Test-Retest Reliability of a Parkinson's Disease Monitoring System

Dustin A. Heldman[1], Alberto J. Espay[2], Peter A. LeWitt[3], Joseph P. Giuffrida[1]

[1]Great Lakes NeuroTechnologies Inc., Cleveland, OH, USA [2]University of Cincinnati Neuroscience Institute, Department of Neurology, Gardner Center for Parkinson's disease and Movement Disorders, University of Cincinnati, Cincinnati, OH, USA. [3]Departments of Neurology, Henry Ford Hospital and Wayne State University School of Medicine, West Bloomfield, MI USA.

Introduction

Large-scale clinical trials of new symptomatic and neuroprotective treatments of Parkinson's disease (PD) often involve dozens of sites and thousands of patients. Outcome measures include clinical assessments completed at weekly or monthly intervals, which can suffer from biases, placebo effects (subject and investigator), limited resolution, and poor intra- and inter-rater reliability. Enhancing the reliability and sensitivity of motor assessments required to demonstrate therapeutic efficacy of selected interventions is an unmet need of PD clinical trials. The objective of this study is to determine the reliability and sensitivity of a motion-sensor system for quantifying PD motor deficits compared to clinical ratings.

Experimental Methods

Eighteen PD patients (13 males; mean age 63.1±8.4 years, range: 44-76) with subthalamic nucleus deep brain stimulation (DBS) performed an automated motor assessment using the Kinesia™ (Great Lakes NeuroTechnologies, Cleveland, OH) portable kinematic system (Figure 1). Kinesia, which outputs motor scores on a 0-4 scale with 0.1 resolution, has been previously described and validated for assessing tremor, bradykinesia, and dyskinesia associated with PD.

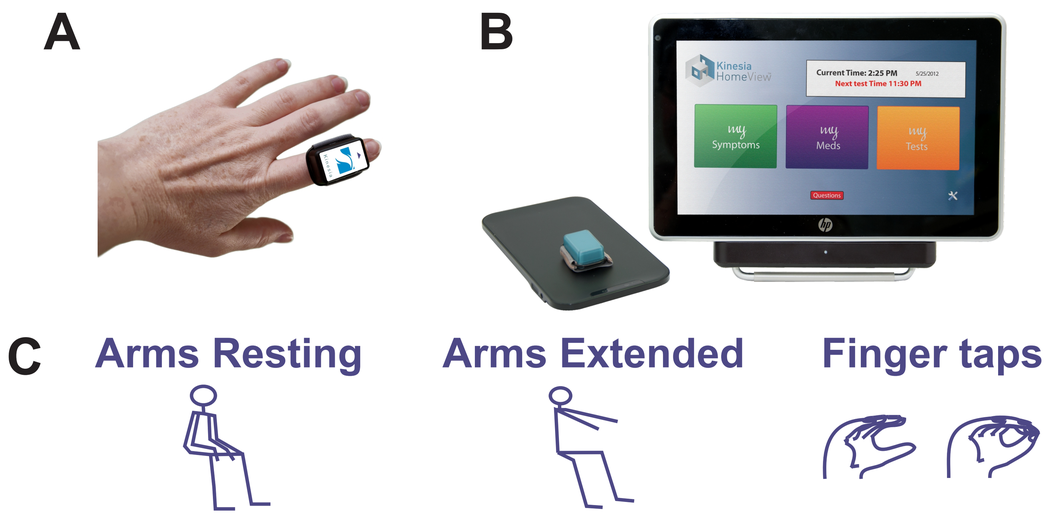

Figure 1. Kinesia includes a wireless finger-worn motion-sensor unit (**A**) and a touch screen tablet PC with inductive charge (**B**). The automated motor assessment included three 15-second tasks (**C**). The first two tasks were assessments of rest and postural tremor, while the third task involved repetitive finger tapping to evaluate bradykinesia (speed), hypokinesia (amplitude), and dysrhythmia (rhythm).

Deep Brain Stimulation

The sequence of tasks was performed three separate times (to ascertain test-retest reliability) with DBS turned off and at each of ten DBS stimulation amplitudes (0.9V below the predetermined optimal stimulation amplitude increasing in 0.1V increments through optimal stimulation amplitude) aimed at yielding small changes in treatment response (Figure 2).

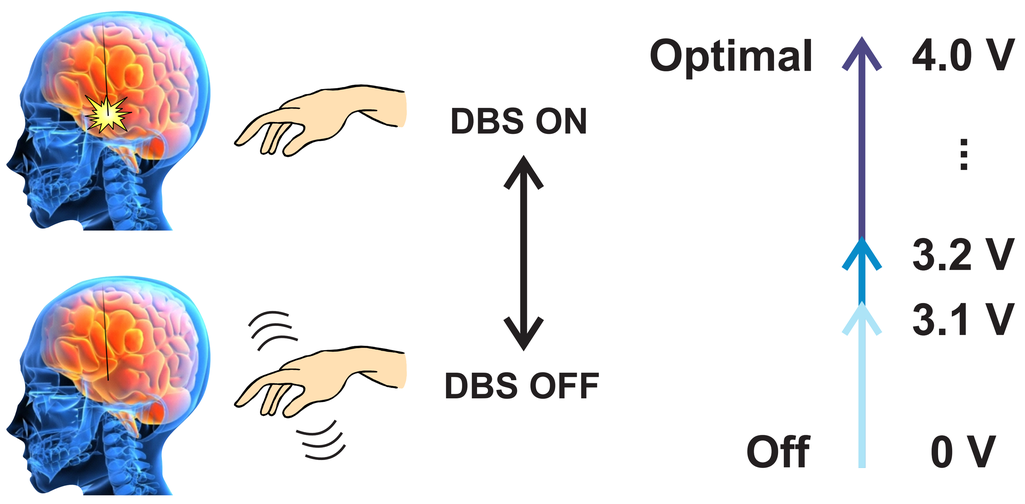

Figure 2. DBS voltage output amplitude was slowly adjusted to modulate the severity of parkinsonism during testing sessions. In this example the subject's previously-determined optimal stimulation amplitude was 4.0 V. Therefore, after the baseline (0 V) assessment, DBS amplitude was set to 3.1 V, then 3.2 V, etc., all the way up to 4.0 V, with automated motor assessments performed at each stimulation amplitude.

Each task performance was video-recorded for subsequent blinded clinical rating according to Unified Parkinson's Disease Rating Scale (UPDRS) and Modified Bradykinesia Rating Scale (MBRS) criteria. Test-retest reliability was calculated as intraclass correlation (ICC) and sensitivity was calculated as minimal detectable change (MDC) for each stimulation amplitude.

Reliability and Sensitivity

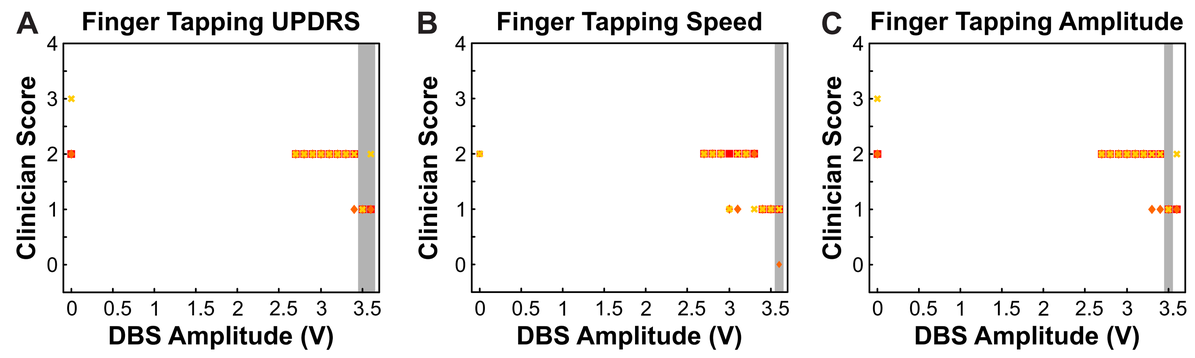

Figure 3. DBS Modulation. Clinician UPDRS – Part 3 score for finger tapping (**A**), Clinician MBRS ratings of finger tapping speed (**B**), Clinician ratings of MBRS finger tapping amplitude (**C**), Kinesia ratings of finger tapping speed (**D**), and, Kinesia ratings of finger tapping amplitude (**E**). 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 red squares, orange diamonds, and yellow x'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.

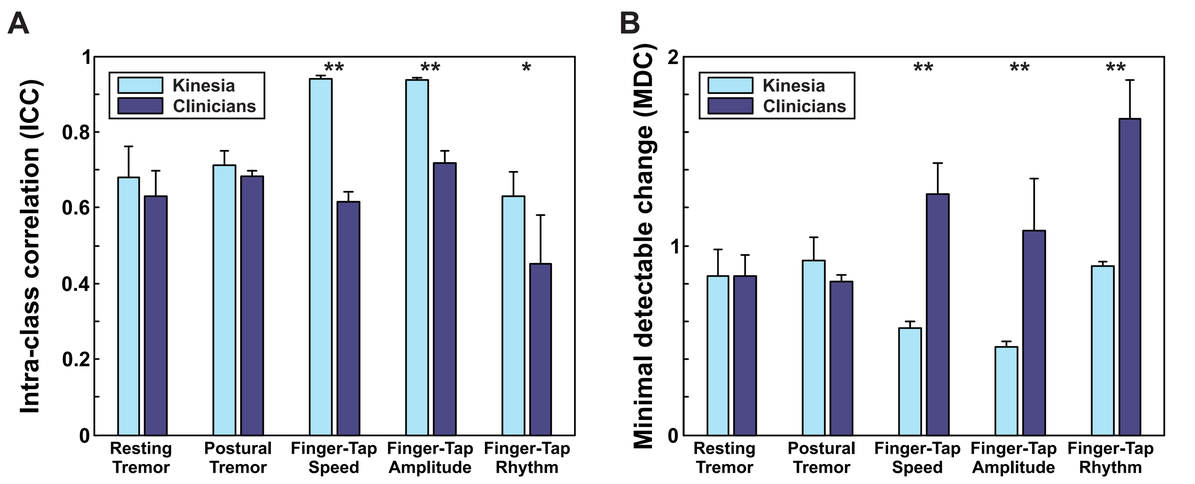

Figure 4. Intraclass correlation and minimal detectable change. The average ICCs (A) and MDCs (B) are plotted for the Kinesia and clinician scores. The metrics were calculated separately for each clinician and averaged together. Error bars correspond to the standard deviation across each combination of repetitions. *p<0.05, **p<0.0001. ICC: intraclass correlation; MDC: minimal detectable change.

Table 1. Comparison of UPDRS- versus Kinesia-based sample size calculations

	Clinician ICC	Kinesia ICC	Percent fewer subjects	Number of subjects based on Clinicians	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

Kinesia would reduce the number of subjects required for a clinical trial by up to 35% compared to trials using clinician UPDRS scores (numbers in the right-most column assume 100 subjects would be required if clinician scores were used).

Conclusions

- Motion sensors can provide increased test-retest reliability and sensitivity over clinical assessments.
- The increased sensitivity and reliability afforded by motion sensors over clinical assessments **can decrease the number of subjects, shorten the duration, and lower the costs** required detect significant outcomes in clinical trials.
- Home-based motion sensor monitoring can improve temporal resolution in addition to score resolution.

This work was supported by NIH/NINDS 1R43NS074627-01A1.