

<https://doi.org/10.1038/s41531-025-01079-9>

Towards scalable screening for the early detection of Parkinson's disease: validation of an iPad-based eye movement assessment system against a clinical-grade eye tracker



Jamie Koerner¹, Erin Zou², Jessica A. Karl³, Cynthia Poon³, Leo Verhagen Metman³, Charles G. Sodini^{1,4,5}, Vivienne Sze^{1,5,6}, Fabian J. David^{7,8} & Thomas Heldt^{1,4,6,8} ✉

Early detection and monitoring of Parkinson's disease (PD) remain challenging, highlighting the need for accessible, cost-effective tools. Saccadic eye movement abnormalities are promising noninvasive biomarkers for PD screening and monitoring. Here, we present an iPad-based system that uses a deep learning algorithm to extract saccade metrics and validate these metrics against the clinical-grade EyeLink 1000 Plus. Twenty-five participants (10 with PD, 15 controls) completed pro-saccade, anti-saccade, memory-guided-saccade, and self-generated-saccade tasks. Relative to the EyeLink, the iPad system achieved average subject-level errors of 2 ms for latency and 0.7° for amplitude in pro-, anti-, and memory-guided saccades, and 0.003 s⁻¹ for inter-saccadic rate and 1.6° for amplitude in self-generated saccades. A review of 22 studies on PD-related saccadic impairments established benchmarks for clinically meaningful effects. The iPad-based system meets or exceeds these benchmarks, supporting its use as a scalable and cost-effective tool for screening and monitoring PD.

Parkinson's disease (PD) is the fastest-growing and second most common neurodegenerative disease in the world^{1,2}. Despite advances in understanding the pathogenesis and treatment of PD, early detection remains a major challenge. By the time cardinal motor symptoms of resting tremors, muscle rigidity, and bradykinesia appear, a large fraction of dopaminergic neurons are already lost, with estimates ranging from 30 to 70%^{3–6}. It is conceivable that by the time PD is currently detectable, neuronal degeneration has already surpassed a critical threshold—potentially explaining the failure of all disease-modifying therapies tested to date. Early detection is therefore essential for identifying cohorts more likely to benefit from such therapies.

At present, there are no definitive early detection tools. The tools proposed for early detection are the same tools used to confirm a diagnosis of PD, such as neuroimaging^{7–10}, skin biopsies^{7,8,10}, and spinal fluid assays^{7–10}. While these methods are invaluable to confirm a diagnosis of PD, their

utility for early detection is less than ideal. The primary issue is that they cannot be applied on a scale—i.e., for large-scale or population-level screening—needed for effective early detection, given that they are expensive, invasive, time-consuming, and require access to specialized equipment, assays, and trained personnel. A promising alternative involves the quantitative assessment of eye movements.

The complex neural circuits that govern eye movements—encompassing brainstem nuclei, the basal ganglia, cerebellum, and cortical regions—are well studied¹¹ and are known to be disrupted in PD¹². These abnormalities can be quantified noninvasively using video-oculography¹³. In PD, saccadic movements (rapid shifts of the eye to refocus gaze) are particularly affected: patients often exhibit impaired saccade initiation and reduced saccade amplitude, especially when saccades are generated volitionally (internally cued) rather than reflexively (externally cued)^{14,15}. Notably, abnormalities can

¹Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA, USA. ²Chicago College of Osteopathic Medicine, Midwestern University, Downers Grove, IL, USA. ³Department of Neurology, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA. ⁴Institute for Medical Engineering and Science, MIT, Cambridge, MA, USA. ⁵Microsystems Technology Laboratories, MIT, Cambridge, MA, USA. ⁶MIT Research Laboratory of Electronics, MIT, Cambridge, MA, USA. ⁷Department of Physical Therapy and Human Movement Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA. ⁸These authors contributed equally: Fabian J. David, Thomas Heldt ✉e-mail: thomas@mit.edu



emerge early in the disease^{14,15}, potentially aiding with early detection. Moreover, different types of abnormalities may reflect pathology in specific brain regions, suggesting that they could help differentiate PD from other neurodegenerative disorders¹⁶. Finally, some abnormalities have been shown to correlate with overall disease progression^{14,17}, making them potentially beneficial for monitoring disease progression¹⁴.

In laboratory settings, researchers typically use high-end eye-tracking devices to measure eye movements. During testing, participants perform specific visual tasks while an eye tracking system—such as the EyeLink®1000 Plus—records their gaze. These high-end trackers achieve high precision and accuracy by employing infrared illumination (Infrared illumination offers several benefits for eye tracking, such as creating a starker contrast between the iris and pupil compared to visible light.) and high-speed, high-resolution cameras. While they are less invasive and more feasible for repeated assessments than other diagnostic tools, such as neuroimaging, skin biopsies, or lumbar punctures (to collect cerebrospinal fluid), they remain bulky, expensive, and require considerable expertise to operate, limiting their clinical use.

To address the limitations of traditional laboratory-based systems, consumer-grade devices—such as smartphones, tablets, and virtual reality (VR) headsets—have emerged as more accessible platforms for eye movement assessment. For instance, BrainEye¹⁸ employs a smartphone-based app to display stimuli and record eye movements for neurological screening, while NeuroSync¹⁹ utilizes a VR-based system for concussion diagnosis. These devices are portable, relatively inexpensive, and user-friendly, making them ideal for large-scale screening for early detection and at-home monitoring of disease progression. However, smartphone screens are too small to allow the large eye movements necessary to elicit abnormalities, and VR headsets are not always intuitive for patients and clinicians. Tablets like the iPad overcome both of these limitations and are therefore our preferred device. In our previous work²⁰, we developed an iPad app that displays visual stimuli and simultaneously records participants using the front-facing camera to analyze eye movements. In that study, we collected over 235,000 eye movements from 80 participants, demonstrating the feasibility of our approach for assessing neurological disorders. Notably, none of these consumer-grade approaches have yet demonstrated concurrent validity with clinical-grade laboratory-based eye trackers like the EyeLink.

Consumer devices typically have limited frame rates (e.g., 60 frames per second (fps) for front-facing cameras) compared to systems like the EyeLink 1000 Plus (≥ 1000 fps). Additionally, the absence of infrared illumination and lower camera resolution may degrade the quality of gaze estimates. Such inaccuracies could obscure genuine PD-related abnormalities or progression trends, underscoring the need for validation studies before these tools are adopted clinically.

Our work makes two key contributions. First, we introduce and validate an iPad-based system for eye movement assessment by directly comparing it with the clinical-grade EyeLink 1000 Plus. Our approach leverages an iPad app that displays visual stimuli while simultaneously recording participants' faces using the front-facing camera, coupled with a deep learning model for gaze estimation. To assess the system's validity, participants complete a battery of saccade tasks (pro-saccade, anti-saccade, memory-guided saccade, self-generated saccade) while their eye movements are recorded simultaneously by both the iPad and the EyeLink. We then compare temporal and spatial eye movement measures extracted from the recordings to quantify the iPad's measurement error relative to the EyeLink at the saccade level, subject level, and by disease status. Second, we contextualize our findings by reviewing the literature on expected saccade changes in PD (relative to healthy controls) across various disease stages, thereby estimating the measurement accuracy required for the iPad to be effective for eye movement assessment in PD. The iPad's accuracy and precision—particularly in its temporal measurements—suggest that it is an effective tool for early PD screening and tracking disease progression.

Results

Saccade-level iPad performance

To assess the iPad's validity, we first compare measurement differences between the EyeLink and iPad on a saccade-by-saccade basis. Figure 1a displays Bland-Altman plots and histograms for the latency and amplitude measures, pooling data from the pro-saccade (PS), anti-saccade (AS), and memory-guided-saccade (MGS) tasks and each participant, while Fig. 1b displays the same for the self-generated-saccade (SGS) task. The limits of agreement (black) in the Bland-Altman plots suggest that 95% of saccade-level differences for the PS, AS, and MGS tasks fall between -16.6 ms and 15.4 ms for latency and -2.6° to 3.5° for amplitude, and the means (purple) indicate that the iPad has an overall bias of -0.6 ms in latency and 0.5° in amplitude. For the SGS task, 95% of saccade-level differences fall between -0.124 s⁻¹ and 0.102 s⁻¹ for instantaneous primary saccade rate (IPSR) and -5.6° and 5.6° for amplitude. There are no apparent trends in the differences across the range of measurements for any of the measures. The histograms of the saccade-level differences are approximately normally distributed. Note that negative latencies indicate inhibitory errors in the MGS task (i.e., a primary saccade occurring after the peripheral stimulus appears but before the fixation target disappears).

Figure 2a displays saccade-level error magnitudes (i.e., the absolute value of the errors) for the PS, AS, and MGS tasks and each measure. Similar results are depicted in Fig. 2b for the SGS task. The data indicate that, for the PS, AS, and MGS tasks, the iPad's measurements deviate from those of the EyeLink by 5.2 – 6.2 ms for latency and 1.1 – 1.4° for amplitude on average. In the SGS task, the iPad's measurements deviate from those of the EyeLink by 0.033 s⁻¹ for IPSR and 1.8° for amplitude on average. Supplementary Figs. S2–S5 additionally depict saccade-level results for each individual participant, including a side-by-side comparison of EyeLink and iPad saccade-level measurements (Figs. S2 and S3) as well as the difference of these measurements (Figs. S4 and S5).

Subject-level iPad performance

Although saccade-level comparisons provide important insights, our primary interest lies in subject-level statistics because an individual's eye movements are generally evaluated using average measurements across the saccades in each task. Figure 3a displays subject-level error magnitudes—defined as the absolute value of the average of saccade-level differences—for the PS, AS, and MGS tasks and each measure, along with their means, standard deviations, the 95% confidence intervals (CI) of the means, and statistical significance. Similar results are depicted in Fig. 3b for the SGS task. Table 1 summarizes these results. The data indicate that, consistently across the PS, AS, and MGS tasks, the iPad's measurements deviate from those of the EyeLink by ~ 2.0 ms (95% CI: 1.0 – 3.0 ms) for latency and 0.7° (95% CI: 0.4 – 1.0°) for amplitude on average. Individual subjects' errors are distributed around these averages with a standard deviation of about 2.0 ms for latency and 0.6° for amplitude. In the SGS task, the iPad's measurements deviate from those of the EyeLink by ~ 0.003 s⁻¹ (95% CI: 0.002 – 0.005 s⁻¹) for IPSR and 1.6° (95% CI: 0.90 – 2.20°) for amplitude on average, with a standard deviation of about 0.004 s⁻¹ for IPSR and 1.5° for amplitude. These mean error magnitudes are statistically significant ($p < 0.001$); their clinical significance will be discussed in "Discussion". Supplementary Figs. S6 and S7 additionally present subject-level Bland-Altman plots for each task and measure.

iPad error convergence across trials

In the PS, AS, and MGS tasks, each subject completes 40 trials (see "Eye movement tasks"). A natural question that arises is how many trials are necessary for the iPad error to stabilize. The solid line in Fig. 4 shows the mean absolute deviation from the full subject-level error as a function of the number of trials used. Note that for each subject, trial-level (i.e., saccade-level) errors were averaged across tasks at each trial index. The vertical lines mark the number of trials required to reach 50% (blue) and 90% (green) convergence toward the full subject-level error. Convergence is defined relative to the error at trial 1, and quantified as $1 - \frac{E_n}{E_1}$, where E_n is the

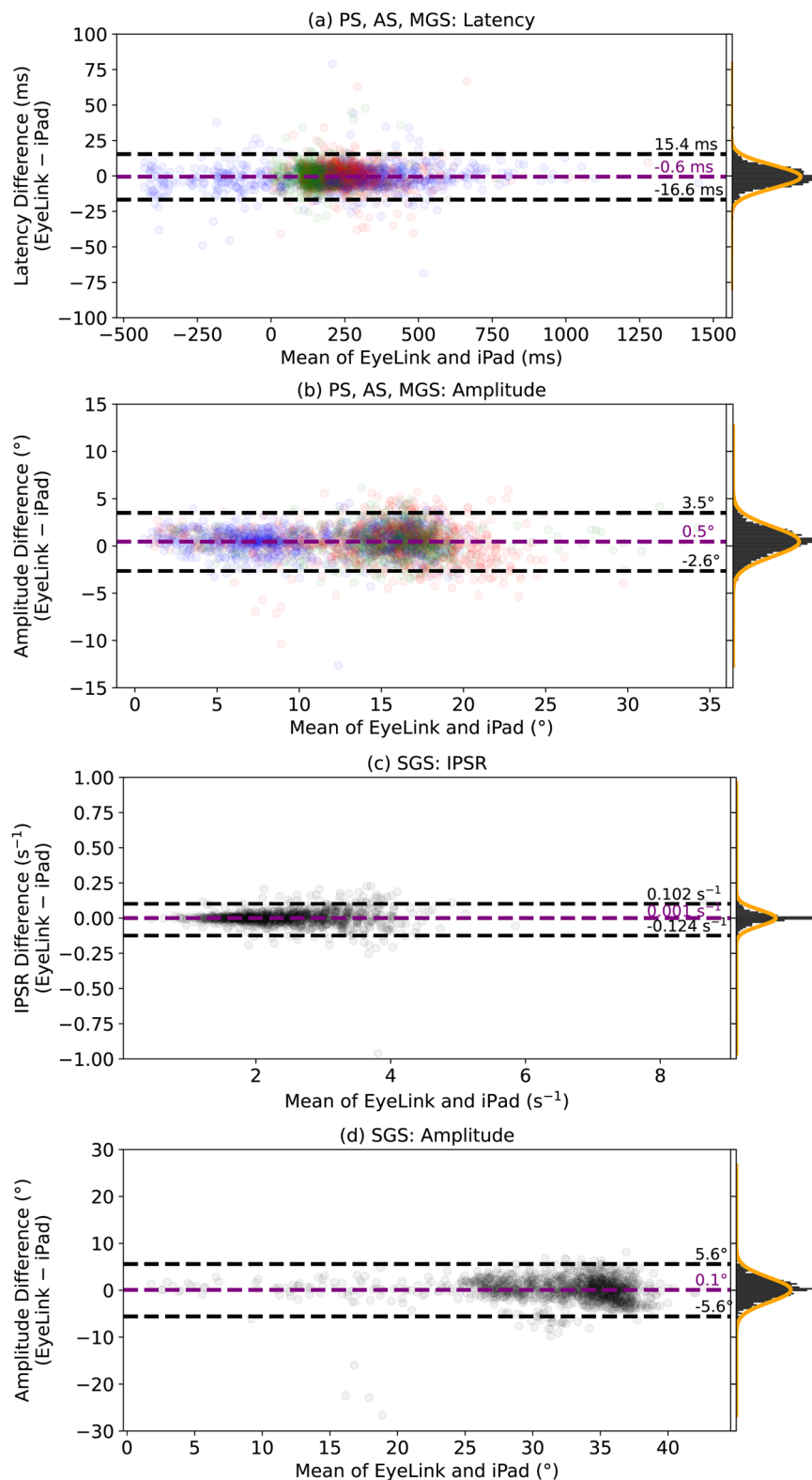


Fig. 1 | Bland-Altman plots and histograms of saccade-level differences (EyeLink – iPad). Data pooled across PS, AS, and MGS tasks and participants for **a** latency and **b** amplitude. Data pooled across participants for the SGS task for **c** IPSR (instantaneous primary saccade rate) and **d** amplitude. In all panels, the purple lines indicate the mean differences. The black lines, estimated using linear mixed-effects

models, represent the 95% limits of agreement. Orange lines show superimposed normal distributions. Note that negative latencies indicate inhibitory errors in the MGS task (i.e., a primary saccade occurring after the peripheral stimulus appears but before the fixation target disappears).

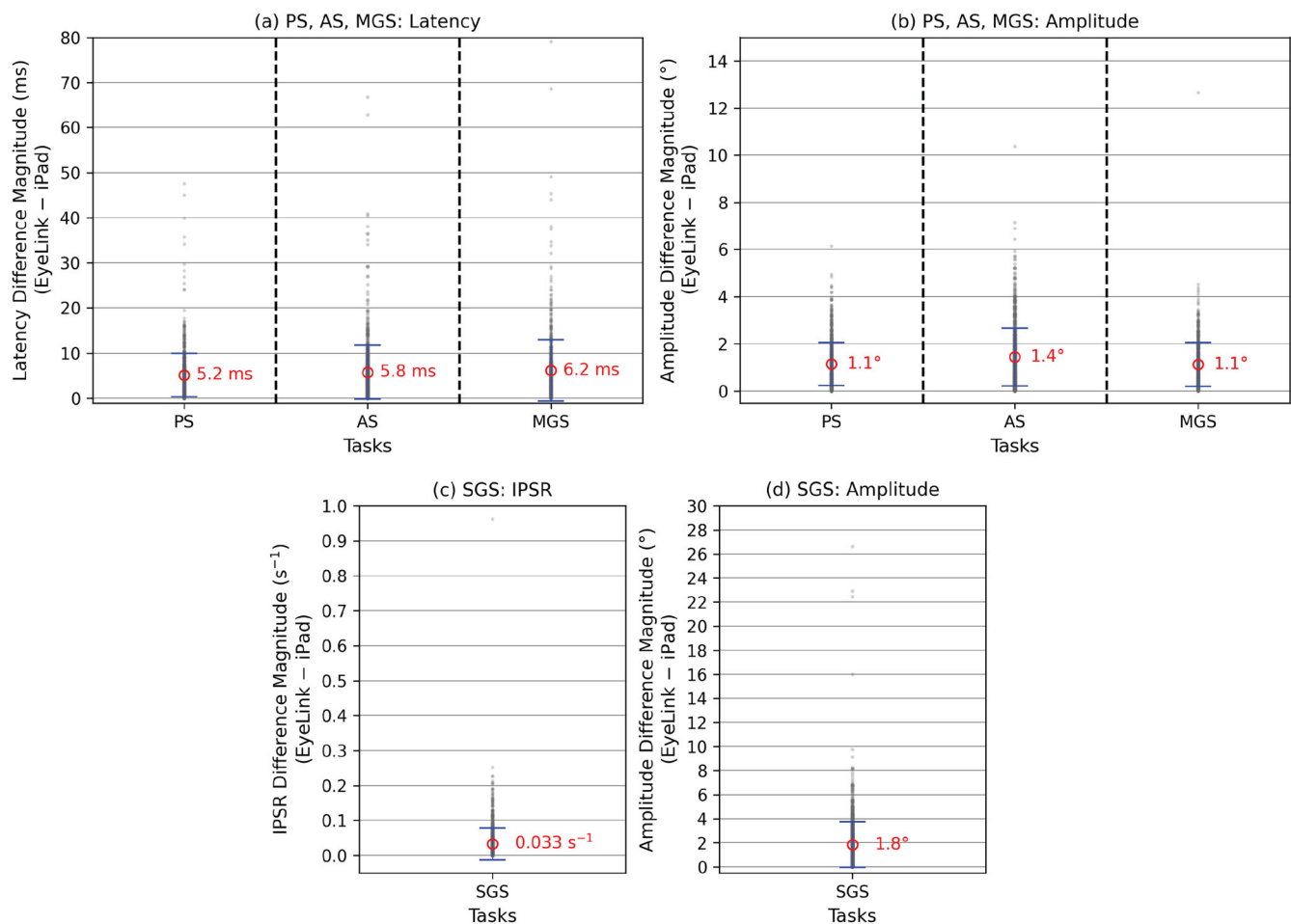


Fig. 2 | Saccade-level iPad error magnitudes (EyeLink – iPad). For the PS, AS, and MGS tasks (x-axis) for **a** latency and **b** amplitude. For the SGS task for **c** IPSR, and **d** amplitude.

average deviation when using the first n trials. Both measures require only 2–3 trials to reach 50% convergence. The iPad's latency error reaches 90% convergence after 18 trials, whereas the amplitude error requires 33 trials.

iPad performance by disease status

Another critical question is whether the iPad's performance differs between healthy control (HC) and PD groups. Variations could arise from factors such as involuntary head movements, tremors, or abnormal pupil dynamics—all common in PD^{21,22}. Figure 5a compares subject-level error magnitudes (pooled across the PS, AS, and MGS tasks) by disease status. Similar results are depicted in Fig. 5b for the SGS task. For all metrics, the iPad generally shows similar mean error magnitudes in both groups. Although differences in standard deviations are apparent for amplitude in the pooled PS, AS, and MGS tasks, and for IPSR and amplitude in the SGS task, these differences disappear after removing outlier subjects (one PD outlier near 0.0175 s^{-1} for IPSR, two HC outliers in the pooled PS, AS, and MGS tasks with amplitude errors of about 3° and one PD outlier in the SGS task with an amplitude error of almost 6.5°).

Expected saccade changes in PD

To fully evaluate the iPad's potential for early detection and disease monitoring, any measurement error must be considered in the context of the saccade changes typically observed in PD—both relative to healthy individuals and over the course of the disease. To this end, we reviewed the literature on saccade changes reported at various disease stages. It should be noted, however, that existing research does not yet clarify how early in PD

eye movements become affected and how these abnormalities evolve over time—limitations that necessarily constrain our analysis. Moreover, the saccadic system is inherently noisy, meaning PD-related impairments may take time to manifest regardless of the measurement method used. Recognizing these uncertainties, our goal is simply to examine the saccade changes reported in prior studies across different disease stages. In “Discussion”, we use this information to evaluate the clinical significance of the measurement errors introduced by the iPad. To review saccade changes in PS, AS, and MGS tasks, we focused on one reflexive saccade task (PS) and one volitional saccade task (AS), with an emphasis on AS given its common use in PD research. Because the SGS task differs substantially from the other tasks, we also reviewed studies reporting SGS results in PD, though it should be noted that, in contrast to the PS and AS tasks, few studies have examined the SGS task in this context.

We extracted relevant information from 20 studies^{16,23–41} that investigated saccades in individuals with PD and HC during PS or AS tasks (Figs. 6 and 7). These studies were selected from¹⁷, which conducted a meta-analysis of AS in PD. We included only those studies from¹⁷ that reported motor scores (UPDRS-III) for the PD subjects and results for at least one of the following measures: PS or AS latency, and PS or AS amplitude or gain. For studies measuring gain (saccade amplitude over target amplitude), we derived amplitude values using the average of the stated target amplitudes. Figure 6 displays the reported differences in saccade latency and amplitude between PD and HC participants, while Fig. 7 shows these differences against UPDRS-III scores and includes a fitted trend line. The UPDRS-III score for each study represents the average score across all participants

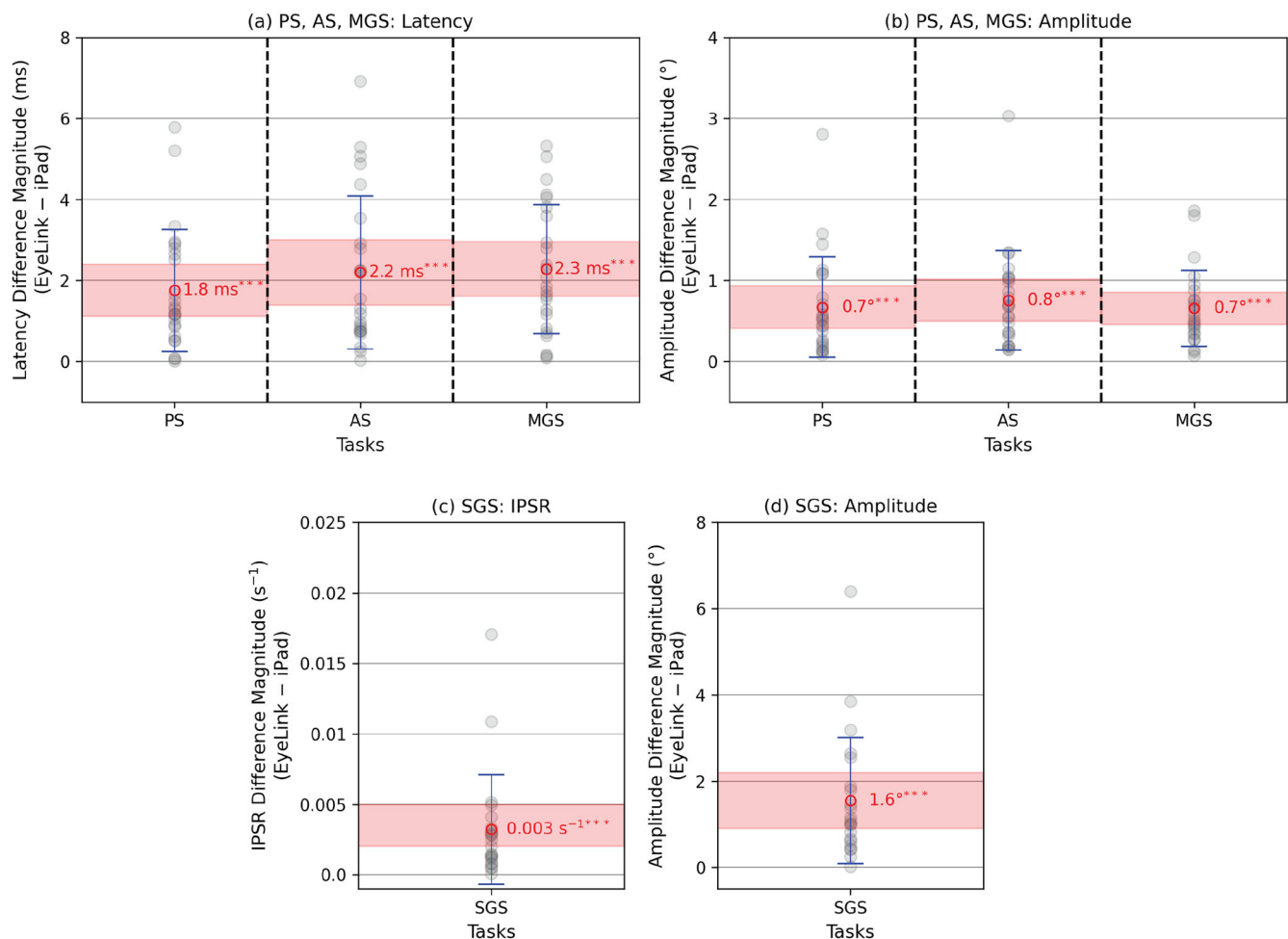


Fig. 3 | Subject-level iPad error magnitudes (EyeLink – iPad). For the PS, AS, and MGS tasks (x-axis) for **a** latency and **b** amplitude. For the SGS task for **c** IPSR, and **d** amplitude. Each sample represents the iPad's average error magnitude for one subject for that particular task (i.e., the absolute value of the average of the task's

saccade-level differences). The means and their CIs are highlighted in red. Standard deviations are shown in blue. Statistical significance (***) denotes $p < 0.001$ is based on one-sample t -tests when the Shapiro–Wilk test for normality was nonsignificant, and Wilcoxon signed-rank tests otherwise.

(within each subgroup or condition). Note that the same study may appear multiple times if it examined different PD subgroups (e.g., with or without dementia), conditions (e.g., on vs. off medication or with vs. without DBS), or task paradigms (e.g., gap vs. overlap). Statistically significant results ($p < 0.05$) are highlighted in red.

Studies capable of statistically differentiating PD and HC cohorts typically observed latency differences of at least 20 ms in the PS task and 60 ms in the AS task (Fig. 6). Similarly, statistically significant amplitude differences of at least -0.8° in the PS task and -1.8° in the AS task have been reported. Although Fig. 6 should be interpreted cautiously due to methodological variations among studies, it suggests that a latency error of only a few milliseconds in iPad measurements is likely acceptable—especially for AS, where PD–HC differences are larger—while even an amplitude error of a few tenths of a degree (e.g., 0.5°) could be problematic. Although we cannot define precise requirements for early PD screening, this analysis provides insight into the expected saccade changes in PD relative to HC, thereby offering a rough bound on the permissible error for the iPad to reliably distinguish between these groups.

Figure 7 illustrates how these PD–HC differences in eye movement measures vary with disease severity, as approximated by UPDRS–III scores. The fitted lines are consistent with the consensus that volitional saccades (AS) are more strongly affected by PD than reflexive saccades (PS) and that these impairments worsen over time. Taken at face value, the AS-task trend lines indicate that for each one-point increase in UPDRS–III score, latency

and amplitude differences between PD and HC increase on average by ~ 3.3 ms and 0.1° , respectively. Given that one longitudinal study⁴² found that UPDRS–III scores increase by an average of 2.40 points per year in de novo PD, these differences would grow by roughly 8 ms in latency and 0.24° in amplitude each year. These estimates provide a rough baseline for the permissible error if the iPad is to be used for tracking PD progression over time.

For the SGS task, we entered search terms (“Parkinson” OR “Parkinson’s”) AND (“saccades”) AND (“self-generated” OR “self-paced” OR “alternating”) into Google Scholar. We screened the first 40 search results for fulfillment of the following inclusion criteria: Full-length original research articles that reported results of a human PD group that performed the SGS task using any eye-tracking device. Only three studies were identified. One of these was excluded because the task instructions differed from the standard SGS task (subjects in that study were instructed to stay fixated on each stationary target for 1 s upon reaching it). Both of the two remaining studies^{43,44} reported temporal metrics (saccade rate or number of saccades) and spatial metrics (amplitude or gain) for the primary saccades. For studies reporting the number of primary saccades, we divided that number by the stated task duration to approximate IPSR. For studies reporting gain, we derived amplitude values using the stated target amplitudes. Figure 8 displays the reported differences in IPSR and amplitude between PD and HC participants. Note that the same study may appear multiple times if it examined different PD subgroups. Statistically significant results ($p < 0.05$) are highlighted in red.

The studies reported statistically significant differences between PD and HC cohorts of at least -0.47 s^{-1} in IPSR and at least -1.4° in amplitude (Fig. 8). Given the limited number of studies involving the SGS task, these findings should be interpreted with caution. Nevertheless, they suggest that an IPSR error in the iPad measurements of about 0.1 s^{-1} or less would likely be acceptable, whereas an amplitude error of just a few tenths of a degree (e.g., 0.5°) could already be problematic, similar to the AS task.

Discussion

The goal of this study was to evaluate the validity of an iPad-based eye movement assessment tool for PD. We compared the iPad’s temporal and spatial measurements to those of the clinical-grade reference EyeLink system. In interpreting our results, we draw on our review of the expected changes in PD, which estimated how eye movement measures differ between PD and HC groups and how these differences evolve over time (see “Expected saccade changes in PD”). When referencing our review, we

emphasize volitional saccades. As noted in “Introduction,” volitional saccades generally exhibit more pronounced abnormalities than reflexive saccades in PD, consistent with Figs. 6 and 7. A consistent finding is the impaired (slower) initiation and hypometria of volitional saccades^{14,15}. By contrast, reflexive saccades appear only subtly impaired¹⁵, suggesting that these movements are largely mediated by circuits less affected by the disease. Consequently, the quantitative assessment of volitional saccades offers a more promising avenue for the early detection of PD and the tracking of disease progression.

We found that 95% of saccade-level measurements from the iPad fall within $\pm 17\text{ ms}$ for latency—roughly the duration of a single frame at 60 fps—and within $\pm 0.1\text{ s}^{-1}$ for IPSR, relative to the EyeLink values. These saccade-level errors are consistent across the range of measured values and follow approximately normal distributions centered near zero, indicating that random positive and negative deviations tend to cancel out on average. As a result, averaging across saccades substantially reduces the error: the average subject-level error magnitude is only about 2 ms for latency and 0.003 s^{-1} for IPSR, with similarly low standard deviations. This represents a threefold reduction in error for latency (2 ms vs. 6 ms) and a tenfold reduction for IPSR (0.003 s^{-1} vs. 0.03 s^{-1}), compared to saccade-level error magnitudes. Although subjects completed 40 trials in the PS, AS, and MGS tasks, our convergence analysis shows that only about half that number is needed for the iPad’s latency error to reach 90% convergence toward the subject-level average. We also found that the iPad’s performance in measuring temporal metrics is consistent across both tasks and disease status.

Although our primary focus is on subject-level averages (i.e., across saccades), saccade-level data remain important for several applications. These include assessing within-subject variability, tracking trends across trials, and modeling saccadic dynamics such as the main sequence. In contrast, subject-level errors are more relevant for between-group comparisons, population-level research, or clinical assessments where a single summary metric per subject is required. Thus, the appropriate error level depends on the application: saccade-level precision may be critical for understanding fine-grained behavior or training dynamics, while subject-level accuracy supports reliable comparisons across groups or robust clinical evaluation of individuals.

Table 1 | Subject-level iPad error magnitudes

Latency (ms)				IPSR (s^{-1})
	PS	AS	MGS	SGS
Mean	1.76	2.20	2.28	0.003
SD	1.51	1.89	1.59	0.004
Mean CI 95%	[1.12, 2.39]	[1.40, 3.00]	[1.61, 2.95]	[0.002, 0.005]
P	<0.001 ^a	<0.001 ^a	<0.001	<0.001 ^a
N	24	24	24	22
Amplitude ($^\circ$)				Amplitude ($^\circ$)
	PS	AS	MGS	SGS
Mean	0.70	0.76	0.66	1.55
SD	0.62	0.62	0.47	1.47
Mean CI 95%	[0.41, 0.93]	[0.50, 1.02]	[0.46, 0.85]	[0.90, 2.20]
P	< 0.001 ^a	< 0.001 ^a	< 0.001 ^a	< 0.001 ^a
N	24	24	24	22

^aBased on the Wilcoxon signed-rank test.

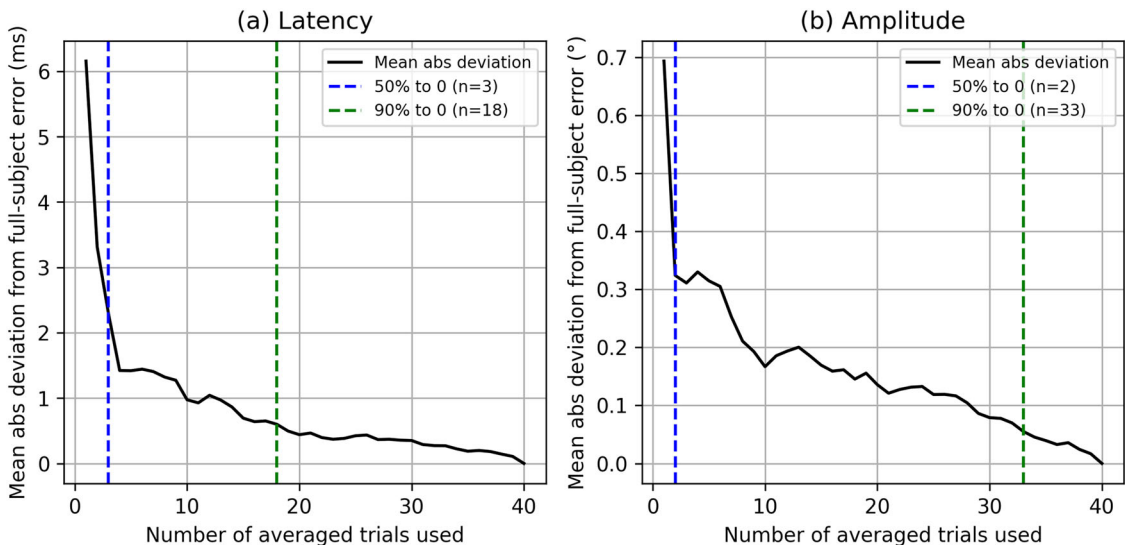


Fig. 4 | iPad error convergence across trials for the PS, AS, and MGS tasks. Convergence for **a** latency and **b** amplitude. For each subject and measure, trial-level (i.e., saccade-level) errors were first averaged across tasks at each trial index to produce a single per-trial error sequence. The curve shows the mean absolute deviation from the full subject-level error as a function of the number of trials used.

Vertical lines mark the trial counts at which the average error reached 50% (blue) and 90% (green) convergence toward the full subject-level error. Convergence was quantified as $1 - \frac{E_n}{E_1}$, where E_n is the mean deviation after n trials and E_1 is the deviation using only the first trial.

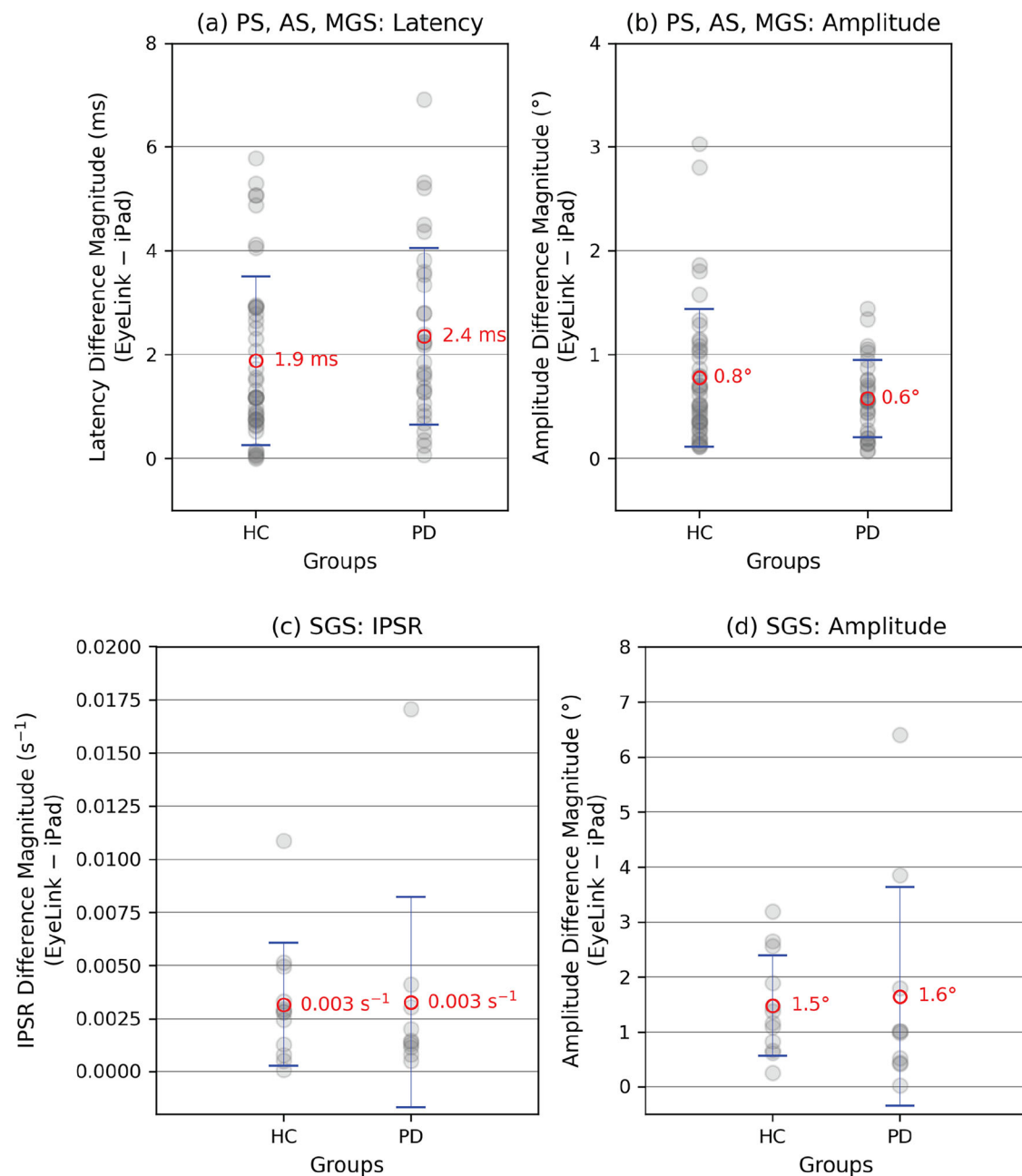


Fig. 5 | Subject-level iPad error magnitudes for the two subject groups (x-axis). Pooled across the PS, AS, and MGS tasks for **a** latency and **b** amplitude. For the SGS task for **c** IPSR and **d** amplitude. Each sample represents the iPad's average error

magnitude for one subject in that particular group. The group means and standard deviations are shown in red and blue, respectively.

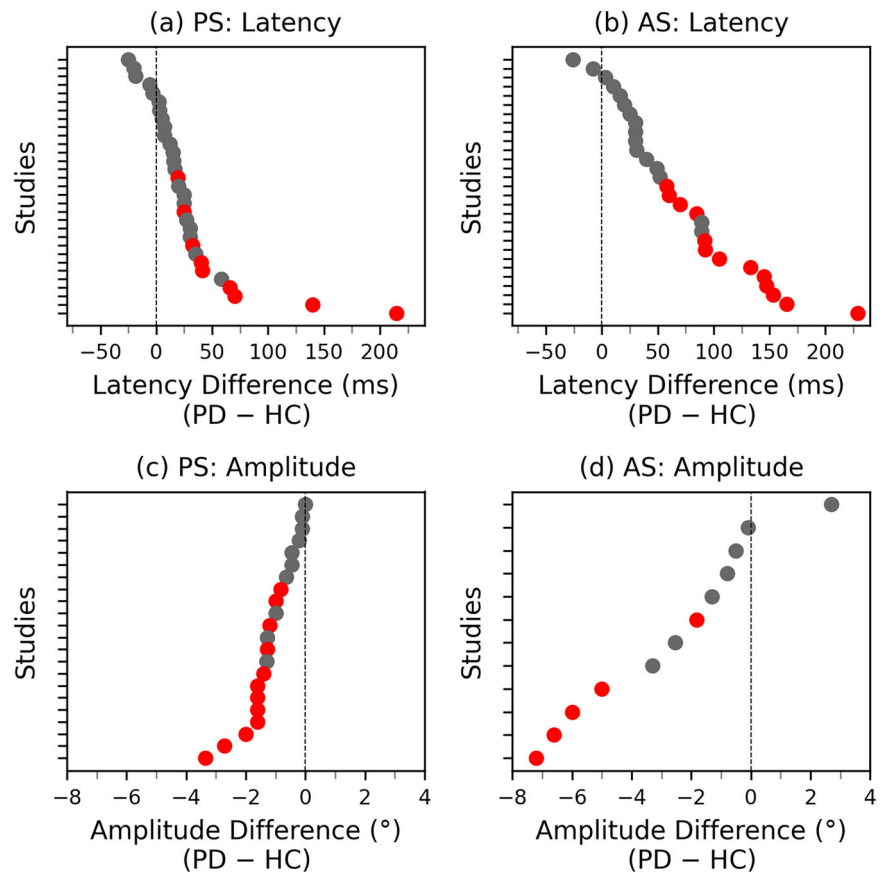
Because statistically significant latency impairments in PD are on the order of ≥ 60 ms in AS tasks (Expected Saccade Changes in PD) and IPSR impairments of $\geq 0.47 s^{-1}$ have been reported in the SGS task, the iPad appears capable of detecting such differences—especially when saccade-level measurements are averaged to reduce noise. Determining how early the iPad can detect PD remains challenging, as this depends on the timing and magnitude of temporal abnormalities in the disease process. Nonetheless, the iPad's subject-level accuracy and precision render it competitive with current clinical-grade eye trackers for early PD detection.

For monitoring disease progression, our literature review suggests that the latency gap between PD and HC grows by about 8 ms per year (Expected Saccade Changes in PD). In principle, the iPad may be sensitive enough to detect changes of this magnitude. Moreover, since latency continues to rise

with PD progression, the iPad's consistent performance across the entire latency range indicates that its measurements remain reliable over time. Although the limited number of studies on the SGS task precludes a detailed analysis of SGS changes across PD stages, the iPad's subject-level error and standard deviation ($\approx 0.003 s^{-1}$) are more than two orders of magnitude smaller than the smallest reported IPSR impairment ($\approx 0.47 s^{-1}$), suggesting that the iPad is likely capable of detecting SGS changes over time. Overall, the iPad's low cost, portability, and ease of use, along with its robust performance in temporal measurements, make it a promising tool for early detection and disease tracking in PD.

For amplitude, we found that 95% of saccade-level measurements from the iPad fall within -2.6° to 3.5° for the pooled PS, AS, and MGS tasks and $\pm 5.6^\circ$ for the SGS task of the EyeLink. The average saccade-level error

Fig. 6 | PD–HC (Parkinson’s disease–healthy control) differences extracted from prior work. For **a** PS latency, **b** AS latency, **c** PS amplitude, and **d** AS amplitude. Red indicates the difference is statistically significant ($p < 0.05$). The same study can appear multiple times if it examined different sub-groups, conditions, or task paradigms.



magnitude is about 1.0–1.5° for the PS, AS, and MGS tasks and 1.8° for the SGS task. Similar to the temporal measurements, these saccade-level errors partially cancel out when averaged. At the subject level, the mean error magnitude in the PS, AS, and MGS tasks is ~0.7°, with a similar standard deviation, consistent across tasks. In the SGS task, the mean and standard deviation of the subject-level error magnitudes are approximately equal to 1.6°. Importantly, the iPad’s performance in measuring amplitude is consistent across both tasks and disease status. Our convergence analysis suggests it takes considerably more trials (33) for the iPad’s amplitude error to converge to the subject-level error compared to latency (18).

The iPad’s performance in spatial measurements is not as strong as that in temporal measurements, given that statistically significant PD–HC amplitude differences as small as –1.8° in the AS task and –1.4° in the SGS task have been reported (see “Expected saccade changes in PD”). While the iPad may be capable of detecting such differences, it will likely struggle in early-stage PD or for fine-grained progression tracking, where changes could be as small as ~0.24° per year in the AS task (see “Expected saccade changes in PD”). Nonetheless, as amplitude abnormalities become more pronounced in later disease stages, the iPad’s amplitude measurements could still offer useful additional insight—especially over longer intervals when cumulative changes are more evident. Moreover, the iPad’s consistent performance across the entire amplitude range suggests it remains a viable tool even as hypometria worsens over time.

Despite these promising findings, several limitations should be noted. First, our experiments were conducted in a controlled laboratory environment, so it remains unclear how the iPad would perform in less-controlled clinical or home settings where variations in lighting, head movements, and patient support could compromise data quality. Second, although we postulate that the iPad could aid in early PD detection and tracking of disease progression, determining the precise measurement accuracy required for these applications is challenging due to methodological differences among

studies, a lack of data on eye movement abnormalities in the preclinical and prodromal stages of PD, and a scarcity of longitudinal research. Third, our evaluation focused on saccade tasks only and considered only one temporal and spatial measure per task; further validation using additional tasks, alternative measures, and assessments in other neurodegenerative disorders is needed to fully establish the iPad’s utility as an alternative to established eye tracking tools such as the EyeLink. Fourth, while our cross-sectional data allowed us to assess validity, we did not evaluate the iPad’s test-retest reliability. Establishing reliability is important because it demonstrates that repeated measurements yield consistent results, thereby ensuring that observed differences reflect true variation in the subject rather than measurement noise. Fifth, because we compared the iPad’s measurements solely to those of the EyeLink—and the EyeLink itself is not entirely error-free—the true accuracy of the iPad relative to an absolute ground truth remains somewhat uncertain. Finally, our study reflects only a snapshot of the iPad’s current capabilities; future improvements in hardware (e.g., higher-resolution, faster cameras) and software (e.g., more accurate gaze estimation models) may further reduce measurement errors and enhance performance.

Our findings demonstrate that the iPad achieves high accuracy and precision in measuring the temporal aspects of saccades, with subject-level errors well below the PD–healthy control differences reported in the literature. This performance makes the iPad well-suited for detecting temporal impairments in Parkinson’s disease (PD), even in its early stages, and for tracking these impairments over time. While spatial (amplitude) measurements are less accurate and precise, they may still be useful in later stages of PD, when hypometric saccades become more pronounced. Combined with its low cost, portability, and user-friendly interface, these results highlight the iPad’s promise as a scalable and accessible tool for clinical screening and remote monitoring of PD. Its ease of use makes it particularly well-suited for deployment in outpatient clinics, home settings, and

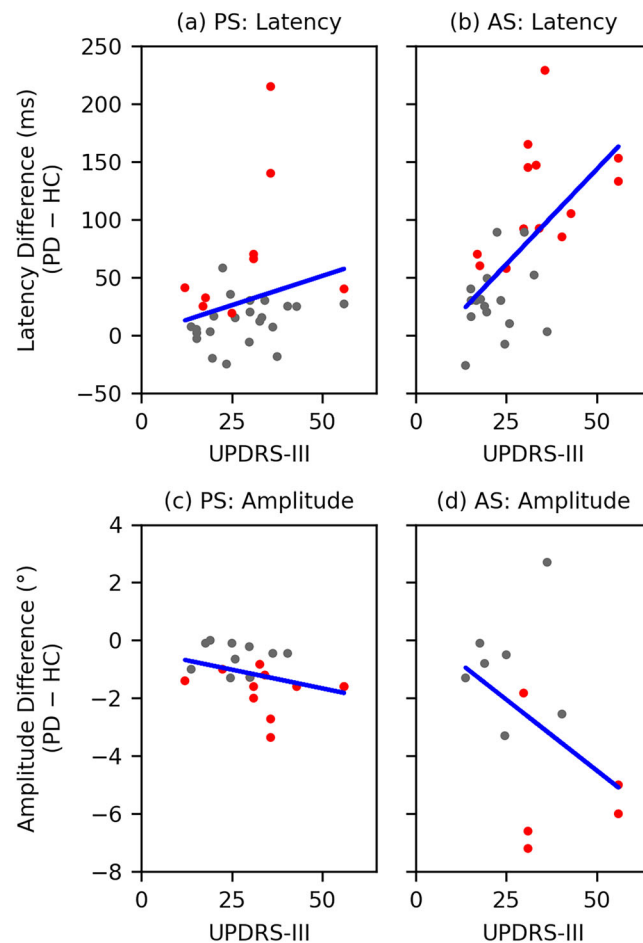


Fig. 7 | PD–HC differences extracted from prior work plotted against UPDRS-III scores. For **a** PS latency, **b** AS latency, **c** PS amplitude, and **d** AS amplitude. Each point represents a group average. The blue line represents the line of best fit. Red

indicates the difference is statistically significant ($p < 0.05$). The same study can appear multiple times if it examined different subgroups, conditions, or task paradigms.

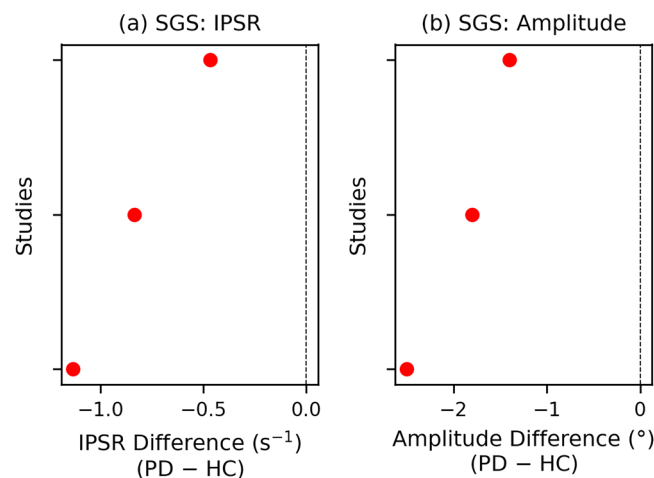


Fig. 8 | PD – HC differences extracted from prior work for the SGS task. For **a** IPSR and **b** amplitude. Red indicates the difference is statistically significant ($p < 0.05$). The same study can appear multiple times if it examined different PD subgroups.

large-scale screening programs. To build on this work, future studies should evaluate the iPad's performance in real-world environments, assess its test-retest reliability, and explore its applicability across a wider range of eye movement tasks and neurological conditions.

Continued improvements in hardware and gaze estimation algorithms could further reduce measurement error, strengthening the case for using portable consumer devices in the early diagnosis and ongoing monitoring of neurodegenerative diseases.

Methods

Experimental setup

Our setup included three main components: (1) an iPad, which serves as the primary device under evaluation; (2) the EyeLink 1000 Plus (SR Research), a clinical-grade arm-mounted eye tracking system; and (3) a separate computer to manage integration and synchronization between the two systems. Because our goal was to validate the iPad's performance, the iPad was central to the experiment. It controlled the visual stimuli—displayed on its own screen—and recorded participants via its front-facing camera.

Given the relatively low resolution of the iPad's front-facing camera, video quality is strongly affected by the distance between the participant and the device. In clinical or home settings, we generally position participants about 30 cm from the iPad. By contrast, the EyeLink camera operates optimally at a distance of 40–70 cm. Because the EyeLink camera was located beneath the iPad, we could accommodate both setups by seating participants in a height-adjustable chair with their eyes level to the iPad's center and their head resting on a chin rest for distance stabilization. This arrangement placed them ~35 cm from the iPad camera and ~55 cm from the EyeLink camera. Adjustments to the chair distance were made as needed to optimize EyeLink signal quality. The average distance to the iPad across participants was 34.2 cm, ranging from 31 to 40 cm.

Clinical-grade eye tracking system

We selected the EyeLink 1000 Plus as our reference system because it is widely recognized for its high accuracy and precision in video-based eye tracking^{45–48}, which makes it a reference system for research and clinical applications. Its core components include a high-resolution, high-speed infrared camera and an infrared illuminator. In remote mode—which allows for some head movement—the system is reported to achieve an accuracy of 0.25–0.5° and a precision of 0.05°, at a binocular sampling rate of 1000 Hz⁴⁹. Head movement is tracked via a sticker placed on the participant's forehead⁴⁹.

The EyeLink relies on a dedicated host computer that connects directly to the camera for data collection and hardware control. Typically, the monitor of a separate computer connected to the host computer serves as the display for visual stimuli. However, in this study, we substituted that monitor with our iPad. Before each eye movement task, we performed standard EyeLink procedures, including calibration and validation, which required participants to fixate on known targets on the iPad screen. During each task, the EyeLink calculates binocular gaze positions on the iPad screen in real time and stores them, along with timestamps, on the host computer.

iPad-based system

In our setup, the iPad serves as both the camera and display. We developed an iOS app (using Swift and SwiftUI) that presents the visual stimulus on the iPad screen, records participants via the front-facing RGB camera, and keeps the camera and display synchronized. During each eye movement task, the app saves the recorded video, corresponding stimulus coordinates, and nanosecond-resolution timestamps. The app also provides real-time feedback about a participant's posture and distance from the iPad, ensuring optimal video quality. Maintaining a controlled distance is crucial for standardization across tasks and participants because sitting farther away changes the visual angle required to reach the target. The distance is measured using the iPad's TrueDepth camera.

We employ a deep learning model⁵⁰ designed for unconstrained, “in-the-wild” environments. This model computes an average gaze vector for each video frame, which we map to screen coordinates using the iPad camera's intrinsic parameters⁵¹. Before each task, participants complete a brief calibration procedure, fixating on known points on the iPad screen to refine the model for their individual gaze patterns.

For this work, we used a 13-inch (7th-generation) iPad Pro, whose front-facing camera supports 1080p recording at 60 fps. We selected this iPad because its landscape-oriented front-facing camera allows the device to be used in landscape mode, maximizing screen width. At a 35 cm distance from the iPad, the display spans 41.4° of visual angle horizontally (about 21.2° to each

side of the center). Because most saccade studies focus on horizontal eye movements⁵², typically testing target amplitudes between 10° and 20°, this setup adequately accommodates established amplitude ranges in the literature.

Integration and synchronization of the two systems

We integrated and synchronized the iPad-based and EyeLink systems using WebLink (SR Research)—a software tool running on a separate computer (the WebLink computer)—and a communication link between WebLink and the iPad app. WebLink enables the EyeLink system to operate with an external display, allowing us to use the iPad screen for stimulus presentation while simultaneously recording with both the iPad's and EyeLink's cameras. WebLink also facilitates communication between the EyeLink host computer and external devices, permitting us to send timestamp data from the iPad to the host computer for synchronization.

To synchronize the two systems, our iPad app initiates a socket connection to WebLink at the start of each experiment. During the experiment, the app transmits the iPad's current timestamp on every display refresh cycle (~16 ms). WebLink receives this timestamp and records it—along with the EyeLink timestamp at the moment of receipt—on the host computer. The WebLink computer connects to both the host computer and the iPad via direct Ethernet to minimize communication delays. We measured the round-trip time between the iPad and host computer through the WebLink computer to be under 1 ms. By sending timestamps continuously rather than just at the beginning of the session, we correct for clock drift and can accurately determine the offset between the iPad and EyeLink clocks, ensuring synchronization of the two data streams.

Eye movement tasks

In this study, we consider four reflexive and volitional saccade tasks (Fig. 9): the Pro-Saccade (PS) task (reflexive)⁵³, the Anti-Saccade (AS) task (volitional)⁵³, the Memory-Guided-Saccade (MGS) task (volitional)⁵³, and the Self-Generated-Saccade (SGS) (or self-paced-saccade) task (volitional)⁵³. The PS and AS tasks both used the gap paradigm⁵³.

The PS task (Fig. 9a) elicits reflexive saccades through a suddenly appearing target stimulus. Each trial begins with a central fixation point that the subject must look at. A new stimulus then appears on either the left or right side of the screen, and the subject is asked to look at it as quickly and accurately as possible.

The AS task (Fig. 9b) requires response inhibition and volitional saccades. Similar to the PS task, each trial begins with a central fixation point, followed by a peripheral stimulus appearing on the left or right side of the screen. However, the subject must not look at this stimulus but instead direct their gaze away from it. Specifically, the subject is asked to look at the mirror-image location on the opposite side of the screen as quickly and accurately as possible.

The MGS task (Fig. 9c) demands response inhibition, spatial working memory, and volitional saccades. Each trial begins with the subject fixating on a central target. A peripheral stimulus then briefly flashes on one side of the screen, but the subject must maintain fixation on the central target. After a delay (retention period), the fixation target disappears, signaling the subject to look at the remembered location of the earlier flash (the saccade target) as quickly and accurately as possible.

The SGS task (Fig. 9d) demands repeated volitional saccades guided by internal cues, without external visual cues triggering their initiation. Specifically, subjects must alternate their gaze between two stationary targets on the screen as quickly as possible for a fixed period of time.

Participants

This study was approved by the Institutional Review Boards at both Northwestern University and MIT, and informed consent was obtained from all participants prior to participation. Twenty-seven individuals initially participated; however, two were excluded—one for repeatedly falling asleep and another for failing to produce a clean EyeLink gaze signal—resulting in 25 participants included in the analysis. For four of these participants, specific task recordings were excluded either due to unsaved

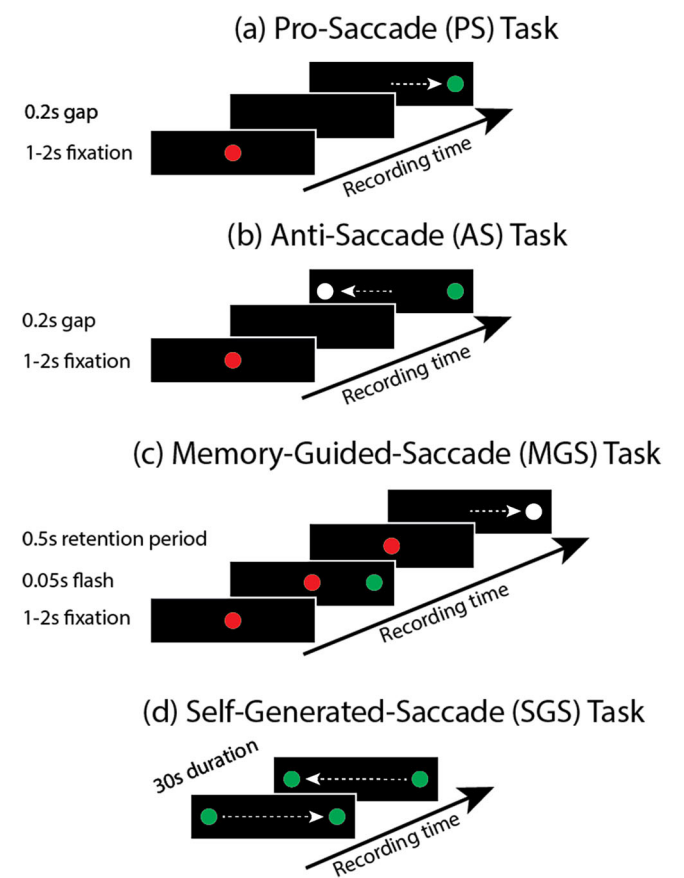


Fig. 9 | The four eye movement tasks considered in this study, along with their timing parameters. In the **a** PS, **b** AS, and **c** MGS tasks, the central fixation target is shown in red, the peripheral stimulus in green, and the saccade target is shown in white when it differs from the peripheral stimulus (AS and MGS tasks). In the **d** SGS task, the two continuously visible saccade targets are shown in green. The correct saccade direction is represented by the white arrow.

EyeLink data from technical issues or an inability to obtain a clean gaze signal. This led to the exclusion of one PS, AS, and MGS task recording, and three SGS task recordings. The final subject pool consisted of 10 individuals with PD and 15 healthy controls (HC). Including both HC and individuals with PD allowed us to examine whether disease status impacts the validity of iPad-based measurements.

PD participants were recruited from the Parkinson’s Disease & Movement Disorders Center at Northwestern University. Each participant was examined by a movement disorders neurologist and deemed eligible if they: (1) satisfied the United Kingdom PD Society Brain Bank clinical diagnostic criteria⁵⁴; (2) had normal or corrected-to-normal visual acuity; (3) presented without specific eye-related abnormalities (e.g., blepharospasms, double vision, eyelid-opening apraxia); (4) had no additional neurological comorbidities; and (5) could understand and complete the eye movement tasks during intake. Table 2 details the average demographics and relevant clinical information. HC participants met the same criteria, except that they reported no history of neurological disorders. Note that because the focus of this work is a comparison of the EyeLink and iPad (and not to diagnose or track disease severity), no further narrowing or characterization of the clinical cohort was performed.

Experiments

Each subject completed a single experiment consisting of a randomized sequence of the PS, AS, MGS, and SGS tasks. Each task was preceded by the calibration and validation procedures discussed in “Clinical-grade eye

Table 2 Subject demographics and clinical characteristics		
	HC	PD
Sex (M/F)	4/11	8/2
Age (mean ± SD, years)	59.5 ± 13.5	61.9 ± 5.4
MDS-UPDRS Part III (mean ± SD)	N/A	21.0 ± 8.5
Wearing Glasses (Y/N)	1/14	3/7

MDS-UPDRS Part III: Motor examination component of the Movement Disorders Society - Unified Parkinson’s disease rating scale.

tracking system” and “iPad-based system.” We used WebLink’s default 5-point calibration and validation scheme for external displays. The mean validation accuracy of the EyeLink system across subjects was 0.72°, ranging from 0.47° to 0.95°, with a standard deviation of 0.11°.

All stimuli were presented on a black background, and their shape—a combination of a bullseye and crosshair—has been shown to optimize fixation stability⁵⁵. The central fixation target appeared in red, while the peripheral stimulus, as well as the two stationary targets in the SGS task, appeared in green. Note that none of our participants reported red-green color blindness. Both the central fixation target and the peripheral stimulus had a diameter of 100 pixels, or about 1 cm, corresponding to a visual angle of about 1.6° at a viewing distance of 35 cm.

The PS, AS, and MGS tasks each consisted of 40 trials, with left and right stimulus positions randomized and counterbalanced. Figure 9 outlines the timing parameters for these tasks. The SGS task lasted 30 s. At 35 cm from the iPad, the saccade target amplitude was 17.2° in the PS and AS tasks, 34.4° in the SGS task, and 8.6° for the near target and 17.2° for the far target in the MGS task.

Eye movement measure extraction

Both the iPad and the EyeLink systems produce a gaze position signal for each eye movement task—a sequence of point-of-gaze pixel coordinates, one per video frame. On the iPad, this signal is automatically synchronized with the displayed stimulus because the video-frame timestamps and display-frame timestamps share the same clock. By contrast, the EyeLink signal requires post-hoc synchronization, achieved via the clock-offset calculations described in “Integration and synchronization of the two systems”.

We then convert the pixel coordinates for both point-of-gaze and stimulus into degrees of visual angle using distance measurements provided by the iPad when the participant’s head was on the chin rest. Figure 10 shows an example of raw iPad and EyeLink position signals for a typical PS trial. Both signals are subsequently filtered for noise reduction with a technique specifically designed for time series containing saccades and fixations⁵⁶.

We detect saccade onsets and offsets using the conventional 30%/s velocity threshold⁵³. The velocity signal is derived from the filtered position data using the central finite difference method.

To mitigate the iPad’s limited sampling rate (60 fps), we fit a curve to each detected saccade in the iPad position signal for the PS, AS, and MGS tasks using the saccade model formula introduced in⁵⁷, upsample the curve to 1000 Hz, and finally reapply velocity estimation and saccade detection. As shown in Supplementary Fig. S1, this curve-fitting approach improves the iPad’s latency estimates.

Eye movement measures aim to capture temporal and spatial features of eye movements to reveal potential abnormalities. Temporal metrics—such as saccade latency (i.e., visual reaction time) and saccade rate—assess impairments in saccade initiation, whereas spatial metrics—such as amplitude, gain, or accuracy—quantify hypometria (undershooting the target) or hypermetria (overshooting the target).

For each task, we compute temporal and spatial measures for the primary saccades. In the PS, AS, and MGS tasks, the primary saccade of a trial is defined as the first saccade with an amplitude exceeding 2° after the onset of the peripheral stimulus, to avoid capturing small fixational saccades or square wave jerks⁵⁸. In the SGS task, a primary saccade is defined as the

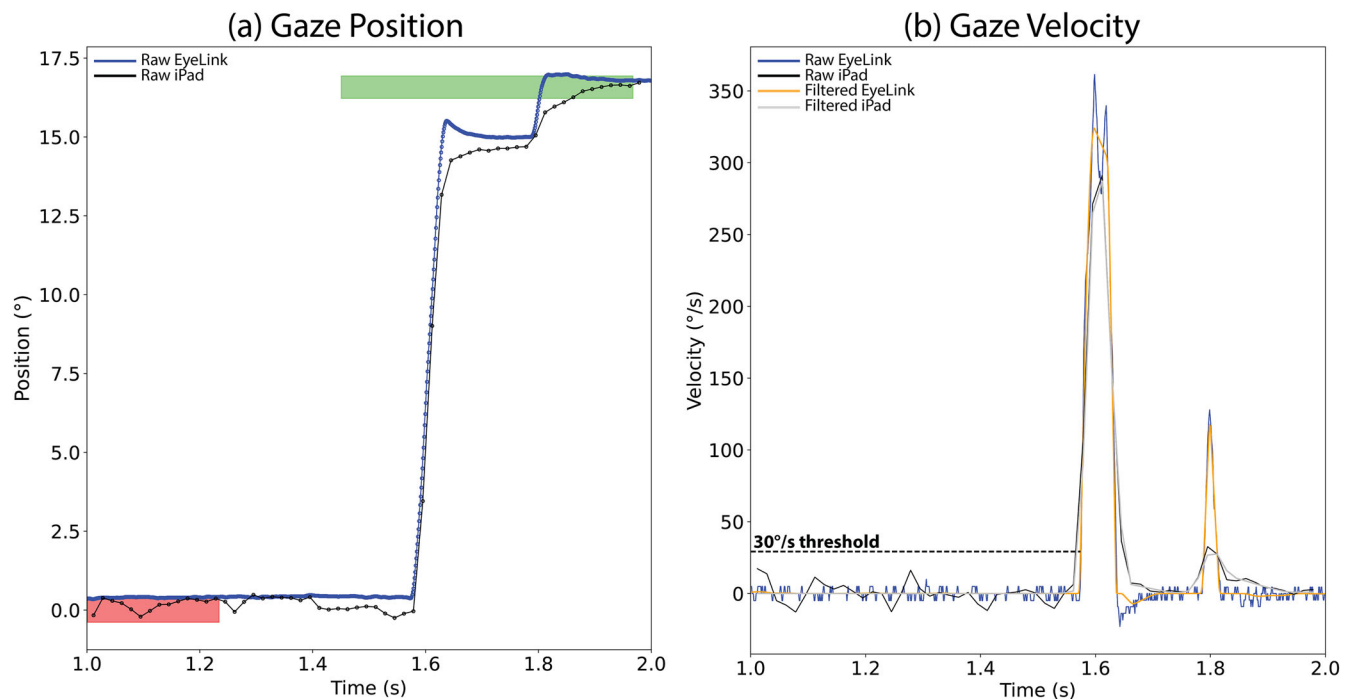


Fig. 10 | Gaze position and velocity of the EyeLink and iPad systems. **a** Gaze position and **b** the corresponding velocity signals produced by the iPad and EyeLink for a PS trial. The red and green horizontal bars represent the central fixation target and peripheral stimulus, respectively. Saccades are detected using a 30°/s velocity threshold.

first saccade with an amplitude exceeding 2° in the opposite direction of the previous primary saccade.

We compute latency for the PS, AS, and MGS tasks, defined as the time difference between the onset of the primary saccade and the stimulus onset. For the SGS task, we compute the instantaneous primary saccade rate (IPSR), defined as the reciprocal of the duration between two successive primary saccade onsets. We compute amplitude for each task, defined as the position difference in degrees of visual angle between the offset and onset of the primary saccade.

Trial and measurement exclusion criteria

We excluded any PS, AS, or MGS trial in which a participant blinked within 100 ms of the peripheral stimulus onset, blinked during the interval between the peripheral stimulus onset and the primary saccade (in the PS and AS tasks), or blinked between the fixation target offset (the cue) and the primary saccade (in the MGS task). In the SGS task, we excluded the IPSR measurement of the first primary saccade following a blink. Blink onsets and offsets were detected by calculating the eye aspect ratio⁵⁹ for each frame in the iPad's video recording, which provides a measure of eye closure. Additionally, PS, AS, or MGS trials were excluded if the participant did not produce a saccade or if signal loss in the EyeLink data rendered the primary saccade unusable. For the SGS task, if signal loss in the EyeLink data rendered a primary saccade unusable, the IPSR measurement of the subsequent primary saccade was excluded.

Statistical analysis

Statistical analysis was conducted at both the saccade and subject level, where subject level refers to averages computed over the saccade-level metrics of each individual. For Bland-Altman plots (Bland-Altman plots⁶⁰ are used to assess agreement between two measurement methods by plotting the difference between paired measurements against their mean, allowing visualization of systematic bias and limits of agreement.) of saccade-level differences between the EyeLink and iPad, we fit a linear mixed-effects model with a random intercept for subject to estimate the 95% limits of

agreement, accounting for repeated measures within subjects across all pooled saccades. The Bland-Altman plots are accompanied by histograms, which allow visual assessment of the normality of the differences. To assess the statistical significance of subject-level mean differences between the EyeLink and iPad, we used one-sample *t*-tests when the Shapiro-Wilk test for normality was nonsignificant ($p \geq 0.05$), and Wilcoxon signed-rank tests otherwise.

Data availability

The saccade metrics analyzed in this study are available from the corresponding author upon reasonable request.

Code availability

The code is not publicly available but may be made available to qualified researchers upon reasonable request.

Received: 30 April 2025; Accepted: 15 July 2025;

Published online: 08 August 2025

References

1. Dorsey, E. R., Sherer, T., Okun, M. S. & Bloem, B. R. The emerging evidence of the Parkinson pandemic. *J. Parkinson's Dis.* **8**, S3–S8 (2018).
2. Steinmetz, J. D. et al. Global, regional, and national burden of disorders affecting the nervous system, 1990–2021: a systematic analysis for the global burden of disease study 2021. *Lancet Neurol.* **23**, 344–381 (2024).
3. Cheng, H.-C., Ulane, C. M. & Burke, R. E. Clinical progression in parkinson disease and the neurobiology of axons. *Ann. Neurol.* **67**, 715–725 (2010).
4. Ross, G. W. et al. Parkinsonian signs and substantia nigra neuron density in decedents elders without PD. *Ann. Neurol.* **56**, 532–539 (2004).
5. Dauer, W. & Przedborski, S. Parkinson's disease: mechanisms and models. *Neuron* **39**, 889–909 (2003).

6. Lang, A. E. & Lozano, A. M. Parkinson's disease. *N. Engl. J. Med.* **339**, 1130–1143 (1998).
7. Tolosa, E., Garrido, A., Scholz, S. W. & Poewe, W. Challenges in the diagnosis of parkinson's disease. *Lancet Neurol.* **20**, 385–397 (2021).
8. Zarkali, A., Thomas, G. E., Zetterberg, H. & Weil, R. S. Neuroimaging and fluid biomarkers in parkinson's disease in an era of targeted interventions. *Nat. Commun.* **15**, 5661 (2024).
9. Yamashita, K. Y., Bhoopatiraju, S., Silverglate, B. D. & Grossberg, G. T. Biomarkers in parkinson's disease: a state of the art review. *Biomark. Neuropsychiatry* **9**, 100074 (2023).
10. Mahlknecht, P., Seppi, K. & Poewe, W. The concept of prodromal parkinson's disease. *J. Parkinson's. Dis.* **5**, 681–697 (2015).
11. Pouget, P. The cortex is in overall control of 'voluntary' eye movement. *Eye* **29**, 241–245 (2015).
12. Antoniadou, C. A. & Sperling, M. Eye movements in parkinson's disease: from neurophysiological mechanisms to diagnostic tools. *Trends Neurosci.* **47**, 71–83 (2024).
13. Jung, I. & Kim, J.-S. Abnormal eye movements in Parkinsonism and movement disorders. *J. Mov. Disord.* **12**, 1 (2019).
14. Anderson, T. J. & MacAskill, M. R. Eye movements in patients with neurodegenerative disorders. *Nat. Rev. Neurol.* **9**, 74–85 (2013).
15. Bronstein, A. M., Anderson, T., Kaski, D., MacAskill, M. & Shaikh, A. *Oculomotor and Visual-Vestibular Disturbances in Parkinson's Disease*, 115–129 (Cambridge University Press, Cambridge, 2022).
16. Waldthaler, J., Tsitsi, P. & Svenningsson, P. Vertical saccades and antisaccades: complementary markers for motor and cognitive impairment in parkinson's disease. *npj Parkinson's. Dis.* **5**, 11 (2019).
17. Waldthaler, J. et al. Antisaccades in parkinson's disease: a meta-analysis. *Neuropsychol. Rev.* **31**, 628–642 (2021).
18. Braineye. <https://braineye.com/>. Accessed: March 20, 2025.
19. Neurosync. <https://www.neurosync.health/>. Accessed: March 20, 2025.
20. Lai, H.-Y., Saavedra-Peña, G., Sodini, C. G., Heldt, T. & Sze, V. App-based saccade latency and directional error determination across the adult age spectrum. *IEEE Trans. Biomed. Eng.* **69**, 1029–1039 (2021).
21. Abusrair, A. H., Elsekaily, W. & Bohlega, S. Tremor in parkinson's disease: from pathophysiology to advanced therapies. *Tremor Other Hyperkinetic Mov.* **12**, 29 (2022).
22. Tsitsi, P. et al. Pupil light reflex dynamics in parkinson's disease. *Front. Integr. Neurosci.* **17**, 1249554 (2023).
23. Briand, K. A., Strallow, D., Hening, W., Poizner, H. & Sereno, A. B. Control of voluntary and reflexive saccades in parkinson's disease. *Exp. Brain Res.* **129**, 38–48 (1999).
24. Bonnet, C. et al. Eye movements in ephedrone-induced Parkinsonism. *PLoS ONE* **9**, e104784 (2014).
25. Barbosa, P. et al. Saccadic direction errors are associated with impulsive compulsive behaviours in parkinson's disease patients. *J. Parkinson's. Dis.* **9**, 625–630 (2019).
26. Cameron, I. G. et al. Impaired executive function signals in motor brain regions in parkinson's disease. *Neuroimage* **60**, 1156–1170 (2012).
27. Crevits, L., Versijpt, J., Hanse, M. & De Ridder, K. Antisaccadic effects of a dopamine agonist as add-on therapy in advanced parkinson's patients. *Neuropsychobiology* **42**, 202–206 (2000).
28. Gorges, M. et al. The association between alterations of eye movement control and cerebral intrinsic functional connectivity in parkinson's disease. *Brain Imaging Behav.* **10**, 79–91 (2016).
29. Harsay, H. A., Buitenveg, J. I., Wijnen, J. G., Guerreiro, M. J. & Ridderinkhof, K. R. Remedial effects of motivational incentive on declining cognitive control in healthy aging and parkinson's disease. *Front. Aging Neurosci.* **2**, 144 (2010).
30. Hanuška, J. et al. Eye movements in idiopathic rapid eye movement sleep behaviour disorder: high antisaccade error rate reflects prefrontal cortex dysfunction. *J. Sleep. Res.* **28**, e12742 (2019).
31. Hood, A. J. et al. Levodopa slows prosaccades and improves antisaccades: an eye movement study in parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* **78**, 565–570 (2007).
32. Van Koningsbruggen, M. G., Pender, T., Machado, L. & Rafal, R. D. Impaired control of the oculomotor reflexes in parkinson's disease. *Neuropsychologia* **47**, 2909–2915 (2009).
33. Lemos, J. et al. Distinct functional properties of the vertical and horizontal saccadic network in health and parkinson's disease: An eye-tracking and fMRI study. *Brain Res.* **1648**, 469–484 (2016).
34. Lu, Z., Buchanan, T., Kennard, C., FitzGerald, J. J. & Antoniadou, C. A. The effect of levodopa on saccades—Oxford quantification in Parkinsonism study. *Parkinsonism Relat. Disord.* **68**, 49–56 (2019).
35. Mosimann, U. P. et al. Saccadic eye movement changes in parkinson's disease dementia and dementia with lewy bodies. *Brain* **128**, 1267–1276 (2005).
36. Nemanich, S. T. & Earhart, G. M. Freezing of gait is associated with increased saccade latency and variability in parkinson's disease. *Clin. Neurophysiol.* **127**, 2394–2401 (2016).
37. Ranchet, M. et al. Pupillary response to cognitive workload during saccadic tasks in parkinson's disease. *Behav. Brain Res.* **327**, 162–166 (2017).
38. Rivaud-Pechoux, S., Vidailhet, M., Brandel, J. & Gaymard, B. Mixing pro- and antisaccades in patients with Parkinsonian syndromes. *Brain* **130**, 256–264 (2007).
39. Visser, F., Bour, L., Lee, Y., Ten Brinke, T. & Van Rootselaar, A. Eye movement abnormalities in essential tremor versus tremor dominant parkinson's disease. *Clin. Neurophysiol.* **130**, 683–691 (2019).
40. Walton, C. C. et al. Antisaccade errors reveal cognitive control deficits in parkinson's disease with freezing of gait. *J. Neurol.* **262**, 2745–2754 (2015).
41. Machner, B. et al. Eye movement disorders are different in parkinson-linked and idiopathic early-onset PD. *Neurology* **75**, 125–128 (2010).
42. Holden, S. K., Finseth, T., Sillau, S. H. & Berman, B. D. Progression of MDS-UPDRS scores over five years in de novo parkinson disease from the parkinson's Progression Markers Initiative cohort. *Mov. Disord. Clin. Pract.* **5**, 47–53 (2018).
43. DeJong, J. D. & Jones, G. M. Akinesia, hypokinesia, and bradykinesia in the oculomotor system of patients with parkinson's disease. *Exp. Neurol.* **32**, 58–68 (1971).
44. Pinkhardt, E. H. et al. Eye movement impairments in parkinson's disease: possible role of extradopaminergic mechanisms. *BMC Neurol.* **12**, 1–8 (2012).
45. Spitzer, L. & Mueller, S. Using a test battery to compare three remote, video-based eye-trackers. In *Proc. 2022 Symposium on Eye Tracking Research and Applications*, 1–7 (Association for Computing Machinery, 2022).
46. Ehinger, B. V., Groß, K., Ibs, I. & König, P. A new comprehensive eye-tracking test battery concurrently evaluating the pupil labs glasses and the Eyelink 1000. *PeerJ* **7**, e7086 (2019).
47. Dalmaijer, E. Is the low-cost eyetribe eye tracker any good for research? *Peer J. PrePrints* <https://doi.org/10.7287/peerj.preprints.585v1> (2014).
48. Aziz, S., Lohr, D. J. & Komogortsev, O. Synchroneyes: a novel, paired data set of eye movements recorded simultaneously with remote and wearable eye-tracking devices. In *Proc. 2022 Symposium on Eye Tracking Research and Applications*, 1–6 (Association for Computing Machinery, 2022).
49. SR Research. Eyelink*1000 plus eye tracker brochure. <https://www.sr-research.com> (2016).
50. Park, S. et al. Few-shot adaptive gaze estimation. In *Proc. IEEE/CVF International Conference on Computer Vision*, 9368–9377 (IEEE, 2019).
51. Hartley, R. & Zisserman, A. *Multiple View Geometry in Computer Vision*, 2nd edn. Camera Models. (Cambridge University Press, 2004).
52. Irving, E. L. & Lillakas, L. Difference between vertical and horizontal saccades across the human lifespan. *Exp. Eye Res.* **183**, 38–45 (2019).

53. Leigh, R. J. & Zee, D. S. *The Neurology of Eye Movements* (Oxford University Press, Oxford, 2015).
54. Daniel, S. & Lees, A. Parkinson's disease society brain bank, London: overview and research. *J. Neural Transm. Suppl.* **39**, 165–172 (1993).
55. Thaler, L., Schütz, A. C., Goodale, M. A. & Gegenfurtner, K. R. What is the best fixation target? the effect of target shape on stability of fixational eye movements. *Vis. Res.* **76**, 31–42 (2013).
56. Dai, W., Selesnick, I., Rizzo, J.-R., Rucker, J. & Hudson, T. Detection of normal and slow saccades using implicit piecewise polynomial approximation. *J. Vis.* **21**, 8–8 (2021).
57. Dai, W., Selesnick, I., Rizzo, J.-R., Rucker, J. & Hudson, T. A parametric model for saccadic eye movement. In *2016 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 1–6 (IEEE, 2016).
58. Munoz, M. J. et al. Medication adversely impacts visually-guided eye movements in parkinson's disease. *Clin. Neurophysiol.* **143**, 145–153 (2022).
59. Soukupová, T. & Čech, J. Real-time eye blink detection using facial landmarks. In *Proc. 21st Computer Vision Winter Workshop, Rimske Toplice, Slovenia*, 1–8 (2016).
60. Bland, J. M. & Altman, D. G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **327**, 307–310 (1986).

Acknowledgements

This study was funded, in part, by Analog Devices, Inc. through a partnership with the MIT Medical Electronic Device Realization Center, and by the MIT Aging Brain Initiative, the Northwestern University Department of Physical Therapy and Human Movement Sciences, and the Northwestern Medicine Enterprise Data Warehouse. The funders played no role in study design, data collection, analysis, and interpretation of data, or the writing of this manuscript.

Author contributions

J.K., F.J.D., and T.H. conceived and organized the research project. J.K., E.Z., J.A.K., C.P., and L.V.M. executed the research project. C.G.S., V.S., F.J.D., and T.H. supervised the research project. J.K., F.J.D., and T.H. designed the statistical analysis; J.K. executed it; J.A.K., C.P., L.V.M., C.G.S., V.S., F.J.D., and T.H. reviewed and critiqued the statistical analysis.

J.K. wrote the first draft of the manuscript. All authors reviewed, critiqued, and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41531-025-01079-9>.

Correspondence and requests for materials should be addressed to Thomas Heldt.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025