Research Square

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

Oculomotor and Pupillary Biomarkers for Detecting and Staging Parkinson's Disease

Devin L. McCaslin
mccaslin@med.umich.edu
University of Michigan
Alex Kiderman
Spryson America, Inc.
Alexandr Braverman
Spryson America, Inc.
Laura C. Gibson
Spryson America, Inc.
Stiven Roytman
University of Michigan
Taylor Brown
University of Michigan
Jaimie Barr
University of Michigan
Giulia Carli
University of Michigan
Batoul Berri
University of Michigan
Kevin Kerber
The Ohio State University
Nicolaas I. Bohnen
University of Michigan

Research Article

Keywords: Parkinson's disease, oculomotor, saccades, pupillometry, video-oculography, diagnostic accuracy

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11485034/v1

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

Additional Declarations: Competing interest reported. Alex Kiderman, PhD, Alexandr Braverman, PhD, and Laura C. Gibson, PhD, are affiliated with Spryson America, Inc., the manufacturer of the goggle system. The other authors declare no competing interests.

Abstract

Purpose

To evaluate to what degree multimodal oculomotor, vestibular, reaction time, and cognitive (OVRT-C) metrics differentiate Parkinson's disease (PD) from healthy aging and to determine their association with clinical severity.

Methods

This is a cross-sectional diagnostic-accuracy study, 42 participants with established PD and 24 healthy controls completed a 23-test OVRT-C battery using portable video-oculography between 2023 and 2024. Group differences, age and sex-adjusted logistic regression, and cross-sectional associations with Hoehn and Yahr severity were examined. A parsimonious multivariable model was evaluated with leave-one-out cross-validation (LOOCV).

Results

PD was characterized by reduced saccadic accuracy, greater hypometria, impaired smooth pursuit and vergence, increased fixation-instability/nystagmus measures, and altered pupillary responses. The final multiple regression model combined vertical saccadic accuracy, horizontal and vertical gaze nystagmus counts in darkness, and predictive-saccade accuracy (AUC 0.975, 95% CI 0.917–0.995; sensitivity 0.950; specificity 0.870). Performance decreased with LOOCV (AUC 0.940, 95% CI 0.850–0.980; sensitivity 0.875; specificity 0.826). Horizontal saccade latency, vertical smooth-pursuit gain, and pupillary measures were associated with greater cross-sectional Hoehn and Yahr severity.

Conclusions

OVRT-C-derived oculomotor and pupillometric metrics discriminated participants with established PD from healthy controls, whereas vestibular measures contributed little to diagnostic classification. These findings support prospective validation of OVRT-C metrics as candidate biomarkers for PD and for cross-sectional disease characterization but will require rigorous validation.

Introduction

Parkinson's disease (PD) is a heterogeneous neurodegenerative disorder characterized by the cardinal motor features bradykinesia, rigidity, tremor, gait dysfunction, postural instability, and numerous non-motor symptoms (Poewe et al. 2017; Wüllner et al. 2023). Oculomotor abnormalities are also common but are less routinely quantified. Saccades may be hypometric or delayed (Shaikh and Ghasia 2019; Terao et al. 2013; White et al. 1983; Zhou et al. 2022), smooth pursuit may show reduced gain and corrective saccades (White et al. 1983; Frei 2021), vergence can be impaired (Hanuška et al. 2015), and pupillary responses may be altered (Giza et al. 2011). These measures may provide objective information about distributed cortical, basal ganglia, brainstem, cerebellar, and autonomic networks affected by PD (Shaikh and Ghasia 2019; Terao et al. 2013; White et al. 1983; Pinkhardt et al. 2012).

There is compelling evidence supporting the association between oculomotor abnormalities and PD severity. Changes in saccadic eye movements have been reported even in early stages of PD and tend to become more apparent with increasing motor disability (Terao et al. 2013; Zhou et al. 2022; Waldthaler et al. 2019). Patients may demonstrate prolonged saccadic latency, reduced amplitude and accuracy, increased hypometria, and greater reliance on corrective saccades (Shaikh and Ghasia 2019; Terao et al. 2013; White et al. 1983; Zhou et al. 2022; Pinkhardt et al. 2012; Waldthaler et al. 2019). Smooth pursuit may also be disrupted, with reduced gain and increased saccadic intrusions (Pinkhardt et al. 2012; Waldthaler et al. 2021), while vergence (Hanuška et al. 2015) and higher-order eye movements such as antisaccades (Waldthaler et al. 2019;

Waldthaler et al. 2021) can be abnormal. These wide-ranging findings illustrate that PD can affect oculomotor control across both reflexive and volitional tasks.

Oculomotor and pupillary measures may provide complementary information in the characterization of PD. For example, the antisaccade task requires suppression of a reflexive gaze shift toward a target and generation of an eye movement in the opposite direction. Patients with PD have been reported to show increased error rates and prolonged latencies (Waldthaler et al. 2019; Waldthaler et al. 2021). PD is also associated with abnormalities of the pupillary light reflex, including altered constriction amplitude, velocity, and latency (Giza et al. 2011). Taken together, these oculomotor measures provide a means of sampling motor, executive, brainstem, and autonomic networks that are not typically evaluated quantitatively during a standard neurologic examination.

PD remains a clinical diagnosis based on parkinsonism together with supportive features, exclusion criteria, and red flags (Tolosa et al. 2021; Hughes et al. 1992). Ancillary imaging and biomarker testing can be informative, but dopaminergic imaging does not reliably distinguish PD from atypical degenerative parkinsonism (Tolosa et al. 2021; Cummings et al. 2014), and fluid or tissue alpha-synuclein assays remain evolving tools (Coughlin and Irwin 2023). There is a need for objective, non-invasive, portable, scalable, and cost-effective diagnostic markers that aid in earlier and more accurate detection of PD. Accordingly, portable quantitative physiologic measures may be useful for studying disease-related network dysfunction, but their diagnostic role must be established through staged diagnostic-accuracy research (Sackett and Haynes 2002).

The Oculomotor, Vestibular, Reaction Time, and Cognitive (OVRT-C) battery integrates eye movements, fixation stability/nystagmus, pupillary responses, reaction time, and visuospatial measures in a brief portable assessment. This investigation evaluated OVRT-C metrics ability to discriminate individuals with established PD from known healthy controls and whether selected metrics are associated cross-sectionally with H&Y severity. The investigations design represents an early-phase test of discrimination rather than an assessment of clinical diagnostic utility. We hypothesized that individuals with mild-to-moderate PD would exhibit a distinct, quantifiable profile of deficits across domains that correlates with disease severity. Specifically, the study aimed to (1) identify which OVRT-C metrics are significantly impaired in PD patients compared to age-matched healthy controls, and (2) determine whether the overall pattern of OVRT-C results could serve as a reliable biomarker for diagnosing PD and correlating this with disease severity measures.

Methods

Participants

The University of Michigan Institutional Review Board approved the study (HUM00296601). Forty-two participants with established PD (H&Y stages 1–4) and 24 healthy controls were recruited from University of Michigan neurology and audiology clinics. Participants were 61 to 83 years of age (mean age: 70.7 ± 4.9·SD), of whom 61.9% were male, and twenty-four healthy controls (HC), aged 64 to 88 years (mean age: 72.9 ± 6.6·SD), of whom 66.7% were male, were recruited from neurology and audiology clinics at the University of Michigan (Table 1).

Table 1
Participant demographics

Group	N	Age, mean +/- SD (range), y	Male, n (%)
PD	42	70.7 +/- 4.9 (61–83)	26 (61.9)
Healthy controls	24	72.9 +/- 6.6 (64–88)	16 (66.7)
Total	66	71.2 +/- 6.3 (61–88)	42 (63.6)
PD, Parkinson's disease; SD, standard deviation.			

All participants were able to provide informed consent and had no diagnosis of dementia. PD participants met UK Parkinson's Disease Society Brain Bank criteria (Sackett and Haynes 2002). Controls had no parkinsonian motor signs or history of

neurodegenerative disease. Exclusion criteria included MoCA < 18, benign paroxysmal positional vertigo, major medical illness preventing participation, untreated psychiatric illness, stroke, or focal brain lesions on brain MRI or contraindications for MRI or PET imaging (e.g., metal implants, claustrophobia).

OVRT-C assessment

Participants completed comprehensive clinical testing using an immersive display as well as motor, cognitive, audiometric, and OVRT-C assessments during one session. The assessment used the Spryson Dx200 system (formerly Neurolign Dx100/I-Portal Portable Assessment System), an FDA-cleared Class II video-oculography device with binocular infrared cameras (100–250 frames/s). The 23-test battery assessed horizontal and vertical gaze and spontaneous nystagmus with and without fixation, optokinetic responses, subjective visual vertical/horizontal, reflexive and predictive saccades, smooth pursuit, vergence, antisaccades, memory-guided and self-paced saccades, visual/auditory reaction time, and pupillary light reflexes. Fixation-instability measures included square-wave jerks and nystagmus fast-phase counts, slow-phase velocity was summarized as average and peak slow-phase velocity. Total testing time was approximately 14.2 min. Calibration was performed before testing using horizontal and vertical calibration targets; recordings were reviewed for adequate tracking before metric extraction.

The I-Portal platform presents controlled visual stimuli while continuously recording binocular eye position. This permits assessment of neuroanatomic pathways involved in gaze holding, saccade generation, smooth pursuit, vergence, and pupillary responses. Reaction-time responses were recorded with a hand-held response pad. The standardized battery was designed to sample both lower-order and higher-order functions while remaining brief enough for use in participants with neurologic disease.

Vestibular-related components included horizontal and vertical gaze testing and spontaneous nystagmus assessment with and without visual fixation, which quantified nystagmus direction, slow-phase velocity, fast-phase events, and square-wave jerks. Horizontal optokinetic testing at 20 degrees/s and 60 degrees/s measured slow-phase eye velocity, gain, and directional asymmetry. Subjective visual vertical and horizontal assessed error in perceived line orientation relative to true vertical or horizontal. Additional tests included random horizontal and vertical saccades, predictive and self-paced saccades, antisaccades, memory-guided saccades, horizontal and vertical smooth pursuit at different frequencies, vergence pursuit and step responses, visual and auditory reaction time, and pupillary light reflex responses to sequential flashes of increasing intensity.

Clinical severity measures

PD severity was characterized with the modified H&Y scale (Hoehn and Yahr 1967). MDS-UPDRS Part III quantified motor examination findings; Parts I and II characterized non-motor and motor experiences of daily living, respectively (Goetz et al. 2008). Disease-severity analyses were cross-sectional. Because only five participants were H&Y 1–1.5, moderate (H&Y 2–2.5; n = 24) and severe (H&Y 3–4; n = 13) groups were compared for stage analyses.

Statistical analysis

To determine whether differences between group means were statistically significant a Student's t-test was applied under a two-tailed hypothesis. Differences were considered significant at $p \leq 0.05$. To calculate the corresponding statistics, the TTEST2 function of MATLAB software (The MathWorks, Inc. USA, version R2015b) is used, assuming unknown and unequal variances (known as the Behrens-Fisher problem) with the Satterthwaite approximation for the effective degrees of freedom. For metrics where normality was questionable or sample sizes were small, we also applied the non-parametric Wilcoxon rank-sum test (equivalent to the Mann-Whitney U-test) to assess whether two independent samples came from the same population.

To evaluate the predictive power of each OVRT-C metric while controlling for age and gender, univariate logistic regression models (logit link) were fitted using the MATLAB FITGLM function with maximum likelihood estimation. PD was coded as 1 and HC as 0, with males coded as 1 and females as 0. Regression coefficients and corresponding p-values were obtained (significant at $p \leq 0.05$). Diagnostic performance was assessed using the area under the receiver operating characteristic curve (AUC) and Somers' D statistic ($= 2 \cdot AUC - 1$), which provides an ordinal measure of association between observed and predicted outcomes.

A logistic regression model combining multiple OVRT-C metrics was developed in two steps: first, candidate metrics were selected based on prior findings; second, a stepwise procedure identified the final metrics providing the best model fit. Multicollinearity among correlated metrics was addressed by model extension and selection of a parsimonious set of predictors. Model performance was evaluated using AUC, Somers' D, sensitivity (true positive rate), specificity (true negative rate), and 95% confidence intervals with a cutoff probability of 0.5. Leave-one-out cross-validation (LOOCV) was performed to examine potential overfitting. This multivariable model provides a basic diagnostic tool for distinguishing PD from HC subjects.

Due to the large number of correlated OVRT-C metrics (47 selected for analysis), factor analysis was used to reduce dimensionality and group metrics into unobserved factors. Metrics were selected based on univariate regression results, significance, and AUC contributions. Only one metric was retained for measures assessed in both eyes and directions. The Varimax method was applied for orthogonal factor rotation using the MATLAB FACTORAN function. Two to three uncorrelated factors were then used as predictors in the multiple logistic regression model.

Beyond distinguishing PD from HC subjects, we identified OVRT-C metrics associated with disease severity. Univariate logistic regression models were applied to determine metrics that differentiate between moderate and severe PD cases, as classified by the H&Y scale. Figure 1 shows the distribution of PD patients according to their H&Y stage.

Results

Participants and individual OVRT-C metrics

Table 2 presents the mean and median OVRT-C metrics for PD and HC subjects, with p-values from Student's t-tests and Wilcoxon rank-sum tests. Several OVRT-C measures differed significantly between PD and HC subjects, including measures of saccadic accuracy, fixation stability, smooth pursuit, vergence, and predictive and antisaccadic performance. Extended results are provided in Supplementary Tables S.1 and S.2.

Table 2
Comparison of OVRT-C Metrics of PD and HC Subjects using Descriptive Statistics and Student's t and Wilcoxon Rank-Sum (WRS) Tests for Two Independent Samples

Tests	Metrics	PD				HC				t-test	WRS
		N	Med	Mean	SD	N	Med	Mean	SD	p	p
SN horizontal	light off—number of SWJ	41	8.00	9.88	6.72	24	4.00	6.38	5.6	0.03	0.05
SN vertical	fixation—ASPV (abs)	41	0.00	0.36	0.52	24	0.00	0.10	0.3	0.01	0.01
	PSPV (abs)	41	0.00	0.67	1.04	24	0.00	0.14	0.3	0.00	0.01
	light off—number of SWJ	42	0.00	1.50	2.30	24	0.00	0.46	0.9	0.01	0.05
GH horizontal	light off—number of SWJ	42	6.00	8.07	5.64	24	5.00	5.21	3.4	0.01	0.03
GH vertical	light off—ASPV (abs)	42	0.28	0.47	0.82	24	0.00	0.12	0.3	0.01	0.00
	PSPV (abs)	42	0.91	1.14	1.33	24	0.00	0.16	0.3	0.00	0.00
	number of beats	42	3.00	3.29	3.22	24	0.00	0.71	1.6	0.00	0.00
GV vertical	fixation—number of beats	41	2.00	3.51	4.56	24	0.00	1.46	2.6	0.02	0.01
	light off—PSPV (abs)	41	0.91	1.60	1.82	23	0.00	0.75	1.4	0.04	0.01
	number of beats	41	4.00	4.37	4.56	24	0.00	1.04	1.7	0.00	0.00
Auditory RT	mean latency (msec)	40	238.7	243.7	50.4	24	200.6	220.0	39.9	0.04	0.06
Visual RT	mean latency (msec)	40	293.5	327.3	106.0	23	246.7	260.5	39.0	0.00	0.00
Saccade & RT	(saccadic) latency grand mean	40	0.22	0.23	0.07	24	0.26	0.26	0.05	0.15	0.03
(SRT)	(saccadic) accuracy grand mean	39	78.4	76.6	17.1	24	89.3	86.1	15.5	0.03	0.01
SVV	error mean (abs)	40	2.82	3.22	2.48	24	1.30	1.89	2.4	0.04	0.00
SVH	error mean (abs)	39	2.12	2.63	2.24	24	1.42	1.50	1.3	0.01	0.04
Vergence	eye correlation inward	42	-0.78	-0.23	0.84	24	-0.94	-0.8	0.4	0.00	0.03

Tests	Metrics	PD				HC				t-test	WRS
		N	Med	Mean	SD	N	Med	Mean	SD	p	p
pursuit (VP)	near point convergence angle	42	-1.84	-1.54	0.95	24	-2.29	-2.2	0.5	0.00	0.03
Vergence	time constant— left inward	41	0.58	0.69	0.33	23	0.42	0.48	0.1	0.00	0.02
step (VS)	gain— left inward	40	1.98	1.75	0.81	23	2.36	2.21	0.7	0.02	0.02
	mean inward correlation	41	0.25	0.17	0.78	23	-0.80	-0.61	0.5	0.00	0.00
	% of saccade— left inward	41	27.7	28.0	14.3	23	17.6	18.3	10.6	0.00	0.01
Saccade —	accuracy grand mean (%)	37	81.2	80.9	10.3	24	89.6	90.5	8.8	0.00	0.00
horizontal (SH)	accuracy % of undershoot — LL	40	50.0	49.0	31.6	23	14.3	24.6	25.4	0.00	0.00
Saccade —	latency grand mean (sec)	40	0.25	0.26	0.05	23	0.25	0.24	0.02	0.02	0.13
vertical	accuracy grand mean (%)	39	76.0	74.3	14.9	23	97.8	95.4	13.1	0.00	0.00
(SV)	accuracy % of undershoot — LD	41	37.5	42.7	28.5	23	12.5	14.0	19.1	0.00	0.00
SPH, 0.1Hz	velocity gain	40	0.91	0.88	0.15	24	0.98	0.96	0.07	0.00	0.01
	saccadic component (%)	41	37.1	37.9	14.1	24	34.3	30.0	12.2	0.02	0.04
SPV, 0.1Hz	velocity gain — up	41	0.84	0.81	0.19	24	0.95	0.9	0.1	0.00	0.01
	position gain	41	1.00	0.99	0.10	24	1.04	1.0	0.1	0.02	0.03
Self-Paced	saccades per second	38	1.52	1.44	0.46	24	1.78	1.67	0.54	0.10	0.04
Saccade	accuracy mean	38	0.56	0.59	0.32	24	0.90	0.81	0.28	0.01	0.00
Predictive	accuracy grand mean	41	66.0	65.4	15.4	24	80.6	80.7	13.5	0.00	0.00

Tests	Metrics	PD				HC				t-test	WRS
		N	Med	Mean	SD	N	Med	Mean	SD	p	p
saccades (PS)	predictive accuracy mean	37	0.48	0.51	0.27	24	0.67	0.69	0.19	0.00	0.00
Anti-	% of pro-saccade errors	42	71.4	68.0	22.0	24	53.6	50.8	27.8	0.01	0.01
saccades	latency grand mean (stim. disp.)	42	0.40	0.43	0.14	24	0.34	0.4	0.1	0.04	0.05
(AS)	accuracy grand mean (stim. disp.)	42	77.0	88.2	74.5	24	125.6	132.2	54.4	0.01	0.01
	anti-saccades with error: latency	42	0.33	0.39	0.19	22	0.29	0.3	0.1	0.02	0.06

Saccadic Abnormalities

PD patients demonstrated pronounced saccadic hypometria. In the vertical saccade (SV) test, the mean number of saccade with undershoot was: 14% in healthy controls and 42.7% in PD patients. The distribution of these measurements is illustrated in Fig. 2. Using a clinically meaningful threshold of > 20% undershoot: 70.7% of PD patients (29/41) exceeded this threshold vs 21.7% of healthy controls (5/23) exceeded this threshold. In addition to hypometria, PD patients exhibited reduced accuracy of both main and final saccades in horizontal and vertical tasks. Horizontal Saccade (SH) Test. Main saccade accuracy < 80% was observed in: 4.2% of healthy controls (1/24) vs 43.2% of PD patients (16/37). Vertical Saccade (SV) Test. Main saccade accuracy < 80% was observed in: 8.7% of healthy controls vs 56.4% of PD patients (Fig. 2).

Other oculomotor abnormalities

Smooth pursuit performance also differed significantly between participants with PD and healthy controls in both horizontal and vertical tracking tasks. A reduction in velocity gain - defined as the ratio of slow-phase eye velocity to target velocity - was observed in PD, particularly for rightward pursuit during the slow-speed (0.1 Hz) horizontal test and upward pursuit during the slow-speed (0.1 Hz) vertical test. In addition to reduced gain, the saccadic component during slow-speed horizontal smooth pursuit differed between cohorts, with participants with PD demonstrating greater reliance on corrective saccades. Predictive and antisaccade tasks likewise demonstrated abnormalities in accuracy, latency, and error rate, consistent with impairment across both lower- and higher-order oculomotor tasks.

Fixation-instability measures also differed. Nonzero peak slow-phase velocity during vertical gaze in darkness was present in 31/42 (73.8%) PD participants and 6/24 (25.0%) controls. Vertical-gaze peak slow-phase velocity in darkness was associated with PD status (beta = 2.168, $p < 0.001$; AUC 0.79; Somers' D 0.59).

Multivariable discrimination of established PD

Multivariable logistic regression was used to determine whether combinations of OVRT-C measures improved discrimination between participants with PD and healthy controls. Model development began with horizontal and vertical saccadic accuracy, followed by testing whether pupillary amplitude and nonredundant fixation-instability and predictive-saccade measures added additional information. Pupillary amplitude did not contribute independently when saccadic accuracy was included ($p = 0.10$).

Logistic regression analysis (Model 1) identified vertical main saccade accuracy as one of the strongest discriminative metrics between PD and healthy individuals: AUC = 0.88, Somers' D = 0.77. These findings indicate substantial impairment in saccadic amplitude scaling and motor programming in PD, particularly for vertical eye movements, and support saccadic accuracy as a potential biomarker for disease classification. The final parsimonious model retained vertical saccadic accuracy, vertical nystagmus fast-phase count during horizontal gaze in darkness, horizontal nystagmus fast-phase count during vertical gaze in darkness, and predictive-saccade accuracy (Table 3).

Table 3
Multiple Logistic Regression Models: PD vs. HC Subjects

Model 1	**Tests**	**Metrics**		**Estimate**	**p-value**	**N obs.**
	Intercept			**28.14**	**0.00**	
Saccade horizontal	**accuracy mean— RL**		**-0.18**	**0.02**		
Saccade vertical	**accuracy mean— LD**		**-0.12**	**0.00**	**60**	
Model 2	Intercept			25.85	0.01	
	Saccade horizontal	accuracy mean— RL		-0.16	0.05	
	Saccade vertical	accuracy mean— LD		-0.12	0.00	
	Pupillary light reflex	amplitude: max sens X (left eye)		0.09	0.10	59
Model 3	Intercept			50.70	0.02	
	Saccade horizontal	accuracy mean— RL		-0.23	0.06	
	Saccade vertical	accuracy mean— LD		-0.24	0.04	
	Gaze horizontal—vertical	light off— number of beats		0.94	0.01	
	Gaze vertical—horizontal	light off— number of beats		1.25	0.04	
	Predictive saccades	accuracy grand mean		-0.15	0.04	59
Model 4	Intercept			22.83	0.00	
	Saccade vertical	accuracy mean— LD		-0.17	0.01	
	Gaze horizontal—vertical	light off— number of beats		0.77	0.01	
	Gaze vertical—horizontal	light off— number of beats		0.83	0.04	
	Predictive saccades	accuracy grand mean		-0.14	0.02	63

† The table entries are maximum likelihood estimates of the regression coefficients with the corresponding P-values and the number of observations used.

The initial two-variable model used horizontal final saccade accuracy and vertical saccade accuracy and achieved an AUC of 0.907. Adding pupillary light-reflex amplitude produced only a small increase in discrimination and the pupillary term was not independently significant when the two saccade measures were present. A broader model then incorporated nonredundant gaze-in-darkness and predictive-saccade measures. Horizontal final saccade accuracy became marginal (p = 0.06) and was removed, yielding the final four-predictor model, with regression coefficients presented in Table 3 and diagnostic performance

summarized in Table 4. This sequence was used to make transparent how the larger set of significant univariate findings was reduced to a smaller multivariable model.

Table 4
Diagnostic Statistics for Multivariate Logistic Regression Models

	Statistics	All Observations			Leave-One-Out Cross-Validation		
		Estimate	95% CI		Estimate	95% CI	
			Lower	Upper		Lower	Upper
Model 1	AUC	0.907	0.779	0.965	0.868	0.706	0.947
	Somers' D	0.814	0.558	0.931	0.737	0.411	0.895
	Sensitivity	0.919	0.789	0.976	0.865	0.714	0.949
	Specificity	0.739	0.533	0.898	0.739	0.521	0.889
Model 2	AUC	0.925	0.804	0.977	0.874	0.718	0.953
	Somers' D	0.850	0.607	0.954	0.749	0.437	0.906
	Sensitivity	0.889	0.742	0.971	0.889	0.743	0.971
	Specificity	0.826	0.610	0.955	0.826	0.615	0.952
Model 3	AUC	0.983	0.931	0.997	0.896	0.735	0.972
	Somers' D	0.966	0.862	0.995	0.792	0.470	0.943
	Sensitivity	0.944	0.818	1.000	0.917	0.779	0.974
	Specificity	0.913	0.722	1.000	0.826	0.619	0.952
Model 4	AUC	0.975	0.917	0.995	0.940	0.850	0.980
	Somers' D	0.950	0.835	0.990	0.880	0.701	0.961
	Sensitivity	0.950	0.834	1.000	0.875	0.732	0.951
	Specificity	0.870	0.667	0.963	0.826	0.619	0.950

‡ Bootstrap method with 5000 replicates used to calculate the 95% confidence intervals (95% CI) using the perfcurve function in MATLAB.

The final four-predictor model (Model 4; Tables 3 and 4), which included vertical saccadic accuracy, gaze-evoked nystagmus measures during horizontal and vertical gaze in darkness, and predictive-saccade accuracy, had an AUC of 0.975 (95% CI 0.917–0.995), sensitivity of 0.950 (95% CI 0.834–1.000), and specificity of 0.870 (95% CI 0.667–0.963). With LOOCV, the AUC decreased to 0.940 (95% CI 0.850–0.980), sensitivity to 0.875 (95% CI 0.732–0.951), and specificity to 0.826 (95% CI 0.619–0.950). This reduction suggests that the apparent development sample performance is likely optimistic.

Cross-sectional disease severity

Saccadic accuracy clearly separated controls from PD but did not further separate moderate from severe PD. Other metrics were associated with cross-sectional severity. Severe PD was associated with longer horizontal saccade latency (beta = 24.5, p = 0.03; AUC 0.76; Somers' D 0.52), lower position and velocity gain during 0.1-Hz vertical smooth pursuit, and altered pupillary light reflex measures, including longer latency, greater constriction velocity, and reduced amplitude. These findings represent between-person associations with H&Y stage and should not be interpreted as longitudinal progression within individuals.

The distribution of H&Y stages provided representation across mild, moderate, and more advanced disease, although the small number of participants at H&Y 1-1.5 limited formal analysis of the earliest stages. Among the larger moderate and severe groups, the measures associated with stage were not simply the same measures that best separated PD from controls. This observation is clinically and biologically plausible: a metric may become abnormal relatively early and then show limited additional change, whereas another measure may track the broader burden of disease more closely. The present cross-sectional design cannot determine which of these patterns reflects true within-person progression, but it identifies candidate measures for longitudinal study.

The severity analysis identified a somewhat different pattern from the PD-versus-control comparison. Measures that strongly distinguished established PD from healthy aging did not necessarily distinguish moderate from more severe disease. Horizontal saccade latency increased with severity, while slow-speed vertical smooth-pursuit velocity and position gain decreased. Pupillary light-reflex measures were also associated with H&Y stage: participants in the more severe group demonstrated longer latency, increased constriction velocity at selected stimulus intensities, and reduced amplitude. This divergence suggests that metrics useful for detecting the presence of established disease may not be identical to those most informative about disease stage.

Discussion

This early-phase diagnostic-accuracy study found a broad pattern of oculomotor and pupillary abnormalities in established PD. The clearest signals involved saccadic hypometria and reduced accuracy, fixation instability/nystagmus in darkness, smooth-pursuit and vergence abnormalities, and pupillary responses. A parsimonious multivariable model strongly separated participants with known PD from known healthy controls, but its AUC decreased from 0.975 to 0.940 under LOOCV. Given the modest single-center sample and case-control design, the high discrimination should be regarded as hypothesis-generating rather than evidence of clinical diagnostic performance.

The relationship to previous reports is important when interpreting the current findings. Previous studies have described hypometric and delayed saccades (Shaikh and Ghasia 2019; Terao et al. 2013; White et al. 1983; Zhou et al. 2022; Pinkhardt et al. 2012; Waldthaler et al. 2019), impaired smooth pursuit (White et al. 1983; Frei 2021), abnormal antisaccades (Waldthaler et al. 2019; Waldthaler et al. 2021), and vergence dysfunction in PD (Hanuška et al. 2015). The horizontal and vertical saccadic accuracy abnormalities observed here have been shown independently in the literature (Shaikh and Ghasia 2019; Terao et al. 2013; Zhou et al. 2022; Waldthaler et al. 2019), but the present study differs in acquiring these measures together within a standardized multimodal battery. Rather than evaluating a single eye-movement task in isolation, the OVRT-C approach samples several related but physiologically distinct functions during one brief session. This design allowed direct comparison of which measures were most informative for PD-versus-control discrimination and which were more closely associated with cross-sectional severity.

The principal findings are consistent with the concept that PD produces a multidimensional oculomotor profile rather than a single disease-specific eye-movement abnormality. Saccadic accuracy and undershoot were among the most informative individual measures, but fixation-instability, predictive-saccade, smooth-pursuit, vergence, and pupillary measures also contributed to the overall pattern. Importantly, the final model combined measures sampling different physiologic functions rather than multiple variants of the same metric. This may explain why the combined model discriminated the two known groups more strongly than individual measures alone.

The saccadic findings are consistent with prior reports of hypometric and delayed eye movements in PD (Shaikh and Ghasia 2019; Terao et al. 2013; White et al. 1983; Zhou et al. 2022; Pinkhardt et al. 2012; Waldthaler et al. 2019). Dysfunction of cortical eye fields, superior colliculus, brainstem saccade generators, cerebellum, and basal-ganglia-thalamocortical loops provides a plausible substrate for these abnormalities (Shaikh and Ghasia 2019; Terao et al. 2013; White et al. 1983; Pinkhardt et al. 2012). Antisaccade and predictive-saccade abnormalities additionally implicate executive control mechanisms (Terao et al. 2013; Waldthaler et al. 2019; Waldthaler et al. 2021). Pupillary abnormalities in PD may reflect dysfunction within autonomic

pathways (Giza et al. 2011). The convergence of these measures supports OVRT-C as a tool for quantifying distributed network dysfunction rather than as a single-pathway test.

The oculomotor deficits observed in PD likely reflect dysfunction across a distributed network that includes the dorsolateral prefrontal cortex, frontal and parietal eye fields, visual cortex, superior colliculus, pontine and midbrain saccade generators, cerebellum, thalamus, and basal ganglia (Shaikh and Ghasia 2019; Terao et al. 2013; White et al. 1983; Pinkhardt et al. 2012). These structures interact through cortico-striatal-thalamic and brainstem pathways to regulate the timing, scaling, and inhibition of eye movements (Shaikh and Ghasia 2019; Terao et al. 2013; White et al. 1983). Dopamine depletion associated with degeneration of the substantia nigra can disrupt these circuits and produce abnormal saccadic patterns (Shaikh and Ghasia 2019; Terao et al. 2013; Pinkhardt et al. 2012). Early changes in reflexive saccades may arise in part from altered inhibitory basal-ganglia output to the superior colliculus, whereas abnormalities in predictive and antisaccade tasks additionally engage frontal executive networks (Terao et al. 2013; Waldthaler et al. 2019; Waldthaler et al. 2021).

The overlap between oculomotor-control circuits and cognitive and autonomic networks provides a plausible framework for the broader OVRT-C findings. Frontal cortical and brainstem systems that modulate saccades and smooth pursuit also participate in attention and executive control (Shaikh and Ghasia 2019; Terao et al. 2013; White et al. 1983; Pinkhardt et al. 2012; Waldthaler et al. 2019; Waldthaler et al. 2021), while pupillary responses are influenced by central autonomic pathways (Giza et al. 2011). Thus, the battery may be sensitive to network-level dysfunction distributed across several systems affected by PD. The present data do not establish the anatomic source of any individual abnormality, but the pattern is consistent with prior physiologic studies of eye movements and pupillometry in PD (Shaikh and Ghasia 2019; Terao et al. 2013; White et al. 1983; Zhou et al. 2022; Frei 2021; Hanuška et al. 2015; Giza et al. 2011; Pinkhardt et al. 2012; Waldthaler et al. 2019; Waldthaler et al. 2021).

Several measures were also associated with H&Y severity. Horizontal saccade latency, vertical smooth-pursuit gain, and pupillary responses differed between moderate and severe PD. Previous studies have similarly reported relationships between selected oculomotor abnormalities and motor or cognitive disease severity (Terao et al. 2013; Zhou et al. 2022; Waldthaler et al. 2019). Because the study was cross-sectional, these results indicate associations with severity, not sensitivity to progression or treatment response. Longitudinal testing is required to determine within-person responsiveness. The present design corresponds to an early phase of diagnostic research: it asks whether a test distinguishes people with established disease from people known to be healthy (Sackett and Haynes 2002).

It does not establish performance among patients presenting with undifferentiated parkinsonism, prodromal symptoms, or competing neurologic diagnoses. The very high development-sample AUC further argues for caution because case-control sampling, variable selection, and model fitting in a modest sample can inflate apparent performance. Independent prospective validation should therefore precede claims of clinical diagnostic or monitoring utility.

Clinical significance and future study

OVRT-C testing can be completed rapidly, in less than 15 minutes, and does not rely on speech-based responses or complex limb-motor assessments. The battery therefore provides a practical means of quantifying oculomotor, fixation-stability, pupillary, reaction-time, and related functions within a single assessment. The present findings support further investigation of these measures as candidate physiologic biomarkers for PD and cross-sectional disease severity. They do not, however, establish that OVRT-C improves diagnosis in clinical practice. Sensitivity to prodromal disease, longitudinal progression, treatment response, and differentiation of PD from atypical parkinsonian disorders must be established in appropriately designed prospective studies before routine clinical use can be recommended (Tolosa et al. 2021; Hughes et al. 1992; Sackett and Haynes 2002).

Longitudinal follow-up of the present cohort will help determine whether the identified metrics remain stable or change predictably with disease progression or treatment. Multicenter studies are needed to establish age-stratified reference ranges and validate the classification model in independent and more diverse populations. Future work should also evaluate OVRT-C

in patients presenting with undifferentiated parkinsonism rather than only established PD and healthy controls (Tolosa et al. 2021; Hughes et al. 1992). Combining OVRT-C metrics with molecular imaging, wearable gait sensors, or digital speech markers may be combined to generate a composite biosignature for differentiation of PD from atypical parkinsonian disorders.

Limitations and future directions

This study has several notable strengths. It used a comprehensive battery that captured oculomotor, vestibular-related, pupillary, reaction-time, and cognitive domains within a single brief session. The study also applied a statistical pipeline that included cross-validation and yielded high sensitivity and specificity for distinguishing patients with PD from healthy controls.

Several limitations should also be considered. The sample was modest and recruited at a single center, which may limit generalizability, and the control group was smaller than the PD group. The design was an early-phase diagnostic-accuracy study of individuals with known PD compared with individuals known to be healthy controls; therefore, both independent validation and later-phase diagnostic-accuracy studies are needed to better assess clinical utility (Sackett and Haynes 2002). Finally, the accuracy and feasibility of eye tracking in individuals with severe blepharospasm, pronounced tremor, or other conditions that interfere with stable eye recording require further evaluation.

Conclusions

OVRT-C metrics, particularly saccadic accuracy, fixation-instability/nystagmus measures, and pupillary responses, strongly discriminated participants with established PD from healthy controls and several measures were associated with cross-sectional H&Y severity. These findings support further investigation of OVRT-C metrics as candidate biomarkers. Prospective multicenter studies in clinically relevant populations, with prespecified models, calibration assessment, and external validation, are required before clinical diagnostic or monitoring use can be recommended.

Declarations

Competing Interests

Alex Kiderman, PhD, Alexandr Braverman, PhD, and Laura C. Gibson, PhD, are affiliated with Spryson America, Inc., the manufacturer of the goggle system. The other authors declare no competing interests.

Funding

This research was supported by National Institutes of Health Grants R01AG073100, S10OD034250-01, and P50NS123067.

Ethics approval

The study was performed in accordance with the Declaration of Helsinki and was approved by the University of Michigan Institutional Review Board (approval no. HUM00296601).

Informed consent

All participants provided written informed consent before participating in the study.

Author Contribution

D.L.M. contributed to conceptualization, data curation, formal analysis, methodology, validation, visualization, original manuscript drafting, and manuscript review and editing. A.K. contributed to conceptualization, methodology, resources, software, validation, and manuscript review and editing. A.B. contributed to formal analysis, data curation, methodology, and manuscript review and editing. L.C.G. contributed to methodology, validation, and manuscript review and editing. S.R., T.B., and

J.B. contributed to data curation, investigation, validation, and manuscript review and editing. G.C. contributed to data curation, formal analysis, investigation, methodology, validation, and manuscript review and editing. B.B. contributed to investigation, methodology, validation, and manuscript review and editing. K.K. contributed to conceptualization, methodology, and manuscript review and editing. N.I.B. contributed to conceptualization, funding acquisition, methodology, project administration, resources, supervision, visualization, original manuscript drafting, and manuscript review and editing. All authors reviewed and approved the final manuscript.

Data Availability

The datasets generated and analyzed during the current study are available from the senior author (Dr. Nicolaas I. Bohnen) upon reasonable request. The MATLAB code used for the statistical analyses is publicly accessible in a GitHub repository.

References

1. Coughlin DG, Irwin DJ (2023) Fluid- and biopsy-based biomarkers in Parkinson's disease. Neurotherapeutics 20:932–954. https://doi.org/10.1007/s13311-023-01379-z
2. Cummings JL, Fine MJ, Grachev ID, Jarecke CR, Johnson MK, Kuo PH et al (2014) Effective and efficient diagnosis of parkinsonism: the role of dopamine transporter SPECT imaging with ioflupane I-123 injection (DaTscan™). Am J Manag Care 20(5 Suppl):S97–S109
3. Frei K (2021) Abnormalities of smooth pursuit in Parkinson's disease: a systematic review. Clin Park Relat Disord 4:100085. https://doi.org/10.1016/j.prdoa.2020.100085
4. Giza E, Fotiou D, Bostantjopoulou S, Katsarou Z, Karlovasitou A (2011) Pupil light reflex in Parkinson's disease: evaluation with pupillometry. Int J Neurosci 121:37–43. https://doi.org/10.3109/00207454.2010.526730
5. Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P et al (2008) Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. Mov Disord 23:2129–2170. https://doi.org/10.1002/mds.22340
6. Hanuška J, Bonnet C, Rusz J, Sieger T, Jech R, Rivaud-Péchoux S et al (2015) Fast vergence eye movements are disrupted in Parkinson's disease: a video-oculography study. Parkinsonism Relat Disord 21:797–799. https://doi.org/10.1016/j.parkreldis.2015.04.014
7. Hoehn MM, Yahr MD (1967) Parkinsonism: onset, progression and mortality. Neurology 17:427–442. https://doi.org/10.1212/WNL.17.5.427
8. Hughes AJ, Daniel SE, Kilford L, Lees AJ (1992) Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. J Neurol Neurosurg Psychiatry 55:181–184. https://doi.org/10.1136/jnnp.55.3.181
9. Pinkhardt EH, Jürgens R, Lulé D, Heimrath J, Ludolph AC, Becker W et al (2012) Eye movement impairments in Parkinson's disease: possible role of extradopaminergic mechanisms. BMC Neurol 12:5. https://doi.org/10.1186/1471-2377-12-5
10. Poewe W, Seppi K, Tanner CM, Halliday GM, Brundin P, Volkmann J et al (2017) Parkinson disease. Nat Rev Dis Primers 3:17013. https://doi.org/10.1038/nrdp.2017.13
11. Sackett DL, Haynes RB (2002) The architecture of diagnostic research. BMJ 324:539–541. https://doi.org/10.1136/bmj.324.7336.539
12. Shaikh AG, Ghasia FF (2019) Hypometric saccades in Parkinson's disease: slow, maladaptive and hypometric. Prog Brain Res 249:81–94. https://doi.org/10.1016/bs.pbr.2019.05.001
13. Terao Y, Fukuda H, Ugawa Y, Hikosaka O (2013) New perspectives on the pathophysiology of Parkinson's disease as assessed by saccade performance: a clinical review. Clin Neurophysiol 124:1491–1506. https://doi.org/10.1016/j.clinph.2013.01.021
14. Tolosa E, Garrido A, Scholz SW, Poewe W (2021) Challenges in the diagnosis of Parkinson's disease. Lancet Neurol 20:385–397. https://doi.org/10.1016/S1474-4422(21)00030-2

15. Waldthaler J, Tsitsi P, Svenningsson P (2019) Vertical saccades and antisaccades: complementary markers for motor and cognitive impairment in Parkinson's disease. NPJ Parkinsons Dis 5:11. https://doi.org/10.1038/s41531-019-0083-7
16. Waldthaler J, Stock L, Student J, Sommerkorn J, Dowiasch S, Timmermann L et al (2021) Antisaccades in Parkinson's disease: a meta-analysis. Neuropsychol Rev 31:628–642. https://doi.org/10.1007/s11065-021-09489-1
17. White OB, Saint-Cyr JA, Tomlinson RD, Sharpe JA (1983) Ocular motor deficits in Parkinson's disease II. Control of the saccadic and smooth pursuit systems. Brain 106:571–587. https://doi.org/10.1093/brain/106.3.571
18. Wüllner U, Borghammer P, Choe CU, Csoti I, Falkenburger B, Gasser T et al (2023) The heterogeneity of Parkinson's disease. J Neural Transm (Vienna) 130:827–838. https://doi.org/10.1007/s00702-023-02635-4
19. Zhou MX, Wang Q, Lin Y, Xu Q, Wu L, Chen YJ et al (2022) Oculomotor impairments in de novo Parkinson's disease. Front Aging Neurosci 14:985679. https://doi.org/10.3389/fnagi.2022.985679