Published in final edited form as:
Parkinsonism Relat Disord. 2014 June ; 20(6): 590–595. doi:10.1016/j.parkreldis.2014.02.022.

Clinician Versus Machine: Reliability and Responsiveness of Motor Endpoints in Parkinson's Disease

Dustin A. Heldman, Ph.D.[1,*], **Alberto J. Espay, M.D., M.Sc.**[2], **Peter A. LeWitt, M.D., M.Med.Sc.**[3], and **Joseph P. Giuffrida, Ph.D.**[1]

[1]Great Lakes NeuroTechnologies Inc., Cleveland, Ohio, USA

[2]UC Neuroscience Institute, Department of Neurology, Gardner Center for Parkinson's disease and Movement Disorders, University of Cincinnati, Cincinnati, Ohio, USA

[3]Departments of Neurology, Henry Ford Hospital and Wayne State University School of Medicine, West Bloomfield, Michigan, USA

Abstract

Background—Enhancing the reliability and responsiveness of motor assessments required to demonstrate therapeutic efficacy is a priority for Parkinson's disease (PD) clinical trials. The objective of this study is to determine the reliability and responsiveness of a portable kinematic system for quantifying PD motor deficits as compared to clinical ratings.

Methods—Eighteen PD patients with subthalamic nucleus deep brain stimulation (DBS) performed three tasks for evaluating of resting tremor, postural tremor, and finger-tapping speed, amplitude, and rhythm while wearing a wireless motion-sensor unit (Kinesia) on the more-affected index finger. These tasks were repeated three times with DBS turned off and at each of 10 different stimulation amplitudes chosen to yield small changes in treatment response. Each task performance was video-recorded for subsequent clinician rating in blinded, randomized order. Test-retest reliability was calculated as intraclass correlation (ICC) and sensitivity was calculated as minimal detectable change (MDC) for each DBS amplitude.

Results—ICCs for Kinesia were significantly higher than those for clinician ratings of finger-tapping speed ($p<0.0001$), amplitude ($p<0.0001$), and rhythm ($p<0.05$), but were not significantly different for evaluations of resting or postural tremor. Similarly, Kinesia scores yielded a lower MDC as compared with clinician scores across all finger-tapping subscores ($p<0.0001$), but did not differ significantly for resting and postural tremor.

Conclusions—The Kinesia portable kinematic system can provide greater test-retest reliability and sensitivity to change than conventional clinical ratings for measuring bradykinesia, hypokinesia, and dysrhythmia in PD patients.

[*]**All correspondence to:** Dustin A. Heldman, Ph.D. Great Lakes NeuroTechnologies Inc. 10055 Sweet Valley Dr. Cleveland, OH 44125 P: 216-446-2437; dheldman@glneurotech.com.

Keywords

Parkinson's disease; UPDRS; Kinesia; tremor; bradykinesia

Introduction

The design of clinical trials to demonstrate efficacy of new symptomatic and neuroprotective treatments of Parkinson's disease (PD) encounters a substantial challenge for reliable quantification of small changes. With the advent of automated, computerized systems providing precise measurement of motion [1–9], the question of “man versus machine” offers renewed implications for designing and conducting clinical trials. Conventional clinical trial outcome measures come from clinician ratings carried out at out-patient visits or from patient-completed home diaries [10]. Such clinical assessments have limitations imposed by various forms of bias, placebo effect (both subject and investigator), limited resolution, and poor intraand inter-rater reliability [2,11–13]. Similarly, patient-completed diaries can yield unreliable data due to inadequate compliance, recall bias, or faulty self-assessment [14].

In recent years, automated computerized motion-sensor systems (e.g., body-worn inertial sensors) have become widely used in clinical trials. These systems offer inexpensive, objective, and quantitative measures that can be repeated at multiple time points [15]. Many of these systems also permit home monitoring, thus enabling the recording of motor fluctuations throughout the day and in their typical settings [4,15–17]. Advanced signal processing algorithms are able to discriminate tremor and dyskinesia from voluntary movements enacted during activities of daily living [3,18]. Motion sensor systems have also shown promise at yielding biomarkers for differentiating disease states from controls based on analyzed measurements of gait and balance [19–21]. These systems have also provided input useful for biofeedback training [22,23]. However, while measurements made by automated computerized motion-sensor systems lack many of the problems that interfere with clinical assessment, their data is potentially subject to contamination of motor endpoints by extraneous non-targeted motor phenomena such as dyskinesias, gravitational effects, and limitations of sensor resolution [18,24]. Additionally, motion sensor platforms are not standardized and vary in calibration, assessment procedures, and processing algorithms. Therefore, clinical trial sites generally would need to use the same type of equipment.

We examined the test-retest reliability and responsiveness (sensitivity to change) of a motion-sensor-based PD monitoring system as compared with rating scores by experienced movement disorder specialists. We studied PD subjects who had undergone successful subthalamic nucleus deep-brain stimulation (DBS) therapy, since the adjustment of stimulation output provided an adjustable means for modulating the severity of parkinsonian features during testing sessions.

Methods

Subject Recruitment

Eighteen subjects (13 males; mean age 63.1±8.4 years, range: 44–76) meeting criteria for levodopa-responsive idiopathic PD and having undergone bilateral subthalamic nucleus DBS implants were recruited. The range of tolerable DBS simulation parameters (adjusted in this instance for experimental purposes) provided an opportunity for gradually modulating parkinsonian severity. In this manner, it was possible to model a range of parkinsonian motor deficit states among a relatively small number of subjects. The clinical testing was carried out at the University of Cincinnati College of Medicine (Cincinnati, OH) and at Henry Ford Hospital (West Bloomfield, MI) under the purview of their respective institutional review boards and in accordance with the Declaration of Helsinki (2008). All study subjects provided signed informed consent prior to their participation.

Experimental Methods

Subjects had received stable and optimized oral medication and DBS treatment regimens prior to their evaluation (mean baseline DBS output voltage: 3.1V ± 0.74V, range: 1.9V–4.3V). Testing was initiated at least 30 minutes after turning off each subject's DBS implantable pulse generator (IPG), a time point typically when effects of stimulation have substantially abated [25]. The subjects wore a wireless portable kinematic system (Kinesia, Great Lakes NeuroTechnologies Inc., Cleveland, OH) on the most distal portion of the index finger of the more parkinsonian hand. Subjects then completed an automated motor assessment, which included three 15-second tasks that were each repeated three times (to ascertain test-retest reliability). In each sequence, the first two tasks were assessments of rest and postural tremor, while the third task involved repetitive finger tapping as quickly and big as possible to evaluate bradykinesia (slowed speed), hypokinesia (diminished amplitude), and dysrhythmia (poor rhythm maintenance). Kinesia, which outputs motor scores on a 0–4 scale with 0.1 resolution, has been validated for scoring the tasks that were performed in the current study [1,2]. Each task-performing hand was videorecorded using a standardized close-up format for subsequent clinical rating.

The baseline motor assessment was performed with DBS turned off. Stimulation ipsilateral to the hand wearing the motion-sensor unit remained off throughout the protocol and all DBS voltage adjustments were made to the contralateral electrode. Next, the voltage output amplitude on the IPG was set to 0.9V below the subject's previously-determined optimal setting and the automated motor assessment was again performed. The amplitude on the DBS IPG was then increased sequentially in steps of 0.1V, with the three repetitions of the automated motor assessment performed at each amplitude level until reaching the subject's previously-determined optimal stimulation amplitude. All other stimulation parameters remained constant. In total, each subject performed the three repetitions of the three motor tasks at each of eleven DBS voltage settings (that is, with the IPG turned off and at ten stimulation amplitudes). Other than the 30-minute washout period after turning DBS off at the start, subjects performed the tasks shortly after DBS voltage was adjusted to the next setting. As the three repetitions of each of the three tasks were performed sequentially (with

only a few seconds in between each task to adjust the video camera), we assumed that parkinsonian state did not change substantially within each voltage output increment.

Clinician Scoring

The video-recordings of each subject's hand movements for the three repetitions at each of the eleven DBS voltage amplitudes were separated by task and then randomized for placement onto a webserver for subsequent online scoring by two experienced movement disorder neurologists (AJE and PAL). Although each of the study participants was a patient under the care of one of the two clinician raters, approximately half of the subjects were not known to each of the raters. To ensure blinded ratings, the videos were cropped to show only the participant's hand during each task, making it impossible for a rater to know which patient was being evaluated. Raters were also blinded to DBS settings. Rest and postural tremor were rated according to the Unified Parkinson's Disease Rating Scale (UPDRS) [26] criteria (0–4 integer scale; higher numbers are worse). The finger-tapping task was rated by UPDRS as well as by the modified bradykinesia rating scale (MBRS), which independently scores speed, amplitude, and rhythm (0–4 integer scale; higher numbers are worse) [27].

Reliability and Sensitivity Analysis

Test-retest reliability (or *consistency*) of clinician and Kinesia scores was calculated by intraclass correlation coefficient (ICC) [12,13]. *Responsiveness*, the minimum amount of true change that can be captured by a scale or instrument, was measured as the minimal detectable change (MDC) for clinician and Kinesia scores (lower MDC, higher sensitivity) using the following equation:

$$MDC = 1.65 \times SD\sqrt{2\,(1-r)} \quad (1)$$

where 1.65 reflects the 90% confidence interval, SD is the standard deviation, √2 represents uncertainty introduced by using measurements at two time-points, and *r* is the coefficient of the test-retest reliability (in this case, ICC) [12,28,29]. In clinical trials, scores beyond the MDC are generally attributable to an intervention effect rather than measurement error [12,28]. Both the ICC and MDC were calculated for scores between repetitions 1 and 2, 2 and 3, and 1and 3 across all same-condition assessments and compared across modalities using Student's t-test. Tukey's HSD (“honestly significant difference”) test was used to determine at which stimulation voltages the motor response showed significant change from baseline. Tukey's HSD test was chosen to correct for the experiment-wise error rate from making multiple comparisons [30].

Sample size calculations

Variations in ICC (reliability) and MDC (responsiveness) of scales (or of any measuring instrument) affect the sample size of clinical trials powered to find significant differences between an intervention and placebo. Specifically, higher ICCs and lower MDCs increase the power to achieve statistical significance using smaller samples. We used the model described by Perkins et al. [31] to ascertain the effects of ICCs on sample size for Kinesia

and clinician scores for a hypothetical clinical trial in which an *a priori* power analysis determined 100 subjects would be necessary to detect a significant change.

Results

The ICCs for Kinesia assessments were significantly greater than those for the clinician rating of finger tapping speed $p<0.0001$), amplitude $p<0.0001$), and rhythm $p<0.05$), but not significantly different for the scores evaluating rest and postural tremor (Figure 1A). Similarly, Kinesia scores yielded a lower MDC as compared with clinician scores across all finger-tapping subscores ($p<0.0001$), but did not differ significantly for resting and postural tremor (Figure 1B). Bland-Altman plots [32,33] are shown to give a graphical representation of the test-retest reliability for Kinesia and the clinicians for the finger-tapping subscores (Figure 2). Neither Kinesia nor clinician scores showed significant systematic bias; however, the 95% limits of agreement were smaller for Kinesia than for clinicians.

Based on the higher ICCs for Kinesia-derived measurements compared to those of the clinicians, using the model described by Perkins et al. [31], we calculated how the use of Kinesia to measure change to an intervention could permit reduction in sample size requirement (Table 1).

Kinesia was also capable of capturing gradual changes in parkinsonian severity in response to increasing stimulation voltage output. At an individual level (Figure 3), significant changes in response to DBS, as determined by Tukey's HSD test, were detected at lower stimulation voltages by use of Kinesia recordings than with the clinician UPDRS or MBRS scores.

Discussion

The higher-resolution Kinesia portable kinematic system was able to detect changes in response to small adjustments in DBS output voltages, from lower stimulation amplitudes and with greater reliability and sensitivity than from clinicians rating the same subjects. Although the changes in finger tapping and tremor severities observed as IPG output was gradually increased may be an artificial way to manipulate parkinsonian severity and could be due to a number of factors (e.g., medication wearing off, DBS after effects), the observations made permitted a relatively small number of subjects to demonstrate a discrepancy between the reliability and sensitivity of changes between expert clinician raters versus Kinesia across tasks.

While several studies have shown relatively high test-retest reliability for the UPDRS motor section (UPDRS-Part III) as a whole [12], this widely-used method of clinical assessment has been somewhat problematic for clinical trials. The required presence of a trained professional to carry out ratings imposes costs in addition to requiring subjects to attend clinics for assessment. Also, multi-center clinical trials that utilize several clinicians at different sites can increase variability of ratings, resulting in decreased sensitivity to changes (responsiveness). Finally, the discrete nature of the UPDRS (a semi-quantitative scoring scheme using integer increments from 0 to 4) is unable to capture subtle changes that might occur during the slow progression of PD or during a clinical trial testing symptomatic or

neuroprotective interventions [11,34,35]. Studies with possible neuroprotective drugs have generally targeted patients afflicted with relatively mild features of PD, in whom the UPDRS scores for several motor endpoints may fall between 0 and 1. In addition to low sensitivity, a poor-reliability scale reduces the signal-to-noise ratio in some clinical trials; the result is a requirement for studies with increased sample sizes and long durations in order to separate the disease-modifying versus symptomatic effects between interventions.

While reliability of the UPDRS-Part III as a whole may be adequate for some studies, certain rating items, such as those related to bradykinesia, suffer from poor intra- and inter-rater reliability [36–38]. In an attempt to improve on precision of rating, for example, the Movement Disorder Society (MDS) revision of the UPDRS scoring guidelines specify measured tremor amplitude ranges that correspond to the scores [26]. However, even if precise judgment of tremor amplitude by visual inspection were actually possible, converting a wide range of tremor amplitudes to an integer score, as required for utilizing the MDS-UPDRS Part III, greatly reduces the resolution of this assessment. The scoring of other UPDRS Part III items (finger tapping, hand movements, and pronation-supination) tends to be even less reliable since raters must account for several attributes of movement: speed, amplitude, rhythm, hesitations, freezing, and decrementing, all of which are combined into a single score. The MBRS was found to be more sensitive than the UPDRS in identifying how different aspects of bradykinesia respond to dopaminergic medication [39,40]. However, the MBRS showed similar inter- and intra-rater reliability to those of UPDRS scores [2], highlighting the need for a more sensitive assessment of bradykinesia such as an instrumented rather than a clinician-dependent measurement. Kinesia recordings were shown previously to provide a more sensitive measure than clinician scores for capturing improvements in amplitude and rhythm resulting from dopaminergic medication treatment [40]. The results of the present study demonstrate that the test-retest reliability and sensitivity of Kinesia are comparable to those of expert clinicians for assessing rest and postural tremor, but are significantly better than clinician ratings for evaluating speed, amplitude, and rhythm of finger tapping.

Most studies examining test-retest reliability for PD assessments have examined the UPDRS Part III as a whole (ICC = 0.9) [12] or in grouped subscales (e.g., bradykinesia, rest tremor, with ICCs ranging from 0.63–0.92 [41]) rather than examining individual UPDRS items. Utilizing only total UPDRS scores as an outcome measure may mask small but meaningful changes in individual clinical signs that might be quite sensitive to therapeutic interventions. Additionally, some clinical features of PD are represented by multiple UPDRS items describing different body parts (e.g., rest tremor) while others are represented solely by a single score (e.g., speech); this situation leads to a differential weighting of UPDRS motor scores. Since individual UPDRS items are restricted to integer-based resolution (0, 1, 2, 3, or 4), each encompassing a wide range of severities, the ability to gauge subtle changes is limited and the efficacy of potential neuroprotective agents or mild symptomatic benefits may be missed. Figure 3 illustrates this matter. With clinician-rated UPDRS scores (Figure 3A), the ratings in response to DBS are manifested as binary “on” or “off” states. The MBRS (Figures 3B and 3C) improves upon this rating limitation by separating speed and amplitude; however, significant changes in response to DBS are still not detected until the IPG voltage output is close to the previously-determined optimal stimulation voltage. Only

the highly sensitive Kinesia rating scores (Figures 3D and 3E) make it evident that measurement of clinical signs can show subtle, but significant, changes in response to the gradations of DBS voltage. We acknowledge that these slight changes may not be observed by the patient. However, the ability to capture small changes in parkinsonian state may be of critical value in determining true efficacy of promising therapies [11].

In any rating system, measurement errors can manifest as inconsistencies caused by a study participant's physical or mental condition, variations in the testing procedure, or tester error (by either clinician or machine) [12]. In the present study, variability from moment to moment in clinical features of parkinsonism or measurement inaccuracy (device or clinician) may have contributed to both the ICC and MDC. It is important to take both sources of error into account when making comparisons across studies.

An important implication of available rating methods for PD is the effect they have on the magnitude of observations needed for demonstrating outcomes of therapeutic interventions in clinical trials. As illustrated in Table 1 for bradykinesia, increasing the reliability of the outcome measurement can reduce the required number of subjects to demonstrate a statistically significant outcome. Computerized digital systems can also improve the overall efficiency of clinical trials by permitting at-home assessments [4,15–17]. This option can minimize patient burden for participation in clinical trials and would provide real-time data accessibility for review and analysis.

Acknowledgments

This work was supported by NIH/NINDS under grant 1R43NS074627-01A1. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Full Financial Disclosures of all Authors for the Past Year: Dr. Heldman has received grant support from NIH/NINDS 1R43NS074627-01A1, NIH/NINDS 2R44NS065554-04A1, NIH/NINDS 1R43NS07765201A1, NIH/NINDS 1R43NS081902-01, and NIH/NIA 5R44AG03470803 and compensation from Great Lakes NeuroTechnologies for employment.

Dr. Espay is supported by the K23 career development award (NIMH, 1K23MH092735); has received grant support from CleveMed/Great Lakes NeuroTechnologies, and the Michael J. Fox Foundation for Parkinson's Research; personal compensation as a consultant/scientific advisory board member for Solvay (now Abbvie), Chelsea Therapeutics, TEVA, Impax, Merz, Solstice Neurosciences, Eli Lilly, and USWorldMeds; royalties from Lippincott Williams & Wilkins and Cambridge; and honoraria from Novartis, UCB, TEVA, the American Academy of Neurology, and the Movement Disorders Society. He serves as Associate Editor of Movement Disorders and Frontiers in Movement Disorders, and on the editorial board of The European Neurological Journal.

Dr LeWitt has served on scientific advisory boards or as a consultant for Adamas, Civitas, Concit, Depomed, Impax, Intec, Ipsen, Knopp Biosciences, Merck, Merz, NeuroDerm, ProStrakan, Prothera, Teva, UCB, USWorldMeds, and XenoPort, and has received speaker honoraria from Ipsen, Lundbeck, Teva, USWorldMeds, and the Movement Disorder Society. He is compensated as Editor-in-Chief of *Clinical Neuropharmacology* and is not compensated for serving on the editorial board of Journal of Neural Transmission, Translational Neurodegeneration, and *Journal of Parkinson's Disease*. The Parkinson's Disease and Movement Disorders Program that Dr. LeWitt directs has received clinical research grant support (for conducting clinical trial and other research) from Adamas, Addex, Biotie, Civitas, Great Lakes NeuroTechnologies, The Michael J. Fox Foundation for Parkinson's Research, Merck, Merz, UCB, USWorldMeds, and XenoPort.

Dr. Giuffrida has received grant support from NIH/NINDS 1R43NS074627-01A1, NIH/NIA 5R44AG03394704, NIH/NIMHD 5R44MD00404904, NIH/NINDS 1R44 NS07356101A1, NIH/NINDS 1R43NS07605201A1, and NIH/NIA 7R44AG03352004 and compensation from Great Lakes NeuroTechnologies for employment.

References

1. Giuffrida JP, Riley D, Maddux B, Heldman DA. Clinically deployable Kinesia technology for automated tremor assessment. Mov Disord. 2009; 24(5):723–30. [PubMed: 19133661]
2. 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. Mov Disord. 2011; 26(10):1859–63. [PubMed: 21538531]
3. Heldman DA, Jankovic J, Vaillancourt DE, Prodoehl J, Elble RJ, Giuffrida JP. Essential tremor quantification during activities of daily living. Parkinsonism Relat Disord. 2011; 17(7):537–42. [PubMed: 21570891]
4. Mera TO, Heldman DA, Espay AJ, Payne M, Giuffrida JP. Feasibility of home-based automated Parkinson's disease motor assessment. J Neurosci Methods. 2012; 203(1):152–6. [PubMed: 21978487]
5. Mera TO, Burack MA, Giuffrida JP. Objective motion sensor assessment highly correlated with scores of global levodopa-induced dyskinesia in Parkinson's disease. J Park Dis. 2013; 3(3):399–407.
6. Hoffman, JD.; McNames, J. 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. EMBC; 2011. Objective measure of upper extremity motor impairment in Parkinson's disease with inertial sensors; p. 4378-4381.
7. Caligiuri MP, Tripp RM. A portable hand-held device for quantifying and standardizing tremor assessment. J Med Eng Technol. 2004; 28(6):254–62. [PubMed: 15513743]
8. Zampieri C, Salarian A, Carlson-Kuhta P, Aminian K, Nutt JG, Horak FB. The instrumented timed up and go test: potential outcome measure for disease modifying therapies in Parkinson's disease. J Neurol Neurosurg Psychiatry. 2010; 81(2):171–6. [PubMed: 19726406]
9. Patel S, Lorincz K, Hughes R, Huggins N, Growdon J, Standaert D, et al. Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors. IEEE Trans Inf Technol Biomed. 2009; 13(6):864–73. [PubMed: 19846382]
10. Hauser RA, Deckers F, Lehert P. Parkinson's disease home diary: Further validation and implications for clinical trials. Mov Disord. 2004; 19(12):1409–13. [PubMed: 15390057]
11. Ahlskog JE, Uitti RJ. Rasagiline, Parkinson neuroprotection, and delayed-start trials. Neurology. 2010; 74(14):1143–1148. [PubMed: 20368634]
12. Steffen T, Seney M. Test-Retest Reliability and Minimal Detectable Change on Balance and Ambulation Tests, the 36-Item Short-Form Health Survey, and the Unified Parkinson Disease Rating Scale in People With Parkinsonism. Phys Ther. 2008; 88(6):733–46. [PubMed: 18356292]
13. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res Natl Strength Cond Assoc. 2005; 19(1):231–40.
14. Papapetropoulos S (Spyros). Patient diaries as a clinical endpoint in Parkinson's disease clinical trials. CNS Neurosci Ther. 2012; 18(5):380–7. [PubMed: 22070400]
15. Maetzler W, Domingos J, Srulijes K, Ferreira JJ, Bloem BR. Quantitative wearable sensors for objective assessment of Parkinson's disease. Mov Disord. 2013; 28(12):1628–37. [PubMed: 24030855]
16. Pulliam CL, Eichenseer SR, Goetz CG, Waln O, Hunter CB, Jankovic J, et al. Continuous in-home monitoring of essential tremor. Parkinsonism Relat Disord. 2014; 20(1):37–40. [PubMed: 24126021]
17. Zampieri C, Salarian A, Carlson-Kuhta P, Nutt JG, Horak FB. Assessing mobility at home in people with early Parkinson's disease using an instrumented Timed Up and Go test. Parkinsonism Relat Disord. 2011; 17(4):277–80. [PubMed: 20801706]
18. Keijsers NL, Horstink MW, Gielen SC. Movement parameters that distinguish between voluntary movements and levodopa-induced dyskinesia in Parkinson's disease. Hum Mov Sci. 2003; 22(1): 67–89. [PubMed: 12623181]
19. Horak FB, Mancini M. Objective biomarkers of balance and gait for Parkinson's disease using body-worn sensors. Mov Disord. 2013; 28(11):1544–51. [PubMed: 24132842]

20. Spain RI, St. George RJ, Salarian A, Mancini M, Wagner JM, Horak FB, et al. Body-worn motion sensors detect balance and gait deficits in people with multiple sclerosis who have normal walking speed. Gait Posture. 2012; 35(4):573–8. [PubMed: 22277368]
21. Mancini M, Horak FB, Zampieri C, Carlson-Kuhta P, Nutt JG, Chiari L. Trunk accelerometry reveals postural instability in untreated Parkinson's disease. Parkinsonism Relat Disord. 2011; 17(7)::557–62. [PubMed: 21641263]
22. Nanhoe-Mahabier W, Allum JH, Pasman EP, Overeem S, Bloem BR. The effects of vibrotactile biofeedback training on trunk sway in Parkinson's disease patients. Parkinsonism Relat Disord. 2012; 18(9):1017–21. [PubMed: 22721975]
23. Rochester L, Baker K, Hetherington V, Jones D, Willems A-M, Kwakkel G, et al. Evidence for motor learning in Parkinson's disease: Acquisition, automaticity and retention of cued gait performance after training with external rhythmical cues. Brain Res. 2010; 1319:103–11. [PubMed: 20064492]
24. Elble RJ. Gravitational artifact in accelerometric measurements of tremor. Clin Neurophysiol Off J Int Fed Clin Neurophysiol. 2005; 116(7):1638–43.
25. Temperli P, Ghika J, Villemure J-G, Burkhard PR, Bogousslavsky J, Vingerhoets FJG. How do parkinsonian signs return after discontinuation of subthalamic DBS? Neurology. 2003; 60(1):78–81. [PubMed: 12525722]
26. 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. Mov Disord. 2008; 23(15):2129–70. [PubMed: 19025984]
27. Kishore A, Espay AJ, Marras C, Al-Khairalla T, Arenovich T, Asante A, et al. Unilateral versus bilateral tasks in early asymmetric Parkinson's disease: Differential effects on bradykinesia. Mov Disord. 2007; 22(3):328–33. [PubMed: 17216641]
28. Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. Phys Ther. 2006; 86(5):735–43. [PubMed: 16649896]
29. Donoghue D, Stokes EK. How much change is true change? The minimum detectable change of the Berg Balance Scale in elderly people. J Rehabil Med. 2009; 41(5):343–6. [PubMed: 19363567]
30. Quinn, GP.; Keough, MJ. Experimental Design and Data Analysis for Biologists. Cambridge University Press; 2002. p. 199-200.
31. Perkins DO, Wyatt RJ, Bartko JJ. Penny-wise and pound-foolish: the impact of measurement error on sample size requirements in clinical trials. Biol Psychiatry. 2000; 47(8):762–6. [PubMed: 10773186]
32. Altman DG, Bland JM. Measurement in Medicine: The Analysis of Method Comparison Studies. J R Stat Soc Ser Stat. 1983; 32(3):307–17.
33. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet. 1986; 327(8476):307–10.
34. Elm JJ, Goetz CG, Ravina B, Shannon K, Wooten GF, Tanner CM, et al. A responsive outcome for Parkinson's disease neuroprotection futility studies. Ann Neurol. 2005; 57(2):197–203. [PubMed: 15668964]
35. Kieburtz K. Issues in neuroprotection clinical trials in Parkinson's disease. Neurology. 2006; 66(10 suppl 4):S50–S57. [PubMed: 16717252]
36. Bennett DA, Shannon KM, Beckett LA, Goetz CG, Wilson RS. Metric properties of nurses' ratings of parkinsonian signs with a modified Unified Parkinson's Disease Rating Scale. Neurology. 1997; 49(6):1580–7. [PubMed: 9409350]
37. Camicioli R, Grossmann SJ, Spencer PS, Hudnell K, Anger WK. Discriminating mild parkinsonism: Methods for epidemiological research. Mov Disord. 2001; 16(1):33–40. [PubMed: 11215590]
38. Siderowf A, McDermott M, Kieburtz K, Blindauer K, Plumb S, Shoulson I. Test-retest reliability of the unified Parkinson's disease rating scale in patients with early Parkinson's disease: results from a multicenter clinical trial. Mov Disord. 2002; 17(4):758–63. [PubMed: 12210871]

39. Espay AJ, Beaton DE, Morgante F, Gunraj CA, Lang AE, Chen R. Impairments of speed and amplitude of movement in Parkinson's disease: a pilot study. Mov Disord. 2009; 24(7):1001–8. [PubMed: 19230031]

40. Espay AJ, Giuffrida JP, Chen R, Payne M, Mazzella F, Dunn E, et al. Differential response of speed, amplitude, and rhythm to dopaminergic medications in Parkinson's disease. Mov Disord. 2011; 26(14):2504–8. [PubMed: 21953789]

41. Metman LV, Myre B, Verwey N, Hassin-Baer S, Arzbaecher J, Sierens D, et al. Test-retest reliability of UPDRS-III, dyskinesia scales, and timed motor tests in patients with advanced Parkinson's disease: an argument against multiple baseline assessments. Mov Disord. 2004; 19(9): 1079–84. [PubMed: 15372601]

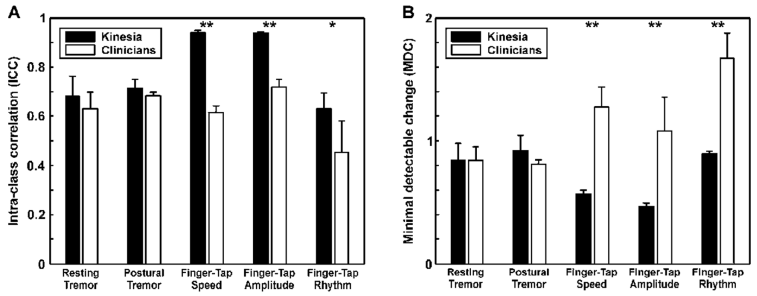

Figure 1. Intraclass correlation and minimal detectable change

The average ICCs (A) and MDCs (B) are plotted for the Kinesia and clinician scores. The metrics were calculated separately for each clinician and averaged together. Error bars correspond to the standard deviation across each combination of repetitions. *p<0.05, **p<0.0001. ICC: intraclass correlation; MDC: minimal detectable change.

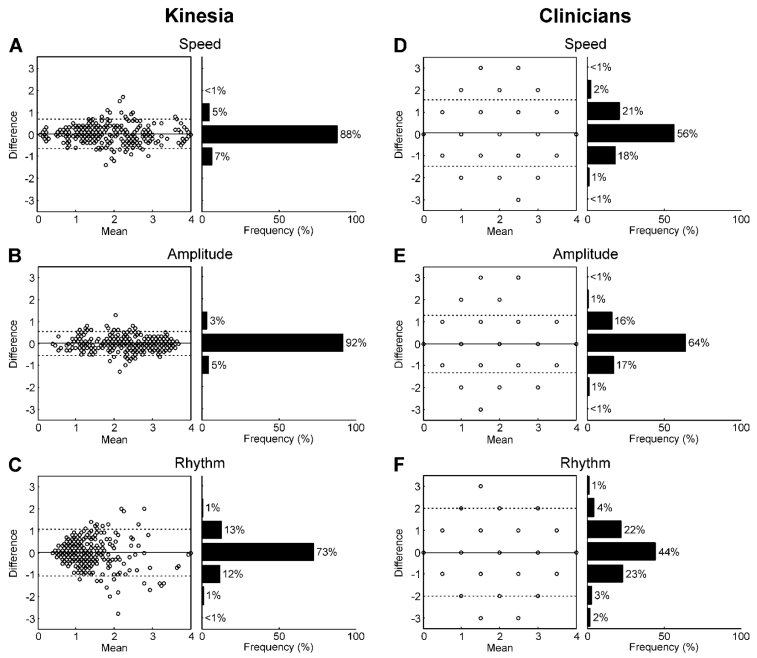

Figure 2.

Bland-Altman plots are shown for the finger-tapping subscores given by Kinesia (**A–C**) and both clinicians (**D–F**). The difference is plotted versus the mean of repetition 1 and 2, 2 and 3, and 1 and 3. The solid line indicates the average difference and the dotted lines indicate the 95% limits of agreement. Frequency histograms are shown to the right of each plot since several comparisons have the same value.

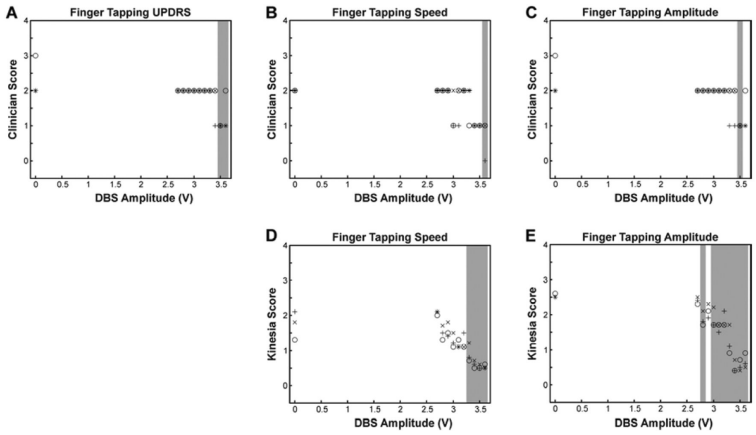

Figure 3. Sensitivity assessment

Clinician UPDRS – Part 3 score for finger tapping (**A**), Clinician MBRS ratings of finger tapping speed (**B**), Clinician ratings of MBRS finger tapping amplitude (**C**), Kinesia ratings of finger tapping speed (**D**), and, Kinesia ratings of finger tapping amplitude (**E**). Scores are plotted for a single subject with DBS turned off (0V) and voltage amplitude gradually increased to its established optimal setting during the three repetitions of the finger tapping task. The x's, +'s, and O's correspond to scores from the first, second, and third repetitions of ratings. The gray shading on each plot indicates voltage amplitudes that result in scores significantly different from baseline at the alpha = 0.05 significance level using Tukey's HSD (“honestly significant difference”) test.

Table 1

Comparison of UPDRS- versus Kinesia-based sample size calculation according to ICC differences

Parkinsonian feature	Clinician ICC	Kinesia ICC	Percent fewer subjects	Number of subjects based on Clinician	Number of subjects based on Kinesia
Rest Tremor	0.63	0.68	7.5	100	93
Postural Tremor	0.68	0.71	3.9	100	96
Speed	0.62	0.94	34.6	100	65
Amplitude	0.72	0.94	23.3	100	77
Rhythm	0.45	0.63	28.3	100	72

Based on ICC differences and the model described in Perkins et al. [31], Kinesia could reduce the number of subjects required for a clinical trial compared to trials using clinician UPDRS scores (numbers in the two right-most columns assume 100 subjects would be required if clinician scores were used).