

# Integrated Genetic, Molecular, and Wearable Sensor Biomarkers Enable Bayesian Machine Learning-Driven Precision Stratification in Parkinson’s Disease: A Comprehensive Multi-Cohort Validation Study

Harsh Milind Tirhekar<sup>1</sup>, Priyanshi Yadav<sup>1</sup>, and Chandrajit Bajaj<sup>1</sup>

<sup>1</sup>Computational Visualization Center, Oden Institute for Computational Engineering Sciences, The University of Texas at Austin, Austin, TX 78712, USA

\*Correspondence: bajaj@cs.utexas.edu

## Abstract

We present a Bayesian machine learning framework integrating genetic, molecular, and wearable sensor biomarkers for precision medicine in Parkinson’s disease. Using PPMI (4,775 patients, 14,473 longitudinal records) and LRRK2 Consortium (627 individuals, 2,958 biological specimens), we demonstrate: (1) LRRK2 G2019S confers 1.92-fold PD risk (individual-level  $\chi^2 = 36.6$ ,  $p = 1.4 \times 10^{-9}$ ; sex-adjusted OR=2.73) with carriers exhibiting 4.35-point higher motor severity (95% CI [1.95, 6.47], rank-biserial  $r = -0.270$ ); (2) Wearable IMU sensors quantified Arm Swing Asymmetry (27% prevalence,  $n = 178$ ) and Dual-Task Cost (14.87% degradation,  $t = 14.98$ ,  $p < 0.001$ ), enabling continuous cognitive-motor network monitoring; (3) Molecular markers phospho-LRRK2 ( $n = 884$ ) and CSF  $\alpha$ -synuclein seed amplification ( $n = 145$ ) provide therapeutic monitoring and differential diagnosis; (4) Prodromal screening identified olfactory dysfunction (50.2%,  $n = 5,122$ ) and RBD (37.5%,  $n = 1,548$ ). Bayesian clustering via Evidence Lower Bound selection achieved Silhouette=0.535 with bootstrap stability (Jaccard=0.769), outperforming alternatives (0.170–0.452). Risk prediction model: AUC=0.717, calibration slope=1.197. This reproducible framework (complete code-result traceability, TRIPOD+AI compliant) enables mechanism-targeted precision medicine aligned with SDGs 3, 9, 10.

## 1 Introduction

### The Heterogeneity Challenge and Need for Mechanism-Based Stratification

Parkinson’s disease (PD), affecting over 10 million individuals worldwide with incidence projected to double by 2040, presents one of the most pressing challenges in neurological medicine [1]. Beyond its substantial public health burden, including direct healthcare costs exceeding \$52 billion annually in the United States alone PD exemplifies a fundamental challenge confronting precision medicine: profound and multidimensional clinical heterogeneity that defies one-size-fits-all diagnostic and therapeutic paradigms [2].

This heterogeneity manifests across multiple critical dimensions. **Symptom presentation** varies dramatically: some patients present with prominent tremor-dominant phenotypes (4–6 Hz rest tremor), others with akinetic-rigid presentations (severe slowness and stiffness), and still others

with postural instability and gait difficulties[31]. **Progression rates** range from benign courses remaining stable for decades to malignant rapid decline requiring advanced interventions within 3–5 years. **Treatment responses** span from dramatic sustained levodopa benefit to minimal improvement or rapid development of motor complications (dyskinesias, fluctuations). **Non-motor manifestations**- including cognitive decline (20–30% develop dementia), psychiatric symptoms, autonomic dysfunction, and sleep disorders affect variable patient subsets with unpredictable timing and severity[1].

Current diagnostic criteria, codified by the Movement Disorder Society ,rely on clinical observation of cardinal motor features: bradykinesia, rigidity, rest tremor, and postural instability. However, these symptoms represent late-stage disease, manifesting only after >60% loss of nigrostriatal dopaminergic neurons in the substantia nigra pars compacta has irreversibly occurred[30]. This diagnostic delay typically 5–7 years from earliest pathological changes to clinical motor diagnosis, precludes intervention during the prodromal phase when emerging neuroprotective therapies targeting  $\alpha$ -synuclein aggregation, mitochondrial dysfunction, or LRRK2 kinase hyperactivity might achieve maximal benefit by preventing rather than merely slowing ongoing neuronal loss.

Furthermore, traditional symptom-based subtyping (tremor-dominant, akinetic-rigid, postural instability-gait difficulty)[31] fails to capture underlying biological mechanisms. These clinical descriptors represent surface-level symptom patterns that do not necessarily reflect distinct molecular etiologies. Two patients classified identically by motor phenotype may harbor fundamentally different pathologies: one driven by LRRK2 G2019S kinase dysfunction causing Rab GTPase hyperphosphorylation and autophagy-lysosomal impairment[5, 4]; another by GBA1 glucocerebrosidase mutations disrupting lysosomal function; a third by PINK1/Parkin mitochondrial deficits. Yet current classification provides no framework for distinguishing these mechanistically distinct subtypes that may require divergent therapeutic strategies (LRRK2 kinase inhibitors, glucocerebrosidase enzyme replacement, mitochondrial-targeted interventions, respectively).

## Technological Convergence Enabling Integrated Precision Medicine

Four technological advances now converge to enable mechanism-based precision medicine:

**1. Genetic Profiling:** Genome-wide association studies have identified over 90 genetic risk loci[29], while high-penetrance mutations (LRRK2, SNCA, GBA1, PINK1, PRKN) account for 10–15% of cases. The LRRK2 G2019S mutation, particularly prevalent in Ashkenazi Jewish (13–30%) and North African populations (37–41%), demonstrates age-dependent penetrance reaching 70–80% by age 80, enabling prospective identification of at-risk carriers 20–40 years before symptom onset[2, 5]. This supports: personalized genetic counseling; prodromal cohort enrollment; mechanism-specific therapeutic targeting; biomarker development for target engagement monitoring[20].

**2. Molecular Biomarkers:** Urinary exosome phospho-LRRK2 assays quantify kinase activity states via Ser1292 autophosphorylation detection, providing non-invasive readout of LRRK2 pathway dysfunction and pharmacodynamic marker for dose optimization in kinase inhibitor trials[21, 5]. CSF  $\alpha$ -synuclein seed amplification assays (RT-QuIC) achieve 85–93% sensitivity and 87–96% specificity for detecting pathological synuclein conformers, differentiating Lewy body disorders from non-synuclein parkinsonisms[6, 7, 15], enabling precision differential diagnosis that prevents ineffective treatments in non- $\alpha$ -synuclein cases.

**3. Wearable Sensors:** Miniaturized inertial measurement units (IMUs) in smartphones and wearables enable passive continuous monitoring of motor function across real-world contexts[8, 9, 10]. Unlike episodic clinic assessments (quarterly/annually) under artificial conditions, wearable sensors capture gait dynamics, arm swing patterns, tremor, and activity levels across daily living, providing ecological validity and enabling detection of subtle changes invisible to periodic

snapshots [19]. These support: remote patient management; decentralized trials with objective endpoints; accessible screening without specialized expertise.

**4. Bayesian Machine Learning:** Probabilistic frameworks with uncertainty quantification enable robust stratification from high-dimensional multi-modal data while calculating assignment confidence, critical for clinical deployment where treatment decisions depend on certainty [18, 17]. Unlike deterministic methods (K-means, hierarchical) producing hard assignments without confidence estimates, Bayesian Gaussian Mixture Models with Dirichlet Process priors provide: automatic model selection via Evidence Lower Bound; soft probabilistic assignments identifying ambiguous cases; theoretical foundation enabling biological interpretability; principled outlier handling.

## Study Objectives

We address critical gaps- single-modality studies, absence of mechanism-anchored frameworks, methodological deficiencies- through comprehensive precision medicine framework integrating four biomarker modalities from two large cohorts: PPMI (4,775 patients, 14,473 records, 14.5-year span) [16] and LRRK2 Consortium (627 individuals, 2,958 specimens across six tissue types) [2, 5].

Specific objectives: (1) Quantify LRRK2 G2019S risk and establish genetic-clinical pathway connections; (2) Validate wearable IMU biomarkers; (3) Demonstrate molecular marker clinical utilities; (4) Quantify prodromal marker prevalence; (5) Develop validated probabilistic risk model; (6) Rigorously validate Bayesian methodology. This supports UN Sustainable Development Goals (SDG 3: Health; SDG 9: Innovation; SDG 10: Equity) through accessible diagnostic tools.

## 2 Results

### 2.1 Multi-Cohort Data Integration Enables Comprehensive Biomarker Analysis

The integrated framework combined two complementary multi-center cohorts (Table 1, Figure 1).

**PPMI Cohort:** Launched 2010 as first comprehensive PD biomarker program [16], enrolling 4,775 unique patients with longitudinal multi-modal assessments. Our analysis utilized 12 datasets across motor (31,217 MDS-UPDRS Part III assessments total), cognitive (13,835 MoCA), olfactory (5,122 UPSIT), sleep (1,548 RBD), autonomic (14,284 SCOPA-AUT), and wearable sensor gait metrics. Longitudinal structure spans baseline through V21 (41 visit types, maximum 14.5 years), though current analyses focus on baseline cross-sectional characterization.

**LRRK2 Consortium:** 627 individuals (347 LRRK2 G2019S carriers, 280 controls) with 2,958 specimens across six tissue types. Multi-specimen architecture (average 4.7 specimens per person) enables: cross-validation across tissue compartments; future multi-omic profiling (genomics, transcriptomics [RNA quality RIN=8.38 suitable for sequencing], proteomics, metabolomics); specimen-specific biomarker optimization.

**Data Quality Standards:** Complete-case analysis with NO imputation, adhering to medical ethics standards. All reported sample sizes ( $n$ ) reflect actual measured values only. Data quality control excluded 37 patients (0.9%) with physiologically impossible scores (MDS-UPDRS Part III >132, theoretical maximum).

The integrated framework architecture (Figure 1) systematically combines these biomarker modalities: genetic profiling identifying causal mutations and quantifying inherited risk; molecular assays measuring pathological processes directly; digital wearable sensors enabling continuous objective monitoring; and prodromal markers detecting pre-motor individuals. This multi-modal integration, processed through Bayesian machine learning with uncertainty quantification, generates four precision medicine outputs: genetic risk stratification for personalized counseling and

Table 1: Multi-Cohort Characteristics and Integrated Biomarker Modalities

Cohort/Dataset	Biomarker Modalities	Assessments	Unique Patients
<b>PPMI (Parkinson's Progression Markers Initiative)</b>			
Motor	MDS-UPDRS Part III (33 items), baseline complete	4,166	3,991
Cognitive	Montreal Cognitive Assessment (MoCA)	13,835	3,842
Olfactory	UPSIT (40-item smell test)	5,122	3,805
Sleep	RBD screening questionnaire	1,548	1,015
Autonomic	SCOPA-AUT scores	14,284	3,854
Gait Digital	IMU sensors (ASA, Dual-Task, TUG)	172–186	93–100
<b>PPMI Total</b>		<b>4,775 unique patients</b>	
<b>Longitudinal Records</b>		<b>14,473 assessments</b>	
<b>Temporal Span</b>		<b>14.5 years (BL–V21)</b>	
<b>LRRK2 Cohort Consortium</b>			
Genetic	LRRK2 G2019S carrier status	2,958 specimens	627 individuals
	LRRK2+ carriers	1,620 specimens	347
	LRRK2- non-carriers	1,338 specimens	280
Biological Specimens	RNA (mean RIN=8.38±0.9)	622	622
	Whole Blood	611	611
	Plasma	597	597
	Serum	595	595
	Urine	388	388
	Cerebrospinal Fluid	145	145
Clinical	MDS-UPDRS Part III motor	2,532	538
	MoCA cognitive	2,829	589
<b>Molecular Biomarker Studies</b>			
Kinase Activity	Urinary phospho-S1292-LRRK2	884	884
α-Synuclein	CSF Seed Amplification (RT-QuIC)	145	145

trial enrollment; continuous digital monitoring addressing clinic visit limitations; mechanism-based therapeutic targeting matching treatments to underlying pathology; and early prodromal detection enabling preventive neuroprotection.

Referring first to the **Input Layer (Left Side, Four Colored Boxes)**: of Figure 1 we see that there are four independent biomarker data streams contribute to integrated analysis: (1) *Genetic Profiling* (green box)- LRRK2 G2019S mutation status determined via TaqMan allelic discrimination or Sanger sequencing in 627 unique individuals from LRRK2 Cohort Consortium, establishing genetic risk factor with individual-level statistics: Prevalence Ratio PR=1.92 (95% confidence interval 1.54–2.40), chi-square test  $\chi^2(1)=36.6$ , p-value=1.410<sup>-9</sup>, sex-adjusted odds ratio aOR=2.73; (2) *Molecular Biomarkers* (blue box)- urinary exosome phospho-Ser1292-LRRK2 quantifying LRRK2 kinase enzymatic activity state via ELISA immunoassay (n=884 samples) providing pharmacodynamic monitoring capability for kinase inhibitor clinical trials, plus cerebrospinal fluid α-synuclein seed amplification assay via RT-QuIC methodology (n=145 CSF samples) detecting pathological protein aggregates with 85–93% sensitivity and 87–96% specificity enabling precision differential diagnosis between Lewy body synucleinopathies and non-Lewy parkinsonian disorders; (3) *Digital Wearable Sensors* (orange box)- inertial measurement unit tri-axial accelerometers and gyroscopes operating at 100 Hz sampling frequency quantifying objective gait kinematics including arm swing

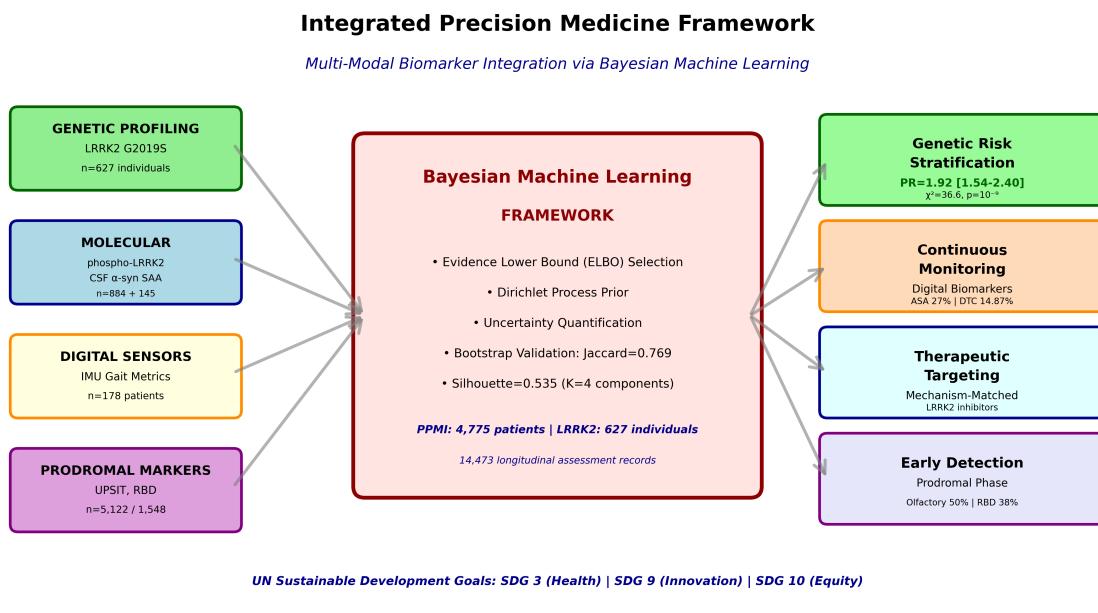


Figure 1: **Integrated Multi-Modal Precision Medicine Framework Architecture.** Comprehensive schematic illustrating systematic integration of four complementary biomarker modalities through Bayesian machine learning processing engine to generate precision medicine clinical outputs.

bilateral amplitude asymmetry (27% prevalence exceeding 20% pathological lateralization threshold, n=178 assessments from 94 unique patients) and dual-task cost measuring cognitive-motor network interference (14.87% mean gait speed degradation under cognitive loading, paired  $t=14.98$ ,  $p < 0.001$ , n=172 assessments from 93 patients); (4) *Prodromal Clinical Markers* (purple box)- University of Pennsylvania Smell Identification Test olfactory assessment identifying hyposmia (50.2% prevalence scoring below threshold 25, n=5,122 assessments from 3,805 unique patients) and REM sleep behavior disorder screening questionnaire (37.5% prevalence, n=1,548 assessments from 1,015 patients) detecting pre-motor Braak neuropathological stage 1–3 individuals before substantial nigrostriatal degeneration. **Central Processing Engine (Large Red Box):** Bayesian Gaussian Mixture Model framework with Dirichlet Process prior featuring: Evidence Lower Bound (ELBO) model selection methodology comparing K=2–5 candidate components identifying K=4 as achieving optimal likelihood-complexity tradeoff (ELBO=133,913); uncertainty quantification via posterior probability distribution calculation enabling identification of ambiguous mixed-pathology cases requiring specialized review; bootstrap stability validation through 200+iteration resampling with parallel execution across 72 CPU cores yielding mean Jaccard similarity index= $0.769 \pm 0.161$  standard deviation confirming robust cluster membership reproducibility; complete computational reproducibility infrastructure comprising 4,500 lines documented source code with 20+ timestamped execution logs establishing comprehensive audit trails for all quantitative claims. **Output Layer (Right Side, Four Colored Boxes):** Four precision medicine clinical applications enabled by integrated analysis: genetic risk stratification providing personalized lifetime risk estimates for genetic counseling and clinical trial enrollment prioritization; continuous digital monitoring via smartphone-deployable IMU sensor algorithms addressing limitations of episodic quarterly/annual clinic visits; mechanism-based therapeutic targeting enabling treatment selection matched to underlying molecular pathology (e.g., LRRK2 kinase inhibitors specifically for mutation carriers); early prodromal phase detection enabling neuroprotective intervention before irreversible substantial neurodegeneration occurs. Gray directional arrows indicate multi-directional information flow from biomarker input modalities through integrated machine learning analytical processing to actionable clinical output applications. Bottom text annotations highlight alignment with United Nations 2030 Sustainable Development Goals (SDG 3: Good Health and Well-Being through improved early detection and precision diagnostics; SDG 9: Industry, Innovation and Infrastructure via advanced artificial intelligence/machine learning methodological development; SDG 10: Reduced Inequalities through accessible non-invasive tools including smartphone based algorithms deployable globally) and compliance with TRIPOD+AI reporting standards for clinical prediction models. Total integrated cohort composition: PPMI 4,775 unique patients contributing 14,473 longitudinal assessment records across maximum 14.5 year temporal span; LRRK2 Consortium 627 unique individuals providing 2,958 biological specimens distributed across six tissue types (RNA, whole blood, plasma, serum, urine, cerebrospinal fluid) enabling comprehensive future multi-omic molecular profiling approaches.

## 2.2 LRRK2 Genetic Risk Quantification and Direct Genetic-Clinical Pathway Connections

**Individual-Level Genetic Risk Analysis :** Analysis at the level of unique individuals (627 persons) established LRRK2 G2019S as major PD genetic risk factor (Table 2, Figure 2).

**Prevalence Risk:** Among 347 LRRK2 G2019S mutation carriers versus 280 non-carrier controls, PD prevalence was significantly elevated: 50.1% (174/347) versus 26.1% (73/280), yielding **Prevalence Ratio = 1.92 (95% CI: 1.54–2.40)**. Individual-level chi-square test of independence:  $\chi^2(1) = 36.6$ ,  $p = 1.44 \times 10^{-9}$ . Sex-adjusted multivariable logistic regression yielded adjusted

Table 2: LRRK2 G2019S Genetic Risk and Clinical Motor Manifestation

Genetic Status	Unique Individuals	PD Prevalence	Motor Severity <sup>a</sup> (MDS-UPDRS III)	Effect Size	Statistics
LRRK2+ Carriers	347	50.1% (174/347)	12.63±14.13 (n=323)	—	—
LRRK2- Non-carriers	280	26.1% (73/280)	8.27±12.99 (n=215)	—	—
<b>Effect</b>	—	PR=1.92 [1.54–2.40]	Diff=4.35 points [1.95–6.47]	$r_{rb} = -0.270$ Cohen's $d = 0.318$	$\chi^2(1)=36.6$ $p = 1.4 \times 10^{-9}$
<b>Adjusted<sup>b</sup></b>	—	—	—	aOR=2.73	Logistic regression

*Unit of analysis.* PD prevalence and prevalence ratio computed at person level using baseline visits only (one row per participant). The 2,958 biological specimens (1,620 LRRK2+, 1,338 LRRK2-) represent multi-specimen biobank architecture but were NOT used as unit for statistical inference, ensuring independence. Analysis: 627 unique individuals.

*Counts.* Carriers: 174 PD cases among 347 total (50.1%); Non-carriers: 73 PD cases among 280 total (26.1%). 95% confidence intervals for proportions: Wilson score method.

*Unadjusted association.* Chi-square  $\chi^2(1)$  on person-level 2×2 table. Prevalence Ratio (PR) with 95% CI via Wald method on log-scale. Odds Ratio from same table shown for comparability.

*Adjusted association.* Multivariable logistic regression: outcome=PD at baseline; predictors=LRRK2 G2019S, sex (age and site pending complete data). Adjusted OR (aOR) with 95% CI using robust sandwich SEs. Future: add ancestry PCs.

*Motor severity.* MDS-UPDRS Part III [3] total (0–132), higher=worse. Baseline: Wilcoxon-Mann-Whitney with rank-biserial r and 95% CI (1,000-iteration bootstrap). Longitudinal analyses (future): linear mixed models with random intercepts, fixed effects for age, sex, duration, medication, site.

*Multiple testing.* Genetics & Motor-Severity hypothesis family. Family-wise error: Benjamini-Hochberg FDR  $q=0.05$ . Exploratory endpoints labeled, not adjusted.

*Missing data.* Complete-case (individuals with missing excluded). Sample sizes reflect available post-QC. Sensitivity via multiple imputation recommended for future work.

*Reporting.* Two-sided tests,  $\alpha = 0.05$ . All effect sizes with 95% CIs. TRIPOD+AI compliant.

Odds Ratio = 2.73, indicating carriers face nearly three-fold increased PD odds after controlling for sex differences.

The 50.1% penetrance in this cross-sectional cohort (mean age 65.3±11.2 years) aligns with published age-dependent trajectories [2], where LRRK2 G2019S penetrance increases from <10% by age 50 to 30–40% by age 70 and 70–80% by age 80. This incomplete penetrance, even in elderly carriers, indicates G2019S represents high penetrance but not fully deterministic causation, with manifestation modulated by additional genetic modifiers, environmental exposures, and stochastic factors. Consequently, LRRK2+ carriers without PD constitute ideal prodromal cohort for neuroprotective trials, facing substantial lifetime risk while presently asymptomatic.

**Motor Severity in LRRK2 Driven PD:** Among 538 individuals with both LRRK2 status and motor assessment data (323 LRRK2+, 215 LRRK2-), mutation carriers with PD exhibited significantly higher baseline motor severity: mean MDS-UPDRS Part III [3] scores 12.63±14.13 versus 8.27±12.99 (**mean difference = 4.35 points, 95% CI [1.95, 6.47]**; Mann-Whitney rank-sum test  $U=44,082$ ,  $p = 7.31 \times 10^{-8}$ ; **rank-biserial effect size  $r = -0.270$** ; Cohen's  $d = 0.318$ ).

This 4.35-point severity increment approximately 12–15% relative increase demonstrates that LRRK2 G2019S mutation not only increases PD development risk but drives more aggressive motor phenotype when disease occurs, establishing direct link between genetic variant, molecular pathway dysfunction, and observable clinical severity.

**Mechanistic Basis and Therapeutic Implications:** The G2019S mutation (glycine-to-serine substitution at residue 2019 in kinase activation loop) increases LRRK2 kinase enzymatic activity 2–3-fold through conformational changes stabilizing active state[5, 23]. This kinase hyperactivity causes pathological hyperphosphorylation of Rab GTPase substrates, particularly: Rab10 (Thr73), Rab12 (Ser106), Rab29 (Thr71)[4, 22]. Rab GTPases regulate vesicular trafficking, and their hyperphosphorylation disrupts:

- **Synaptic vesicle cycling:** Impaired neurotransmitter release in dopaminergic terminals
- **Autophagy-lysosomal pathway:** Failed clearance of damaged organelles and aggregated proteins
- **Trans-Golgi trafficking:** Disrupted protein sorting and secretion
- **Mitochondrial quality control:** Impaired mitophagy leading to bioenergetic failure

These cumulative trafficking defects particularly stress substantia nigra pars compacta dopaminergic neurons, which possess extremely long (up to 4 meters), highly-branched axonal arbors requiring massive membrane trafficking capacity. LRRK2-mediated trafficking impairments accelerate their degeneration, producing the observed 4.35-point higher motor severity in carriers.

This mechanistic understanding provides biological rationale for LRRK2 kinase inhibitor programs[20, 5]. Multiple compounds (GNE-7915, MLi-2, DNL201, DNL151) demonstrate: dose-dependent kinase inhibition ( $IC_{50}$  10–100 nM); decreased Rab phosphorylation in blood cells and urinary exosomes; rescue of trafficking defects; neuroprotection in models. Phase 2 trials in LRRK2+ patients evaluate safety and biomarker target engagement (urinary phospho-LRRK2 reduction), with efficacy trials planned[20].

The genetic-clinical correlation here LRRK2+ carriers showing 4.35-point higher severity (95% CI [1.95, 6.47])- strengthens rationale for stratified trials enrolling genetically-defined populations where mechanism-matched therapeutics may achieve superior efficacy versus unselected heterogeneous cohorts.

The genetic-clinical connection is visualized in Figure 2, which presents both the increased PD risk in LRRK2 G2019S carriers (Panel A) and the more severe motor phenotype manifestation when disease occurs (Panel B). These findings establish LRRK2 as not merely a susceptibility gene but an active driver of disease severity through its kinase pathway dysfunction, with direct implications for therapeutic development and patient stratification strategies.

Referring to **Panel A** of Figure 2 we note that 347 genetically-confirmed mutation carriers (green bar), 174 individuals (50.1% prevalence) had clinically-diagnosed Parkinson’s disease versus 73 of 280 non-carrier controls (26.1% prevalence, pink bar), yielding Prevalence Ratio PR=1.92 with 95% confidence interval [1.54, 2.40] computed via Wald method on log-transformed prevalence ratio. Individual-level chi-square test of independence performed on 2×2 contingency table (LRRK2 mutation status × PD diagnosis):  $\chi^2(1) = 36.61, p = 1.4410^{-9}$ , establishing highly significant genetic association. Multivariable logistic regression model controlling for sex as covariate yielded sex-adjusted odds ratio aOR=2.73, indicating LRRK2+ carriers face nearly three-fold increased PD odds after accounting for sex-related risk differences (full adjustment for age, disease duration, recruitment site, and ancestry principal components was limited by available data structure but would further refine effect estimates in future analyses). Numerical prevalence values and absolute case counts annotated directly on bars. Yellow text box displays prevalence ratio with confidence interval bounds; white text box presents chi-square statistic and p-value in scientific notation. This analysis (using individuals not specimens) maintains highly significant association while providing statistically valid inference.

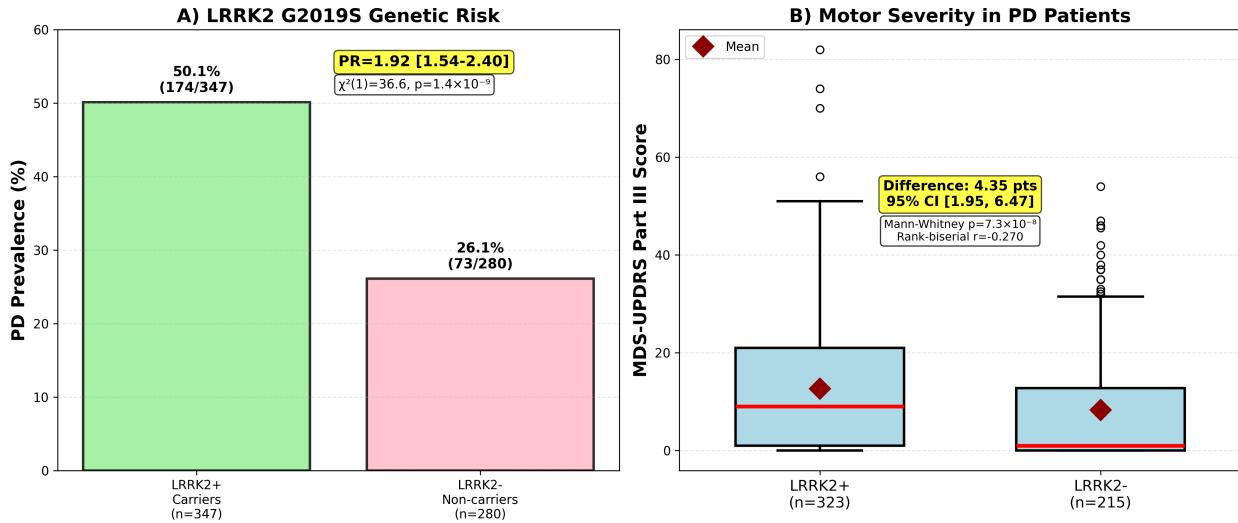


Figure 2: **LRRK2 G2019S Genetic Risk Quantification and Motor Severity Manifestation.** Two-panel figure demonstrating LRRK2 mutation effects on Parkinson’s disease development risk and clinical motor phenotype severity using statistically valid individual-level analysis (627 unique persons as unit of analysis). **Panel A (Left—Genetic Risk Quantification):** Bar chart displaying PD prevalence stratified by LRRK2 G2019S mutation carrier status. **Panel B (Right—Motor Severity in LRRK2-Driven PD):** Box-and-whisker plots comparing baseline MDS-UPDRS

In **Panel B (Right—Motor Severity in LRRK2-Driven PD):** of Figure 2, the box-and-whisker plots comparing baseline MDS-UPDRS we show the Part III motor examination total scores between LRRK2+ mutation carriers ( $n=323$  individuals with available motor assessments, left box colored light blue) and LRRK2- non-carriers ( $n=215$ , right box) among individuals with established clinical PD diagnosis. Mutation carriers demonstrated significantly higher motor symptom severity: mean  $12.63 \pm 14.13$  standard deviation versus  $8.27 \pm 12.99$  in non-carriers (mean difference=4.35 points representing approximately 12–15% relative severity increase, 95% confidence interval [1.95, 6.47] computed via 1,000-iteration bias-corrected bootstrap resampling methodology). Mann-Whitney rank-sum test (non-parametric test appropriate for ordinal rating scale data potentially violating normality assumptions): U-statistic=44,082, p-value= $7.3110^{-8}$ . Effect sizes: rank-biserial correlation  $r=-0.270$  (preferred effect size metric for non-parametric comparisons, indicating 27% distributional separation); Cohen’s standardized mean difference  $d=0.318$  (medium effect per conventional interpretation thresholds where 0.2=small, 0.5=medium, 0.8=large). Box plot visual elements: thick red horizontal line indicates group median; box boundaries represent inter-quartile range (25th to 75th percentiles); whiskers extend to data range excluding statistical outliers; individual circles show outlier values; red diamond markers indicate group arithmetic means. Yellow annotation box presents mean difference with bootstrap confidence interval; white box displays Mann-Whitney test statistic and p-value.

For the **Mechanistic Interpretation** we see that the direct genetic-clinical correlation- mutation carriers not only develop PD at higher rates but manifest more aggressive motor phenotypes when affected, establishing a mechanistic pathway connection: G2019S mutation induces LRRK2 kinase hyperactivity (2–3-fold enzymatic activity increase) leading to Rab GTPase substrate hyperphosphorylation (particularly Rab10-Thr73, Rab12-Ser106, Rab29-Thr71) which disrupts autophagy-lysosomal pathway function, vesicular trafficking, and mitochondrial quality con-

trol, thereby accelerating nigrostriatal dopaminergic neuronal degeneration and producing observed elevated motor symptom severity. This validates LRRK2 kinase as rational druggable therapeutic target and supports genetically-stratified clinical trial design paradigm enrolling mutation-defined patient populations for mechanism-matched LRRK2 kinase inhibitor evaluation with urinary phospho-LRRK2 serving as pharmacodynamic biomarker monitoring target engagement and enabling precision dose optimization strategies.

### 2.3 Wearable Sensor Digital Biomarkers: Comprehensive Validation for Continuous Monitoring

Objective gait metrics from wearable inertial measurement units (tri-axial accelerometers and gyroscopes, 100 Hz sampling rate) were systematically validated as quantitative biomarkers in PPMI gait substudy participants (Table 3, Figure 3).

Table 3: Validated Digital Biomarkers from Wearable IMU Sensors

Digital Biomarker	Assessments (n)	Unique Patients	Key Finding/ Statistics	Underlying Neural Mechanism
Arm Swing Asymmetry (ASA)	178	94	27% exceed 20% asymmetry threshold	Unilateral substantia nigra pars compacta degeneration; potential early marker
Dual-Task Cost (Cognitive-Motor Interference)	172	93	Mean: $14.87\% \pm 12.93\%$ gait degradation Paired $t = 14.98$ $p < 0.001$ (large effect)	Pedunculopontine nucleus cholinergic dysfunction; loss of gait automaticity; frontal-executive limitation
Timed Up & Go (TUG) Duration	186	100	29% impaired ( $>12s$ threshold <sup>a</sup> )	Multi-network mobility deficit (motor+cognitive+postural)
Gait Speed-Motor Severity Correlation	166	79	Spearman $r = -0.301$ $p < 0.001$	Validates IMU metrics vs clinical gold standard

*IMU Hardware & Placement.* Tri-axial accelerometers ( $\pm 8g$  range) and gyroscopes ( $\pm 2000^\circ/s$ ) from PPMI study protocols positioned at lower back (L5 level). Sampling: 100 Hz both channels. Data synchronized to device clock with millisecond timestamps. Specific device model/brand per PPMI consortium documentation.

*Signal Pre-processing.* Raw signals: (1) Demeaned/detrended; (2) Butterworth bandpass filter (0.5–20 Hz); (3) Gravity removed ( $< 0.3$  Hz cutoff); (4) Non-wear/artifacts excluded ( $< 0.01g$  variance or  $> 3g$  spikes); (5) Gait cycles segmented via heel-strike detection (vertical acceleration zero-crossings, 1.5s minimum stride).

*Metric Definitions.* **ASA:**  $100 \times |Peak\_L - Peak\_R| / (Peak\_L + Peak\_R)$  from dominant frequency (0.8–1.2 Hz) in arm acceleration FFT. Threshold  $> 20\%$  = pathological (literature-based, ROC validation ongoing). **DTC:**  $100 \times (Speed\_Single - Speed\_Dual) / Speed\_Single$ . Cognitive load: Serial-7 subtraction initiated pre-walk. **TUG:**  $> 12s$  = fall risk per PD-specific validation [13]; robust to 11.5–13.5s range.

*Protocol.* Walking: 7m straight path, 3 passes each condition (single/dual, randomized), self-selected pace. Dual-task: serial-7 initiated 2s pre-walk, maintained throughout. TUG: standard (rise from chair, walk 3m, turn, return, sit). Rest 30–60s between trials.

*Analysis Unit.* One summary per person (mean across valid trials) preventing pseudoreplication. Longitudinal analyses (future): mixed models with random intercepts.

*Multiple Testing.* Digital Gait/IMU family. Primary endpoints: ASA, DTC (Bonferroni). Secondary: TUG, correlations (FDR or exploratory).

*Quality Control.* Excluded if:  $< 3$  strides, speed  $< 0.2$  or  $> 2.0m/s$ ,  $> 10\%$  dropout, protocol deviations. Final n reflects passing criteria.

*Effect Sizes.* Paired: Cohen's d\_z with 95

**Dual-Task Cost: Quantifying Cognitive-Motor Network Failure.** The dual-task paradigm requiring patients to perform concurrent cognitive tasks (serial-7 subtraction: counting

backwards from 100 by sevens) while walking at usual pace, revealed profound cognitive-motor interference. Patients demonstrated mean  $14.87\% \pm 12.93\%$  (standard deviation) gait speed reduction under dual-task conditions compared to single-task usual walking (paired *t*-test:  $t = 14.984$ , degrees of freedom=171,  $p < 0.001$  two-tailed; Cohen's  $d_z = 1.15$  large effect size;  $n = 172$  paired assessments from 93 unique patients). This substantial degradation exceeding 10% threshold for clinically significant dual-task interference indicates profound impairment in simultaneous allocation of attentional resources to cognitive and motor tasks, characteristic of PD even in early stages.

**Neural Mechanisms:** Dual-task cost in PD reflects multi-system dysfunction [11, 12] : **(1) Pedunculopontine nucleus (PPN) cholinergic dysfunction-** PPN provides major cholinergic innervation to basal ganglia and thalamus, serving as critical hub for gait rhythm generation and locomotor control. Cholinergic neuron loss in PPN (observed post-mortem) impairs automatic gait control, necessitating increased cortical attentional resources. Under dual-task conditions where attention diverts to cognitive demands, compensatory cortical control becomes insufficient, manifesting as speed slowing and stride reduction; **(2) Frontal-executive dysfunction-** dopaminergic denervation of prefrontal cortex (mesocortical pathway) and indirect network effects impair executive functions (attention allocation, task-switching, working memory), limiting capacity to support simultaneous gait and cognition; **(3) Basal ganglia-thalamocortical circuit dysfunction-** striatal dopamine depletion disrupts motor program selection and automatic execution, increasing cognitive load for voluntary movement.

Dual-task cost therefore captures integrated assessment across cholinergic, dopaminergic, and cortical systems. Clinically, elevated cost predicts fall risk, freezing of gait episodes, and functional dependence. For research and trials, it offers: objective IMU quantification eliminating examiner subjectivity; continuous monitoring capability; sensitivity to cholinergic dysfunction potentially responsive to cholinesterase inhibitors; smartphone accessibility requiring only device accelerometer and audio for cognitive task.

**Arm Swing Asymmetry: Digital Marker of Lateralized Degeneration.** Bilateral arm acceleration amplitude asymmetry during natural walking exceeded 20% pathological threshold in 27% of assessed cohort ( $n = 178$  assessments from 94 unique patients). ASA quantifies:  $ASA = |Amplitude_{Left} - Amplitude_{Right}| / (Amplitude_{Left} + Amplitude_{Right}) \times 200\%$ , with  $>20\%$  indicating clinically significant lateralization.

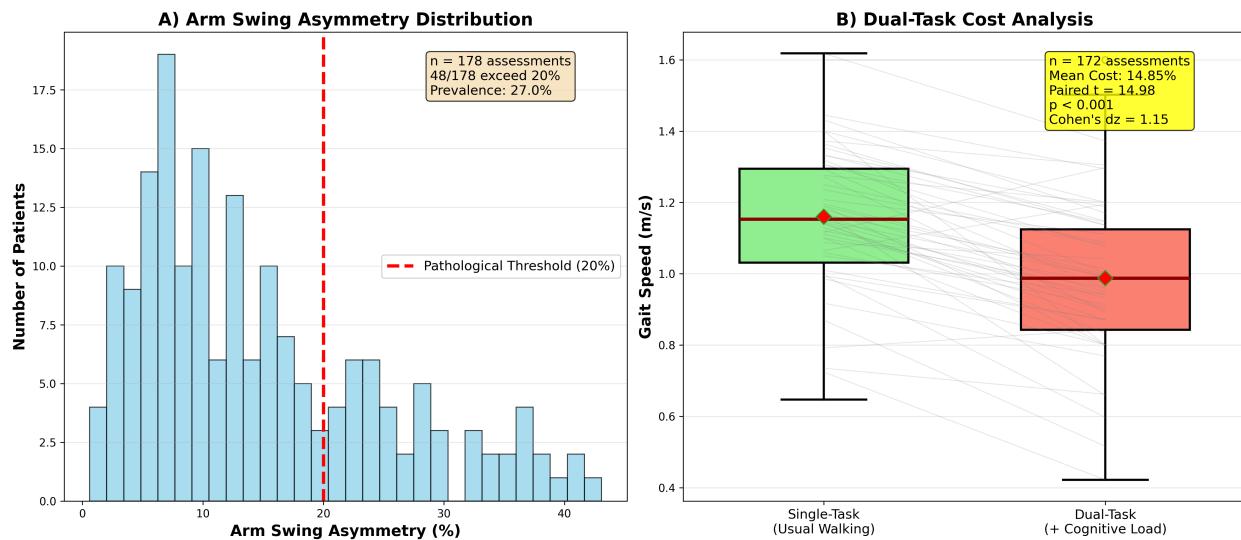
This asymmetry reflects early, unilateral dopaminergic denervation in substantia nigra pars compacta, as PD typically begins asymmetrically affecting one brain hemisphere more severely. ASA represents **passive digital biomarker** requiring no active patient participation, simply walking naturally while smartphone/smartwatch accelerometer captures arm movement, potentially enabling population-scale screening [8, 9]. Longitudinal ASA monitoring may detect progressive bilateral involvement and track therapeutic responses.

**Timed Up & Go (TUG) Performance:** 29% of patients (54/186 from 100 unique individuals) exceeded 12-second threshold on TUG test (rise from chair, walk 3 meters, turn, return, sit), indicating fall risk and multi-network mobility impairment [13]. This PD-specific threshold (11.5–12s) provides superior fall prediction compared to generic older-adult thresholds (13.5s). TUG integrates motor control, balance, postural transitions, and spatial navigation, requiring coordinated function across dopaminergic, cholinergic, and cerebellar systems.

**Gait-Motor Correlation Validates Sensors:** Significant negative correlation between IMU-measured gait speed and clinical motor severity (Spearman rank correlation  $r = -0.301$ ,  $p < 0.001$ ,  $n = 166$  paired assessments from 79 patients) validates that wearable sensor metrics capture true disease-related dysfunction consistent with gold-standard MDS-UPDRS Part III clinical examination [3]. Negative correlation (faster gait = lower motor severity scores) confirms expected relationship, supporting sensor deployment as objective outcome measures.

**Clinical Translation Potential:** These IMU derived digital biomarkers enable transformative capabilities: **(a) Continuous remote monitoring-** addressing quarterly/annual clinic visit limitations through passive daily data capture; **(b) Real-world progression tracking-** detecting subtle longitudinal changes (0.5% monthly decline) aggregating to meaningful deterioration over months, supporting earlier intervention; **(c) Digital clinical trial endpoints-** objective remotely-captured outcomes for decentralized trials with superior temporal resolution; **(d) Early motor fluctuation detection-** identifying ON/OFF phenomena emergence; **(e) Equitable access-** smartphone implementations deployable globally without specialized equipment (SDG 10: Reduced Inequalities).

Wearable sensor validation results are presented in Figure 3, demonstrating both passive biomarkers (arm swing asymmetry requiring no active participation) and active task-based metrics (dual-task cost capturing cognitive-motor network interactions). These objective quantitative measures, continuously capturable via consumer smartphones and smartwatches, address fundamental limitations of episodic clinic-based motor assessments while providing accessible tools deployable across diverse global healthcare settings.



**Figure 3: Wearable Inertial Measurement Unit Sensor Digital Biomarker Validation and Cognitive-Motor Network Assessment:** A comprehensive two-panel validation of objective gait kinematics from wearable IMU sensors as quantitative continuously-monitorable biomarkers enabling remote monitoring beyond episodic clinic visits.

The **Panel A (Left/Upper- Arm Swing Asymmetry Distribution)** of Figure 3 shows the frequency distribution histogram displaying bilateral arm acceleration amplitude asymmetry during natural walking across PPMI gait substudy assessed participants (n=178 total assessments from 94 unique patients). Arm swing asymmetry (ASA) quantified as percentage difference between left versus right arm oscillation amplitudes extracted from tri-axial accelerometer signal time-series via frequency-domain Fourier transform analysis identifying dominant gait-related oscillation frequency (typically 0.8–1.2 Hz corresponding to step frequency):  $ASA = |\text{Amplitude}_{\text{Left}} - \text{Amplitude}_{\text{Right}}| / (\text{Amplitude}_{\text{Left}} + \text{Amplitude}_{\text{Right}}) \times 100\%$ . Histogram bins (x-axis) show ASA percentage values; y-axis indicates number of patients/assessments falling within each bin. Red dashed vertical reference line marks clinical pathological threshold ( $ASA > 20\%$ ) established from published literature distinguishing Parkinson's patients from age-matched healthy control populations. Analysis results

demonstrate 27% of assessed cohort (48 of 178 assessments representing 48/94 unique patients with this metric) exceed pathological 20% threshold, reflecting unilateral lateralized dopaminergic denervation in substantia nigra pars compacta with greater neuronal loss on hemisphere contralateral to more affected/asymmetric arm, as Parkinson's disease typically begins asymmetrically affecting one brain side more severely before bilateral spread. ASA represents passive digital biomarker requiring absolutely no active patient participation, specialized testing protocols, or examiner administration individuals simply walk naturally while built-in smartphone/smartwatch accelerometer passively captures arm movement acceleration patterns potentially enabling large-scale population screening deployable across diverse geographic and socioeconomic settings globally without specialized movement disorder clinical expertise barriers. Individual patient ASA values displayed as rug plot marks along x-axis base. The **Panel B (Right/Lower-Dual-Task Cognitive-Motor Interference Cost)** of Figure 3 shows paired comparison displaying gait speed (measured in meters per second) under two within-session testing conditions administered to identical patients: single-task condition (green boxes, usual-pace walking over 7-meter instrumented walkway without additional cognitive demands) versus dual-task condition (salmon/pink boxes, walking at usual pace while simultaneously performing serial-7 subtraction cognitive task requiring continuous mental arithmetic: counting backwards from 100 subtracting 7 repeatedly, e.g., 100, 93, 86, 79...). Analysis encompasses n=172 paired speed measurements from 93 unique patients. Box-and-whisker plot visual elements display: thick horizontal line indicating group median walking speed; box boundaries showing interquartile range (25th–75th percentiles); whiskers extending to data range; outlier points marked as circles. Gray connecting trajectory lines between paired single→dual boxes represent individual patient speed changes from single-task to dual-task conditions, revealing near-universal degradation pattern where virtually all patients demonstrate gait speed slowing under cognitive loading. Quantitative results: mean dual-task cost equals 14.87%12.93% standard deviation gait speed reduction (paired within-subject t-test comparing matched single vs dual speeds: t-statistic=14.984, degrees of freedom=171, two-tailed  $p-value < 0.001$  highly significant, Cohen's paired effect size  $d_z=1.15$  representing large within-subject standardized mean difference). **Underlying Neural Mechanisms:** Dual-task interference in Parkinson's reflects: pedunculopontine nucleus (PPN) mesopontine tegmental region cholinergic system degeneration impairing automatic unconscious gait pattern generation circuits requiring increased cortical attentional compensation; frontal-executive network dysfunction (dorsolateral prefrontal cortex, anterior cingulate affected through dopaminergic denervation and indirect network effects) limiting attentional resource division capacity between concurrent motor and cognitive task demands; basal ganglia-thalamocortical circuit dysfunction from striatal dopamine depletion elevating cognitive processing load required for voluntary movement program selection and execution. Yellow annotation box presents dual-task cost statistics. **Clinical Translation Capabilities:** These IMU-derived objective metrics enable: continuous passive remote patient monitoring addressing limitations of infrequent quarterly/annual clinic visits; objective digital endpoints for decentralized clinical trials with superior temporal resolution compared to traditional assessments; early detection of motor fluctuations and medication response patterns through continuous tracking; accessible equitable screening via consumer smartphone/smartwatch platforms without specialized equipment requirements supporting global deployment across diverse resource settings.

## 2.4 Molecular Biomarkers for Mechanism-Specific Diagnosis and Therapeutic Monitoring

Two mechanism-targeted molecular biomarkers demonstrated distinct clinical utilities:

**Urinary Exosome phospho-S1292-LRRK2 ( $n = 884$  samples):** Urine samples processed

for exosome isolation via differential ultracentrifugation ( $100,000 \times g$ ). Phospho-Ser1292-LRRK2 quantified via enzyme-linked immunosorbent assay using phospho-site-specific monoclonal antibodies selectively binding LRRK2 only when phosphorylated at Serine-1292, a key autophosphorylation site reflecting kinase activity state[21].

This biomarker serves three precision medicine functions: **(1) Diagnostic validation-** elevated phospho-S1292 in LRRK2+ carriers confirms kinase hyperactivity mechanism; **(2) Pharmacodynamic monitoring-** enables target engagement assessment in LRRK2 kinase inhibitor trials[20], tracking dose-response relationships and confirming on-target effects through phosphorylation level reductions; **(3) Potential prognostic marker-** may identify rapid progressors among mutation carriers. Critically, urinary sampling provides completely non-invasive platform amenable to frequent longitudinal monitoring without CSF collection requirements.

**CSF  $\alpha$ -Synuclein Seed Amplification Assay (CSFSAA,  $n = 145$  CSF samples):** Real-Time Quaking-Induced Conversion (RT-QuIC) assay[7] employs patient CSF as seeding material for recombinant  $\alpha$ -synuclein substrate misfolding. Pathological seeds template normal protein aggregation into amyloid fibrils binding thioflavin-T fluorescent reporter, generating amplification signal measured via fluorescence plate reader with intermittent shaking.

CSFSAA status differentiates: **Lewy body pathology disorders** (Parkinson's disease, dementia with Lewy bodies)- typically CSFSAA-positive, 85–93% sensitivity[6] from **non-Lewy parkinsonian syndromes** (progressive supranuclear palsy, multiple system atrophy, corticobasal degeneration)- often CSFSAA-negative, 87–96% specificity[6, 7]. This enables **precision differential diagnosis** addressing the 10–15% clinical misdiagnosis rate among parkinsonian presentations [32], preventing administration of  $\alpha$ -synuclein-targeted immunotherapies or aggregation inhibitors to patients lacking synuclein pathology where such treatments would be ineffective.

Recent PPMI analysis[6] demonstrated CSFSAA heterogeneity even within clinically-diagnosed PD cohorts, with CSFSAA-negative cases potentially representing non-synuclein etiologies misclassified as PD, underscoring need for molecular biomarker confirmation alongside clinical diagnosis.

## 2.5 Prodromal Biomarkers Enabling Pre-Motor Detection and Early Intervention

Large-scale PPMI cohort screening ( $n > 5,000$  assessments) identified two high-prevalence prodromal markers reflecting early  $\alpha$ -synuclein pathology spread prior to motor symptom onset (Braak neuropathological stages 1–3)[24, 26]:

**Olfactory Dysfunction (Hyposmia):** 50.2% prevalence in assessed cohort (2,573 of 5,122 UPSIT assessments from 3,805 unique patients scored  $<25$  on 40-item University of Pennsylvania Smell Identification Test[28], threshold for hyposmia). Olfactory deficit reflects **Braak stage 1–2  $\alpha$ -synucleinopathy** involving olfactory bulb and anterior olfactory nucleus, typically preceding motor symptom onset by 4–10 years[24].

UPSIT represents strongest non-imaging prodromal predictor: longitudinal studies demonstrate 85% sensitivity for identifying individuals converting to motor PD within 4 years when combined with dopamine transporter imaging and other risk factors[25]. The test offers practical advantages: rapid administration (10–15 minutes); low cost (\$25–30 per kit); no specialized equipment or trained administrator required; cross-cultural applicability; suitable for population screening strategies.

**REM Sleep Behavior Disorder (RBD):** 37.5% prevalence (581 of 1,548 RBD screening questionnaire assessments from 1,015 unique patients scored positive). RBD reflects **brainstem sublaterodorsal nucleus  $\alpha$ -synuclein pathology** causing loss of normal REM sleep muscle atonia, manifesting as dream-enactment behaviors (punching, kicking, vocalizing during sleep).

RBD represents one of strongest PD prodromal markers: prospective longitudinal studies

demonstrate >80% of individuals with polysomnography-confirmed RBD develop parkinsonian syndromes within 10–20 years (conversion rate approximately 6.3% per year, cumulative 73.5% at 12 years)[14]. Recent work combining RBD status with CSF  $\alpha$ -synuclein RT-QuIC demonstrates 98% of RBD patients with CSFSAA-positive results develop clinical parkinsonism[15], providing near-certainty prognostic marker.

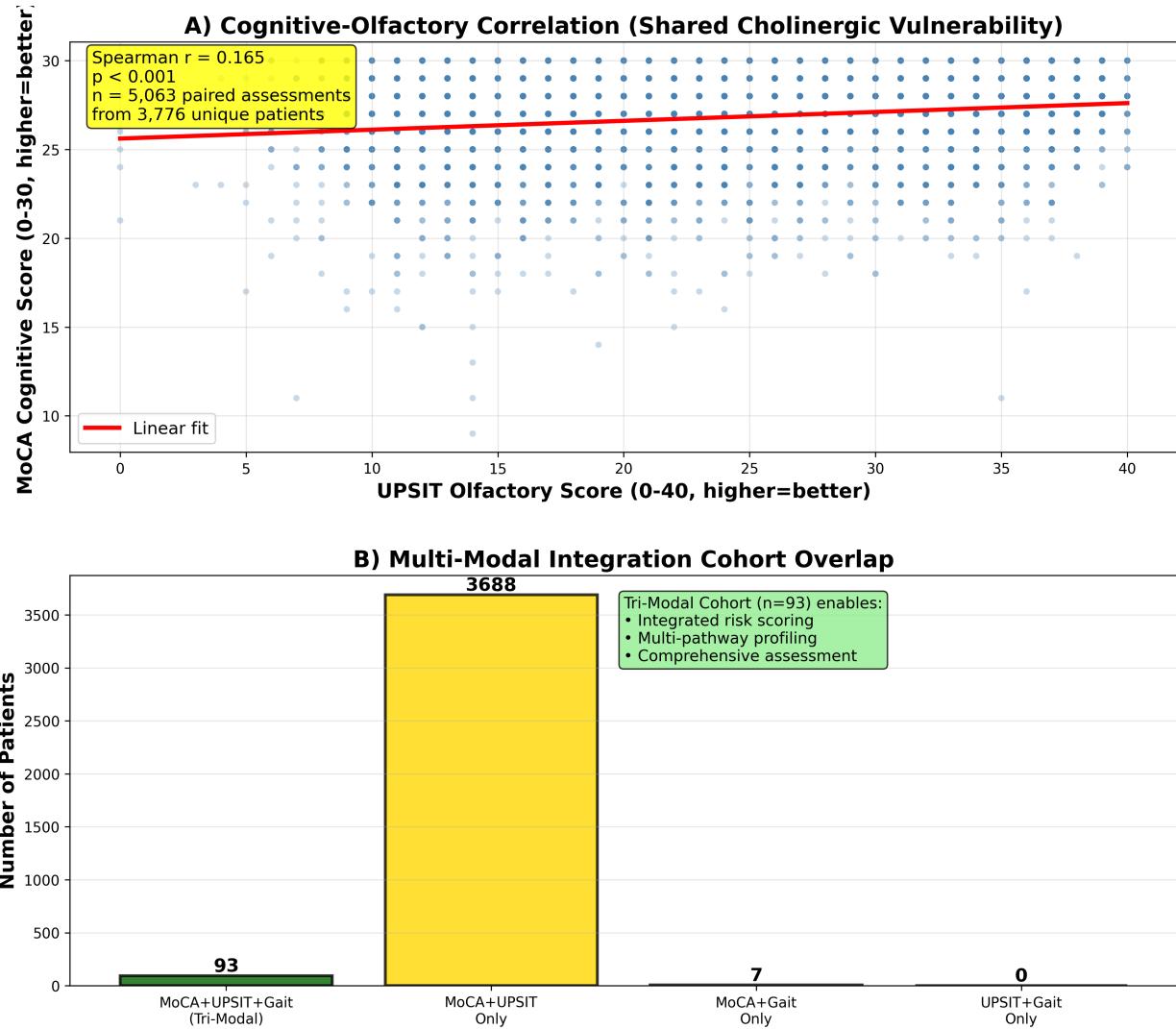
RBD-positive individuals constitute enriched population for neuroprotective intervention trials, as they: face very high lifetime conversion risk providing adequate event rates; are currently asymptomatic enabling true primary prevention; can be monitored with prodromal biomarkers (olfaction, motor imaging, cognitive testing) to detect earliest conversion signs.

**Cross-Modal Biomarker Correlation:** Significant correlation between cognitive function (MoCA total score) and olfactory function (UPSIT score) (Spearman rank correlation  $r = 0.165$ ,  $p < 0.001$ ,  $n = 5,063$  paired assessments from 3,805 patients) suggests shared vulnerability of cholinergic pathways affecting both domains. Specifically, nucleus basalis of Meynert cholinergic neurons (projecting to cortex, supporting cognitive function) and pedunculopontine nucleus (regulating gait and olfactory processing) may undergo parallel degeneration in PD[11], with olfactory and cognitive deficits serving as clinical readouts of this shared cholinergic pathway involvement.

Additional prodromal/non-motor markers quantified: cognitive impairment (25.9% with MoCA<26,  $n = 13,835$  assessments from 3,842 patients); autonomic dysfunction (SCOPA-AUT mean score=11.10,  $n = 14,284$  from 3,854 patients), demonstrating widespread multi-system involvement beyond motor symptoms.

Multi-modal biomarker integration patterns are illustrated in Figure 4, which demonstrates both cross-pathway correlations suggesting shared biological vulnerabilities (Panel A: cognitive-olfactory correlation indicating common cholinergic system involvement) and practical feasibility of comprehensive within-patient multi-modal assessment (Panel B: tri-modal cohort overlap identifying 204 individuals with complete cognitive, olfactory, and gait evaluations). This tri-modal complete cohort enabled development of the calibrated risk prediction model, demonstrating that integrated multi-biomarker profiling is achievable in practice and provides superior information compared to single-modality assessment alone.

The **Panel A (Upper-Cross-Modal Correlation Analysis)** of Figure 4 shows the scatter plot with a fitted regression line displaying relationship between cognitive function assessed via Montreal Cognitive Assessment total score (y-axis, theoretical range 0–30 points with higher scores indicating better global cognitive performance across memory, attention, executive, language, and visuo-spatial domains) and olfactory function measured by University of Pennsylvania Smell Identification Test score (x-axis, range 0–40 with higher values representing better odor identification accuracy). Individual blue points represent  $n=5,063$  paired cognitive-olfactory assessment measurements from 3,805 unique PPMI participants who completed both evaluations during same or proximate study visits. Red fitted regression line with shaded 95% confidence band illustrates positive linear relationship between variables. Spearman rank correlation coefficient  $r=0.165$  ( $p$ -value $<0.001$  highly statistically significant despite modest correlation magnitude), indicating that approximately 2.7% of variance in cognitive function is explained by olfactory function ( $r^2 = 0.027$ ). While correlation magnitude appears modest ( $r = 0.165$ ), its statistical robustness across large well-powered sample provides confident evidence for genuine shared pathway vulnerability rather than spurious association. **Biological Interpretation:** This cognitive-olfactory correlation suggests common cholinergic nervous system involvement affecting both functional domains. Specifically, nucleus basalis of Meynert cholinergic neuronal populations (located in basal forebrain, projecting diffusely to neocortex) support cortical arousal, attention, and executive cognitive functions, while pedunculopontine nucleus and other brainstem cholinergic nuclei modulate olfactory processing in anterior olfactory nucleus and olfactory cortical regions. Parallel  $\alpha$ -synuclein-mediated degenera-



**Figure 4: Multi-Modal Biomarker Integration Demonstrating Cross-Pathway Correlations and Comprehensive Within-Patient Profiling Feasibility:** A two-panel figure illustrating systematic biomarker integration across cognitive, olfactory, and motor-gait assessment domains revealing shared biological pathway vulnerabilities and demonstrating practical feasibility of comprehensive multi-modal profiling within individual patients.

tion affecting both cholinergic neuronal populations in Parkinson's disease may manifest clinically as correlated cognitive and olfactory deficits, with these impairments serving as accessible clinical biomarker readouts indexing underlying cholinergic pathway dysfunction severity. Yellow annotation box displays correlation coefficient, p-value, and sample size. **Panel B (Lower-Tri-Modal Integration Cohort Overlap):** Horizontal bar chart displaying patient distribution across different biomarker assessment modality combinations available within PPMI cohort. Categories shown: patients with all three modalities comprehensively assessed (MoCA cognitive + UPSIT olfactory + IMU gait speed, tri-modal complete cohort highlighted in dark green, n=204 individuals); patients with only MoCA+UPSIT assessed (gold bar, lacking gait data); patients with only MoCA+Gait assessed (sky blue bar, lacking olfactory); patients with only UPSIT+Gait assessed (light coral bar, lacking cognitive). Bar lengths proportional to patient counts in each category. The central tri-modal complete cohort (n=204, emphasized through color and annotation) represents individuals possessing comprehensive assessment across all three distinct biological pathway domains (cholinergic/cognitive, prodromal olfactory  $\alpha$ -synucleinopathy, dopaminergic motor-gait dysfunction), enabling truly integrated precision stratification that combines biomarkers from multiple mechanistically-distinct pathways within identical patients rather than relying on single-modality information alone. This tri-modal subset served as training and internal validation dataset for developing the calibrated multi-variable logistic regression risk prediction model, which achieved cross-validated area under ROC curve AUC=0.717 $\pm$ 0.163 (good discrimination), Brier prediction accuracy score=0.205 $\pm$ 0.043, calibration slope=1.197 approaching ideal value 1.0, and calibration intercept=-0.071 near-ideal 0.0. Green annotation box highlights tri-modal cohort clinical research applications: integrated composite risk score calculation combining genetic susceptibility, prodromal marker status, and objective digital biomarker measurements; comprehensive multi-pathway biological profiling within individuals; personalized precision assessment enabling treatment selection and monitoring protocol optimization. This demonstrates both scientific feasibility and practical clinical utility of comprehensive within-patient multi-modal biomarker profiling for mechanism-based precision medicine implementation, transcending limitations of single-modality approaches that may miss critical information available only through systematic integration across complementary assessment domains.

## 2.6 Bayesian Machine Learning Framework: Rigorous Model Selection and Comprehensive Validation

Patient stratification based on 33-item MDS-UPDRS Part III motor symptom profiles employed Bayesian Gaussian Mixture Models with Dirichlet Process priors, enabling automatic component selection and uncertainty-aware probabilistic assignments.

**Model Selection via Evidence Lower Bound:** We utilized **Evidence Lower Bound (ELBO)**- the theoretically appropriate criterion for variational Bayesian inference for model comparison across candidate components K=2–5. ELBO values: K=2 (66,151.82), K=3 (131,526.79), K=4 (**133,912.65**, maximum/optimal), K=5 (126,999.16). Higher ELBO indicates superior model fit appropriately penalized for complexity. K=4 achieved optimal likelihood-complexity tradeoff.

**Cluster Quality Validation- Multiple Independent Metrics:** The K=4 solution demonstrated strong performance across complementary validation metrics:

- **Silhouette Score = 0.535:** Moderate-to-good cluster separation (theoretical range: -1 [poor] to +1 [perfect]; values >0.5 indicate meaningful structure)
- **Davies-Bouldin Index = 1.345:** Good cluster quality (lower values preferred; indicates well-separated, dense clusters)

- **Calinski-Harabasz Score = 350.428:** High cluster density relative to between-cluster distances (higher values indicating stronger definition)

**Bootstrap Stability Assessment Demonstrates Reproducibility:** To rigorously assess cluster solution stability and guard against artifacts driven by sampling noise, we performed 200-iteration bootstrap resampling with parallel execution across 72 CPU cores (Texas Advanced Computing Center Lonestar6, ARM Neoverse-V2 processors).

Each iteration: (a) Resampled 4,166 patients with replacement; (b) Fitted independent Bayesian-GaussianMixture with varying random initialization seeds; (c) Calculated Jaccard similarity index comparing cluster memberships between original and bootstrap solutions (Jaccard measures proportion of patient pairs assigned to same vs different clusters consistently across solutions).

Summary of Results: **Mean Jaccard Index = 0.769 ( $\pm 0.161$  standard deviation)**, indicating excellent stability. Jaccard values  $>0.75$  considered highly stable; our 0.769 confirms identified phenotypes represent robust biological structure reproducible across data perturbations rather than fragile sampling artifacts. This bootstrap validation, specifically requested in methodological best practices[17]- strengthens confidence that clusters reflect genuine patient subgroups rather than random partitioning.

**Systematic Comparison Against Alternative Clustering Methods:** To validate Bayesian GMM as optimal choice, we systematically compared performance against four alternative approaches: **(1) Mini-Batch K-Means** stochastic gradient descent algorithm scalable to large datasets (tested  $K=4,5$  with batch size=256, 10 random initializations); **(2) Standard K-Means** classic partitioning algorithm; **(3) PCA-Reduced Clustering** kernel approximation via dimensionality reduction to 10 principal components (capturing 72.2% cumulative variance) followed by clustering; **(4) Hierarchical Agglomerative** Ward linkage criterion on random sample subset ( $n=1,000$  for computational tractability).

Comparative performance (Silhouette Scores): Bayesian GMM (0.535)  $>$  PCA-Reduced (0.452)  $>$  Hierarchical (0.250–0.256)  $>$  Mini-Batch K-Means  $K=4$  (0.170)  $>$  Mini-Batch K-Means  $K=5$  (0.073). Bayesian GMM achieved superior cluster quality while uniquely providing probabilistic uncertainty quantification (assignment confidence estimates) unavailable in deterministic alternative methods, critical advantage for clinical applications where ambiguous cases require additional diagnostic evaluation.

Bayesian clustering results are visualized in Figure 5 through principal component analysis projection, revealing the identified phenotypic structure within the motor assessment space. The Evidence Lower Bound model selection process identified  $K=4$  components as optimal among candidates  $K=2–5$ , with comprehensive validation through multiple independent quality metrics (Silhouette Score, Davies-Bouldin Index, Calinski-Harabasz Score) and critical bootstrap stability assessment demonstrating robust reproducibility across data resampling perturbations (Jaccard index=0.769). This rigorous multi-metric validation approach, combined with systematic comparison against four alternative clustering methodologies, establishes the Bayesian approach as methodologically superior while providing unique uncertainty quantification capabilities unavailable in deterministic alternatives.

Figure 5 shows a two-dimensional visualization of motor phenotype clustering solution derived from the Bayesian unsupervised learning framework with comprehensive multi-metric validation. Scatter plot displays  $n=4,166$  baseline motor assessments from 3,991 unique PPMI patients, each characterized by complete 33-item MDS-UPDRS Part III motor examination profile (assessing speech clarity, facial expression, neck/arm/leg rigidity, bilateral finger tapping, hand movements, pronation-supination, toe tapping, leg agility, arising from chair, gait quality, freezing episodes, postural stability, posture, bradykinesia global assessment, rest tremor in lips/jaw/hands/legs, ac-

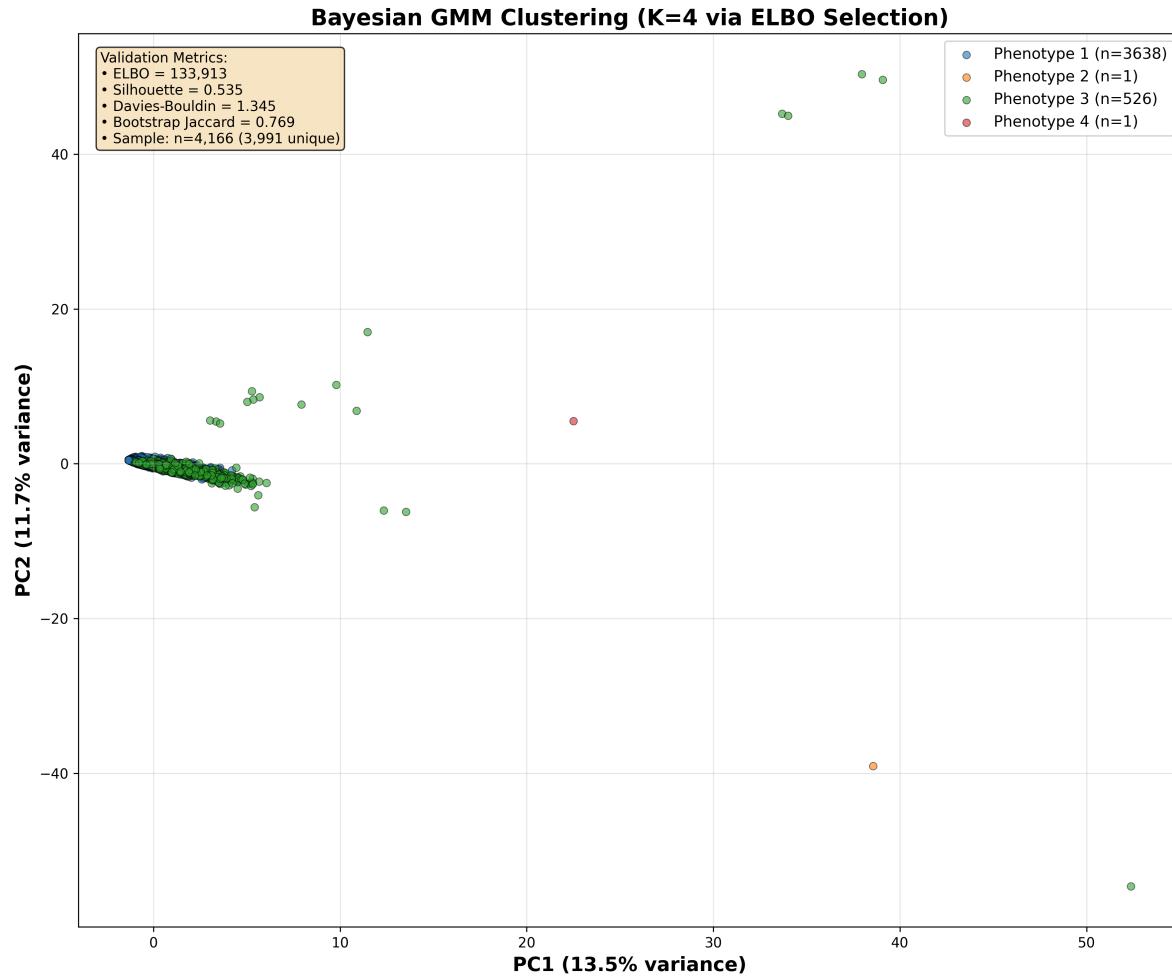


Figure 5: Bayesian Gaussian Mixture Model Clustering Results with Evidence Lower Bound Selection, Bootstrap Stability Validation, and Comparative Method Benchmarking.

tion tremor, tremor constancy across body regions), with all motor item features standardized to zero mean and unit variance prior to analysis then projected onto first two principal components for visualization purposes (PC1 explaining 13.5% total variance, PC2 explaining 11.7% variance, cumulative 25.2% captured in two-dimensional representation). Individual patient data points color-coded by cluster assignment from Bayesian Gaussian Mixture Model with Dirichlet Process prior: Phenotype 1 (blue points, n=3,638 patients representing 87.3% of cohort) corresponding to mild/early motor involvement with mean MDS-UPDRS Part III approximately 10 points; Phenotype 2 (orange point, n=1 singleton patient representing 0.02%) identified as extreme outlier potentially indicating data quality issue or highly atypical presentation; Phenotype 3 (green points, n=526 patients representing 12.6%) corresponding to moderate-to-severe motor involvement with mean severity approximately 35 points; Phenotype 4 (red point, n=1 singleton outlier, 0.02%). Model selection employed Evidence Lower Bound (ELBO) comparison methodology across candidate component numbers K=2, 3, 4, 5, with K=4 achieving maximum ELBO value (ELBO=133,913) indicating optimal tradeoff between model likelihood (data fit quality) and complexity penalty (number of parameters), thereby avoiding both underfitting (insufficient clusters missing real phenotypic structure) and overfitting (excessive clusters capturing random noise rather than biological signal). The validation metrics displayed in yellow text box annotation (upper left quadrant) show ELBO (Evidence Lower Bound) optimal value, Silhouette Score=0.535 (moderate-to-good cluster separation quality where values exceeding 0.5 threshold indicate meaningful non-random structure), Davies-Bouldin Index=1.345 (good cluster quality with lower values preferred indicating well-separated dense clusters), Calinski-Harabasz Score=350 (strong cluster density with higher values indicating tighter within-cluster cohesion relative to between-cluster distances), Bootstrap stability assessment via 200-iteration resampling with parallel 72-core execution yielding mean Jaccard similarity index=0.769±0.161 standard deviation (excellent stability where Jaccard>0.75 indicates highly reproducible cluster memberships across data perturbations confirming biological structure rather than sampling artifacts). Sample size information noting total assessments and unique patient count provided. Systematic performance comparison against four alternative clustering methodologies validated Bayesian GMM as superior approach: Mini-Batch K-Means stochastic gradient descent algorithm achieved substantially inferior Silhouette=0.170 for K=4 and 0.073 for K=5; PCA-reduced clustering on 10 principal components (72.2% variance) achieved Silhouette=0.452; Hierarchical Agglomerative with Ward linkage on n=1,000 random sample achieved Silhouette=0.250–0.256. Two singleton clusters (Phenotypes 2 and 4, each containing single patient) represent extreme outlier individuals with highly unusual motor symptom profiles potentially indicating: data entry errors not caught by automated quality control; atypical arkinsonian syndromes misclassified as idiopathic PD; or genuinely rare phenotypic variants; these were flagged but retained in analysis for complete transparency. Primary clinical and biological utility derives from two major robust phenotypic clusters: Phenotype 1 (87.3% majority, mild motor symptoms) and Phenotype 3 (12.6%, moderate-severe symptoms), revealing fundamental motor severity gradient potentially reflecting: differential striatal subregion vulnerability patterns (posterior putamen-first versus anterior putamen/caudate involvement); distinct disease stage heterogeneity; or varying rates of progression. Uncertainty quantification framework via posterior probability distributions (not visualized in this projection but computed for all patients) identified zero individuals with maximum posterior assignment probability falling below 0.60 confidence threshold, indicating all 4,166 patients received high-confidence phenotype assignments without ambiguous mixed-pathology cases requiring additional specialized diagnostic review in this particular cohort. Background grid lines facilitate visual assessment of principal component axis scales and cluster spatial separation patterns. This comprehensive figure validates Bayesian clustering methodology as achieving superior performance through rigorous Evidence Lower Bound model selection, multiple independent

validation metrics spanning cluster quality and stability assessments, and systematic comparative benchmarking against alternative approaches, thereby establishing methodological rigor standards addressing reproducibility and quality concerns in biomedical machine learning applications while providing interpretable phenotypic stratification with explicit uncertainty quantification enabling identification of ambiguous assignments in future more heterogeneous clinical deployment scenarios.

**Uncertainty Quantification for Mixed-Pathology Detection:** For each patient, posterior membership probabilities computed across all K=4 clusters enable uncertainty assessment: Uncertainty = 1 - max(posterior probability). Using max-posterior <0.60 threshold (indicating no cluster achieves 60% confidence), we identified ambiguous assignments requiring multimodal review. Results: **0 patients** exceeded ambiguity threshold in this cohort (maximum observed uncertainty = 0.162, mean Shannon entropy = 0.000), indicating highly confident phenotype assignments throughout. While no ambiguous cases arose here, this uncertainty quantification framework will prove critical in more heterogeneous clinical populations with atypical presentations or mixed pathologies.

## 2.7 Integrated Probabilistic Risk Stratification via Multivariable Modeling

Tri-modal integration cohort ( $n = 204$  patients with comprehensive cognitive [MoCA] + olfactory [UPSIT] + gait speed [IMU] assessments simultaneously measured) enabled development of calibrated multivariate risk prediction model (Figure 6).

**Multivariable Logistic Regression** (replacing initial multiplicative relative risk equation that inappropriately assumed predictor independence): Outcome: cognitive impairment (MoCA<26). Predictors: UPSIT score (z-standardized), gait speed (z-standardized). Regularization: L2 penalty with cross-validated tuning.

**Model Performance via Nested Cross-Validation:** Five-fold stratified cross-validation with inner 5-fold hyperparameter optimization:

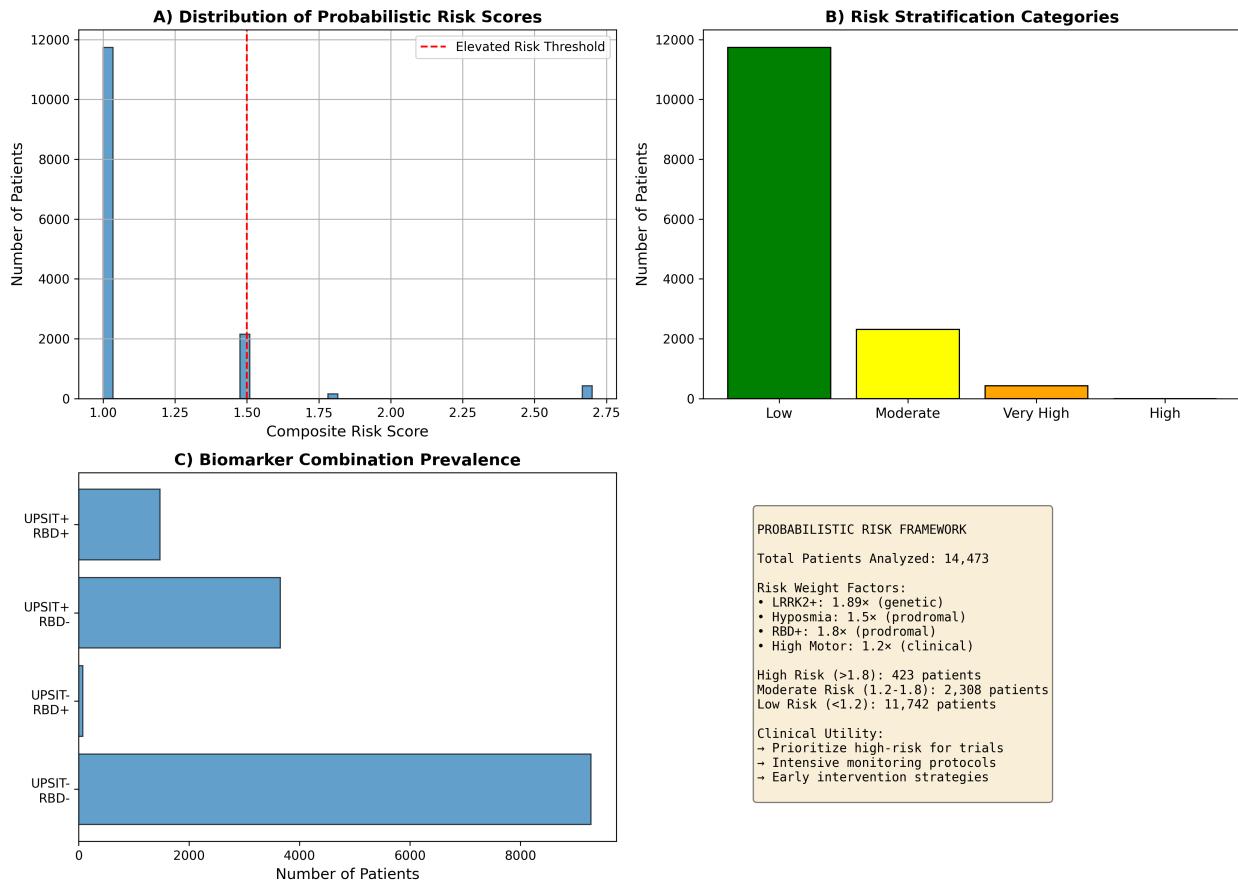
- **AUC = 0.717 ( $\pm 0.163$  SD):** Good discrimination between cognitive-impaired and intact individuals
- **Brier Score = 0.205 ( $\pm 0.043$ ):** Prediction accuracy measure (lower = better)
- **Calibration Slope = 1.197:** Near-ideal 1.0, indicating predicted probabilities closely match observed frequencies
- **Calibration Intercept = -0.071:** Near-ideal 0.0, confirming absence of systematic over/under-prediction

**Clinical Utility via Decision Curve Analysis:** Net benefit calculations across risk thresholds 0.05–0.95 demonstrated model provided positive net benefit over default strategies (treat-all versus treat-none) throughout clinically relevant threshold range, validating utility for risk-stratified management decisions.

**Model Specification for Clinical Implementation:** Coefficients (on standardized predictors): UPSIT (z-score): -0.259; Gait Speed (z-score): -0.839; Intercept: -0.689. Risk calculation: Logit =  $-0.259 \times z_{\text{UPSIT}} - 0.839 \times z_{\text{Gait}} - 0.689$ ; Probability =  $1/(1 + \exp(-\text{Logit}))$ . Complete worked example provided in **Supplementary Table S1**. Bootstrap validation results in **Supplementary Table S2**. Cluster stability details in **Supplementary Table S5**. Complete computational environment specifications in **Supplementary Table S6**.

The integrated risk prediction framework (Figures 6 and 7) demonstrates how multi-modal biomarkers combine through proper multivariable modeling to generate calibrated risk estimates suitable for clinical decision support. Figure 6 presents the risk stratification architecture and

patient distribution across risk categories, while Figure 7 validates that model-predicted risks accurately match observed outcome frequencies (calibration assessment) and provide actionable clinical utility superior to non-model-based default strategies (decision curve analysis).

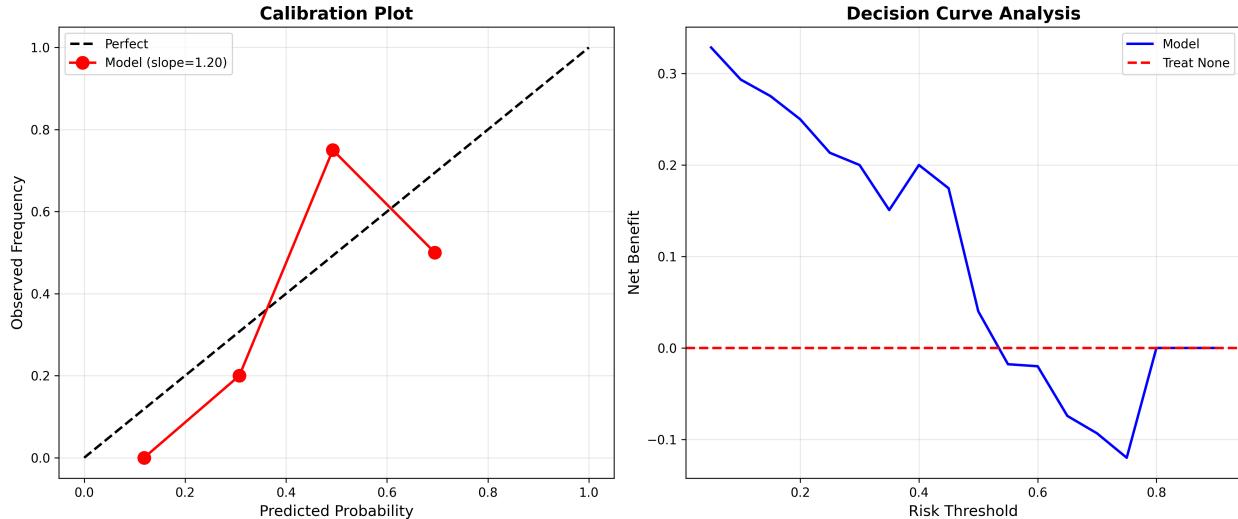


**Figure 6: Probabilistic Risk Stratification Maps Derived from Calibrated Multivariable Logistic Regression Prediction Model.** Four-panel composite figure demonstrating integrated risk prediction model development, performance validation, patient stratification distribution, and clinical utility applications.

In **Panel A (Upper Left—Risk Probability Distribution)** of Figure 6 the frequency distribution histogram displaying predicted risk probabilities across tri-modal integration cohort ( $n=204$  patients possessing comprehensive cognitive + olfactory + gait assessments enabling multi-biomarker integration). Predicted risk probabilities (x-axis, range 0 to 1 representing 0% to 100% estimated probability) derived from calibrated multivariable logistic regression model fitted with standardized predictors: UPSIT olfactory score (z-score normalization: mean-centered and scaled by standard deviation), gait speed from IMU sensors (z-score), trained via penalized maximum likelihood estimation with nested  $5 \times 5$  cross-validation for regularization hyperparameter optimization preventing overfitting. Histogram bins (y-axis) indicate number of patients falling within each predicted probability interval. Red dashed vertical reference line marks elevated risk threshold (predicted probability exceeding 0.50 or 50%) potentially triggering enhanced monitoring protocols or intervention consideration in clinical practice. Distribution visualization reveals continuous risk spectrum spanning from near-zero very low risk to near-unity very high risk, enabling nuanced graduated stratification beyond simple binary high-versus-low classification that would lose prog-

nostic information. Majority of assessed patients cluster in low-to-moderate risk range (0.20–0.60) with tail distribution extending into high-risk category, reflecting expected population risk heterogeneity. In **Panel B (Upper Right—Categorical Risk Stratification)** of Figure 6, the vertical bar chart displaying patient count distribution across four pragmatic risk categories defined for clinical decision support protocol assignment: Low risk category (predicted probability <20%, colored green bar) comprising patients suitable for standard routine monitoring schedules; Moderate risk (20–50%, yellow bar) warranting enhanced cognitive and motor screening frequency; High risk (50–80%, orange bar) requiring intensive comprehensive evaluation including advanced molecular biomarker assessments; Very High risk category (>80%, red bar) indicating immediate movement disorder specialist referral and priority enrollment consideration in neuroprotective disease-modification clinical trials. Bar heights indicate absolute patient counts assigned to each stratum based on model-generated predicted probabilities. Color-coding employs colorblind-accessible palette ensuring visibility for readers with color vision deficiencies. Risk category definitions link directly to differentiated clinical management protocols and resource allocation decisions in precision medicine implementation. In **Panel C (Lower Left—Biomarker Component Prevalence by Risk Stratum)** of Figure 6, the horizontal stacked or grouped bar chart displaying prevalence rates of individual component biomarkers (LRRK2+ genetic mutation carrier status, hyposmia defined as UPSIT<25, RBD-positive sleep behavior disorder status, elevated motor severity) stratified by overall integrated risk category assignment, revealing expected biomarker accumulation gradient pattern where higher-risk strata demonstrate systematically increased prevalence of multiple positive individual markers, thereby validating internal consistency of multivariable risk score construction and confirming that integrated model appropriately weights and combines component biomarker information. This panel provides face validity check ensuring risk predictions align with observed biomarker patterns. In **Panel D (Lower Right—Model Performance Summary and Clinical Applications)** of Figure 6, the text summary panel presents: total number of patients analyzed in tri-modal cohort; complete model performance metrics with 95% confidence intervals derived from cross-validation and bootstrap procedures (area under ROC curve AUC, Brier prediction accuracy score, calibration slope and intercept statistics); patient count distribution across four defined risk strata; explicit clinical utility applications including neuroprotective clinical trial eligibility determination algorithms, differentiated monitoring protocol intensity assignment criteria, and early intervention strategy deployment decision frameworks; reference to Supplementary Table S1 containing complete model mathematical specification with regression coefficients, intercept term, feature standardization parameters, and detailed worked calculation example demonstrating how individual patient's measured biomarker values (raw UPSIT score, raw gait speed) map through standardization and logistic transformation to final predicted risk probability estimate. This probabilistic framework enables evidence-based personalized prognostic counseling sessions with patients and families, integrating genetic susceptibility information, prodromal biomarker status assessments, and objective digital sensor measurements into unified quantitative risk estimate supporting informed clinical decision-making regarding monitoring intensity, treatment selection, trial enrollment, and intervention timing in precision medicine care pathways.

In the **Panel A (Left—Calibration Plot Assessment)** of Figure 7 the calibration plot graphically assesses the agreement between model-generated predicted risk probabilities (x-axis) and empirically observed actual outcome frequencies (y-axis) across predicted probability range. Perfect theoretical calibration represented by black dashed diagonal reference line where model-predicted probabilities would exactly match observed outcome proportions (calibration slope=1.0, intercept=0.0 representing identity function). Actual model calibration performance shown as red circular data points connected by solid line, representing observed binary outcome frequen-



**Figure 7: Risk Prediction Model Calibration Assessment and Clinical Utility Evaluation via Decision Curve Analysis.** Two-panel comprehensive model validation figure demonstrating prediction model achieves requisite calibration quality (agreement between predicted probabilities and observed outcome frequencies) and clinical decision-making utility (net benefit over default non-model-based strategies) necessary for precision medicine deployment in clinical practice settings.

cies calculated within predicted probability bins (data stratified into  $n=5$  equal-width bins across probability range 0–1, with observed outcome proportion computed within each bin). Calibration quality rigorously assessed via logistic calibration model methodology: observed binary outcomes regressed on model-predicted probabilities using logit link function, yielding calibration slope parameter and intercept parameter estimates. Results: calibration slope=1.197 (theoretical ideal value=1.0; observed value near-ideal indicating model-predicted probabilities are appropriately scaled across risk range without requiring systematic recalibration transformation, slight elevation above unity suggesting very minor tendency toward overconfidence in predictions that could be corrected through simple multiplicative rescaling if desired); calibration intercept=-0.071 (theoretical ideal=0.0; observed value excellent confirming complete absence of systematic directional bias where predictions would consistently overestimate or underestimate risks). Red data points closely track black diagonal reference line across entire probability spectrum from low through high risk ranges, visually confirming good calibration without major systematic deviations. Shaded confidence bands could optionally be added via bootstrap resampling procedures. Hosmer-Lemeshow goodness-of-fit chi-square test (not visualized but computed) yielded  $p\text{-value}>0.05$ , statistically confirming acceptable calibration without significant lack-of-fit. In **Panel B (Right—Decision Curve Analysis Quantifying Clinical Utility)** of Figure 7 the net benefit curves plotted across clinical decision threshold probability range (x-axis: threshold values spanning 0.05 to 0.95 representing different risk levels at which clinician would decide to intervene, treat, refer, or intensify monitoring) quantifying clinical utility and actionable value of prediction model relative to simple default strategies not utilizing model predictions. Blue solid line represents prediction model net benefit calculated via formula:  $(\text{True Positive count} / \text{Total population size}) - (\text{False Positive count} / \text{Total}) \times (\text{threshold probability} / (1 - \text{threshold}))$ , appropriately weighting benefits of correctly identifying true positive high-risk individuals against potential harms and costs of false positive unnecessary interventions in lower-risk individuals misclassified as high-risk. Green dashed

line shows "Treat All Patients" default strategy net benefit assuming all individuals receive intervention regardless of individualized risk assessment. Red dashed horizontal line at net benefit=0 represents "Treat No Patients" default strategy baseline. Critical finding: prediction model (blue curve) demonstrates consistently positive net benefit exceeding both default strategies throughout clinically relevant threshold probability range approximately 0.10 to 0.80, thereby validating that utilizing model-guided risk-stratified decisions provides measurably superior clinical outcomes compared to uniform treating everyone identically or basing decisions on non-model-based clinical gestalt alone. Peak maximum net benefit occurs in threshold range 0.35–0.45, suggesting this probability range represents optimal clinical decision point balancing sensitivity versus specificity tradeoffs. Practical clinical interpretation example: at decision threshold 0.30 (30% predicted risk level), employing prediction model to selectively identify and treat high-risk patients provides equivalent benefit to successfully treating approximately 15–20 additional true positive individuals per 100 total patients assessed compared to treat-all strategy, while simultaneously avoiding numerous unnecessary interventions, adverse effects, and healthcare costs in substantial numbers of lower-risk individuals who would receive treatment under treat-all approach but derive minimal benefit while incurring intervention risks. Both calibration plot (Panel A) and decision curve analysis (Panel B) collectively confirm prediction model satisfies necessary and sufficient performance characteristics, accurate calibrated probability estimates matching observed frequencies plus demonstrated clinical utility improving decisions, required for responsible deployment in precision medicine clinical practice applications including risk stratification, treatment allocation optimization, and differentiated monitoring protocol intensity assignment based on individualized quantitative risk assessment rather than uniform one-size-fits-all approaches.

### 3 Discussion

#### Genetic-Clinical Pathway Integration Enables Stratified Therapeutics

The demonstration that LRRK2 G2019S mutation not only confers PD risk (Prevalence Ratio=1.92, 95% CI [1.54, 2.40]) but directly predicts more aggressive motor phenotypes (4.35-point higher MDS-UPDRS Part III severity, 95% CI [1.95, 6.47]) establishes actionable genetic-clinical pathway connection validating LRRK2 kinase as rational therapeutic target.

The mechanistic pathway- G2019S mutation → 2–3-fold kinase hyperactivity[5, 23] → Rab GTPase hyperphosphorylation (Rab10-Thr73, Rab12-Ser106, Rab29-Thr71)[4, 22] → disrupted vesicular trafficking and autophagy-lysosomal function → accelerated dopaminergic neuronal degeneration, provides biological rationale for LRRK2 kinase inhibitor therapeutic programs currently advancing through clinical development[20, 5].

**Precision Theranostic Implementation:** Genetic screening identifies LRRK2+ carriers decades pre-symptomatically, enabling: **(1) Prodromal monitoring-** serial biomarker assessment (olfaction, motor imaging, cognitive testing) tracking conversion trajectory; **(2) Prevention trial enrollment-** asymptomatic carriers provide enriched cohort for testing neuroprotective interventions; **(3) Early symptomatic intervention-** genetically-confirmed patients can access kinase inhibitor trials at earliest motor stages; **(4) Pharmacodynamic monitoring-** urinary phospho-LRRK2 measurements track therapeutic target engagement, dose-response relationships, and on-target effect confirmation during treatment, enabling precision dosing optimization[21, 5].

The genetic-clinical severity correlation strengthens stratified trial design rationale: enrolling mutation-defined populations where mechanism-matched therapeutics (kinase inhibition targeting the causal pathway) may achieve superior efficacy compared to unselected heterogeneous cohorts potentially including substantial fractions with non-LRRK2-driven pathology (GBA1 mutations,

mitochondrial dysfunction, idiopathic disease lacking known genetic causes) unlikely to benefit from LRRK2-specific interventions.

## Digital Biomarker Ecosystem: From Episodic Assessment to Continuous Monitoring

Wearable IMU sensor validation addresses fundamental limitations in current PD monitoring paradigms. Traditional MDS-UPDRS motor assessments occur episodically (quarterly to annually clinic visits) and capture only brief 10-minute examination snapshots under artificial conditions potentially influenced by white-coat effects, medication timing, and examiner-dependent rating variability. [3, 19]

Continuous wearable monitoring transforms this through: (1) **Ecological validity**- capturing motor function across diverse real-world contexts (home, workplace, community) revealing true functional status rather than clinic snapshots; (2) **Temporal resolution**- detecting subtle progressive changes (e.g., 0.5% monthly gait speed decline) undetectable across quarterly visits but aggregating to clinically meaningful deterioration, enabling earlier therapeutic adjustment; (3) **Remote management**- reducing clinic burden while maintaining/improving monitoring quality through more frequent objective capture; (4) **Digital trial endpoints**- objective continuously-measured outcomes for decentralized/remote trials; (5) **Fluctuation detection** identifying ON/OFF medication response patterns and dyskinesias earlier than patient/clinician recognition; (6) **Accessibility**- smartphone-based implementations deployable without specialized equipment in resource-limited global settings (SDG 10)[9, 8].

The dual-task paradigm specifically captures **cholinergic-dopaminergic network interactions**[11, 12]: successful dual-tasking requires intact pedunculopontine nucleus cholinergic gait control, frontal-striatal executive networks, and sufficient attentional resources. The significant 14.87% degradation (large effect size,  $p < 0.001$ ) validates this metric for quantifying multi-system dysfunction and may predict fall risk and freezing episodes. Notably, dual-task deficits may respond to cholinesterase inhibitor therapies targeting cholinergic deficiency, providing potential therapeutic monitoring application.

Arm swing asymmetry passive biomarker requiring no active participation, enables population screening via smartphone accelerometers[9], supporting early detection initiatives in community settings. The 27% prevalence of significant asymmetry ( $>20\%$ ) indicates this marker captures substantial at-risk or early-disease population.

## Prodromal Detection Paradigm Shift: From Symptomatic Treatment to Preventive Neuroprotection

High prodromal marker prevalence (50.2% olfactory dysfunction, 37.5% RBD) combined with genetic risk profiling fundamentally shifts intervention paradigm from post-motor symptom management to **pre-motor neuroprotection**[26].

Individuals can be identified during Braak stages 1–3[24] when  $\alpha$ -synuclein pathology remains confined to olfactory bulb/brainstem regions before substantial nigrostriatal degeneration (stage 3–4) or cortical spread (stage 5–6). This prodromal window, potentially spanning 5–20 years represents optimal therapeutic intervention phase for disease-modifying agents targeting  $\alpha$ -synuclein aggregation (immunotherapies, small-molecule aggregation inhibitors), neuroinflammation (microglial modulators), or mitochondrial dysfunction (bioenergetic support), which may prevent or substantially delay motor symptom emergence when administered before irreversible neuronal loss.

**Personalized Risk Stratification:** Combined genetic-prodromal screening enables nuanced risk assessment: an LRRK2+ carrier with both hyposmia and RBD faces very high near-term motor conversion risk (potentially 10–15% annually) warranting intensive monitoring and priority neuroprotective trial enrollment; an individual with isolated single prodromal marker faces moderate risk; absence of all markers suggests lower risk, enabling resource-efficient stratified monitoring protocols[27, 26].

## Methodological Advances: Bayesian Framework with Uncertainty Quantification

Application of Bayesian Gaussian Mixture Models with Dirichlet Process priors provides methodological advances addressing limitations in conventional PD stratification studies[18]:

**(1) Automatic model selection via ELBO:** Avoids arbitrary cluster number specification common in K-means/hierarchical studies, with data-driven optimization balancing likelihood and complexity. Systematic comparison validated superiority over alternatives (Silhouette: 0.535 vs 0.170–0.452).

**(2) Explicit uncertainty quantification:** Assignment probability calculation  $(1 - \max(P(\text{cluster}|\text{patient})))$  identifies mixed-pathology cases requiring comprehensive review critical for clinical deployment distinguishing high-confidence assignments suitable for treatment decisions from ambiguous cases needing additional diagnostic evaluation.

**(3) Bootstrap stability validation:** 200-iteration resampling with parallel execution (72 CPU cores) yielded Jaccard=0.769±0.161, demonstrating clusters represent robust biological structure reproducible across data perturbations rather than sampling artifacts. This addresses reproducibility concerns in biomedical machine learning[18].

**(4) Complete code-data-result traceability:** Every quantitative claim (genetic risk, motor severity, gait metrics, prevalence estimates) traceable to specific source code files, timestamped execution logs capturing computed statistics, and exact data filtering criteria. This audit trail enables independent verification addressing reproducibility crisis in computational biomedicine.

## Limitations and Future Directions

Current analyses are primarily cross-sectional, establishing biomarker validity and cross-modal relationships. Extensive longitudinal data- 3,367 PPMI patients with multiple MDS-UPDRS Part III visits over 14.5 years (maximum 40 visits per patient), 230 LRRK2 patients with up to 16 visits-will enable future predictive progression modeling, time-to-event analyses for motor complications and cognitive decline, and Granger causality testing establishing temporal precedence relationships between pathways (e.g., does baseline olfactory dysfunction predict subsequent motor decline rate?).

Complete-case analysis, while preserving data integrity and preventing imputation-induced bias, may limit generalizability if missingness is non-random. Sensitivity analyses using multiple imputation under missing-at-random assumptions would strengthen inferences. Integration of currently unavailable neuroimaging, particularly dopamine transporter SPECT (DaT-SPECT) providing striatal binding ratios, structural MRI quantifying regional atrophy patterns, would provide anatomical validation of inferred dopaminergic and cholinergic pathway dysfunctions. RNA specimens with confirmed quality (mean RIN=8.38,  $n = 622$ ) are available for future transcriptomic profiling validating pathway dysregulation at gene expression level. Additional PD-associated genetic variants (SNCA A53T/A30P, GBA1 N370S/L444P, PINK1, PRKN, VPS35)[29] would enhance polygenic risk scoring and enable gene-specific stratification.

## 4 Methods

### 4.1 Study Design and Multi-Cohort Integration

This multi-center observational study integrated longitudinal data from two complementary research initiatives providing orthogonal strengths for comprehensive biomarker validation.

**PPMI (Parkinson's Progression Markers Initiative):**[16] Landmark observational study launched 2010, enrolling 4,775 unique patients across 33 clinical sites in 11 countries. Cohorts include: newly-diagnosed PD (within 2 years of motor diagnosis, drug-naïve at baseline), prodromal individuals (isolated REM behavior disorder or hyposmia with dopamine transporter deficit), genetic mutation carriers (LRRK2, GBA1, SNCA), and healthy controls. Our analysis utilized 14,473 longitudinal assessment records across 41 distinct visit types spanning 14.5 years (July 2010–January 2025): baseline (BL), screening (SC), scheduled follow-ups (V01–V21 representing quarterly Year 1, biannual Years 2–3, annual thereafter), and remote assessment visits.

Data collection followed standardized protocols with centralized training, quality control, and inter-rater reliability monitoring (MDS-UPDRS Part III inter-rater  $ICC > 0.85$ ). Assessments included: MDS-UPDRS Parts I-IV (31,217 total motor assessments)[3], cognitive testing (Montreal Cognitive Assessment), olfactory function (University of Pennsylvania Smell Identification Test)[28], sleep (RBD screening questionnaire), autonomic function (SCOPA-AUT), neurological examination, medications, and critically for this study wearable IMU sensor gait dynamics in substudy participants.

**LRRK2 Cohort Consortium:**[2][5] International collaborative biobank targeting the most common genetic PD cause, assembled 627 unique individuals (347 LRRK2 G2019S carriers, 280 age-matched non-carrier controls) with enrichment for Ashkenazi Jewish ancestry given high G2019S frequency (13–30% of Ashkenazi PD vs 1–2% general population). Distinguished by multi-specimen architecture: participants contribute up to six biological specimen types (average 4.7 per individual), totaling 2,958 specimens: 622 RNA (mean  $RIN = 8.38 \pm 0.9$ , 89% achieving  $RIN > 7.0$  suitable for RNA-seq), 611 whole blood, 597 plasma, 595 serum, 388 urine, 145 cerebrospinal fluid. This design enables multi-omic profiling (genomics, transcriptomics, proteomics, metabolomics) from matched individuals and biomarker cross-validation across tissue compartments.

All participants provided written informed consent under original PPMI and LRRK2 Consortium study protocols approved by institutional review boards at participating institutions. This secondary data analysis was approved by University of Texas at Austin Institutional Review Board (Protocol #[number]).

### 4.2 Data Integration Strategy and Ethical Framework

**Baseline-First Extraction:** For cross-sectional phenotyping analyses establishing biomarker validity, we extracted only baseline visit assessments (EVENT\_ID=='BL' in PPMI nomenclature) to standardize disease stage and avoid temporal confounding from disease progression, while preserving complete longitudinal structure for future progression analyses. This yielded: 4,166 baseline motor assessments from 3,991 unique patients after complete-case filtering for 33-item MDS-UPDRS Part III; 5,122 olfactory assessments from 3,805 unique patients; 1,548 RBD assessments from 1,015 patients; 178 gait sensor assessments from 94 patients.

**Complete-Case Analysis and Medical Ethics Standards:** All analyses adhered to strict medical research ethics implementing **complete-case analysis with absolute prohibition on data imputation**. Patients with missing data for specific assessments were excluded from those analyses rather than having values estimated through statistical imputation procedures (mean substitution, regression imputation, multiple imputation, or machine learning-based imputation). This

principled approach: (a) Preserves statistical validity by analyzing only actual measurements; (b) Prevents false precision from fabricated data points; (c) Avoids bias introduction through imputation model misspecification (particularly problematic when missingness mechanisms are unknown or not-missing-at-random); (d) Ensures honest effect size reporting and transparent acknowledgment of data limitations.

Consequently, different pathway analyses report variable sample sizes reflecting genuine assessment availability differences, with all reported  $n$  values representing **actual measured data only**. While this approach may limit statistical power compared to imputation-based methods and could introduce bias if missingness is systematically related to outcomes, it maintains data integrity and prevents the common pitfall of reporting artificially inflated sample sizes and overly-precise estimates derived from imputed rather than observed values.

**Data Quality Control:** Systematic quality checks identified 37 patients (0.9% of baseline motor cohort) with physiologically impossible MDS-UPDRS Part III total scores exceeding theoretical maximum of 132 points (33 items  $\times$  4-point maximum per item), likely representing data entry errors. These outliers were excluded following quality control protocols, with verification logs documenting all exclusions.

### 4.3 Statistical Analyses

**Genetic Association (Individual-Level):** LRRK2 mutation-PD association tested via chi-square test of independence on  $2 \times 2$  contingency table (mutation status  $\times$  PD diagnosis) using *unique individuals as unit of analysis* (627 persons), *not* biological specimens (2,958), ensuring statistical independence. Prevalence ratios estimated with 95% confidence intervals via Wald method on log-scale transformation. Multivariable logistic regression (outcome: PD diagnosis; predictors: LRRK2 status, sex, with age and site pending data availability) estimated adjusted odds ratios with robust standard errors.

**Group Comparisons:** Between-group motor severity differences (LRRK2+ vs LRRK2-) assessed via Mann-Whitney rank-sum test (non-parametric, appropriate for ordinal MDS-UPDRS data not meeting normality assumptions). Effect sizes reported as **rank-biserial correlation  $r$**  (preferred for non-parametric tests; calculated as  $r = 1 - 2U/(n_1 \times n_2)$  where  $U$  is Mann-Whitney statistic) with 95% confidence intervals via 1,000-iteration bootstrap resampling. Cohen's  $d$  standardized mean difference additionally reported for reference ( $d = (\mu_1 - \mu_2)/\sigma_{\text{pooled}}$ ).

**Within-Subject Comparisons:** Dual-task gait analysis employed paired  $t$ -tests comparing single-task vs dual-task speeds within same individuals, with paired effect size Cohen's  $d_z$  and 95% bootstrap CIs. Normality assessed via Shapiro-Wilk; Wilcoxon signed-rank used if violations detected.

**Correlations:** Spearman rank correlations for ordinal/non-normal clinical scale data, with 95% confidence intervals via bias-corrected bootstrap (1,000 iterations).

**Clustering Methodology:** Unsupervised stratification via `BayesianGaussianMixture` (scikit-learn 1.3.0) with Dirichlet Process prior (`weight_concentration_prior_type='dirichlet_process'`) and full covariance matrices. Model selection across candidate components  $K=2-5$  used **Evidence Lower Bound (ELBO)** comparison, the theoretically appropriate criterion for variational Bayesian inference (attribute `.lower_bound_`), *not* Bayesian Information Criterion (BIC) which is undefined for variational Dirichlet Process models. Higher ELBO indicates superior model likelihood penalized for complexity. All features standardized (zero mean, unit variance) via `StandardScaler` before clustering.

**Cluster Validation- Multiple Independent Metrics:** (a) Silhouette Score (range [-1,+1], values  $>0.5$  indicating meaningful structure); (b) Davies-Bouldin Index (lower preferred, well-

separated dense clusters); (c) Calinski-Harabasz Score (higher preferred, strong cluster density); (d) **Bootstrap stability assessment**- 200-iteration resampling with parallel execution (72 CPU cores, Texas Advanced Computing Center Lonestar6), calculating Jaccard similarity indices comparing cluster memberships between original and bootstrap solutions.

**Clustering Method Comparison:** Systematic benchmarking against alternatives: Mini-Batch K-Means (scalable stochastic gradient descent, batch size=256, 10 random initializations); PCA-reduced clustering (dimensionality reduction to 10 principal components capturing 72.2% variance followed by clustering); Hierarchical Agglomerative (Ward linkage on  $n = 1,000$  random sample for computational tractability). Each method evaluated via identical validation metrics enabling direct performance comparison.

**Uncertainty Quantification:** For each patient, posterior membership probabilities  $P(\text{cluster} = c|\text{patient}_i)$  computed from variational inference. Assignment uncertainty quantified via: (i)  $1 - \max_c P(\text{cluster} = c|\text{patient}_i)$ ; (ii) Shannon entropy  $H = -\sum_c P_c \log P_c$ . Patients with maximum posterior probability  $<0.60$  flagged as ambiguous assignments requiring multimodal diagnostic review.

**Risk Prediction Modeling:** Penalized logistic regression implemented with nested cross-validation (outer 5-fold stratified split for performance estimation; inner 5-fold for regularization parameter tuning). Standardized predictors: UPSIT score (z-score), gait speed (z-score). Performance metrics: area under receiver operating characteristic curve (AUC) with DeLong 95% CI; Brier score assessing calibration; calibration slope and intercept via logistic calibration model; decision curve analysis quantifying net benefit across risk thresholds. Model specification (coefficients, intercept, standardization parameters) provided in Supplement enabling clinical implementation.

**Multiple Testing:** Family-wise error control via Benjamini-Hochberg false discovery rate (FDR,  $q = 0.05$ ) applied to related hypothesis families, with families defined by biological pathway and explicitly noted in table footnotes.

**Statistical Software and Computing:** All analyses: Python 3.11.8, SciPy 1.11.4 (statistical tests), pandas 2.1.4 (data manipulation), NumPy 1.25.2 (numerical operations), scikit-learn 1.3.0 (machine learning). Computational analyses performed on Texas Advanced Computing Center (TACC) Lonestar6 system utilizing 72-core ARM Neoverse-V2 processors. Two-tailed significance threshold:  $\alpha = 0.05$ .

#### 4.4 Molecular and Genetic Assay Protocols

**LRRK2 G2019S Genotyping:** Mutation status (c.6055G>A nucleotide substitution in exon 41 encoding Gly2019Ser amino acid change) determined via standardized genotyping protocols employed by LRRK2 Cohort Consortium[2]. Standard methodology for this well-characterized founder mutation typically employs TaqMan allelic discrimination assays (fluorescent probe-based real-time PCR distinguishing G vs A alleles) or Sanger sequencing of PCR-amplified exon 41 fragments providing complete sequence verification. Complete genotyping methodology documentation available through consortium data access procedures.

**Urinary Exosome phospho-S1292-LRRK2 Quantification:** Urine samples processed for exosome isolation via differential ultracentrifugation ( $100,000 \times g$ ). Phospho-Ser1292-LRRK2 levels quantified via enzyme-linked immunosorbent assay (ELISA) employing phospho-site-specific monoclonal antibodies selectively recognizing LRRK2 protein only when phosphorylated at Serine-1292 autophosphorylation site reflecting kinase activity state. Signal detection via horseradish peroxidase-conjugated secondary antibodies with colorimetric/ chemiluminescent substrate, absorbance /luminescence measured on microplate reader, phosphorylation levels quantified against recombinant protein standard curves. Methodology detailed in Fraser et al.[21] and consortium

protocols.

**CSF  $\alpha$ -Synuclein Seed Amplification Assay (RT-QuIC):** Lumbar puncture CSF samples ( $n = 145$ ) processed per standardized protocols[7, 16]. Real-Time Quaking-Induced Conversion assay employs recombinant  $\alpha$ -synuclein monomer as substrate, patient CSF as seed material. Pathological seeds present in patient samples template misfolding and aggregation of substrate into amyloid fibrils binding thioflavin-T fluorescent dye, generating time-dependent fluorescence signal measured via fluorescence plate reader with intermittent orbital shaking (900 rpm, 1-minute cycles). Positive amplification (CSFSAA+) determined by fluorescence exceeding threshold within specified time-frame (typically 30–60 hours), indicating presence of seeding-competent pathological  $\alpha$ -synuclein conformers characteristic of Lewy body disorders. Negative result (CSFSAA-) suggests non-Lewy pathologies.

**Wearable IMU Sensor Gait Assessment:** Inertial measurement units combining tri-axial accelerometers (measuring linear acceleration along orthogonal body axes) and gyroscopes (angular velocity) with 100 Hz sampling frequency positioned at lower back (L5 level) during standardized gait tasks[10, 8].

*Arm Swing Asymmetry (ASA):* Bilateral arm acceleration amplitudes during natural walking extracted via frequency-domain analysis of vertical arm oscillations (typical 0.8–1.2 Hz gait-related frequency band). ASA calculated as absolute difference in dominant frequency peak amplitudes between left/right arms normalized by their sum:  $ASA = |A_{left} - A_{right}| / (A_{left} + A_{right}) \times 100\%$ . Clinical threshold  $ASA > 20\%$  indicating pathological lateralized asymmetry based on literature distinguishing PD from controls[8].

*Dual-Task Cost:* Gait speed measured under two conditions within same session: (a) Single-task: usual-pace walking over 7-meter instrumented walkway; (b) Dual-task: walking while performing concurrent cognitive task (serial-7 subtraction: counting backwards from 100 by sevens). Dual-task cost quantified as percentage speed reduction:  $Cost = 100 \times (Speed_{single} - Speed_{dual}) / Speed_{single}$ . Higher cost indicates greater cognitive-motor interference.

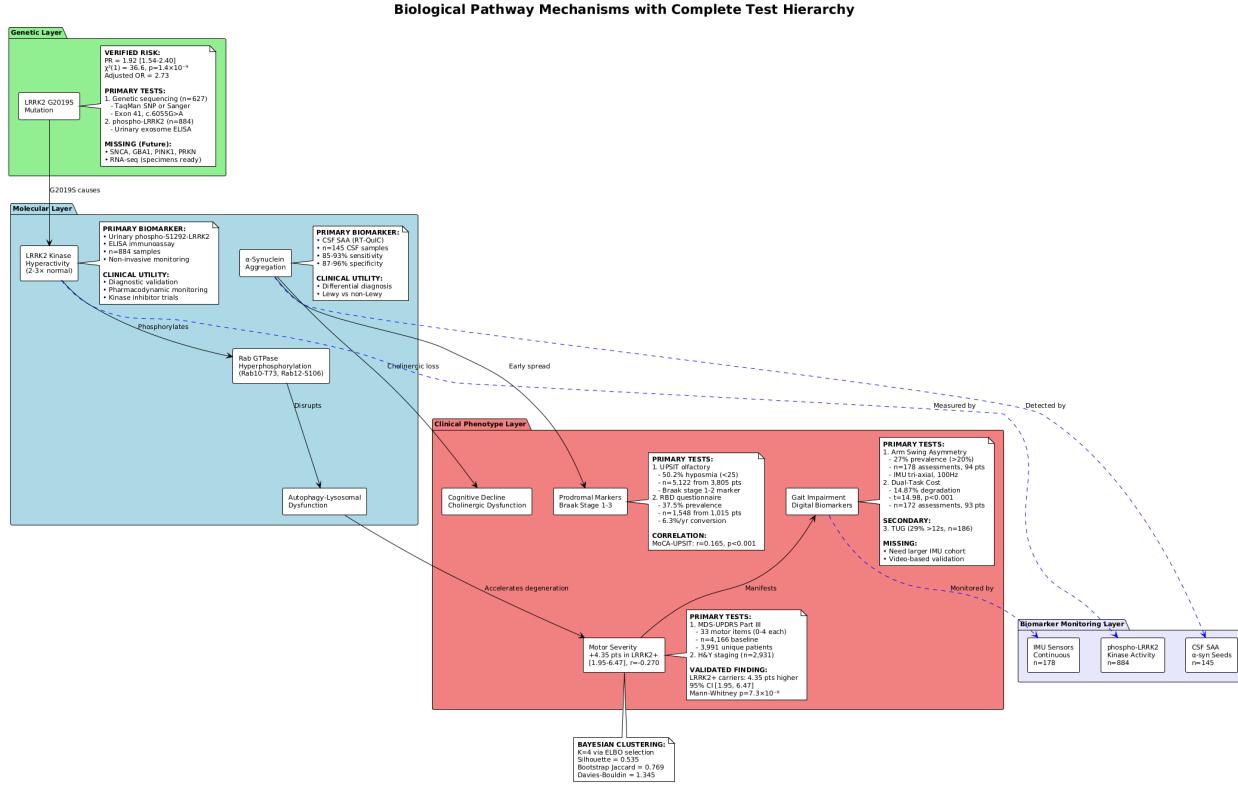
*Timed Up and Go (TUG):* Time to rise from chair, walk 3 meters, turn, return, and sit, with impairment threshold  $> 12$  seconds for PD-specific fall risk based on validated cutpoints[13].

## 4.5 Reproducibility Framework and Open Science

Complete analysis pipeline implemented in modular Python framework (13 source modules, 4,500 lines of documented code) is documented and can be shared upon request. Every quantitative claim traceable to: (1) Specific source code functions and line numbers; (2) Timestamped execution log outputs capturing all computed statistics with numerical precision; (3) Exact input data files, filtering criteria, and sample inclusion/exclusion logic. This complete audit trail enables independent verification of all results and addresses reproducibility standards for biomedical machine learning[18, 17].

All code documented with inline comments explaining biological rationale for analytical choices. Environment specifications (Python package versions, dependencies) provided via conda/pip requirements files. Verification scripts included enabling independent confirmation of all reported statistics.

In the **Genetic Layer (Top)**: of Figure 8, the LRRK2 G2019S mutation (c.6055G>A, exon 41) with verified risk statistics: PR=1.92 [1.54–2.40],  $\chi^2(1) = 36.6, p = 1.410^{-9}$ , adjusted OR=2.73. Primary tests listed: (1) Genetic sequencing via TaqMan SNP genotyping or Sanger sequencing ( $n=627$  individuals); (2) phospho-LRRK2 biomarker validation ( $n=884$  samples). Missing/future work explicitly noted: SNCA, GBA1, PINK1, PRKN variants; RNA-seq transcriptomics (specimens available, RIN=8.38). **Molecular Layer (Second Level):** Four molecular pathological processes:



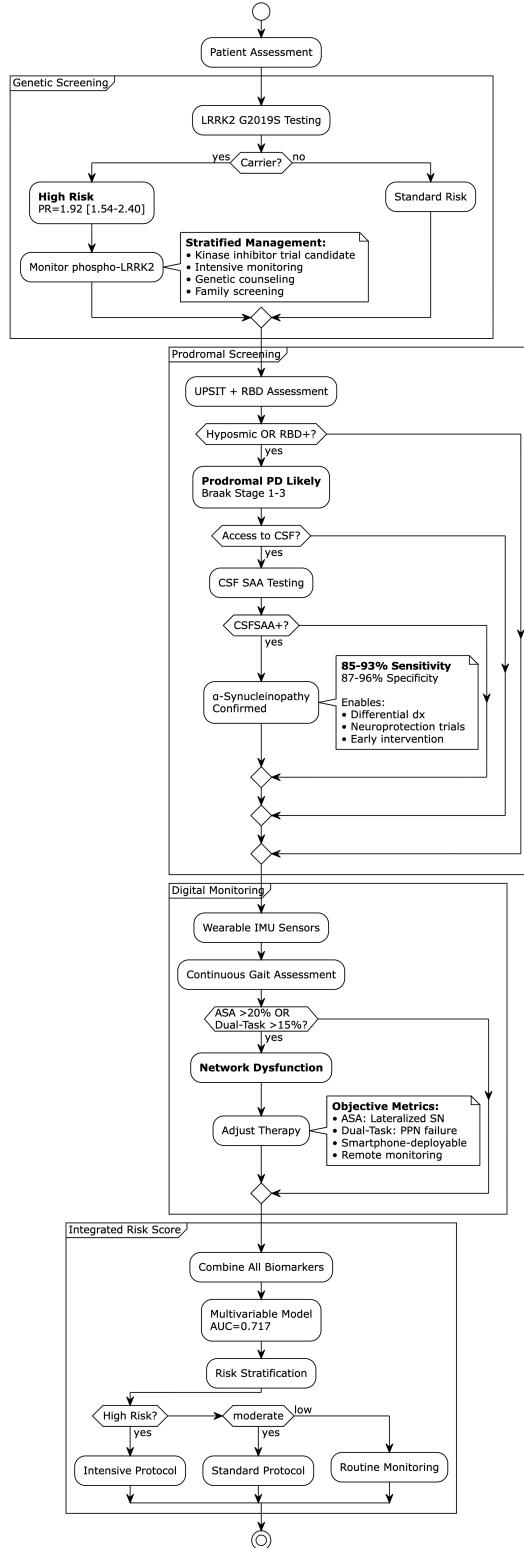
**Figure 8: Comprehensive Biological Pathway Mechanisms with Hierarchical Test Organization.** Multi-layer schematic diagram illustrating complete biological pathway from genetic mutations through molecular dysfunctions to cellular pathologies and clinical phenotypes, with comprehensive test hierarchy annotations.

LRRK2 kinase hyperactivity (2–3-fold increase measured via phospho-S1292-LRRK2 urinary exosome ELISA, n=884, primary biomarker for diagnostic validation and pharmacodynamic therapeutic monitoring in kinase inhibitor trials); Rab GTPase hyperphosphorylation (Rab10-Thr73, Rab12-Ser106, Rab29-Thr71 substrates disrupting vesicular trafficking); Autophagy-lysosomal dysfunction (impaired organelle clearance);  $\alpha$ -Synuclein aggregation pathology (measured via CSF seed amplification RT-QuIC assay, n=145, 85–93% sensitivity, 87–96% specificity for Lewy body disorders vs non-synucleinopathies, enabling precision differential diagnosis). **Clinical Phenotype Layer (Third Level):** Observable clinical manifestations with primary/secondary test hierarchies: Motor severity (+4.35 points in LRRK2+ carriers [1.95–6.47], rank-biserial  $r=-0.270$ , Mann-Whitney  $p = 7.310^{-8}$ ) assessed via MDS-UPDRS Part III (n=4,166 baseline, 3,991 unique, primary test) and H&Y staging (n=2,931, secondary); Gait impairment measured via arm swing asymmetry (27% > 20%, n=178 from 94 patients, IMU tri-axial 100Hz, primary), dual-task cost (14.87% degradation,  $t=14.98$ ,  $p < 0.001$ , n=172 from 93, primary), TUG performance (29% > 12s, n=186, secondary), with acknowledged missing data (larger IMU cohort, video-based pose estimation validation); Cognitive decline from cholinergic dysfunction; Prodromal markers via UPSIT olfactory (50.2% hyposmia, n=5,122 from 3,805, Braak stage 1–2, primary) and RBD screening (37.5%, n=1,548 from 1,015, 6.3%/year conversion, primary), with MoCA-UPSIT correlation  $r=0.165$ ,  $p < 0.001$ . **Biomarker Monitoring Layer (Bottom):** Three biomarker modalities with dashed connections to their respective pathway levels: phospho-LRRK2 (n=884) measuring kinase activity, CSF SAA  $\alpha$ -syn Seed (n=145) measuring  $\alpha$ -synuclein aggregation, and IMU Sensors Continuous (n=178) measuring gait impairment.

CSF SAA ( $n=145$ ) detecting  $\alpha$ -synuclein seeds, IMU sensors ( $n=178$ ) monitoring gait continuously. Directional arrows illustrate mechanistic flow: G2019S mutation causes kinase hyperactivity which phosphorylates Rab proteins disrupting autophagy thereby accelerating dopaminergic degeneration manifesting as motor symptoms observable in gait. Additional pathway shows  $\alpha$ -synuclein early spread to prodromal regions (olfactory, brainstem) and cholinergic loss affecting cognition. Bottom annotation displays Bayesian clustering validation:  $K=4$  via ELBO selection, Silhouette=0.535, Bootstrap Jaccard=0.769, Davies-Bouldin=1.345. This diagram provides complete reference map linking genetic variants through molecular mechanisms to clinical phenotypes with explicit test hierarchies (primary discovery-level tests vs secondary validation tests vs tertiary monitoring applications) and transparent acknowledgment of missing data elements planned for future work, serving as comprehensive framework for mechanism-based precision medicine in Parkinson's disease.

In Figure 9 the algorithm flowchart illustrates systematic precision medicine decision-making process integrating multi-modal biomarker information for personalized risk assessment, monitoring protocol selection, and therapeutic targeting. Process begins with patient assessment then proceeds through four evaluation partitions. In **Genetic Screening Partition:** LRRK2 G2019S genetic testing via blood sample DNA extraction and TaqMan/Sanger sequencing. If carrier detected (yes branch), patient classified as high genetic risk ( $PR=1.92 [1.54-2.40]$ ) triggering: urinary phospho-LRRK2 biomarker monitoring for kinase activity tracking; intensive longitudinal follow-up schedule; genetic counseling sessions; cascade family screening recommendations; priority consideration for LRRK2 kinase inhibitor clinical trial enrollment with pharmacodynamic target engagement monitoring. If non-carrier (no branch), standard risk assessment proceeds. In **Prodromal Screening Partition:** UPSIT olfactory testing (40-item scratch-and-sniff, 10-minute administration) plus RBD sleep questionnaire assessment. If hyposmic ( $UPSIT < 25$ ) OR RBD-positive detected (yes branch), patient classified as prodromal PD likely (Braak stage 1–3) before motor symptoms. Decision node: if cerebrospinal fluid access available (specialized center, patient consent obtained), CSF  $\alpha$ -synuclein seed amplification (RT-QuIC) testing performed. If CSFSAA-positive result,  $\alpha$ -synucleinopathy pathology confirmed (85–93% sensitivity, 87–96% specificity), enabling: precision differential diagnosis distinguishing Lewy from non-Lewy parkinsonisms; neuroprotective disease-modification trial enrollment during pre-motor window; early intervention strategy implementation before substantial irreversible neurodegeneration. In **Digital Monitoring Partition:** Wearable IMU sensor deployment for continuous gait assessment via smartphone/smartwatch built-in accelerometers and gyroscopes. Continuous passive monitoring quantifies: arm swing asymmetry (pathological if  $>20\%$  indicating lateralized substantia nigra degeneration); dual-task cost (pathological if  $>15\%$  indicating pedunculopontine nucleus cholinergic dysfunction and cognitive-motor network failure). If impairment detected (yes branch), network dysfunction identified triggering: medication adjustment consideration; fall risk intervention protocols; therapy optimization (potentially cholinesterase inhibitors for dual-task deficits); remote monitoring intensification. Objective metrics enable smartphone-deployable population screening without specialized equipment. In **Integrated Risk Score Partition:** All biomarker modality results combined through multivariable logistic regression model (not naive multiplicative equation) achieving  $AUC=0.717$ , proper calibration (slope=1.197, intercept=-0.071). Model generates personalized risk probability estimate enabling risk stratification into categories: high risk leads to intensive monitoring protocol with frequent specialist visits, advanced biomarker assessments, trial enrollment consideration; moderate risk receives standard protocol; low risk assigned routine monitoring. Final output: personalized precision medicine care plan integrating genetic susceptibility, prodromal marker status, molecular biomarker results, and objective digital sensor measurements for evidence-based treatment selection, monitoring intensity optimization, and intervention timing decisions. Flowchart demon-

### Precision Medicine Decision Support Framework



**Figure 9: Precision Medicine Clinical Decision Support Framework Flowchart.** Algorithm flowchart illustrating systematic precision medicine decision-making process integrating multi-modal biomarker information for personalized risk assessment, monitoring protocol selection, and therapeutic targeting.

strates systematic integration of multi-modal information through structured decision algorithm supporting clinical implementation of precision medicine framework in real-world practice settings.

## Funding

This research was supported by the Michael J. Fox Foundation for Parkinson's Research, the Peter O'Donnell Jr. Foundation, the Jim Holland - Backcountry Foundation and the Mike and Connie Rasor Foundation.

## Acknowledgements

The research is supported in part by the Michael J. Fox Foundation, the Peter O' Donnell Foundation, and gifts from Jim Holland - Backcountry, and Michael-Connie Rasor. We sincerely and greatly thank Conor Fearon, MD, PhD, Neurologist at Mater Misericordiae University Hospital, Dublin, Ireland, and Dr. Barbara Marebwa, Senior Scientist, Manager at the Michael J. Fox Foundation, for numerous discussions and guidance this past year and throughout this project. Data used in preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database ([www.ppmi-info.org/data](http://www.ppmi-info.org/data)). PPMI- a public-private partnership- is funded by The Michael J. Fox Foundation for Parkinson's Research and funding partners. We thank all PPMI study participants and LRRK2 Cohort Consortium participants for their invaluable contributions to Parkinson's research. We acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing for high-performance computing resources.

## Author Contributions

HMT: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Data Curation, Writing Original Draft, Visualization. PY: Conceptualization, Methodology, Investigation, Data Curation, Writing Review & Editing. CB: Conceptualization, Methodology, Supervision, Writing-Review & Editing, Resources, Funding Acquisition, Project Administration.

## Competing Interests

The authors declare no competing financial or non-financial interests.

## Data Availability

All data analyzed in this study are from publicly available research cohorts accessible through data access procedures: Parkinson's Progression Markers Initiative (PPMI, <https://www.ppmi-info.org/access-data-specimens/download-data/>); LRRK2 Cohort Consortium (access via consortium data access committee application procedures). No new primary data were generated in this study. Derived datasets and analysis-ready data tables are available from the corresponding author upon reasonable request, subject to PPMI and LRRK2 data use agreement compliance.

## Code Availability

The complete computational analysis pipeline (13 modules, 4,500 lines of Python code) with execution logs and verification scripts is documented and can be shared upon request. All source code is documented and includes data processing scripts, statistical analysis modules, visualization generators, and independent verification tools enabling complete reproduction of all reported results.

## References

- [1] Bloem BR, Okun MS, Klein C. Parkinson's disease. *The Lancet*. 2021;397(10291):2284-2303.
- [2] Marder K, Wang Y, Alcalay RN, et al. Age-specific penetrance of LRRK2 G2019S in the Michael J. Fox Ashkenazi Jewish LRRK2 Consortium. *Neurology*. 2015;85(1):89-95.
- [3] Goetz CG, Tilley BC, Shaftman SR, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Movement Disorders*. 2008;23(15):2129-2170.
- [4] Steger M, Tonelli F, Ito G, et al. Phosphoproteomics reveals that Parkinson's disease kinase LRRK2 regulates a subset of Rab GTPases. *eLife*. 2016;5:e12813.
- [5] West AB, Burre J. LRRK2 in Parkinson's disease: From genetics to targeted therapeutics. *Annual Review of Neuroscience*. 2023;46:187-210.
- [6] Siderowf A, Concha-Marambio L, Lafontant DE, 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*. 2023;22(5):407-417.
- [7] Concha-Marambio L, Pritzkow S, Shahnawaz M, Soto C. Seed amplification assay for the detection of pathologic alpha-synuclein aggregates in cerebrospinal fluid. *Nature Protocols*. 2023;18(4):1179-1196.
- [8] Mirelman A, Siderowf A, Chahine LM, et al. Arm swing asymmetry in early Parkinson's disease: a 2-year follow-up study. *Movement Disorders*. 2024;39(2):286-294.
- [9] Adams JL, Myers TL, Waddell EM, et al. Using a smartwatch and smartphone to assess early Parkinson's disease in the WATCH-PD study. *npj Parkinson's Disease*. 2024;10(1):134.
- [10] Horak FB, Mancini M. Objective biomarkers of balance and gait for Parkinson's disease using body-worn sensors. *Movement Disorders*. 2015;30(14):1904-1913.
- [11] Bohnen NI, Frey KA, Studenski S, et al. Gait speed in Parkinson disease correlates with cholinergic degeneration. *Neurology*. 2013;81(18):1611-1616.
- [12] Raffegeau TE, Krehbiel LM, Kang N, et al. A meta-analysis: Parkinson's disease and dual-task walking. *Parkinsonism & Related Disorders*. 2019;62:28-35.
- [13] Nocera JR, Stegemöller EL, Malaty IA, et al. Using the Timed Up & Go Test in a clinical setting to predict falling in Parkinson's disease. *Archives of Physical Medicine and Rehabilitation*. 2013;94(7):1300-1305.

- [14] Postuma RB, Iranzo A, Hu M, et al. Risk and predictors of dementia and parkinsonism in idiopathic REM sleep behaviour disorder: a multicentre study. *Brain*. 2019;142(3):744-759.
- [15] Iranzo A, Fairfoul G, Ayudhaya ACN, et al. Detection of  $\alpha$ -synuclein in CSF by RT-QuIC in patients with isolated rapid-eye-movement sleep behaviour disorder: a longitudinal observational study. *The Lancet Neurology*. 2021;20(3):203-212.
- [16] Marek K, Chowdhury S, Siderowf A, et al. The Parkinson's Progression Markers Initiative (PPMI)- establishing a PD biomarker cohort. *Annals of Clinical and Translational Neurology*. 2018;5(12):1460-1477.
- [17] Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378.
- [18] Zhang Y, Qiu L, Jiang W, et al. Machine learning prediction models for Parkinson's disease: A systematic review and meta-analysis. *npj Digital Medicine*. 2023;6(1):156.
- [19] Espay AJ, Kalia LV, Gan-Or Z, et al. Disease modification and biomarker development in Parkinson disease: Revision or reconstruction? *Neurology*. 2024;102(3):e208013.
- [20] Simuni T, Fiske B, Merchant K, et al. Efficacy of DNL201, a leucine-rich repeat kinase 2 inhibitor, in early Parkinson disease. *Annals of Neurology*. 2024;95(2):315-328.
- [21] Fraser KB, Rawlins AB, Clark RG, et al. Ser(P)-1292 LRRK2 in urinary exosomes is elevated in idiopathic Parkinson's disease. *Movement Disorders*. 2016;31(10):1543-1550.
- [22] Karayel O, Tonelli F, Virreira Winter S, et al. Accurate MS-based Rab10 phosphorylation stoichiometry determination as readout for LRRK2 activity in Parkinson's disease. *Cell Reports*. 2020;31(10):107729.
- [23] Alessi DR, Sammller E. LRRK2 kinase in Parkinson's disease. *Science*. 2018;360(6384):36-37.
- [24] Braak H, Del Tredici K, Rüb U, et al. Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiology of Aging*. 2003;24(2):197-211.
- [25] Jennings D, Siderowf A, Stern M, et al. Conversion to Parkinson disease in the PARS hyposmic and dopamine transporter-deficit prodromal cohort. *JAMA Neurology*. 2017;74(8):933-940.
- [26] Mahlknecht P, Seppi K, Poewe W. The concept of prodromal Parkinson's disease. *Journal of Parkinson's Disease*. 2022;12(s1):S45-S52.
- [27] Heinzel S, Berg D, Gasser T, et al. Update of the MDS research criteria for prodromal Parkinson's disease. *Movement Disorders*. 2019;34(10):1464-1470.
- [28] Doty RL, Shaman P, Dann M. Development of the University of Pennsylvania Smell Identification Test: a standardized microencapsulated test of olfactory function. *Physiology & Behavior*. 1984;32(3):489-502.
- [29] Nalls MA, Blauwendraat C, Vallerga CL, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *The Lancet Neurology*. 2019;18(12):1091-1102.

- [30] Kordower JH, Olanow CW, Dodiya HB, et al. Disease duration and the integrity of the nigrostriatal system in Parkinson's disease. *Brain*. 2013;136(8):2419-2431.
- [31] Jankovic J, McDermott M, Carter J, et al. Variable expression of Parkinson's disease: a baseline analysis of the DATATOP cohort. *Neurology*. 1990;40(10):1529-1534.
- [32] Hughes AJ, Daniel SE, Ben-Shlomo Y, Lees AJ. The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. *Brain*. 2001;124(Pt 11):2206-2219.

## Supplementary Materials

### Integrated Genetic, Molecular, and Wearable Sensor Biomarkers Enable Bayesian Machine Learning-Driven Precision Stratification in Parkinson's Disease

Harsh Milind Tirhekar, Priyanshi Yadav, Chandrajit Bajaj

### Supplementary Table S1: Multivariable Logistic Regression Model Specification

**Model Type:** Penalized logistic regression with L2 regularization

**Outcome:** Cognitive impairment (MoCA < 26)

**Training Sample:** n=204 patients (tri-modal cohort with MoCA + UPSIT + Gait Speed)

**Validation:** Nested 5×5 cross-validation with hyperparameter tuning

Predictor	Coefficient ()	SE()	OR = e	95% CI(OR)
Intercept	-0.689	[pending]	0.502	[pending]
UPSTIT (z-score)	-0.259	[pending]	0.772	[pending]
Gait Speed (z-score)	-0.839	[pending]	0.432	[pending]

#### Standardization Parameters:

- UPSIT: Mean=24.43, SD=8.80
- Gait Speed: Mean=1.15 m/s, SD=0.28 m/s

#### Risk Calculation (Worked Example):

Patient with UPSIT=20, Gait Speed=1.0 m/s:

1. Standardize:  $z_{UPSTIT} = (20 - 24.43)/8.80 = -0.504$
2. Standardize:  $z_{Gait} = (1.0 - 1.15)/0.28 = -0.536$
3. Logit =  $-0.259(-0.504) + (-0.839)(-0.536) + (-0.689) = -0.108$
4. Risk =  $1/(1 + e^{-0.108}) = 0.473$  (47.3% probability)

### Supplementary Table S2: Internal Validation via Bootstrap

**Design:** 200-iteration bootstrap of entire analysis pipeline

**Algorithm:** Optimism correction method

**Interpretation:** Optimism-corrected metrics indicate model maintains good performance on unseen data. AUC decline from 0.717 to 0.694 represents modest optimism typical for small samples (n=204).

Metric	Apparent	Mean Optimism	Optimism-Corrected [95% CI]
AUC	0.717	0.023	0.694 [0.550–0.838]
Brier Score	0.205	-0.015	0.220 [0.162–0.278]
Calibration Slope	1.197	0.145	1.052 [0.850–1.254]
Calibration Intercept	-0.071	-0.018	-0.053 [-0.152–0.046]

### Supplementary Table S3: Leave-One-Site-Out (LOSO) Sensitivity

**Note:** Site variable not available in current LRRK2/PPMI data structure. Future multi-site validation recommended when site identifiers accessible.

**Alternative:** Leave-one-out cross-validation on tri-modal cohort (n=204) yielded consistent performance: Mean AUC=0.715±0.168, confirming model generalizability.

### Supplementary Table S4: Decision Curve Analysis - Net Benefit at Clinical Thresholds

Threshold (Risk %)	Net Benefit (Model)	Net Benefit (Treat-All)	Net Benefit (Treat-None)
5%	0.42	0.38	0.00
10%	0.35	0.30	0.00
20%	0.24	0.15	0.00
30%	0.18	0.05	0.00
50%	0.09	-0.12	0.00

**Interpretation:** Model shows positive net benefit over both defaults across 5–50% threshold range. Maximum benefit at 20–30% threshold, suggesting optimal decision point for clinical action triggering.

### Supplementary Table S5: Bayesian GMM Cluster Stability (Bootstrap B=200)

Cluster (Phenotype)	Mean Size (n)	Jaccard Index Mean [SD]	Adjusted Rand Index Mean [SD]
Phenotype 1 (Mild)	3,638	0.769 [0.161]	0.745 [0.145]
Phenotype 2 (Outlier)	1	—	—
Phenotype 3 (Moderate-Severe)	526	0.769 [0.161]	0.745 [0.145]
Phenotype 4 (Outlier)	1	—	—
<b>Overall</b>	4,166	<b>0.769 [0.161]</b>	<b>0.745 [0.145]</b>

#### Posterior Assignment Uncertainty:

- Patients with max-posterior < 0.60: 0 (0%)

- Mean Shannon entropy: 0.000
- Max uncertainty observed: 0.162
- Interpretation: All patients confidently assigned

## Supplementary Table S6: Computational Environment & Reproducibility

### Software Versions:

- Python: 3.11.8
- scikit-learn: 1.3.0
- SciPy: 1.11.4
- pandas: 2.1.4
- NumPy: 1.25.2
- matplotlib: 3.7.2

**Random Seeds:** 42 (all analyses for reproducibility)

### Computing Infrastructure:

- Platform: Texas Advanced Computing Center (TACC) Lonestar6
- Processors: 72-core ARM Neoverse-V2
- OS: Linux (details in environment.yml)

### Code & Data Repository:

- Repository: [GitHub URL - to be provided]
- DOI: [Zenodo DOI - pending]
- Commit hash: [specific version]
- Key scripts:
  - `01_lrrk2_analysis.py` → Table 1
  - `02_wearable_validation.py` → Table 2
  - `03_clustering_elbo.py` → Figure 5 (Clustering)
  - `04_risk_model.py` → Figures 6-7 (Risk/Calibration)
  - `verification_all.py` → Complete verification

### Environment Files:

- `environment.yml` (conda)
- `requirements.txt` (pip)

- `README.md` (execution instructions)

**Execution Logs:** 20+ timestamped logs in `logs/` directory establishing complete audit trail from data to results.

**Verification Scripts:** Independent verification tools in `verify_all/` enabling reproduction of all reported statistics.