Parkinsonian hand or clinician's eye? Finger tap bradykinesia interrater reliability for 21 experts

Stefan Williams,[1,2] Jane E Alty,[3] David Wong[4] and Samuel D Relton[1]

Abstract

Bradykinesia is the fundamental motor feature of Parkinson's disease. As a clinical sign determined purely by visual judgement, it is vulnerable to interrater variability, but the reliability of humans to detect and measure bradykinesia remains unclear. This is important as it is central to diagnosis, monitoring and research outcomes. We aimed to establish interrater reliability for expert neurologists assessing bradykinesia during the finger tapping test.

21 movement disorder expert neurologists rated finger tapping bradykinesia, using the Unified Parkinson's Disease Rating Scale (MDS-UPDRS) and the Modified Bradykinesia Rating Scale (MBRS), in 133 videos of hands: 73 from 39 people with idiopathic Parkinson's (PwP), 60 from 30 healthy controls. Each neurologist rated 30 randomly-selected videos and was also asked to judge whether the hand in the video was a PwP or a control. Intraclass correlation coefficients (ICC) for absolute agreement and consistency of MDS-UPDRS ratings were calculated, using standard linear and cumulative linked mixed models.

There was only moderate agreement for finger tapping MDS-UPDRS between neurologists, with ICC 0.53 (standard linear model), 0.65 (cumulative linked mixed model). Among control videos, 24% of were rated as bradykinesia (by MBRS subscores), and 53% rated >0 by MDS-UPDRS. PwP or control status was correctly judged in 70% of videos, but those judgements did not strictly follow bradykinesia presence.

Even experts disagree about the level of bradykinesia on finger tapping, and frequently see bradykinesia in the hands of those without neurological disease. Experts appear to judge Parkinsonian tapping using perception beyond a simple definition of bradykinesia.

Bradykinesia is to some extent a phenomenon in the eye of the clinician rather than simply the hand of the PwP.

Author affiliations:

1 Leeds Institute of Health Science, University of Leeds, Leeds LS2 9JT, UK

2 Leeds Teaching Hospitals NHS Trust, Leeds LS1 3EX, UK

3 Wicking Dementia Research and Education Centre, University of Tasmania, Hobart, Tasmania, 7005, Australia.

4 Centre for Health Informatics, University of Manchester, Manchester M13 9PL, UK

Correspondence to: Dr Stefan Williams

Department of Neurology, F Floor, Martin Wing, Leeds General Infirmary, Leeds Teaching Hospitals NHS Trust, Leeds LS1 3EX, UK

stefanwilliams@doctors.org.uk

Running title: Finger tap bradykinesia interrater reliability

Keywords: Parkinson's; finger tapping; bradykinesia; interrater reliability.

Abbreviations: ICC = Intraclass Correlation Coefficient; MBRS = Modified Bradykinesia Rating Scale; UPDRS = Unified Parkinson's Disease Rating Scale; MDS-UPDRS = Movement Disorder Society revision of the Unified Parkinson's Disease Rating Scale.

Introduction

Parkinson's disease is a clinical diagnosis, and at the centre of this is the presence of bradykinesia:

"*slowness of movement AND decrement in amplitude or speed (or progressive hesitations/halts) as movements are continued*".[1]

The Movement Disorder Society (MDS) criteria for a diagnosis of Parkinson's begins with a requirement for 'parkinsonism' - defined as bradykinesia, in combination with either rest tremor, rigidity, or both.[1] Thus, bradykinesia is the *sine qua non* of Parkinson's. Further, assessment of bradykinesia severity is central to measuring disease progression, response to treatment and research outcomes. Despite this fundamental importance, the gold standard test for bradykinesia is a visual judgement made through the eye of an expert clinician, during observation of movements.[1,2]

One of the most common methods to ascertain the presence and severity of bradykinesia in clinical practice is finger tapping, whereby an expert observes the patient repeatedly tapping their index finger against thumb "*as quickly and as big as possible*".[3] This finger tapping test is part of the gold-standard clinical rating scale – the (1987) Unified Parkinson's Disease Rating Scale (UPDRS),[4,5] and its (2008) Movement Disorder Society revision (MDS-UPDRS).[3] In this scale, three elements of finger tapping bradykinesia - speed, amplitude and rhythm - are assessed into a composite score between 0 and 4, **Table 1**. The Modified Bradykinesia Rating Scale (MBRS), developed to rate each component separately also includes a finger tapping item, **Table 2**.[6,7]

Score	**Criteria**
0: Normal	No problems

1: Slight	Any of the following: a) the regular rhythm is broken with one or two interruptions or hesitations of the tapping movement; b) slight slowing; c) the amplitude decrements near the end of the 10 taps.
2: Mild	Any of the following: a) 3 to 5 interruptions during tapping; b) mild slowing; c) the amplitude decrements midway in the 10-tap sequence.
3: Moderate	Any of the following: a) more than 5 interruptions during tapping or at least one longer arrest (freeze) in ongoing movement; b) moderate slowing; c) the amplitude decrements starting after the 1st tap.
4: Severe	Cannot or can only barely perform the task because of slowing, interruptions or decrements.

Table 1.

Score	Speed	Amplitude	Rhythm
0	Normal	Normal	Regular, no arrests or pauses in ongoing movement
1	Mild slowing	Mild reduction in amplitude in later performance, most movements close to normal	Mild impairment, up to two brief arrests / 10 seconds, none lasting > 1 second
2	Moderate slowing	Moderate reduction in amplitude visible early in performance but continues to maintain 50% amplitude through most of the task	Moderate, 3 to 4 arrests / 10 seconds; or 1 or 2 lasting > 1 second
3	Severe slowing	Severe, less than 50% amplitude through most of the task	Severe, 5 or more arrests / 10 seconds; or more than 2 lasting > 1 second
4	Can barely perform the task	Can barely perform the task	Can barely perform the task

Table 2.

Visual judgement as the gold standard evaluation for bradykinesia is problematic. Human assessment of movement is imprecise, with frequent disagreement amongst observers.[8] Bradykinesia is a complex, heterogeneous clinical sign that is difficult to gauge accurately.

This is important because it means that subtle changes of parkinsonism are difficult to measure, which hinders the accuracy of clinical decisions and research outcomes.

For several reasons, a robust estimate of interrater reliability for finger tapping bradykinesia has not been published. First, almost all studies have used very few (between 2 and 5) raters,[5,6,9–12] which is likely to be too few to assess the range of variability in rater judgements. Second, most studies involved clinical raters applying the entire UPDRS motor examination to each participant,[5,9–11,13,14] thus providing additional clinical information that influences the rater's judgement for any specific aspect of the examination. Henderson *et al.*[13] previously demonstrated this effect, showing that there was greater variation in rater scores when finger tapping was assessed in isolation (Kendall's W 0.5-0.6), rather than alongside other clinical assessments (Kendall's W >0.8). Third, most studies used only people with Parkinson's, without control participants that do not have Parkinson's.[5,6,11–15] This artificially avoids the difficult but important distinction between subtle bradykinesia and normal older age movement. Fourth, in some studies, Parkinson's medications are withheld prior to rating. This artificially exaggerates bradykinesia, making differences larger and therefore easier to reliably detect.[6,11–13] Fifth, all but one[15] interrater reliability study used the older (now obsolete) version of UPDRS, [5,6,9–11,13,14,16] not the current MDS-UPDRS, despite major differences in the way they define the grades of bradykinesia severity (e.g. "mild" deficits score grade 1 in UPDRS but grade 2 in MDS-UPDRS), and only two studies have used MBRS.[6,12]

This list of methodological problems likely explains why the published figures for interrater reliability vary widely for finger tapping bradykinesia, for example: Cohen's κ of -0.07 (poor agreement or no agreement),[10,17] κ of 0.47 (fair agreement),[5,17] Kendall's W of 0.87 (almost perfect agreement).[14,17] We aim to address this knowledge gap by comparing 21 expert neurologists' bradykinesia ratings for finger tapping when no other information is given, in people with Parkinson's and also in people without a neurological diagnosis.

Methods

Ethical review

The study was approved by the North of Scotland Research Ethics Committee, United Kingdom Health Research Authority (IRAS project ID 256116).

Finger tapping video

39 people with idiopathic Parkinson's disease (PwP) and 30 healthy controls provided written consent. All PwP had previously been diagnosed by a movement disorder specialist neurologist at Leeds Teaching Hospitals NHS Trust, United Kingdom, according to Movement Disorder Society clinical diagnostic criteria.[1] PwP were subjectively and objectively in the 'on' state at the time of participation (no medications were withheld). One investigator, SW, graded Hoehn and Yahr stage for each participant (but did not score any video for bradykinesia). Healthy controls were recruited from the companions of patients and hospital/university staff. They had no history of Parkinson's or other neurological diagnosis, and were not taking any medication that could induce parkinsonism.

Participants rested their elbow on a chair arm with the forearm lifted at 45°. Each participant was instructed to tap their index finger and thumb together "as quickly and as big as possible" with each hand examined separately, in accordance with MDS-UPDRS instructions. The participants tapped for just over 10 seconds, because the MDS-UPDRS specifies 10 taps while the MBRS specifies 10 seconds.[3,12]

We recorded videos of each hand during the task using a standard smartphone (iPhone SE) placed on a tripod (60 frames per second, 1920x1080 px) under ambient lighting. Only the

hand and part of the forearm were within the video frame. The distance from camera to hand was approximately 1m and digits 1 and 2 were closest to the camera.

One video was discarded because the hand moved outside the video frame, making 137 videos: 77 Parkinson's hands and 60 control hands. Each video was edited to contain 1s prior to tapping onset and 10s of finger tapping.

Clinical rating

We invited 21 consultant neurologists specialising in movement disorders, from a range of leading Parkinson's clinics in the United Kingdom, to each rate 30 videos of finger tapping, selected at random from the set of 137 videos. Each video was rated according to the MDS-UPDRS Item 3.4 Finger Tapping[3] and the MBRS[12] (**Tables 1 and 2)**. The raters undertook the task independently, at separate locations, and were blinded to both PwP / healthy control status and to each other's scores.

Inspired by informal comments made by the first two raters, we added an additional question for the subsequent 19 raters - asking them to judge whether the hand was most likely to be from a control or a PwP participant. This was in recognition that a clinician may judge an element or elements of bradykinesia to be technically present according to the rating scales, but nevertheless form a more overall impression that the tapping is essentially normal.

Outcomes

The primary outcome is the interrater reliability for MDS-UPDRS finger tapping scores, reported as the intraclass correlation coefficient (ICC), which was the basis of the statistical power calculations.

The secondary outcomes were correlation coefficients describing the relationship between the MDS-UPDRS score and each of the three components of the MBRS score, and the accuracy of clinicians in judging PwP from controls using the finger tapping videos.

Statistical analysis

Interrater reliability reflects the variation between more than one rater measuring the same group of participants.[18] We report ICCs for both *absolute agreement* and *consistency*. Absolute agreement concerns the degree to which one rater's score (x) is exactly equal to another's (y), whereas consistency concerns the degree to which x can be related to y plus a systematic error (x + c).

For each ICC, we calculate scores using a standard linear model, which assumes the underlying distribution of UPDRS scores is normally distributed, and a more sophisticated novel approached based upon cumulative linked mixed models (CLMMs), which is more appropriate for dealing with ordinal data. The normal distribution assumption of the first model is clearly incorrect, but allows direct comparison to previous research. Both approaches are two-way random effects models, where each item is assessed by the same set of raters randomly selected from a larger population of raters.

The random effects models consist of a random effect for video number (capturing the tendency of a video to be scored higher or lower than expected), a random effect for rater number (to capture the tendency of a rater to under-/over-rate videos), a fixed effect for whether the video is of a patient or control participant to give a baseline score in each case, and an intercept term. If σ_v^2 denotes the variance of the random effect for video number, σ_r^2 is the variance of the random effect for rater, and σ_ϵ^2 is the variance of the residual error then the agreement ICC is calculated as follows.

$$\frac{\sigma_v^2}{\sigma_v^2 + \sigma_r^2 + \sigma_\epsilon^2}$$

Meanwhile the consistency ICC is calculated as follows.

$$\frac{\sigma_v^2}{\sigma_v^2 + \sigma_\epsilon^2}$$

We fit two models to the data for calculating the ICC. The first uses a normal approximation to the ordinal score as in previous work. Our second model keeps the dependent variable ordinal using a cumulative linked mixed model (CLMM) – essentially fitting a latent normal model with the addition of "cut-points" which split the latent normal distribution into segments corresponding to the dependent ordinal variable.[19]

Whilst this latter CLMM readily gives the variance of the random effects for video numbers and raters, it is not initially clear how to define the residuals, which are required to calculate the ICC. In effect we need to define the "optimal" value in the latent space for each level that the ordinal variable can take. We took the following approach: after fitting the latent normal distribution and cut-points the optimal points were defined as the median of each segment of the normal distribution (calculated using Monte Carlo). With these points defined, the residual can be calculated using the latent value of the fitted model on each data point and the corresponding optimal values.

The study power calculation was done via simulation using the normal approximation to the ordinal variable, based on pilot data with two raters. Based on recruiting 20 raters and covering a variety of different strength ICC values, we determined that giving 30 random videos to each rater allows us to calculate the ICC to within 0.05 in 95% of trials and to within 0.03 in 80% of trials.

Secondary analysis consisted of calculating the three Spearman correlation coefficients of the relationship between the median UPDRS score across all raters, with the each of the median MBRS speed score, amplitude score, and rhythm scores.

Results

Expert neurologists' rating of finger tap bradykinesia in people with Parkinson's and controls

The age, gender and Hoehn and Yahr scores for the participants are given in **Table 3**. The median number of raters per video was 5 (range 1 to 12, interquartile range 3 to 7). In the random selection of 30 videos per rater, 4 videos from the total of 137 were not allocated to any rater, so that the total number of unique hand videos rated was 133. A total of 630 video ratings were made (21 raters, 30 videos each): 325 of these were ratings of PwP videos, and 305 ratings of healthy control videos.

	People with Parkinson's	**Healthy control participants**
Age (Std. Dev.) yrs	68 (9.6)	59 (19.4)
Male/Female	47/26	22/38
Median years since diagnosis	4	n/a
Median H&Y [IQR]	2 [1,3]	n/a
H&Y = 1	32	
H&Y = 1.5	2	
H&Y = 2	12	
H&Y = 2.5	4	

Table 3.

The distribution of MDS-UPDRS finger tapping scores for PwP and control videos rated are shown in **Figure 1**. 53% of control participant videos were given an MDS-UPDRS finger tapping score greater than 0. The distribution of MBRS scores for finger tapping speed, amplitude and rhythm are shown in **Figure 2**. Across both rating scales, scores of grade 1

('slight' impairment by MDS-UPDRS, 'mild' impairment by MBRS) were common in both control videos and PwP videos, and the proportion of videos rated as grade 1 was similar in both groups, between 20% and 40% (**Figure 3**). Thus, the prevalence of slight or mild impairment of tapping was similar in control and PwP videos.

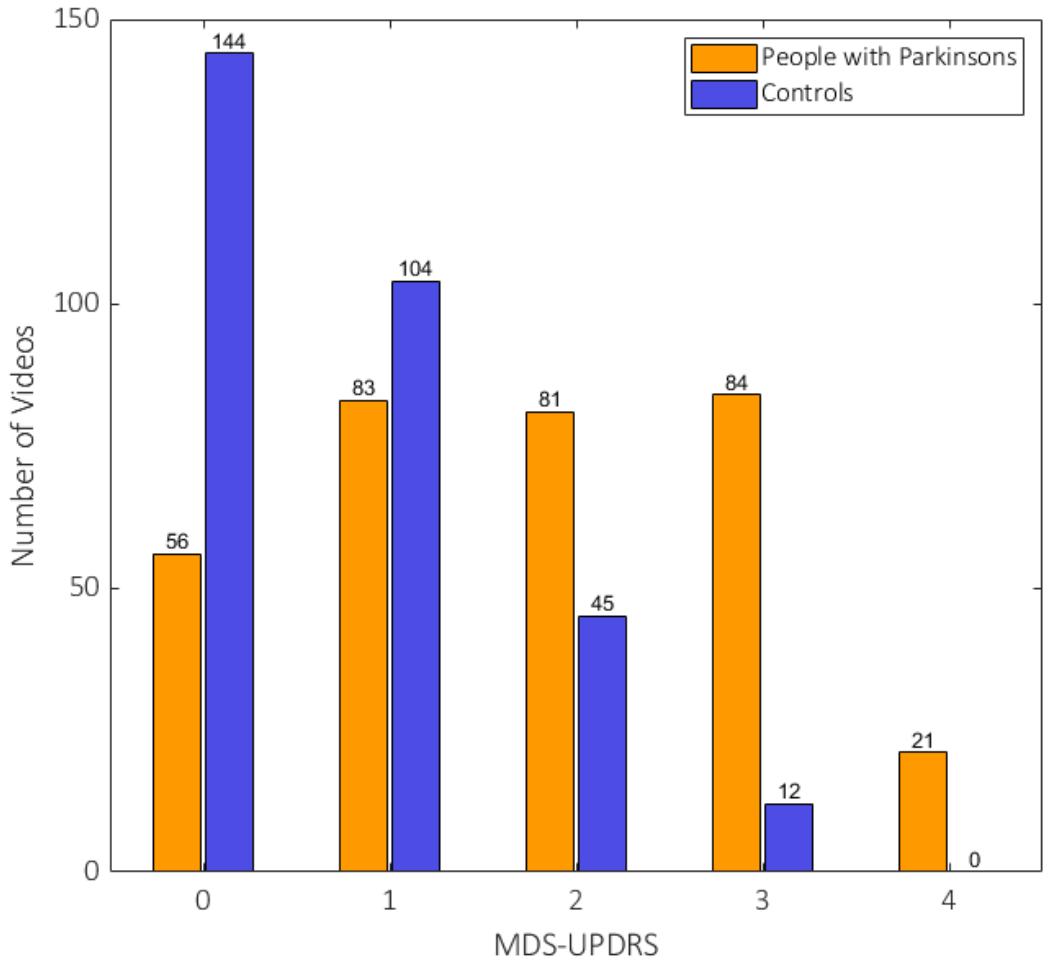

Figure 1.

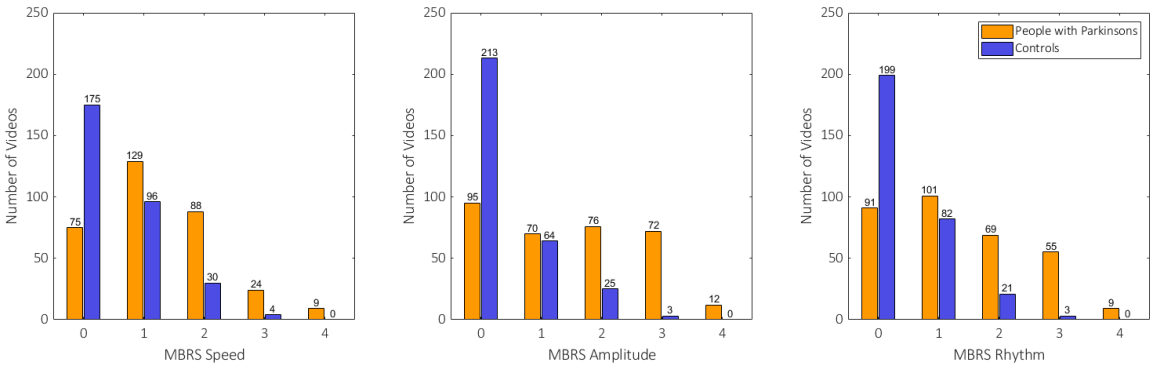

Figure 2.

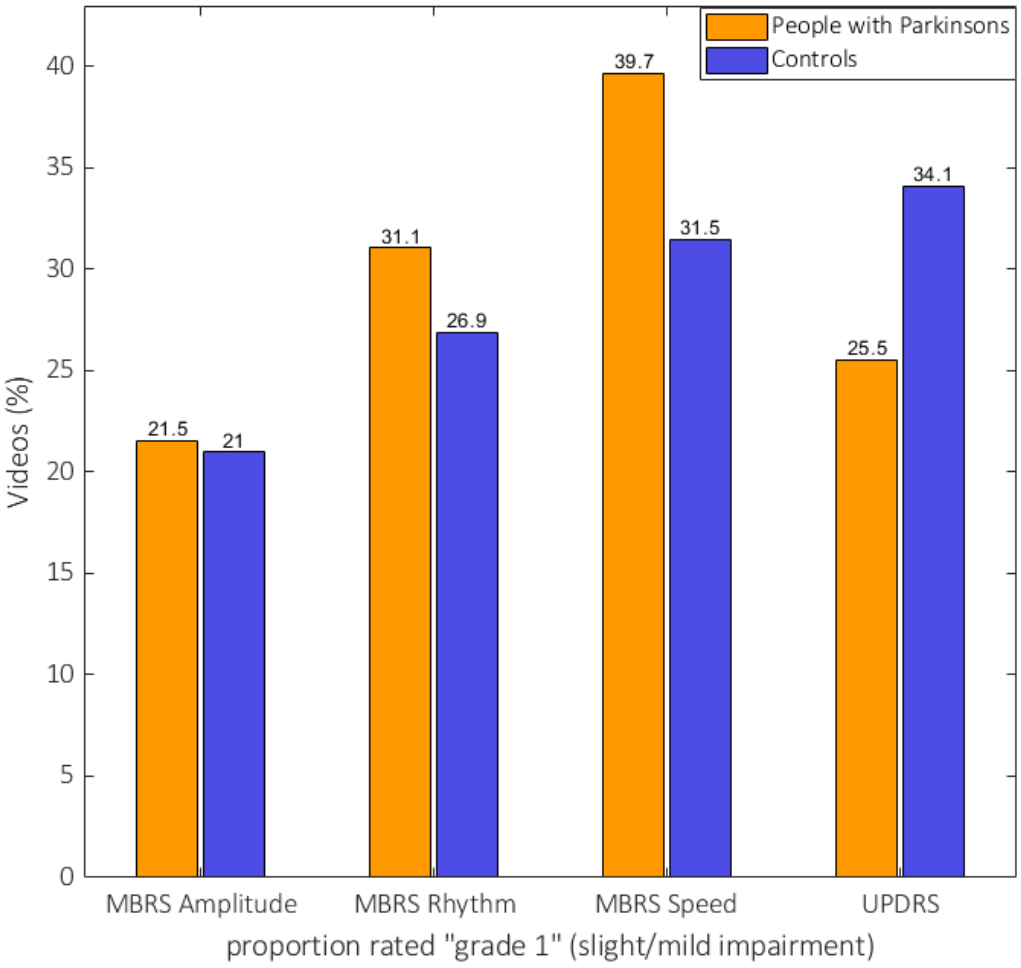

Figure 3.

Bradykinesia is defined as slowness of movement AND decrement in amplitude or speed (or progressive hesitations/halts) as movements are continued.[1] Therefore, the MBRS clinical rating of finger tapping can be used to classify tapping as bradykinesia by those videos for which a rater scored >0 for speed and >0 for one or more of amplitude or rhythm. **Table 4** shows the proportions of videos in PwP and controls (respectively) with impaired speed, rhythm, and amplitude, as well as combinations of those deficits, and the specific combination defined as bradykinesia. Among videos of PwP, 77% were rated as slow, and 64% were rated as bradykinesia (>0 for speed and >0 for one or more of amplitude or rhythm, by MBRS). Among videos of control participants, 43% were rated as slow, and 24% were rated as bradykinesia (>0 for speed and >0 for one or more of amplitude or rhythm, by MBRS). Thus, one in four control participant hand videos were rated as bradykinesia.

	People with Parkinson's	Healthy control participants
Impaired speed	77%	43%
Impaired rhythm	72%	35%
Impaired amplitude	70%	30%
Impaired speed and rhythm	62%	19%
Impaired speed and amplitude	61%	19%
Bradykinesia (Impaired speed + impaired rhythm and/or impaired amplitude)	**64%**	**24%**

Table 4.

Interrater reliability for finger tapping bradykinesia

The intraclass correlation coefficient (ICC) for MDS-UPDRS rating of finger tapping bradykinesia for exact agreement was 0.53 using the normal model ('fair'[20] or 'moderate'[18]) and 0.65 using the cumulative linked mixed model ('good'[20] or 'moderate'[18]). The ICC for consistency (that allows systematic differences) was 0.58 using the normal model ('fair'[20] or 'moderate'[18]), and 0.78 using the cumulative linked mixed model ('good'[20] or 'moderate'[18]).

To assess model calibration for the CLMM, we investigated the predicted values with the original ratings. The CLMM predicts the correct MDS-UPDRS score with 70% accuracy and is accurate to within one point on the five-point MDS-UPDRS finger tapping scale 98% of the time.

Figure 4 shows the variation in clinical ratings. Each point is an individual clinical rating of a video: the x-axis orders the videos by CLMM random effect size, and the y-axis is the clinical MDS-UPDRS rating. The values are jittered in the y-axis for visual clarity. It demonstrates the

considerable variation in movement disorder specialist judgement of individual videos, with disagreement common.

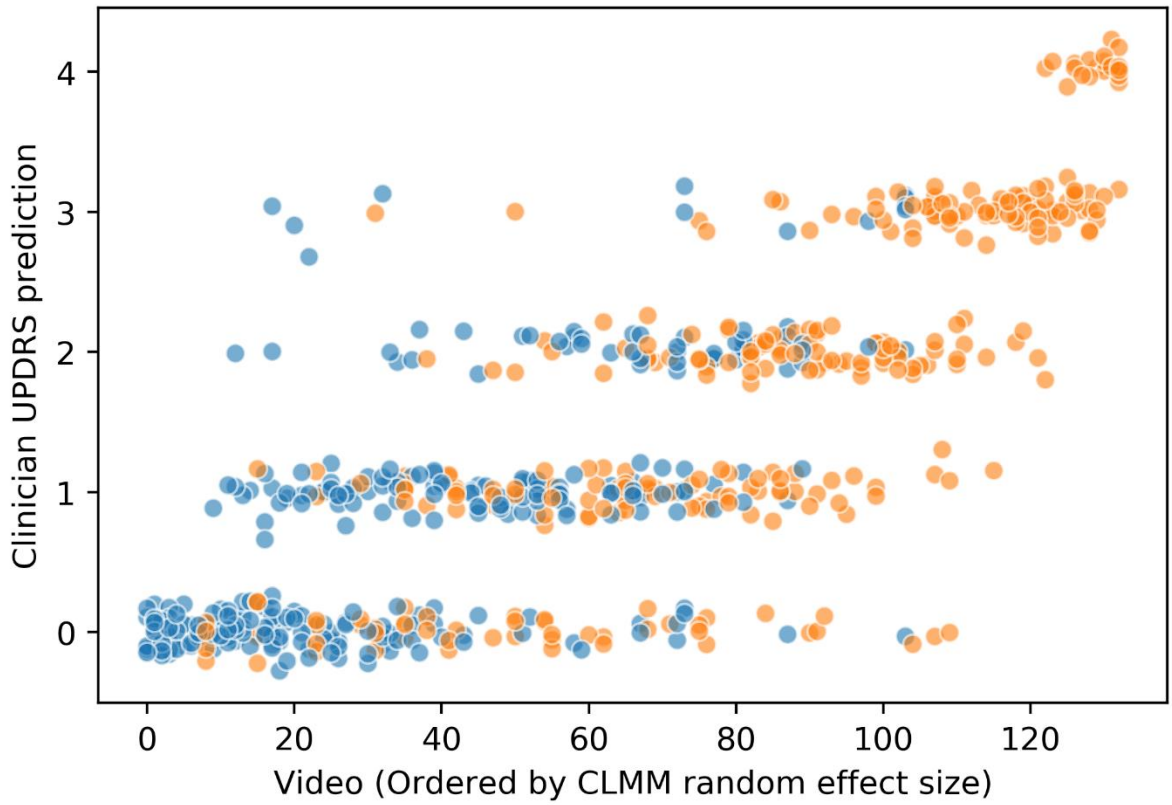

Figure 4

Correlations between finger tapping MDS-UPDRS and individual MBRS elements

Figure 5 shows the correlations between finger tapping ratings by MDS-UPDRS and finger tapping ratings by each of the MBRS subcomponents of speed, amplitude and rhythm. The MDS-UPDRS rating had moderate correlation with each of the ratings for speed, amplitude and rhythm.

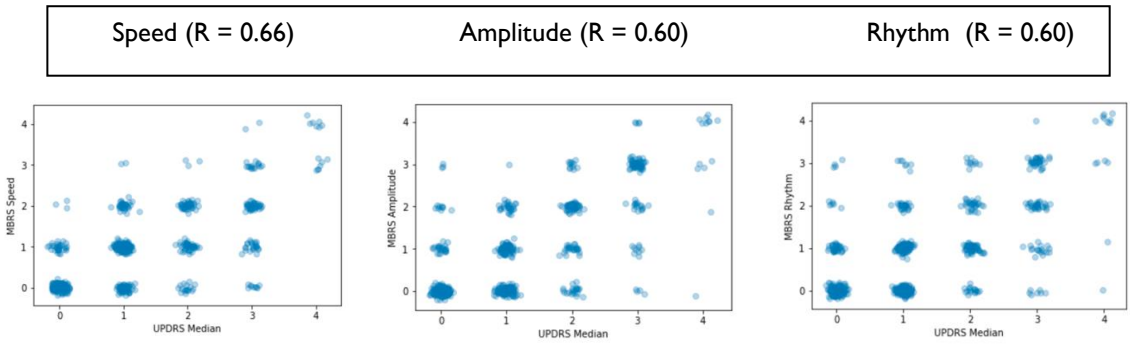

Figure 5.

Overall perception of finger tapping: expert neurologists' judgement of whether PwP or control

The movement disorder specialists correctly judged PwP or control status in 70% (400 of 570) videos, The median number of correct judgements was 20/30 (67%), with a range from 17/30 to 27/30, interquartile range 18.75 to 23.5 (out of 30).

Of those videos judged to show a PwP hand, only 77% were formally judged as showing bradykinesia and of the *correct* PwP guesses 84% were scored as bradykinesia. In other words, the movement disorders specialists' overall perception of PwP or control was not entirely related to the presence or absence of bradykinesia. Among videos correctly judged to show a control hand, 5% were formally judged as showing bradykinesia. Of the *correct* control judgements 3% were scored as bradykinesia.

Discussion

Our study findings show that even experts frequently disagree about the level of bradykinesia on finger tapping, despite clinical examination representing the gold standard for determining

the presence and degree of bradykinesia[1,3]. The 21 movement disorder specialists showed only 'moderate' agreement[18] for MDS-UPDRS finger tapping ratings (ICC=0.53, CLMM-ICC=0.65). Furthermore, the same movement disorder specialists classified one in four healthy control participants as showing bradykinesia on finger tapping (using MBRS sub-scores to match the definition of bradykinesia), and the proportions of participants showing slight or mild abnormalities on MDS-UPDRS and MBRS was similar in PwP and controls. This suggests that finger-tapping bradykinesia is also a non-specific sign and overlaps with changes of movements associated with normal ageing, at least when mild. It is perhaps unsurprising that bradykinesia is difficult to judge. It is a heterogenous clinical sign, and human vision cannot accurately measure and compare movement speed, amplitude and rhythm in isolation, let alone in simultaneous combination.

Our findings are particularly robust because they are based on data collected from a larger number of raters (21) and unique videos (137) than previous studies. Each rater saw 30 videos and the median number of raters per video was 5, but these numbers were based on statistical power calculations, and the random distribution of videos to raters mean that variation among the whole group is well characterised. Another strength of this study is the use of CLMM model, respecting the ordinal nature of MDS-UPDRS scores which has been neglected in previous research. Furthermore, we not only reported MDS-UPDRS finger tap ratings, but also MBRS ratings, which separately score each of tap speed, amplitude and rhythm. Unlike a previous study that suggested clinicians weighted amplitude and rhythm more than speed in UPDRS bradykinesia scores[7], we found similar correlations for all MBRS subscores with MDS-UPDRS (0.60-0.66), suggesting that clinicians do not favour any particular subcomponent of bradykinesia in MDS-UPDRS judgements. In addition, we reported consistency ICCs, which were a little higher than agreement results (ICC=0.58, CLMM-ICC=0.73), but in a five-point scale, consistent inter-rater variation is of little clinical relevance compared with absolute rater agreement.

A previous study of a UPDRS 'teaching tape' supports the idea that finger tapping bradykinesia is difficult to judge[8]. 226 raters were tested in their UPDRS motor scores for 4 PwP (using video recordings). A 'pass' in this test was defined as a score within the 95% confidence

interval of 3 international Parkinson's disease experts for each case. Only 54.6% of raters 'passed' the 4 cases, and of those that failed first time, 70.6% failed finger tapping.

Previous studies of finger tapping interrater reliability by UPDRS grading have reported Kendall's W 0.84 and 0.87,[14] weighted κ of 0.53 to 0.71,[9] 0.72 to 0.86,[11] κ of 0.47, 0.44, -0.07,[5,10] and Kendall's τ of 0.88 and 0.84,[15] while MBRS raters showed Pearson correlations of 0.51, 0.77 and 0.69 respectively [12]. It is difficult to draw conclusions from those results because each protocol had one or more methodological problems. A low number of raters and/or PwP is unlikely to capture rater and finger tapping variability.[5,6,9–12,15] This includes one study in which the total numbers were high but divided into small and non-overlapping subsets of raters and participants.[14] In most protocols, the participants are exclusively PwP – with no 'healthy control' participants – which artificially removes a challenging yet clinically important set of judgements at the lower end of the rating scale.[5,6,11–15] Recording videos with participants 'off' their usual medication artificially exaggerates bradykinesia.[6,11–13,15] Many of the statistical methods previously employed are inappropriate for the ordinal nature of rating scale data,[12] and/or are inappropriate measures of simple correlation rather than rater agreement.[15] Rating the entire UPDRS or UPDRS motor exam provides a wealth of additional information apart from finger tapping, and that is likely to bias the scoring of finger tapping bradykinesia.[5,9–11,13,14]

In relation to the idea that the other UPDRS items may influence finger tap scores, it could perhaps be argued that the influence of a broader assessment is appropriate and reflects clinical practice, in which finger tapping would never be tested in complete isolation. Therefore, difficulty 'seeing' bradykinesia on finger tapping is of little concern. However, busy routine clinics do not involve enough time for the complete UPDRS (a "vast instrument"[13]). In addition, UPDRS bradykinesia items are commonly analysed as a standalone 'bradykinesia' endpoint in trials.[12] Furthermore, clinician scoring of finger tapping is often used as a specific gold standard or ground truth for demonstrating that technological devices 'quantify' bradykinesia.[12,21–55] Most fundamentally, finger tapping bradykinesia is presented in the literature (and in the MDS-UPDRS) as a measure of a specific phenomenon with a specific definition. Finger tapping bradykinesia is not defined as a surrogate for an overall impression. If the latter is to some extent true, then it becomes less clear what bradykinesia actually is, and

less clear that movement disorder specialists are able to define and measure this "cardinal manifestation"[1] of Parkinson's disease.

In our results, 1 in 4 control videos were rated as showing finger tapping bradykinesia (using MBRS subscores). This is consistent with a previous study in which three trained nurses and one movement disorder specialist rated older people with no clinical Parkinson's disease according to a modified UPDRS motor score.[9] They gave 74 out of 75 participants a score greater than 0 (mean score 13.4 out of 127). Of course, the MDS diagnostic criteria for Parkinson's disease are not based on bradykinesia alone, and instead require a combination of clinical features to be present or absent to diagnose Parkinson's disease.[1] However, to some extent this only amplifies the challenge for clinician reliability, because other clinical features such as tremor are also non-specific, and there is considerable evidence that the overall diagnostic assessment of Parkinson's is difficult, with less-than-ideal sensitivity and specificity. This includes misdiagnosis rates of Parkinson's versus Essential tremor of one in three,[56] as well high false positive (17.4-26.1%) and false negative (6.7-20%) rates for the diagnosis of Parkinson's based on video examinations of people with tremor.[57] One postmortem study showed misdiagnosis of Parkinson's in 24 out of 100 cases.[58]

We asked the clinicians to judge whether the hand in the video was most likely to be that of a PwP or a control. Of those videos guessed to show the tapping of a PwP, only 77% were judged to show bradykinesia by the formal definition. This suggests that clinicians are forming an overall impression of finger tapping that does not purely follow the formal definition of bradykinesia: a gestalt perception of finger tapping normality / abnormality beyond the presence / absence of bradykinesia, that is intuitive and based on pattern recognition.[59,60] In support of this idea, a clinicopathological study found that experienced movement disorder specialists showed a higher accuracy than claimed for most clinical diagnostic criteria, for the diagnostic distinction of different forms of Parkinsonism. The authors state that these experts, "may be using a method of pattern recognition for diagnosis that goes beyond any formal set of diagnostic criteria".[61]

In conclusion, a classic sign of a cardinal clinical feature of a common neurological disease - finger tapping bradykinesia - is not easy to reliably see, even for expert eyes. Our findings suggest that bradykinesia is to some extent a phenomenon present in the eye of the clinician rather than simply the hand of the person with Parkinson's.

References

1. Postuma RB, Berg D, Stern M, et al. MDS clinical diagnostic criteria for Parkinson's disease. *Movement Disorders*. 2015;30(12):1591-1601. doi:10.1002/mds.26424

2. Goetz CG, Fahn S, Martinez-Martin P, et al. MDS-UPDRS: The MDS-sponsored Revision of the Unified Parkinson's Disease Rating Scale. 2008;1(414). www.movementdisorders.orginfo@movementdisorders.org

3. Goetz CG, Nyenhuis D, Poewe W, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Movement Disorders*. 2008;23(15):2129-2170. doi:10.1002/mds.22340

4. Fahn S, Elton R. Unified Parkinson's Disease Rating Scale. In: Fahn S, Marsden C, Calne D, Goldstein M, eds. *Recent Developments in Parkinson's Disease, Vol. 2.* NJ. Macmillan Health Care Information; 1987:153-164.

5. Martinez-Martin P, Gil-Nagel A, Gracia LM, et al. Unified Parkinson's disease rating scale characteristics and structure. *Movement Disorders*. 1994;9(1):76-83. doi:10.1002/mds.870090112

6. Kishore A, Espay AJ, Marras C, et al. Unilateral versus bilateral tasks in early asymmetric Parkinson's disease: Differential effects on bradykinesia. *MOVEMENT DISORDERS*. 2007;22(3):328-333. doi:10.1002/mds.21238

7. Heldman DA, Giuffrida JP, Chen R, et al. The modified bradykinesia rating scale for Parkinson's disease: Reliability and comparison with kinematic measures. *Movement Disorders*. 2011;26(10):1859-1863. doi:10.1002/mds.23740

8. Goetz CG, Stebbins GT. Assuring interrater reliability for the UPDRS motor section: Utility of the UPDRS teaching tape. *Movement Disorders*. 2004;19(12):1453-1456. doi:10.1002/mds.20220

9. Bennett DA, Shannon KM, Beckett LA, Goetz GG, Wilson RS. Metric properties of nurses' ratings of parkinsonian signs with a modified Unified Parkinson's Disease Rating Scale. *Neurology*. 1997;49(6):1580-1587. doi:10.1212/WNL.49.6.1580

10. Camicioli R, Grossmann SJ, Spencer PS, Hudnell K, Kent Anger W. Discriminating mild parkinsonism: Methods for epidemiological research. *Movement Disorders*. 2001;16(1):33-40.

11. Post B, Merkus MP, de Bie RMA, de Haan RJ, Speelman JD. Unified Parkinson's Disease Rating Scale motor examination: Are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable? *Movement Disorders*. 2005;20(12):1577-1584. doi:10.1002/mds.20640

12. Heldman DA, Giuffrida JP, Chen R, et al. The modified bradykinesia rating scale for Parkinson's disease: Reliability and comparison with kinematic measures. *Movement Disorders*. 2011;26(10):1859-1863. doi:10.1002/mds.23740

13. Henderson L, Kennard C, Crawford TJ, et al. Scales for rating motor impairment in Parkinson's disease: Studies of reliability and convergent validity. *Journal of Neurology Neurosurgery and Psychiatry*. 1991;54(1):18-24. doi:10.1136/jnnp.54.1.18

14. Rabey J, Bass H, Bonuccelli U, et al. Evaluation of the Short Parkinson's Evaluation Scale: A New Friendly Scale for the Evaluation of Parkinson's Disease in Clinical Drug Trials. *Clinical Neuropharmacology*. 1997;20(4):322-337.

15. Luiz LMD, Marques IA, Folador JP, Andrade AO. Intra and inter-rater remote assessment of bradykinesia in Parkinson's disease. *Neurología*. 2021;(xxxx). doi:10.1016/j.nrl.2021.08.005

16. Palmer JL, Coats MA, Roe CM, Hanko SM, Xiong C, Morris JC. Unified Parkinson's Disease Rating Scale-Motor Exam: Inter-rater reliability of advanced practice nurse and neurologist assessments. *Journal of Advanced Nursing*. 2010;66(6):1382-1387. doi:10.1111/j.1365-2648.2010.05313.x

17. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics*. 1977;33(1):159. doi:10.2307/2529310

18. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012

19. Liddell TM, Kruschke JK. Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*. 2018;79:328-348. doi:10.1016/j.jesp.2018.08.009

20. Cicchetti D v. Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. *Psychological Assessment*. 1994;6(4):284-290.

21. Giovannoni G, van Schalkwyk J, Fritz VU, Lees AJ. Bradykinesia akinesia inco-ordination test (BRAIN TEST): an objective computerised assessment of upper limb motor function. *Journal of neurology, neurosurgery, and psychiatry*. 1999;67(5):624-629.

22. Homann CN, Suppan K, Wenzel K, et al. The Bradykinesia Akinesia Incoordination Test (BRAIN TEST), an objective and user-friendly means to evaluate patients with parkinsonism. *Movement disorders : official journal of the Movement Disorder Society*. 2000;15(4):641-647.

23. Pal PK, Lee CS, Samii A, et al. Alternating two finger tapping with contralateral activation is an objective measure of clinical severity in Parkinson's disease and correlates with PET [18F]-DOPA Ki. *Parkinsonism and Related Disorders*. 2001;7(4):305-309. doi:10.1016/S1353-8020(00)00048-1

24. Tavares ALT, Jefferis GSXE, Koop M, et al. Quantitative measurements of alternating finger tapping in Parkinson's disease correlate with UPDRS motor disability and reveal the improvement in fine motor control from medication and deep brain stimulation. *Movement Disorders*. 2005;20(10):1286-1298. doi:10.1002/mds.20556

25. Papapetropoulos S, Jagid JR, Sengun C, Singer C, Gallo B V. Objective monitoring of tremor and bradykinesia during DBS surgery for Parkinson disease. *Neurology*. 2008;70(15):1244-1249. doi:10.1212/01.wnl.0000308936.27780.94

26. Papapetropoulos S, Katzen HL, Scanlon BK, Guevara A, Singer C, Levin BE. Objective quantification of neuromotor symptoms in Parkinson's disease: implementation of a

portable, computerized measurement tool. *Parkinson's disease*. 2010;2010:760196. doi:10.4061/2010/760196

27. Maetzler W, Ellerbrock M, Heger T, Sass C, Berg D, Reilmann R. Digitomotography in Parkinson's disease: A cross-sectional and longitudinal study. *PLoS ONE*. Published online 2015. doi:10.1371/journal.pone.0123914

28. Lee CY, Kang SJ, Hong SK, Ma HI, Lee U, Kim YJ. A Validation Study of a Smartphone-Based Finger Tapping Application for Quantitative Assessment of Bradykinesia in Parkinson's Disease. *PLOS ONE*. 2016;11(7). doi:10.1371/journal.pone.0158852

29. Kassavetis P, Saifee TA, Roussos G, et al. Developing a Tool for Remote Digital Assessment of Parkinson's Disease. *MOVEMENT DISORDERS CLINICAL PRACTICE*. 2016;3(1):59-64. doi:10.1002/mdc3.12239

30. Mitsi G, Mendoza EU, Wissel BD, et al. Biometric Digital Health technology for Measuring Motor Function in Parkinson's Disease: results from a Feasibility and Patient satisfaction study. *FRONTIERS IN NEUROLOGY*. 2017;8. doi:10.3389/fneur.2017.00273

31. Lalvay L, Lara M, Mora A, et al. Quantitative Measurement of Akinesia in Parkinson's Disease. *Movement Disorders Clinical Practice*. 2017;4(3):316-322. doi:10.1002/mdc3.12410

32. Prince J, Arora S, de Vos M. Big data in Parkinson's disease: using smartphones to remotely detect longitudinal disease phenotypes. *PHYSIOLOGICAL MEASUREMENT*. 2018;39(4). doi:10.1088/1361-6579/aab512

33. Roalf DR, Rupert P, Mechanic-Hamilton D, et al. Quantitative assessment of finger tapping characteristics in mild cognitive impairment, Alzheimer's disease, and Parkinson's disease. *JOURNAL OF NEUROLOGY*. 2018;265(6):1365-1375. doi:10.1007/s00415-018-8841-8

34. Yokoe M, Okuno R, Hamasaki T, Kurachi Y, Akazawa K, Sakoda S. Opening velocity, a novel parameter, for finger tapping test in patients with Parkinson's disease. *Parkinsonism and Related Disorders*. 2009;15(6):440-444. doi:10.1016/j.parkreldis.2008.11.003

35. Costa J, González HA, Valldeoriola F, Gaig C, Tolosa E, Valls-Solé J. Nonlinear dynamic analysis of oscillatory repetitive movements in Parkinson's disease and essential tremor. *Movement Disorders*. 2010;25(15):2577-2586. doi:10.1002/mds.23334

36. Kim JW, Lee JH, Kwon Y, et al. Quantification of bradykinesia during clinical finger taps using a gyrosensor in patients with Parkinson's disease. *MEDICAL & BIOLOGICAL ENGINEERING & COMPUTING*. 2011;49(3):365-371. doi:10.1007/s11517-010-0697-8

37. Stamatakis J, Ambroise J, Cremers J, et al. Finger Tapping Clinimetric Score Prediction in Parkinson's Disease Using Low-Cost Accelerometers. *COMPUTATIONAL INTELLIGENCE AND NEUROSCIENCE*. Published online 2013. doi:10.1155/2013/717853

38. Heldman DA, Espay AJ, LeWitt PA, Giuffrida JP. Clinician versus machine: reliability and responsiveness of motor endpoints in Parkinson's disease. *Parkinsonism & related disorders*. 2014;20(6):590-595. doi:10.1016/j.parkreldis.2014.02.022

39. Lee MJ, Kim SL, Lyoo CH, Lee MS. Kinematic Analysis in Patients with Parkinson's Disease and SWEDD. *JOURNAL OF PARKINSONS DISEASE*. 2014;4(3):421-430. doi:10.3233/JPD-130233

40. Kim JW, Kwon Y, Yun JS, et al. Regression models for the quantification of Parkinsonian bradykinesia. *Bio-medical materials and engineering*. 2015;26 Suppl 1:S2249-58. doi:10.3233/BME-151531

41. Martinez-Manzanera O, Roosma E, Beudel M, Borgemeester RWK, Van Laar T, Maurits NM. A Method for Automatic and Objective Scoring of Bradykinesia Using Orientation Sensors and Classification Algorithms. *IEEE Transactions on Biomedical Engineering*. 2016;63(5):1016-1024. doi:10.1109/TBME.2015.2480242

42. Heldman DA, Urrea-Mendoza E, Lovera LC, et al. App-Based Bradykinesia Tasks for Clinic and Home Assessment in Parkinson's Disease: Reliability and Responsiveness. *Journal of Parkinson's disease*. 2017;7(4):741-747. doi:10.3233/JPD-171159

43. di Biase L, Summa S, Tosi J, et al. Quantitative Analysis of Bradykinesia and Rigidity in Parkinson's Disease. *FRONTIERS IN NEUROLOGY*. 2018;9. doi:10.3389/fneur.2018.00121

44. Agostino R, Curra A, Giovannelli M, Modugno N, Manfredi M, Berardelli A. Impairment of individual finger movements in Parkinson's disease. *MOVEMENT DISORDERS*. 2003;18(5):560-565. doi:10.1002/mds.10313

45. Ling H, Massey LA, Lees AJ, Brown P, Day BL. Hypokinesia without decrement distinguishes progressive supranuclear palsy from Parkinson's disease. *BRAIN*. 2012;135(4):1141-1153. doi:10.1093/brain/aws038

46. Lainscsek C, Rowat P, Schettino L, et al. Finger tapping movements of Parkinson's disease patients automatically rated using nonlinear delay differential equations. *CHAOS*. 2012;22(1). doi:10.1063/1.3683444

47. Krupicka R, Szabo Z, Viteckova S, Ruzicka E. Motion Capture System for Finger Movement Measurement in Parkinson Disease. *RADIOENGINEERING*. 2014;23(2):659-664.

48. Ruzicka E, Krupicka R, Zarubova K, Rusz J, Jech R, Szabo Z. Tests of manual dexterity and speed in Parkinson's disease: Not all measure the same. *Parkinsonism & related disorders*. 2016;28:118-123. doi:10.1016/j.parkreldis.2016.05.009

49. Bologna M, Leodori G, Stirpe P, et al. Bradykinesia in early and advanced Parkinson's disease. *Journal of the Neurological Sciences*. 2016;369:286-291. doi:10.1016/j.jns.2016.08.028

50. Bank PJM, Marinus J, Meskers CGM, de Groot JH, van Hilten JJ. Optical Hand Tracking: A Novel Technique for the Assessment of Bradykinesia in Parkinson's Disease. *MOVEMENT DISORDERS CLINICAL PRACTICE*. 2017;4(6):875-883. doi:10.1002/mdc3.12536

51. Kandori A, Yokoe M, Sakoda S, et al. Quantitative magnetic detection of finger movements in patients with Parkinson's disease. *NEUROSCIENCE RESEARCH*. 2004;49(2):253-260. doi:10.1016/j.neures.2004.03.004

52. Shima K, Tsuji T, Kandori A, Yokoe M, Sakoda S. Measurement and Evaluation of Finger Tapping Movements Using Log-linearized Gaussian Mixture Networks. *SENSORS*. 2009;9(3):2187-2201. doi:10.3390/s90302187

53. Sano Y, Kandori A, Shima K, et al. Quantifying Parkinson's disease finger-tapping severity by extracting and synthesizing finger motion properties. *MEDICAL & BIOLOGICAL ENGINEERING & COMPUTING*. 2016;54(6):953-965. doi:10.1007/s11517-016-1467-z

54. Gao C, Smith S, Lones M, et al. Objective assessment of bradykinesia in Parkinson's disease using evolutionary algorithms: Clinical validation. *Translational Neurodegeneration*. 2018;7(1):1-8. doi:10.1186/s40035-018-0124-x

55. Teo WP, Rodrigues JP, Mastaglia FL, Thickbroom GW. Comparing kinematic changes between a finger-tapping task and unconstrained finger flexion-extension task in patients with Parkinson's disease. *Experimental brain research*. 2013;227(3):323-331. doi:10.1007/s00221-013-3491-7

56. Jain S, Lo SE, Louis ED. Common Misdiagnosis of a Common Neurological Disorder. *Archives of Neurology*. 2006;63(8):1100. doi:10.1001/archneur.63.8.1100

57. Bajaj NPS, Gontu V, Birchall J, Patterson J, Grosset DG, Lees AJ. Accuracy of clinical diagnosis in tremulous parkinsonian patients: a blinded video study. *Journal of Neurology, Neurosurgery & Psychiatry*. Published online January 1, 2010.

58. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: A clinico-pathological study of 100 cases. *Journal of Neurology Neurosurgery and Psychiatry*. 1992;55(3):181-184. doi:10.1136/jnnp.55.3.181

59. Cervellin G, Borghi L, Lippi G. Do clinicians decide relying primarily on Bayesians principles or on Gestalt perception? Some pearls and pitfalls of Gestalt perception in medicine. *Internal and Emergency Medicine*. 2014;9(5):513-519. doi:10.1007/s11739-014-1049-8

60. Vancheri F. Bayesian principles or Gestalt perception for clinical judgment. *Internal and Emergency Medicine*. 2015;10(2):253. doi:10.1007/s11739-014-1133-0

61. Hughes AJ, Daniel SE, Ben-Shlomo Y, Lees AJ. The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. *Brain*. 2002;125(4):861-870. doi:10.1093/brain/awf080