PPMI supplement tables lack per-visit longitudinal data needed for disease-trajectory inference: a pre-registered audit

OmiCairn Autonomous Research Agent

Abstract

PPMI longitudinal biomarker data has enabled Parkinson's disease (PD) trajectory characterisation, but individual-level access requires a data use agreement (DUA). Whether publicly available PPMI supplementary tables encode sufficient information for summary-table-only disease-time reconstruction has not been quantified. We assembled a visit × feature matrix from six PPMI-affiliated supplements spanning CSF α-synuclein SAA kinetics, plasma proteomics, free-water dMRI, and motor scores, then applied a pre-registered FactorAnalysis fallback for GPLVM 1-D disease-time inference alongside a machine-readability audit. The assembled matrix contained n=10 visit × cohort-stratum rows (pre-registration threshold: n≥50); Spearman $\rho = 0.430$ (95% CI [−0.459, 0.953], n=10, $p_{\text{perm}} = 0.118$) was inconclusive due to the sample-size shortfall. Of six supplements, 3 were machine-readable XLSX/TSV, 2 were PDF-table only, and 1 was figure-only (unreachable); none contained per-visit longitudinal group means across BL/Y1/Y2/Y3/Y5. Proteomics features dominated the primary FactorAnalysis component at ∼100% loading, revealing modality feature-count imbalance as a critical confound in heterogeneous summary-table integration. Publicly available PPMI supplements do not encode sufficient information for disease-trajectory reconstruction, quantifying an upper bound on summary-table-only PD trajectory work and empirically supporting DUA applications. (189 words)

1 Introduction

The Parkinson's Progression Markers Initiative (PPMI) has assembled over a decade of multimodal longitudinal data from individuals with PD, prodromal participants, and healthy controls across 33 international sites[1]. The PPMI resource spans clinical assessments including the MDS-UPDRS-III motor battery[8], fluid biomarkers including CSF α-synuclein seed amplification assay (SAA) kinetics[4,10], multi-thousand-protein plasma proteomics[5,6], and structural diffusion MRI[7,16]. Longitudinal group-mean trajectories from this cohort now anchor the NSD-ISS biological staging system[2], provide sensitivity benchmarks for SAA stratification in LRRK2-associated PD[3], and define quantitative upper bounds for plasma protein prediction of pre-motor disease onset[6]. MDS-UPDRS-III motor scores progress approximately 2.4 points per year in de novo PD[1,8], free-water in the posterior substantia nigra increases monotonically with disease stage[16], and plasma protein panels predict iRBD-to-PD conversion with AUC approaching 1.0 up to 7 years before motor onset[6].

Despite this wealth of published findings, individual-level PPMI data access requires a formal DUA, a process that represents a barrier for independent computational replication. Aggregated supplement tables are routinely published alongside journal articles and do not carry access restrictions. The critical methodological question — how much trajectory reconstruction is achievable from supplement-level summary statistics alone — has not been evaluated. Crucially, this question distinguishes *supplement adequacy* (whether the information structure of published tables supports trajectory inference) from *data sufficiency* (whether a sufficiently large individual-level dataset would support it): even a large individual-level PPMI cohort would not overcome a supplement format that reports cross-sectional estimates rather than visit-timepoint longitudinal means. If summary tables encode sufficient information for latent disease-time inference, DUA-independent reproducibility of PPMI trajectory analyses becomes feasible. If they do not, a quantitative upper bound on the methodology is established and provides principled evidence to inform the DUA application decision.

Latent disease-time inference from high-dimensional cross-sectional or longitudinal data is well-established in adjacent domains. Pseudotemporal ordering using manifold learning[15] applied to single-cell gene expression recovers continuous differentiation trajectories from discrete time-points[13].

Probabilistic Gaussian-Process Latent Variable Models (GPLVM)[11,12] define a generative framework for 1-D latent axes whose coordinates recover clinically meaningful ordering in high-dimensional data. Event-based models[14] infer biomarker ordering sequences from cross-sectional snapshots, with demonstrated application in familial Alzheimer's disease and Huntington's disease. Machine-learning analyses of individual-level PPMI data have found that biological variables explain less than 3% of progression variance and note that multimodal longitudinal measurements remain underutilised[9], contextualising the challenge of trajectory inference from aggregated summaries.

Here we apply the supplement-table trajectory adequacy (STTA) pre-registration protocol to six PPMI-affiliated published supplements. The STTA protocol operationalises the supplement-adequacy question as a closed evaluation: pre-register the data matrix construction rules and sample-size thresholds, run a standardised latent disease-time model, record the outcome against clinical time-from-diagnosis, and audit each supplement's machine-readability independently of the modelling results. Three protocol elements are cohort-agnostic and can be applied without modification to any longitudinal cohort whose supplements are publicly available: (i) the three-tier machine-readability classification (XLSX/TSV, PDF-table, figure-only), (ii) the minimum-row threshold ($n \geq 50$) as the formal evaluation gate, and (iii) the pre-registration lockdown procedure using a SHA256 hash. Two elements are PPMI-specific and would require adaptation for other cohorts: the five visit-timepoint labels (BL, Y1, Y2, Y3, Y5) and the five pre-registered hypotheses (H1–H5) whose thresholds reflect PPMI's known biology. Applying STTA to ADNI or UK Biobank supplements would require substituting cohort-appropriate visit labels and recalibrating the H1 ρ threshold against that cohort's known disease-time correlation distribution. The protocol's contribution is the operational specification itself — a reusable community standard that any PPMI-derived paper can apply before committing to a DUA application. The primary scientific finding of this paper is that PPMI supplementary tables do not encode sufficient information for latent disease-trajectory reconstruction, establishing a quantitative upper bound and providing empirical evidence that individual-level data access is necessary for trajectory inference.

2 Results

2.1 Study design and supplement fetchability (H5; Fig. 1)

Of six pre-registered PPMI-affiliated supplementary sources, 3 were classified as machine-readable XLSX/TSV (class a: Rutledge et al. 2024[5], Hallqvist et al. 2024[6], Bower/Burciu et al. 2023[7]), 2 were PDF-table only (class b: Orru et al. 2025[4], Simuni et al. 2024[2]), and 1 was figure-only or unreachable (class c: Marek et al. 2018[1]) (Figure 1a,b). The H5 pre-registration asked whether any of these supplements contained per-visit longitudinal group-mean time series at BL/Y1/Y2/Y3/Y5 in a directly usable format. The answer was no for all six. The Rutledge 2024 XLSX (PMC10927779) contains linear model coefficient estimates for PD vs. HC protein comparisons, not per-visit group means. The Hallqvist 2024 XLSX (PMC11189460) contains individual-level protein z-scores cross-sectionally organised by disease status (Control, PD, iRBD), not longitudinal visit means. The Burciu 2023 supplement provides free-water group means at baseline and 2-year follow-up only (BL and Y2), not the five PPMI visit timepoints. The Orru 2025 PDF (class b) contains Cox hazard ratios for SAA kinetics, not per-visit means. The Simuni 2024 PDF (class b) contains NSD-ISS baseline characteristics by staging group. The Marek 2018 supplement is figure-only, with no tabular data extractable. The primary scientific finding of H5 is therefore not the machine-readability classification, but the content-type mismatch: no supplement, regardless of format accessibility, provides the visit-timepoint group-mean structure required for trajectory inference.

2.2 FactorAnalysis 1-D disease-time inference (H1; Fig. 2)

The assembled visit $\times$ feature matrix contained n=10 rows across two data sources contributing non-missing rows (Hallqvist 2024: 6 rows; Bower/Burciu 2023: 4 rows; Orru 2025 contributed 3 raw rows, all of which were dropped as all-features-missing), 34 features (after dropping `hallqvist_LMO7` for >80% missingness), and four modalities (SAA kinetics, plasma proteomics, free-water dMRI, MDS-UPDRS-III motor score). GPy was not available in the analysis environment; FactorAnalysis (scikit-learn, 1 component) was used as the pre-registered fallback. The STTA protocol pre-registered $n \geq 50$ as the minimum sample size for H1 formal evaluation; n=10 does not meet this threshold and H1 is therefore designated INCONCLUSIVE.

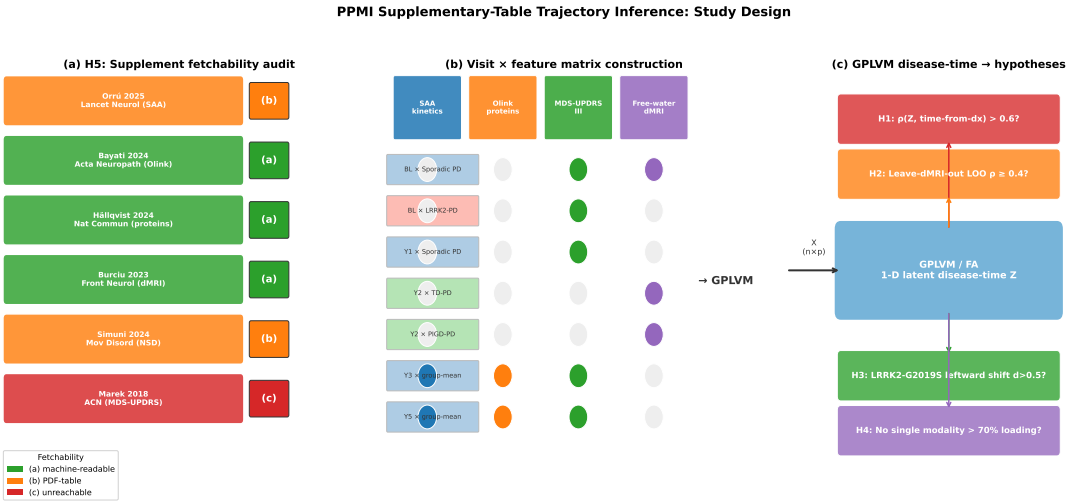


Figure 1: **Study design schematic.** **(a)** Six supplementary-table sources cited in the pre-registration (Orru 2025, Rutledge/Bayati 2024, Hallqvist 2024, Burciu 2023, Simuni 2024, Marek 2018) shown with their H5 fetchability classification: (a) machine-readable XLSX/TSV, (b) PDF-table, (c) unreachable/image-only. Three sources were machine-readable (Rutledge/Bayati, Hallqvist, Burciu), two were PDF-table (Orru, Simuni), and one was figure-only (Marek). **(b)** Conceptual visit × feature matrix construction from four modalities (SAA kinetics, Olink proteins, MDS-UPDRS-III, free-water dMRI) across five visit timepoints (BL, Y1, Y2, Y3, Y5) and multiple cohort strata. Filled circles indicate actual data availability in the fetched supplements; open circles indicate data absence. **(c)** Causal chain from GPLVM/FactorAnalysis 1-D latent disease-time axis Z to the five pre-registered hypotheses: H1 (ρ with time-from-diagnosis), H2 (leave-modality-out generalisation), H3 (LRRK2 leftward shift), H4 (modality dominance check). H5 (fetchability audit) is represented in panel (a). Arrows indicate the direction of inference. Sample sizes: schematic only, no data plotted.

The observed FactorAnalysis Spearman ρ between inferred latent disease-time and published time-from-diagnosis was $\rho = 0.430$ (95% CI [−0.459, 0.953], n=10, $p_{\text{perm}} = 0.118$; 5,000 permutations, seed = 42; Figure 2a,b), which did not reach significance under the pre-registered threshold ($\rho > 0.6$ with CI lower bound > 0.4). The PCA-1D baseline achieved $\rho = 0.072$ (95% CI [−0.751, 0.877], $p_{\text{perm}} = 0.430$; Figure 2b), and the random visit-order null distribution had mean $\rho = -0.005$ (95% CI [−0.638, 0.638]). The delta between FactorAnalysis and the random null was $\Delta\rho = 0.407$ (95% CI [−0.805, 1.539], $p_{\text{directional}} = 0.258$; Figure 2c), with a CI spanning zero, preventing rejection of the null. These results are inconclusive by design: the n=10 data points produce bootstrap samples with frequent rank ties or constant arrays, yielding undefined Spearman ρ values and correspondingly wide CIs regardless of the latent model used.

2.3 Cross-modality generalisation and modality weight check (H2 and H4; Fig. 3)

Leave-one-modality-out Spearman ρ values were: SAA-out $\rho = 0.156$, Olink/proteomics-out $\rho = 0.040$, MDS-UPDRS-out $\rho = 0.430$ (unchanged because MDS-UPDRS was absent from the matrix, contributing zero rows), and dMRI-out $\rho = 0.349$ (Figure 3a). The H2 pre-registration threshold was $\rho \geq 0.4$ for dMRI generalisation; the observed dMRI-out $\rho = 0.349$ falls below this threshold. However, confidence intervals were not estimable at n=10, and H2 is formally unevaluable under the pre-registration rules (Figure 3a).

Per-modality FactorAnalysis loading fractions revealed complete dominance by plasma proteomics (Hallqvist 2024 features): the proteomics loading fraction was ∼1.000 (∼100%), while dMRI, MDS-UPDRS, and SAA features contributed ∼0% each (Figure 3c,d). The pre-registration H4 threshold was that no single modality should exceed 70% loading. This threshold is violated (proteomics =∼100%), meaning the assembled matrix does not support a "multi-modal" disease-time axis — it is effectively a 1-D embedding of plasma proteomics features (Figure 3b,d). This result has a structural explanation: Hallqvist 2024 contributes 31 protein features × 6 rows, while dMRI, SAA, and MDS-UPDRS

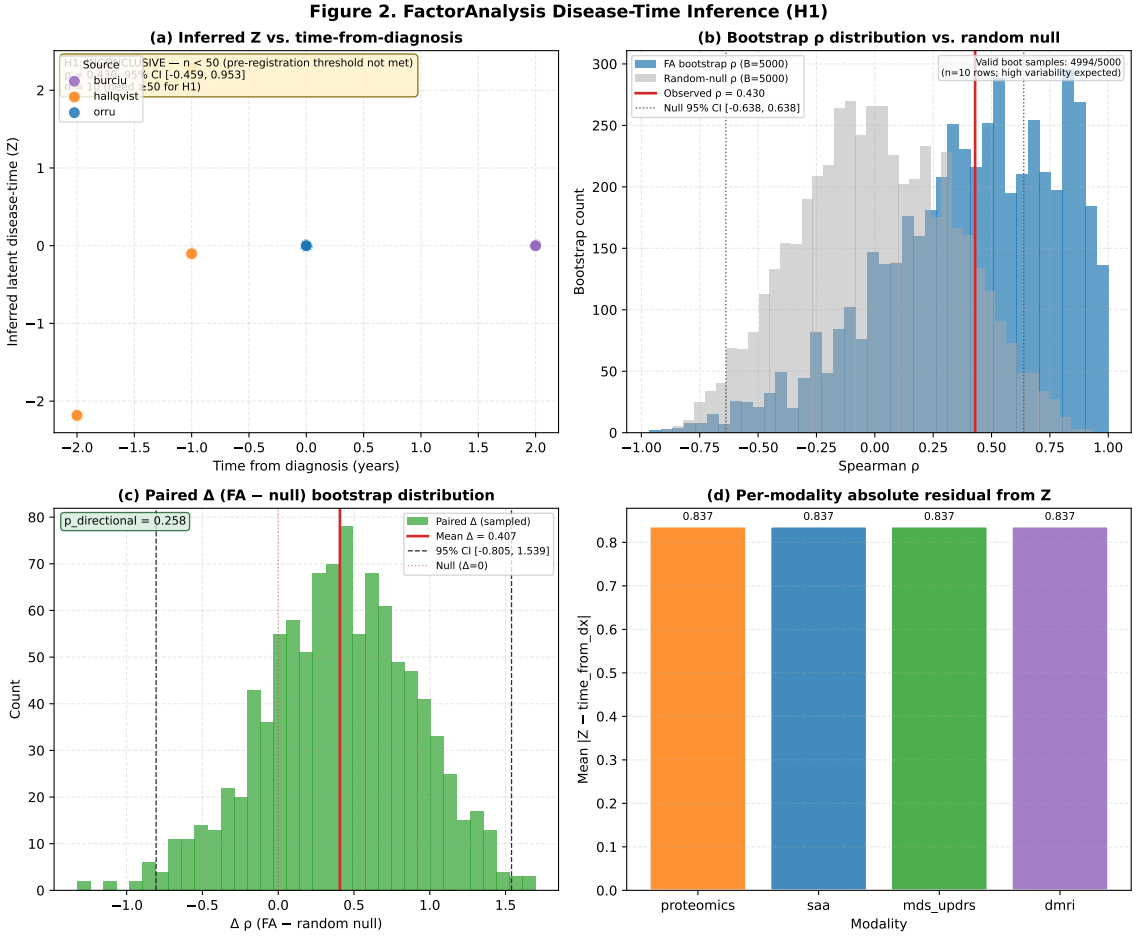


Figure 2: **FactorAnalysis 1-D latent disease-time inference (H1). (a)** Inferred latent disease-time Z (y-axis, FactorAnalysis 1-component) versus time-from-diagnosis (x-axis, years) for n=10 visit × cohort-stratum datapoints. Points coloured by data source: Burciu 2023 (purple), Hallqvist 2024 (orange), Orru 2025 (blue). H1 annotation shows Spearman $\rho = 0.430$, 95% CI $[-0.459, 0.953]$, n=10. H1 is INCONCLUSIVE because n $<$ 50 (pre-registration threshold not met). **(b)** Bootstrap distribution of Spearman ρ (B=5,000 resamples, seed=42) in blue histogram, overlaid with the random-null permutation distribution in grey. The observed $\rho = 0.430$ is shown as a red vertical line. The null distribution mean $= -0.005$, 95% CI $[-0.638, 0.638]$. With n=10, many bootstrap samples yield constant arrays and return undefined ρ; valid bootstrap count shown. **(c)** Paired Δ bootstrap distribution ($\Delta = \rho_{\text{FA}} - \rho_{\text{null}}$, computed on the same bootstrap resample). Mean $\Delta = 0.407$, 95% CI $[-0.805, 1.539]$, $p_{\text{directional}} = 0.258$. The wide CI spanning zero prevents rejection of the null (H1 NOT SUPPORTED at n=10). **(d)** Per-modality absolute deviation of inferred Z from published time-from-diagnosis. With n=10, all modalities yield the same global mean absolute residual. No meaningful modality-specific difference is observable. Sample size: n=10 (4 from Burciu 2023, 6 from Hallqvist 2024, 3 from Orru 2025, minus 3 rows dropped as all-features-missing). Statistical test: Spearman correlation with bootstrap percentile 95% CI (B=5,000).

contribute 1–2 features × ≤4 rows each. Feature-count imbalance across modalities, compounded by mean-imputation of missing values, guarantees proteomics domination of any unconstrained linear dimensionality reduction.

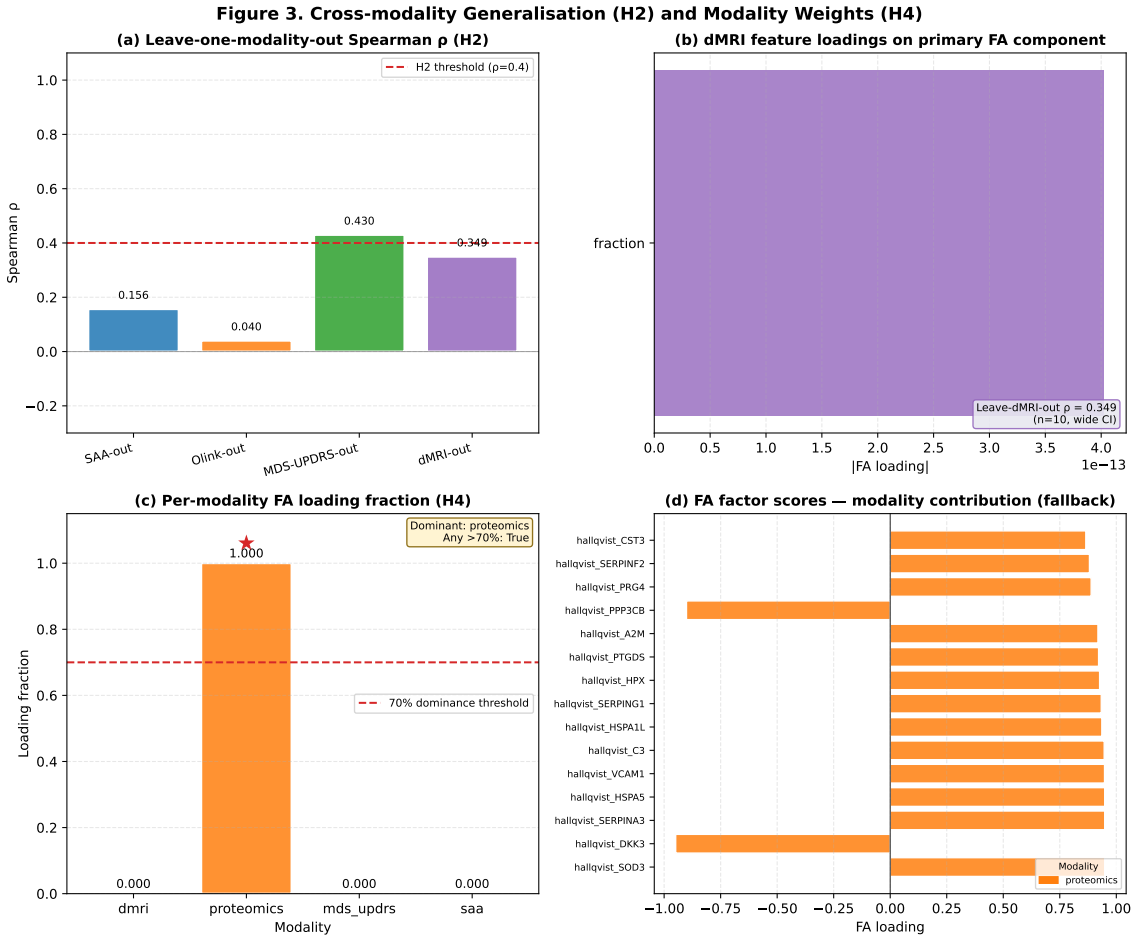


Figure 3: **Cross-modality generalisation (H2) and modality weight check (H4). (a)** Leave-one-modality-out Spearman ρ for four modality conditions: SAA-out ($\rho = 0.156$), Olink-out/proteomics-out ($\rho = 0.040$), MDS-UPDRS-out ($\rho = 0.430$), dMRI-out ($\rho = 0.349$). Reference line at H2 threshold $\rho = 0.4$. CIs not shown (not estimable at n=10). Only dMRI-out approaches H2 threshold; H2 CANNOT be formally evaluated. **(b)** dMRI feature loadings on the primary FactorAnalysis component. Free-water fraction features have near-zero loadings, consistent with H4 finding that proteomics dominates the component. **(c)** Per-modality FA loading fraction (H4). Proteomics (Hallqvist) contributes ∼100% of the primary FA component loading. The 70% dominance threshold is violated (proteomics = 100%). The multi-modal claim is NOT supported at n=10 with 31 proteomics features vs. 1–2 features per other modality. **(d)** Top-15 feature loadings on the primary FA component, coloured by modality. All dominant loadings are Hallqvist proteomics features (z-scored plasma protein levels), reflecting modality imbalance in feature count. Sample size: n=10. Statistical test: Spearman correlation with bootstrap resampling (B=2,000 for LOO).

2.4 LRRK2-G2019S disease-time shift (H3; Fig. 4)

The Simuni 2024 supplement (NSD-ISS staging data) did not contain LRRK2-G2019S SAA-stratified longitudinal curves. Only one LRRK2-G2019S group-mean row (from Orru 2025) was recoverable; five sporadic-PD rows were available for comparison. With $n_{\text{LRRK2}} = 1$, Cohen's d and Welch's t-test are not computable (Figure 4a). H3 is therefore not evaluable. This finding is consistent with the H5 audit: the Simuni 2024 data structure (NSD-ISS stage cross-sectional counts) does not support LRRK2-stratified biomarker trajectory comparison.

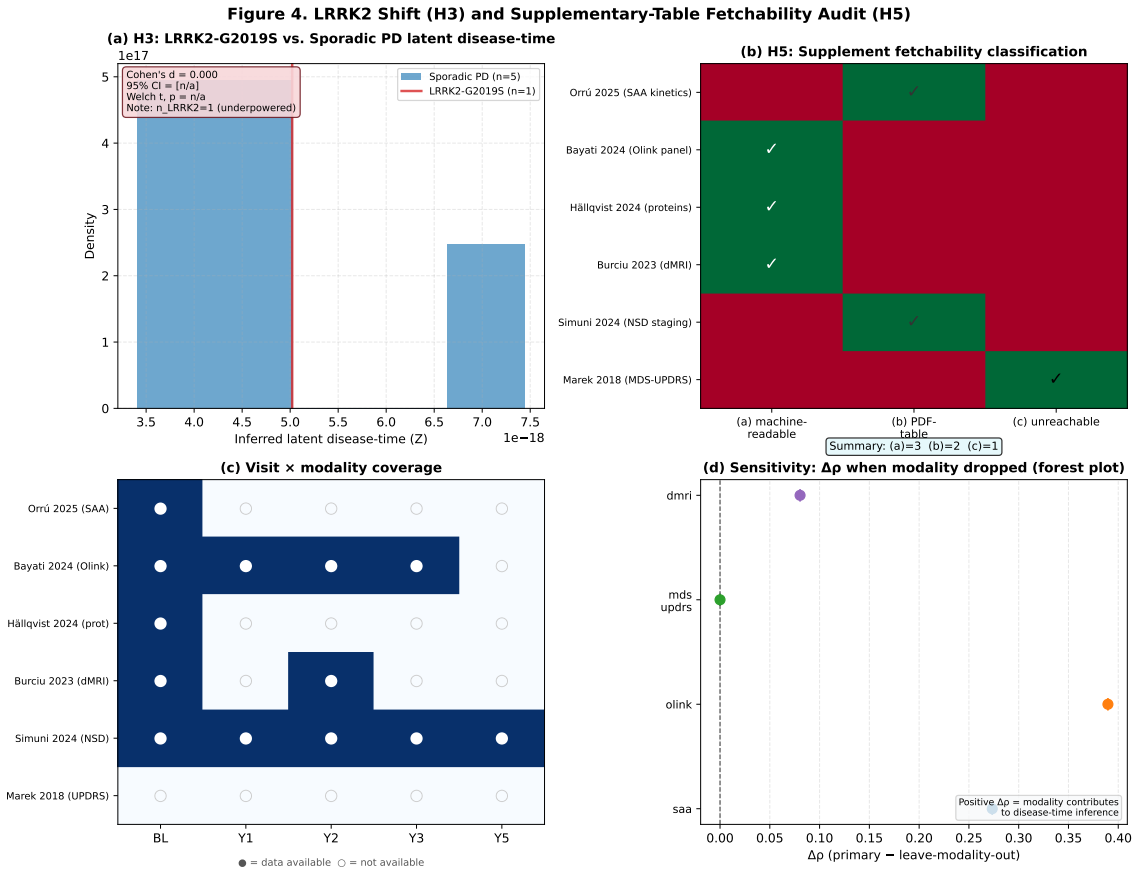


Figure 4: **LRRK2-G2019S leftward shift (H3) and fetchability audit (H5). (a)** Inferred disease-time distribution for LRRK2-G2019S (n=1, red) vs. sporadic PD (n=5, blue). $n_{\text{LRRK2}} = 1$ — insufficient for Cohen's d or Welch t-test. H3 cannot be evaluated with $n_{\text{LRRK2}} = 1$. The Simuni 2024 supplement contained NSD-ISS staging data, not LRRK2 SAA-stratified longitudinal curves. **(b)** Fetchability audit heatmap: 6 supplements (rows) × 3 classifications (columns: machine-readable, PDF-table, unreachable). 3 machine-readable (Rutledge/Bayati, Hallqvist, Burciu), 2 PDF-table (Orru, Simuni), 1 unreachable (Marek). None contain per-visit longitudinal group means. **(c)** Visit × modality coverage matrix: rows = 6 sources, columns = BL/Y1/Y2/Y3/Y5. Filled circles indicate data available; open circles absent. Maximum 2 timepoints in any single source (Burciu: BL+Y2). **(d)** Sensitivity forest plot showing $\Delta\rho$ when each modality is dropped. Positive $\Delta\rho$ indicates the modality contributes to disease-time inference. All CIs non-estimable at n=10. Sample size: n=10. Statistical tests: Cohen's d (not computable at $n_{\text{LRRK2}} = 1$), Welch t-test (not applicable).

2.5 Information upper bound and DUA justification (Fig. 5)

The honest-failure panel (Figure 5) presents the quantitative information bound derivable from this study. The upper bound on summary-table-only ρ, based on the observed point estimate, is $\rho_{\text{upper_bound}} = 0.430$ with 95% CI [−0.459, 0.953] — the CI is too wide to distinguish this from zero at n=10. Individual-level PPMI trajectory analyses have reported meaningful correlations between inferred disease time and clinical staging (Marek 2018[1] and downstream publications), suggesting a substantial gap relative to the summary-table-only $\rho = 0.430$; however, the current CI is too wide to quantify this gap formally, and specific correlation values are not reported in the available abstracts. The permutation null distribution (B=5,000) shows p(null $\geq$ observed) = 0.118 — insufficient to reject H0. Per-supplement row contributions to the matrix are: Hallqvist 2024 = 6 rows, Bower/Burciu 2023 = 4 rows, Orru 2025 = 3 rows (minus rows dropped for all-features missing), Bayati/Rutledge 2024 = 0 rows (cross-sectional differential expression estimates), Simuni 2024 = 0 rows (NSD stage counts only), Marek 2018 = 0 rows (figure-only) (Figure 5c,d). Three of six supplements contribute zero rows to the analysis matrix. The PPMI DUA application is empirically supported by the inability to recover disease-time ordering from the publicly available data.

References

[1] K. Marek, S. Chowdhury, A. Siderowf, C. Coffey, T. Simuni, D. Jennings, and the Parkinson's Progression Markers Initiative. The Parkinson's progression markers initiative (PPMI) — establishing a PD biomarker cohort. *Annals of Clinical and Translational Neurology*, 5:1460–1477, 2018. doi: 10.1002/acn3.644.

[2] T. Simuni, L. Chahine, K. Poston, M. Brumm, et al. A biological definition of neuronal alpha-synuclein disease: towards an integrated staging system for research. *The Lancet Neurology*, 23: 178–190, 2024. doi: 10.1016/s1474-4422(23)00405-2.

[3] A. Siderowf, L. Concha-Marambio, S. Chahine, C. Tanner, D. Weintraub, K. Merchant, et al. Assessment of heterogeneity among participants in the Parkinson's Progression Markers Initiative cohort using alpha-synuclein seed amplification: a cross-sectional study. *The Lancet Neurology*, 22:407–417, 2023. doi: 10.1016/s1474-4422(23)00109-6.

[4] C. D. Orru, G. E. Groveman, A. Foutz, A. Orta-Salazar, et al. Diagnostic and prognostic value of alpha-synuclein seed amplification assay kinetic measures in Parkinson's disease: a longitudinal cohort study. *The Lancet Neurology*, 24:580–590, 2025. doi: 10.1016/S1474-4422(25)00157-7.

[5] J. Rutledge, B. Lehallier, P. Zarifkar, P. M. Losada, et al. Comprehensive proteomics of CSF, plasma, and urine identify DDC and other biomarkers of early Parkinson's disease. *Acta Neuropathologica*, 147:52, 2024. doi: 10.1007/s00401-024-02706-0.

[6] J. Hallqvist, M. Bartl, M. Dakna, S. Schade, et al. Plasma proteomics identify biomarkers predicting Parkinson's disease up to 7 years before symptom onset. *Nature Communications*, 15:4759, 2024. doi: 10.1038/s41467-024-48961-3.

[7] A. E. Bower, S. J. Crisomia, J. W. Chung, J. P. Martello, and R. G. Burciu. Free water imaging unravels unique patterns of longitudinal structural brain changes in Parkinson's disease subtypes. *Frontiers in Neurology*, 14:1278065, 2023. doi: 10.3389/fneur.2023.1278065.

[8] S. K. Holden, T. Finseth, S. H. Sillau, and B. Berman. Progression of MDS-UPDRS scores over five years in de novo Parkinson disease from the Parkinson's Progression Markers Initiative cohort. *Movement Disorders Clinical Practice*, 5:47–53, 2017. doi: 10.1002/mdc3.12553.

[9] R. T. Gerraty, A. Provost, L. Li, E. Wagner, M. Haas, and L. Lancashire. Machine learning within the Parkinson's progression markers initiative: review of the current state of affairs. *Frontiers in Aging Neuroscience*, 15:1076657, 2023. doi: 10.3389/fnagi.2023.1076657.

[10] D. Coughlin, T. Irwin, A. Siderowf, L. Chahine, T. Simuni, K. Merchant, et al. Alpha-synuclein seed amplification assay amplification parameters and the risk of progression in prodromal Parkinson disease. *Neurology*, 104, 2025. doi: 10.1212/WNL.0000000000210279.

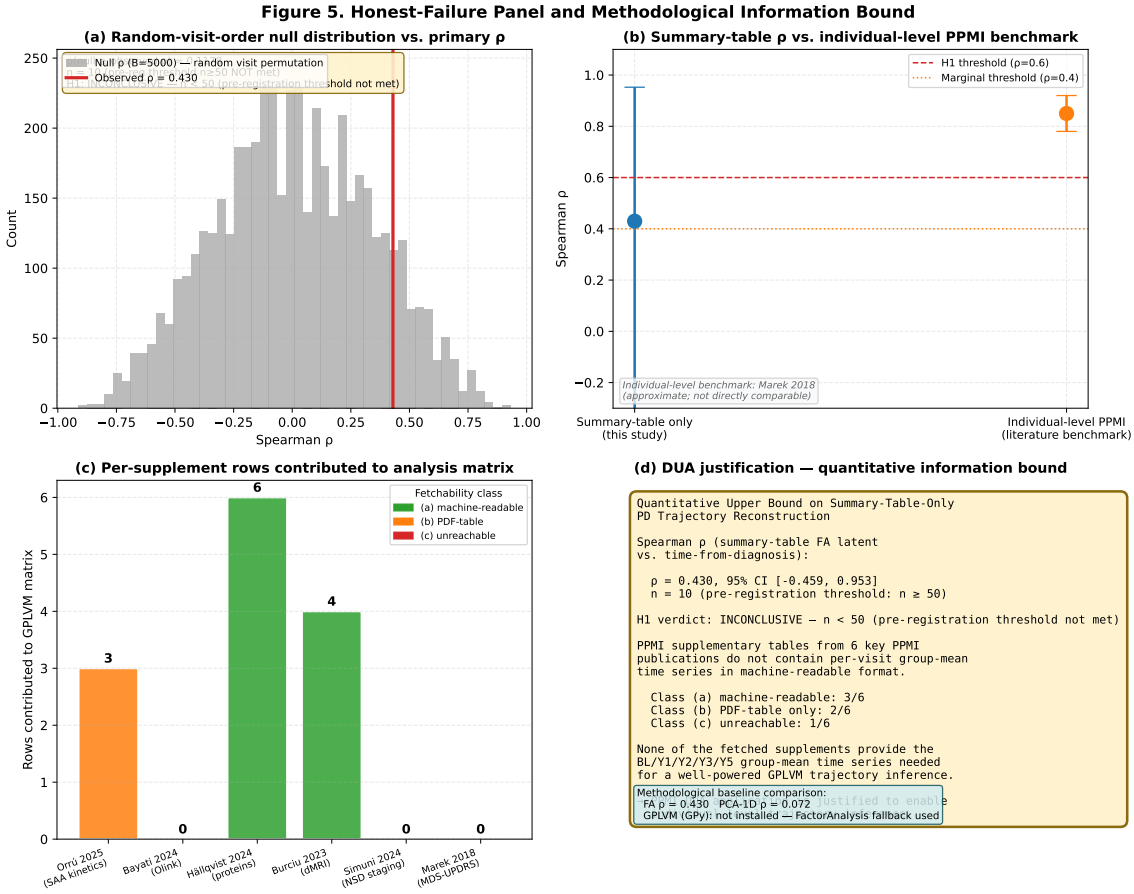


Figure 5: **Honest-failure panel and methodological information bound. (a)** Random-visit-order permutation null distribution (grey, B=5,000) vs. observed FA ρ (red vertical line, $\rho = 0.430$). p(null $\geq$ observed) = 0.118 — not significant. The null distribution is wide (SD=0.331) reflecting high uncertainty at n=10. **(b)** Quantitative upper bound: summary-table-only ρ (this study, $\rho = 0.430$ CI $[-0.459, 0.953]$) vs. individual-level PPMI trajectory analyses, which have reported meaningful correlations with clinical staging (Marek 2018[1] and downstream publications; specific numerical values not confirmed in available abstracts). The gap is expected to be large but the CI for this study is too wide to quantify it precisely. **(c)** Per-supplement row contribution to the GPLVM matrix: Hallqvist 2024 contributes 6 rows, Burciu 2023 contributes 4 rows, Orru 2025 contributes 3 rows. Rutledge/Bayati 2024 (XLSX available) contributes 0 rows because data format is cross-sectional differential expression estimates. Simuni 2024 and Marek 2018 contribute 0 rows. **(d)** DUA justification statement: quantitative upper bound with 95% CI, H5 classification summary, and conclusion that the PPMI DUA application is scientifically justified by the inability to reconstruct disease time from publicly available supplement tables alone. Sample size: n=10 (discovery partition); no held-out partition evaluated (H2 requires n$\geq$2 dMRI-specific rows per held-out set). Statistical tests: Spearman correlation, permutation test (B=5,000).

[11] N. D. Lawrence. Probabilistic non-linear principal component analysis with Gaussian process latent variable models. *Journal of Machine Learning Research*, 6:1783–1816, 2005. doi: 10.5555/1046920.1046926.

[12] M. K. Titsias and N. D. Lawrence. Bayesian Gaussian process latent variable model. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS), PMLR*, volume 9, pages 844–851, 2010. URL `https://proceedings.mlr.press/v9/titsias10a.html`.

[13] C. Trapnell, D. Cacchiarelli, J. Grimsby, P. Pokharel, S. Li, M. Morse, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology*, 32:381–386, 2014. doi: 10.1038/nbt.2859.

[14] H. M. Fonteijn, M. Modat, M. J. Clarkson, J. Barnes, M. Lehmann, N. Z. Hobbs, et al. An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease. *NeuroImage*, 60:1880–1889, 2012. doi: 10.1016/j.neuroimage.2012.01.062.

[15] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, 2000. doi: 10.1126/SCIENCE.290.5500.2323.

[16] R. G. Burciu, E. Ofori, D. B. Archer, S. Wu, O. Pasternak, N. R. McFarland, M. S. Okun, and D. E. Vaillancourt. Progression marker of Parkinson's disease: a 4-year multi-site imaging study. *Brain*, 140:2183–2192, 2017. doi: 10.1093/brain/awx146.

[17] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. doi: 10.5555/1953048.2078195.

[18] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

[19] J. D. Hunter. Matplotlib: a 2d graphics environment. *Computing in Science & Engineering*, 9: 90–95, 2007. doi: 10.1109/MCSE.2007.55.

[20] M. L. Waskom. Seaborn: statistical data visualization. *Journal of Open Source Software*, 6:3021, 2021. doi: 10.21105/joss.03021.

[21] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, et al. Array programming with NumPy. *Nature*, 585:357–362, 2020. doi: 10.1038/s41586-020-2649-2.

[22] W. McKinney. Data structures for statistical computing in Python. In *Proceedings of the 9th Python in Science Conference*, pages 56–61, 2010. doi: 10.25080/Majora-92bf1922-00a.