Enhancing Diagnostic Concordance in Parkinson's Disease: A Neurologist-Centric Analysis Using MediaPipe Hands

Purinat Tikkapanyo,
Narongrit Kasemsap

ABSTRACT

Background: The diagnosis of Parkinson's disease (PD) relies on clinical criteria, particularly bradykinesia, which is subjectively assessed using scales such as UPDRS and MDS-UPDRS, impacting interrater reliability. However, the finger tapping test produces coefficients that vary among clinicians, resulting in reliability ranging from inadequate to exceptional.

Objective: To improve the inter-rater reliability of the finger taps test in PD using MediaPipe Hands.

Method: Three neurologists evaluated a random selection of 94 videos, utilizing MDS-UPDRS for bradykinesia diagnosis. Inter-rater reliability, measured by Krippendorff's alpha, was assessed both before and after neurologists viewed Amplitude-time correlation graphs generated from the videos.

Results: Analysis of 94 participants (47 PD, 47 controls) indicated that PD participants were mostly male (59.6%), with symptoms lasting 1 to 5 years. UPDRS scores for PD participants were 1.9 (right hand) and 2.1 (left hand), while controls scored 0 for both hands. The Amplitude-time correlation graph revealed significant differences in tapping occurrences and amplitude percentages between PD and control groups. Krippendorff's alpha showed moderate-to-good agreement among neurologists for both the right and left hands, with and without the graph.

Conclusion: The amplitude-time correlation graph using MediaPipe Hands did not significantly enhance agreement in our study.

Keywords: Interrater reliability, Computer vision, Parkinson's disease, Parkinsonism, MDS-UPDRS, Amplitude correlation graph, MediaPipe Hands

Purinat Tikkapanyo, Narongrit Kasemsap
Division of Neurology, Department of Medicine Faculty of Medicine,
Khon Kaen University Khon Kaen, Thailand

Corresponding author:
Narongrit Kasemsap, MD.
Division of Neurology, Department of Medicine,
Faculty of Medicine, Khon Kaen University,
Khon Kaen, Thailand, 40002 Email: Naroka@kku.ac.th

รับต้นฉบับ 24 กุมภาพันธ์ 2567, ปรับปรุงต้นฉบับ 1 สิงหาคม 2567, ตอบรับต้นฉบับตีพิมพ์ 26 สิงหาคม 2567

Introduction

Parkinson’s disease (PD) is characterized by the loss of dopaminergic neurons which causes slowed muscular movement and motor rigidity. PD is a progressive and incurable disease and results in extensive utilization of health and community services. The Thailand PD Registry in March 2011 had estimated the prevalence of PD was 126.83/100,000 in urban areas and 90.82/100,000 in rural areas.[1]

Parkinson's disease is clinically defined by the presence of bradykinesia and at least one additional cardinal motor feature (rigidity or rest tremor), as well as additional supporting and exclusionary criteria.[4,5] Bradykinesia is the prerequisite for PD diagnosis [5] which makes accurate identification of bradykinesia pivotal. It can be evaluated by a subjective judgment of several tasks, such as finger tapping, pronation-supination movements, toe tapping, and foot tapping. [6] Clinical ratings scale assessments have been developed to measure bradykinesia, the Unified Parkinson's Disease Rating Scale (UPDRS), and the Movement Disorder Society revision of the UPDRS (MDS-UPDRS).[7] Although, they are based on objective face-to-face examination of a patient, clinical judgment, and some degree of subjectivity are involved. Interrater reliability regarding finger tapping produces coefficients that vary between clinicians, ranging from inadequate to exceptional reliability.[8,10,11,12,13,14] These studies suggest that bradykinesia assessment depends on subjective scoring by a physician, relying on clinical experience and having limited capability in detecting mild bradykinesia.

Despite clinical scales remaining the gold standard for most researchers and regulatory agencies, various technology-based tools have been developed to quantify bradykinesia more objectively. In our study, we employed a computer vision technique called MediaPipe Hands to detect finger tapping in videos and convert it into a graph. The aim was to assess the inter-rater reliability of three neurologists using this tool to grade MDS-UPDRS for the finger taps test.

Patients and Method

The participants were categorized into two distinct groups. Initially, the first group comprised patients who had recently received a diagnosis of idiopathic Parkinson's disease at the Neurological specialty outpatient department of Srinagarind Hospital, Khon Kaen University. The second group consisted of individuals from the healthy population.

The inclusion criteria necessitated that all participants be at least 18 years old prior to the recording date. Conversely, the exclusion criteria comprised individuals who displayed an inability to freely mobilize both hands due to factors such as intense pain, joint contracture, profound muscular debilitation, or the incapacity to execute a finger tapping test for 10 seconds. Furthermore, participants who displayed an inability to adhere to instructions or experienced severe dementia were also excluded from the study.

All participants were recorded performing a finger tapping test for 10 seconds (Figure 1). The video was processed using our tool, a computer vision machine learning based on MediaPipe Hands[9], to detect hand movement. MediaPipe Hands is a GoogleAI project that tracks hand and finger movement using 21 different points on each hand. We measured the distance between landmark number 4 (THUMB_TIP) and 8 (INDEX_FINGER_

TIP) to determine the amplitude of the graph (Figure 2). We compared the amplitudes of each video to the first amplitude and plotted the correlation graph. (Figures 3 and 4)

Figure 1: Illustrative images depicting participants performing the finger tap test, commencing with the largest amplitude (left) and the lowest amplitude (right).

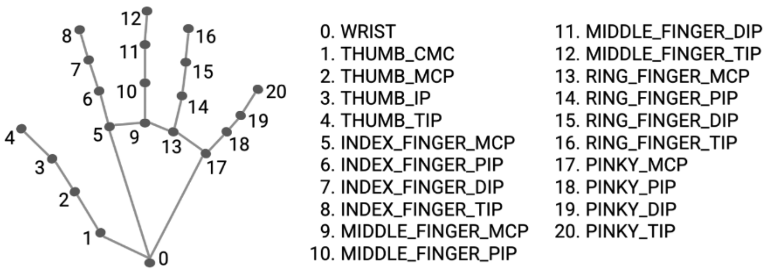

Figure 2: Hand-tracking landmark from MediaPipe [9]

We employed a randomization technique to select videos from a pool of 94 participants, which were subsequently evaluated by a panel of three neurologists.

The raters utilized the MDS-UPDRS part III for the rate severity of bradykinesia through the finger tapping test. [10] Following the initial process, the same clinicians reevaluated the identical videos, this time incorporating the graph. Our study involved a comprehensive analysis of inter-rater agreement in the rating MDS-UPDRS scale of idiopathic Parkinson's disease using the finger tapping assessment, both before and after observing the graph.

For measuring inter-rater reliability Krippendorf's alpha was used to determine if inter-rater reliability showed improvement after seeing the Amplitude-time correlation graph. Krippendorf's alpha was suitable for nominal data for more than 2 raters and also categories.

Table 1: Baseline characteristics of participants

Table 1. Demographic profile of patients with idiopathic Parkinson's disease and controls				
	Overall	PD	Control	P-value
Number of subjects	94	47	47	
Age at study	60.5	63.0	58.0	0.115
Male, n (%)	39 (41.5)	28 (59.6)	11 (23.4)	0.001
Handedness, right, n (%)	81 (86.2)	42 (89.4)	39 (83.0)	0.55
Symptom side				
- Right	15 (41.7)	15 (41.7)		
- Left	19 (52.8)	19 (52.8)		
- Bilateral	2 (5.6)	2 (5.6)		
Hoehn & Yahr stages (median)	2.0 [1.0,2.5]	2.0 [1.0,2.5]		
MDS-UPDRS				
- Right	1.0 (1.2)	1.9 (1.0)	0.0 (0.0)	<0.001
- Left	1.1 (1.2)	2.1 (0.7)	0.0 (0.0)	<0.001
Right				
Number of tap count	21.6 (8.7)	17.7 (8.2)	25.5 (7.3)	<0.001
Amplitude difference 1^{st}-5^{th} tap	4.2 (17.1)	9.5 (18.7)	-1.1 (13.5)	0.002
Amplitude difference 1^{st}-7^{th} tap	7.1 (21.1)	15.0 (24.1)	-0.7 (13.9)	<0.001
Amplitude difference 1^{st}-10^{th} tap	11.8 (25.6)	22.3 (29.5)	1.3 (15.2)	<0.001
Mean amplitude of right side, % (SD)	78.0 (14.7)	70.5 (16.6)	85.6 (6.6)	<0.001
Left				
Number of tap count	21.6 (7.6)	17.7 (5.9)	25.4 (7.2)	<0.001
Amplitude difference 1^{st}-5^{th} tap	2.3 (16.4)	6.5 (18.6)	-1.8 (12.8)	0.014
Amplitude difference 1^{st}-7^{th} tap	6.8 (19.0)	13.7 (22.7)	-0.1 (10.9)	<0.001
Amplitude difference 1^{st}-10^{th} tap	9.8 (21.7)	19.1 (25.6)	0.6 (11.3)	<0.001
Mean amplitude of left side, % (SD)	79.1 (10.5)	74.7 (10.8)	83.5 (8.0)	<0.001

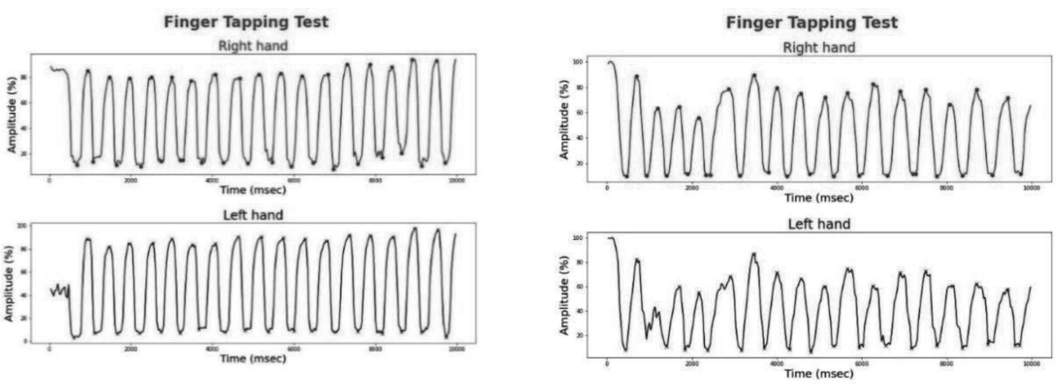

Figure 3: Amplitude-time correlation graph comparing the controlled group (Left) to bradykinesia group (Right).

A: MDS-UPDRS 0 B: MDS-UPDRS 1

C: MDS-UPDRS 2 D: MDS-UPDRS 3

E: MDS-UPDRS 4

Figure 4: Examples of amplitude-time correlation graphs for each participant, determined by the MDS-UPDRS assessment (graded on a scale from 0 to 4), are shown in sections A to E, respectively.

Result

The demographic and clinical characteristics of 94 participants are detailed in Table 1. In the PD group, there was a predominance of male participants (59.6%), while only 23.4 of the control group comprised males. The duration of parkinsonism symptoms ranged from 1 to 5 years, with an average duration of 2 years. Regarding UPDRS scores, the PD group exhibited an average score of 1.9 for the right hand and 2.1 for the left hand, whereas the control group scored 0 for both hands. No significant differences were observed between the PD and control groups concerning the age at the time of the study (63 years [range: 56.0-66.5] for PD group vs. 58 years [range: 47.5-68.5] for the control group, P-value = 0.115) and handedness (89.4% right-sided in the PD group vs. 83% in the control group, P-value = 0.550).

From the amplitude-time correlation graph comparing the two groups, significant differences were observed in the number of tapping occurrences. Specifically, the right hand exhibited 17.7 taps compared to 25.5 taps (P-value<0.001), while the left hand showed 17.7 taps compared to 25.4 taps (P-value<0.001). Additionally, the mean amplitude percentage differed significantly between the right hand (70.5% vs. 85.6%) and the left hand (74.7% vs. 83.5%) (P-value<0.001). Furthermore, when examining the amplitude difference for the right side between the 1st and 5th tap, it was found to be 9.5 compared to -1.1 (P-value 0.002). Similarly, the amplitude difference for the right side between the 1st and 7th tap was 15.0 compared to -0.7 (<0.001), and between the 1st and 10th tap, it was 22.3 compared to 1.3 (<0.001). Likewise, the amplitude difference for the left side between the 1st and 5th tap was 6.5 compared to -1.8 (P-value 0.014). Moreover, for the 1st and 7th tap, the amplitude difference was 13.7 compared to -0.1 (<0.001), and for the 1st and 10th tap, it was 19.1 compared to 0.6 (<0.001), respectively.

After analyzing the reliability for diagnosing using Krippendorff's alpha measures, it was determined that the inter-rater reliability ratings for the pre-seen graph were 0.71, while those for the post-seen graph were 0.71 for each side.

Discussion

The results revealed Krippendorff's alpha values for the pre- and post-seen graphs of the right hand were 0.71 and 0.71, and for the left hand, they were also 0.71 and 0.71, respectively. These values indicate a moderate-to-good level of agreement within both groups. The similarity in values for the Amplitude-time correlation graph in our study suggests that it did not contribute to an enhancement in agreement among neurologists.

In a study by Williams et al., inter-rater reliability among 21 specialists in the movement was measured to be ICC = 0.53, CLMM-ICC = 0.65, indicating a moderate level of agreement for MDS-UPDRS, slightly lower than the agreement observed in our study.[11] In a study by Rabey et al., the inter-rater reliability of finger tapping, as determined by UPDRS grading, reported Kendall's W values of 0.84 and 0.87, indicating good agreement and a higher level of agreement compared to our study.[12] Heldman et al. discovered that the correlation coefficients for inter-rater reliability were 0.72, indicating a moderate-to-good level of agreement.[13] Similarly, Vignoud et al. found an ICC score of 0.792, falling into the category of good agreement.[14] Based on the aforementioned comparable studies, the inter-rater reliability of our baseline falls within the average range observed in all studies, demonstrating a level of agreement that can be described as moderate to good.

Our study exclusively employed video recordings captured by various devices, with smartphones being the predominant choice. These recordings were manipulated using MediaPipe Hands to generate graphical representations. Within the domain of analytical methodologies, the viewpoints of neurologists possess the highest significance in the diagnostic procedure, while artificial intelligence (AI) functions solely as an ancillary support. Consequently, this method exhibits greater potential for seamless integration into real-life medical practices than a fully autonomous AI detection system, encouraging reliance on the diagnostic outcome.

Possible reasons for the lack of improvement in the agreement of the Amplitude-time correlation graph can be enumerated. Firstly, our study focuses on the concordance among raters rather than the accuracy of the diagnosis. The factor of raters, specifically neurologists, may not be sufficiently acquainted with the graph, and familiarizing themselves with it through practical experience could potentially enhance the level of agreement. Furthermore, the neurologists' familiarity with working with patients diagnosed with Parkinson's disease in their clinical practice could also influence both the accuracy and agreement. Secondly, the factor of participants should be considered. It is conceivable that in a certain subset of individuals suffering from Parkinson's disease, tremor could manifest as the primary symptom, in addition to the presence of bradykinesia. This combination of symptoms could have a detrimental impact on the accurate interpretation of visual and graphical data, potentially leading to erroneous conclusions. Thirdly, the assessment of the finger taps test within the MDS-UPDRS part III poses a considerable obstacle for clinicians and gives rise to challenges in terms of visual comparisons between successive grades, such as 0 and 1 or 1 and 2. Within this sequential grading, the graph may not exhibit sufficient dissimilarity to enable determination by the clinician, thus indicating a need for further refinement. Lastly, the factor of analysis must be taken into account. Our examination did not include analyzing how MediaPipe hand tracking performs in high-quality versus low-quality videos, which was a distinction made in a previous study by Islam et al.[15] Notably, variations in pose and hand angle were identified in their research, and it is plausible that these factors could introduce potential implications for the study outcomes.

The future trajectory for diagnosing Parkinson's disease is inclined toward employing AI as a means of enhancing accuracy and achieving consensus among clinicians. Consequently, it may be necessary to assess the comparative accuracy of AI versus human capabilities using the MediaPipe Hand track method. To facilitate this investigation, our evaluators may need to possess a degree of familiarity with the graphical representation in order to discern the disparities that exist across varying levels of contiguity grading.

Conclusion

While it is true that the amplitude-time correlation graph may not enhance agreement among evaluators in the present study, as previously mentioned, the utilization of AI-assisted detection methods for diagnosing Parkinson's disease in the future has the potential to enhance both accuracy and agreement among clinicians.

Reference

1. Bhidayasiri R, Wannachai N, Limpabandhu S, et al. A national registry to determine the distribution and prevalence of Parkinson's disease in Thailand: Implications of urbanization and pesticides as risk factors for Parkinson's disease. Tremor Other Hyperkinet Mov 2013;3:tre-03-157-3495-1. doi:10.7916/D81R6NQZ.

2. Zhao N, Yang Y, Zhang L, et al. Quality of life in Parkinson’s disease: A systematic review and meta-analysis of comparative studies. CNS Neurosci Ther 2021;27:270-9. doi:10.1111/cns.13549.
3. Erro R, Picillo M, Vitale C, et al. The non-motor side of the honeymoon period of Parkinson’s disease and its relationship with quality of life: a 4-year longitudinal study. Eur J Neurol 2016;23:1673-9. doi:10.1111/ene.13106.
4. Kalia LV, Lang AE. Parkinson’s disease. Lancet 2015; 386(9996):896-912. doi:10.1016/S0140-6736(14)61393-3.
5. Postuma RB, Berg D, Stern M, et al. MDS clinical diagnostic criteria for Parkinson’s disease. Mov Disord 2015;30:1591-601. doi:10.1002/mds.26424.
6. Goetz CG, Tilley BC, Shaftman SR, et al. Movement Disorder Society-Sponsored Revision of the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. Mov Disord 2008;23:2129-70. doi:10.1002/mds.22340.
7. Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. Movement Disorders [Internet]. 2008;23:2129–70. Available from: http://doi.wiley.com/10.1002/mds.22340
8. Lewis E, Fors L, Tharion WJ. Interrater and intrarater reliability of finger goniometric measurements. American Journal of Occupational Therapy 2010 1;64:555–61.
9. Google Developers. Hand Landmark Model. Google Developers. https://developers.google.com/mediapipe/solutions/vision/hand_landmarker. Accessed December 26, 2023.
10. Heldman DA, Espay AJ, LeWitt PA, Giuffrida JP. Clinician versus machine: Reliability and responsiveness of motor endpoints in Parkinson’s disease. Parkinsonism Relat Disord 2014;20:590–5.
11. Williams S, Wong D, Alty JE, Relton SD. Parkinsonian hand or clinician’s eye? Finger tap bradykinesia interrater reliability for 21 movement disorder experts. Journal of Parkinson’s Disease 2023;13:1–12.
12. Rabey JM, Bass H, Bonuccelli U, Brooks D, Klotz P, Korczyn AD, Kraus P, Martinez-Martin P, Morrish P, Van Sauten W, Van Hilten B. Evaluation of the Short Parkinson's Evaluation Scale: a new friendly scale for the evaluation of Parkinson's disease in clinical drug trials. Clin Neuropharmacol 1997;20:322-37. PMID: 9260730.
13. Heldman DA, Giuffrida JP, Chen R, Payne M, Mazzella F, Duker AP, et al. The modified bradykinesia rating scale for Parkinson’s disease: reliability and comparison with kinematic measures. Movement Disorders: Official Journal of the Movement Disorder Society [Internet]. 2011;26:1859–63. Available from: https://pubmed.ncbi.nlm.nih.gov/21538531/
14. Vignoud G, Desjardins C, Salardaine Q, Mongin M, Garcin B, Venance L, et al. Video-based automated assessment of movement parameters consistent with MDS-UPDRS III in Parkinson’s Disease. Journal of Parkinson’s Disease 2022;10:1–12.
15. Islam S, Rahman W, Abdelkader A, Lee S, Yang PT, Purks J, et al. Using AI to measure Parkinson’s disease severity at home. npj digital medicine. 2023;6:1-16.