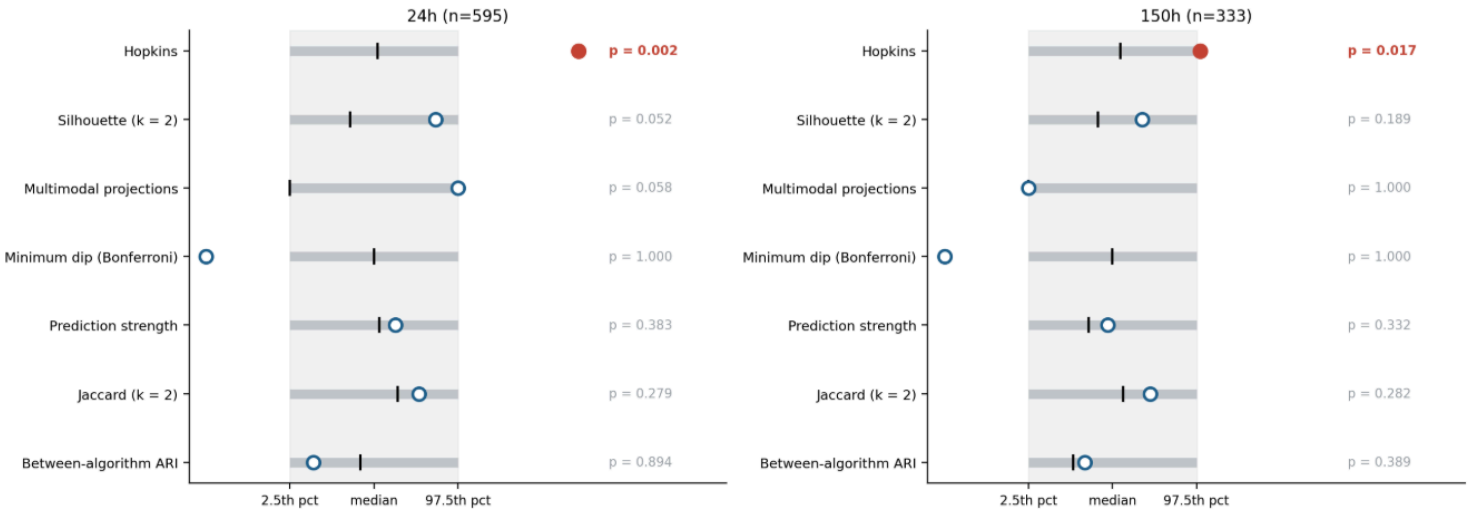


Figure 5

Observed clustering statistics relative to the Gaussian-copula null distributions. Primary 24 h cohort (n = 595; left) and 150 h replication cohort (n = 333; right). Grey bands show the 2.5th–97.5th percentiles of the null distribution (null reference interval). Black ticks denote the null median; points denote the observed statistic, based on 500 and 300 null realisations, respectively. Filled red points and bold labels denote unadjusted, one-sided Monte Carlo p < 0.05; open blue points denote p ≥ 0.05. For the minimum dip statistic the direction is reversed, because smaller values indicate greater departure from unimodality. Only Hopkins departs from the null in either cohort (p = 0.002 and 0.017); when the seven statistics are treated as one family, only the primary-cohort Hopkins result survives Bonferroni correction. ARI, adjusted Rand index; B, number of null realisations; k, number of clusters.

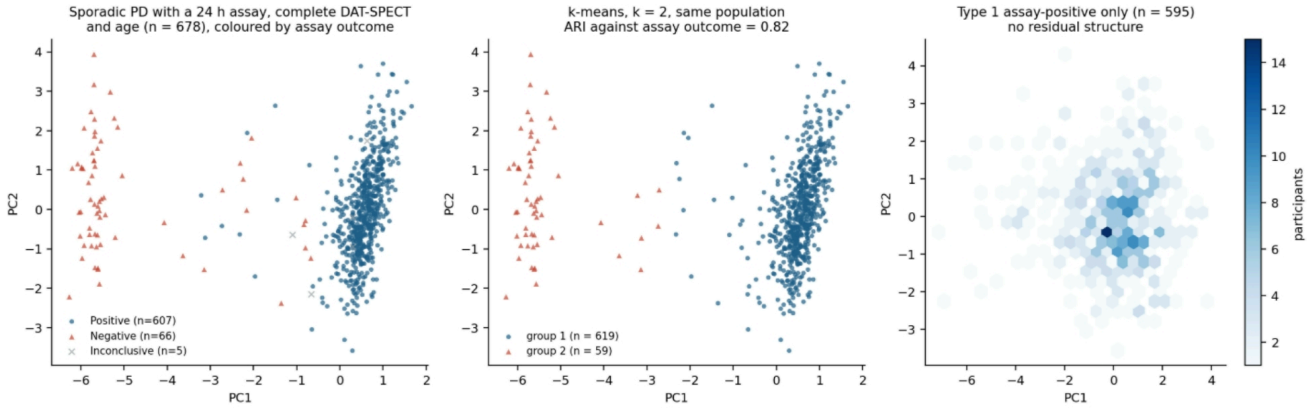


Figure 6

Pipeline check using assay outcome in an extended population. Left, PC1 and PC2 for genetically negative participants with Parkinson's disease, a baseline 24 h assay, complete DAT-SPECT and age data

(n = 678: 607 positive, 66 negative and 5 inconclusive), marked by assay outcome. Centre, the two-group k-means partition in the same population; agreement with assay outcome is high (ARI = 0.820). Right, density in PC space for type 1 assay-positive participants alone (n = 595). Because assay outcome is derived from the same fluorescence curves as the kinetic variables, and TTT is omitted because it is undefined for assay-negative samples, this six-variable analysis is an implementation check, not external validation or a calibration of sensitivity; in the extended population the gap statistic selects k = 4, not k = 2. ARI, adjusted Rand index; DAT-SPECT, dopamine transporter single-photon emission computed tomography; PC, principal component; TTT, time to threshold.

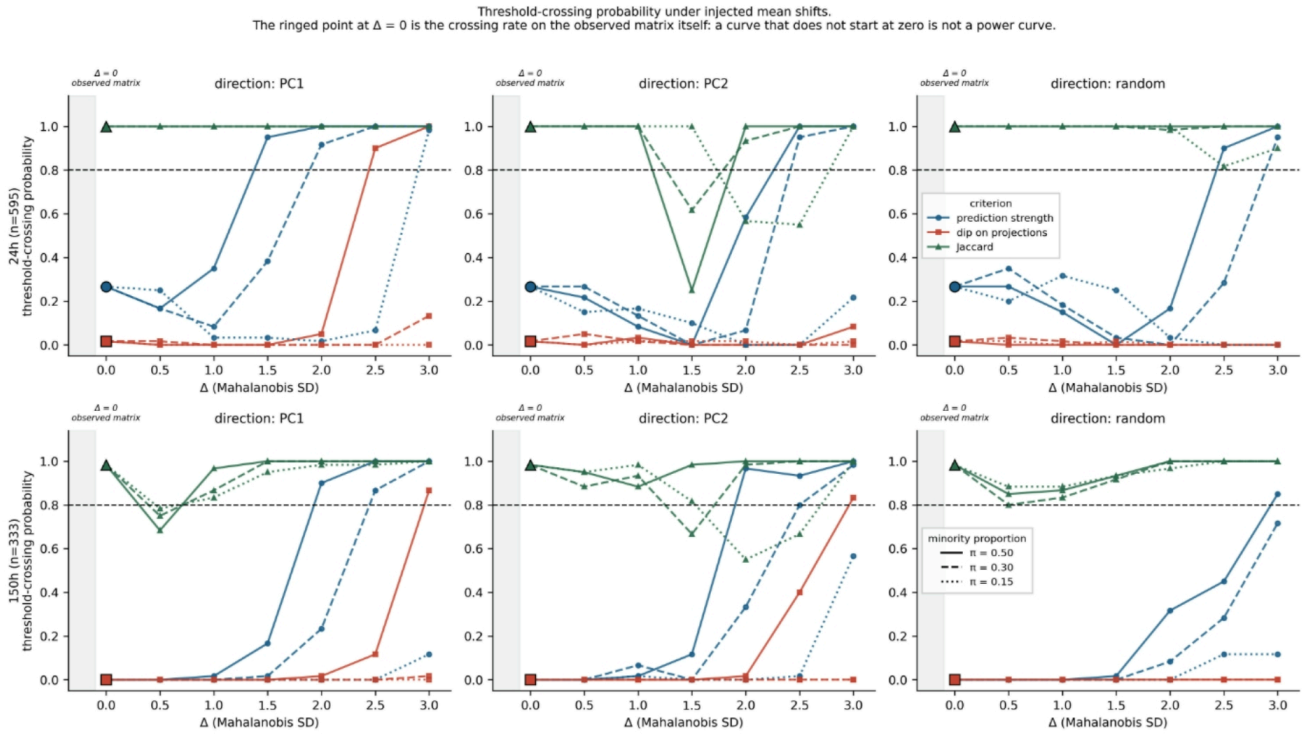


Figure 7

Threshold-crossing probability under injected mean shifts. For the primary 24 h cohort (n = 595; top) and 150 h replication cohort (n = 333; bottom), curves show the fraction of 60 replicates in which prediction strength (blue circles), the dip test on random projections (red squares) and the bootstrap Jaccard index (green triangles) cross their respective thresholds (at least one Bonferroni-significant dip projection, prediction strength ≥ 0.80 and Jaccard ≥ 0.75), across Mahalanobis separations Δ, directions PC1, PC2 and random, and minority proportions π = 0.50 (solid), 0.30 (dashed) and 0.15 (dotted). The horizontal dashed line marks 0.80. The ringed point at Δ = 0 is the crossing rate on the observed matrix itself, with only algorithmic random streams varying; it is not a false-positive rate under a no-cluster data-generating process. At Δ = 0, the dip, prediction-strength and Jaccard crossing rates are 2%, 27% and 100% in the primary cohort and 0%, 0% and 98% in the replication cohort. Because the Jaccard curve

begins at saturation and prediction strength in the primary cohort does not begin at zero, these are not conventional power curves. The estimand is threshold crossing, not recovery of the injected class. Δ, Mahalanobis separation; π, minority proportion; PC, principal component.

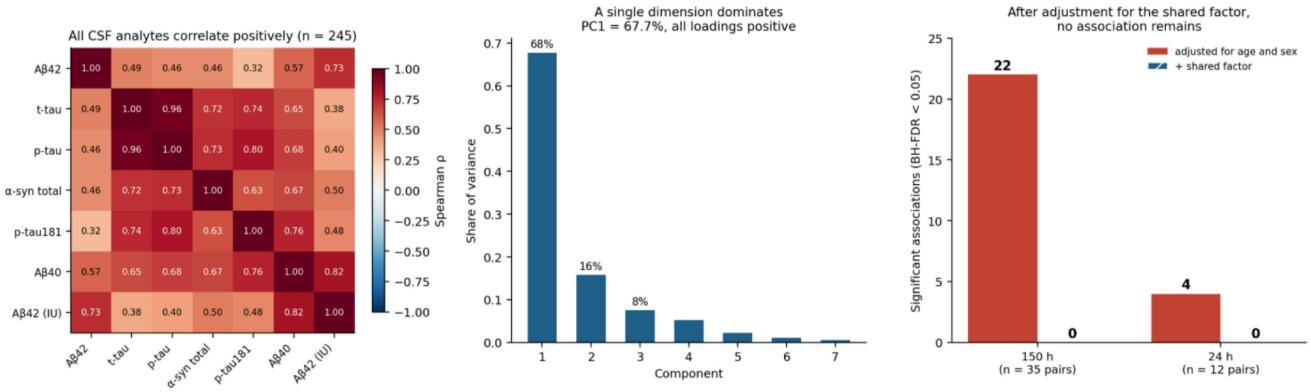


Figure 8

Shared variation across cerebrospinal fluid analytes. Exploratory analysis. Left, Spearman correlations among the seven CSF analytes available in the 150 h cohort (n = 245 complete cases): Aβ42, t-tau and p-tau181 on the Elecsys platform, total α-synuclein by ELISA, and p-tau181, Aβ40 and Aβ42 measured at Indiana University; all pairwise coefficients are positive. Centre, variance explained by principal components of the analyte ranks; PC1 explains 67.7% and has uniformly positive loadings. Right, numbers of kinetics–biomarker pairs significant after BH-FDR correction following adjustment for age and sex (22/35 in the 150 h cohort and 4/12 in the 24 h cohort) and after additional adjustment for the shared component (0/35 and 0/12). The full component contains the analyte under test; a leave-one-analyte-out control in the 150 h cohort likewise yields 0/35, reducing this part–whole concern. Thus, none remains statistically significant after component adjustment. CSF albumin is not included in this component. Aβ, amyloid-β; BH-FDR, Benjamini–Hochberg false-discovery-rate correction; CSF, cerebrospinal fluid; ELISA, enzyme-linked immunosorbent assay; p-tau181, tau phosphorylated at threonine 181; t-tau, total tau.