Parkinsonism and Related Disorders xxx (2014) 1–6

Contents lists available at ScienceDirect

Parkinsonism and Related Disorders

journal homepage: www.elsevier.com/locate/parkreldis

Clinician versus machine: Reliability and responsiveness of motor endpoints in Parkinson's disease

Dustin A. Heldman [a,*], Alberto J. Espay [b], Peter A. LeWitt [c], Joseph P. Giuffrida [a]

[a] *Great Lakes NeuroTechnologies Inc., 10055 Sweet Valley Drive, Cleveland, OH 44125, USA*
[b] *UC Neuroscience Institute, Department of Neurology, Gardner Center for Parkinson's Disease and Movement Disorders, University of Cincinnati, Cincinnati, OH, USA*
[c] *Departments of Neurology, Henry Ford Hospital and Wayne State University School of Medicine, West Bloomfield, MI, USA*

ARTICLE INFO

Article history:
Received 11 November 2013
Received in revised form
12 February 2014
Accepted 24 February 2014

Keywords:
Parkinson's disease
UPDRS
Kinesia
Tremor
Bradykinesia

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 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.

1. 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 outpatient 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 intra- and 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

* Corresponding author. Tel.: +1 216 446 2437.
E-mail address: dheldman@glneurotech.com (D.A. Heldman).

http://dx.doi.org/10.1016/j.parkreldis.2014.02.022
1353-8020/

motion-sensor systems lack many of the problems that interfere with clinical assessment, their data are 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.

2. Methods

2.1. 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.

2.2. Experimental methods

Subjects had received stable and optimized oral medication and DBS treatment regimens prior to their evaluation (mean baseline DBS output voltage: 3.1 ± 0.74 V, range: 1.9–4.3 V). Testing was initiated at least 30 min 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-s 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 video-recorded 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 0.9 V 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 0.1 V, 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 11 DBS voltage settings (that is, with the IPG turned off and at ten stimulation amplitudes). Other than the 30-min 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.

2.3. Clinician scoring

The video-recordings of each subject's hand movements for the three repetitions at each of the 11 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 tremors 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].

2.4. Reliability and sensitivity analysis

Test–retest reliability (or *consistency*) of clinician and Kinesia scores were 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:

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

where 1.65 reflects the 90% confidence interval, SD is the standard deviation, $\sqrt{2}$ 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].

2.5. 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.

3. 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 (Fig. 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 (Fig. 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 (Fig. 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 (Fig. 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.

4. Discussion

The higher-resolution Kinesia portable kinematic system was able to detect changes in response to small adjustments in DBS