

Validity and Responsiveness of At-Home Touch Screen Assessments in Advanced Parkinson's Disease

Mevludin Memedi, Dag Nyholm, Anders Johansson, Sven Pålhagen, Thomas Willows, Håkan Widner, Jan Linder, and Jerker Westin

Abstract—The aim of this study was to investigate if a telemetry test battery can be used to measure effects of Parkinson's disease (PD) treatment intervention and disease progression in patients with fluctuations. Sixty-five patients diagnosed with advanced PD were recruited in an open longitudinal 36-month study; 35 treated with levodopa-carbidopa intestinal gel (LCIG) and 30 were candidates for switching from oral PD treatment to LCIG. They utilized a test battery, consisting of self-assessments of symptoms and fine motor tests (tapping and spiral drawings), four times per day in their homes during week-long test periods. The repeated measurements were summarized into an overall test score (OTS) to represent the global condition of the patient during a test period. Clinical assessments included ratings on unified PD rating scale (UPDRS) and 39-item PD questionnaire (PDQ-39) scales. In LCIG-naïve patients, the mean OTS compared to baseline was significantly improved from the first test period on LCIG treatment until month 24. In LCIG-nonnaïve patients, there were no significant changes in the mean OTS until month 36. The OTS correlated adequately with total UPDRS ($\rho = 0.59$) and total PDQ-39 (0.59). Responsiveness measured as effect size was 0.696 and 0.536 for OTS and UPDRS, respectively. The trends of the test scores were similar to the trends of clinical rating scores but the dropout rate was high. Correlations between OTS and clinical rating scales were adequate indicating that the test battery contains important elements of the information of well-established scales. The responsiveness and reproducibility were better for OTS than for total UPDRS.

Index Terms—Home assessment, levodopa infusion, Parkinson's disease (PD), remote monitoring, telemedicine.

I. INTRODUCTION

MEASURING symptoms- and treatment-related changes in advanced Parkinson's disease (PD) with motor fluctuations is challenging. The current state of the art in clinical trials is to use the unified PD rating scale (UPDRS), different

Manuscript received April 2, 2015; revised July 12, 2015; accepted August 7, 2015. Date of publication August 13, 2015; date of current version November 3, 2015. This work was supported in part by the Swedish Knowledge Foundation, Abbot Product Operations AG, Allschwil, Switzerland, and Nordforce Technology AB, Stockholm, Sweden.

M. Memedi and J. Westin are with the School of Technology and Business Studies, Computer Engineering, Dalarna University, Falun 791 88, Sweden (e-mail: mmi@du.se; jwe@du.se).

D. Nyholm is with the Department of Neuroscience, Neurology, Uppsala University, Uppsala 751 85, Sweden (e-mail: dag.nyholm@neuro.uu.se).

A. Johansson is with the Department of Clinical Neuroscience, Neurology, Karolinska Institutet, Stockholm 14186, Sweden (e-mail: anders.c.johansson@karolinska.se).

S. Pålhagen and T. Willows are with the Department of Neurology, Karolinska University Hospital, Stockholm 171 77, Sweden (e-mail: sven.palhagen@karolinska.se; thomas.willows@karolinska.se).

H. Widner is with the Department of Neurology, Skåne University Hospital, Lund 221 85, Sweden (e-mail: hakan.widner@med.lu.se).

J. Linder is with the Department of Pharmacology and Clinical Neuroscience, Umeå University, Umeå 901 87, Sweden (e-mail: jan.linder@neuro.umu.se).

Digital Object Identifier 10.1109/JBHI.2015.2468088

scales rating health-related quality of life such as the 39-item PD Questionnaire (PDQ-39), and patient diaries recording time in off, in on and in on with dyskinesia. However, scales used to rate fluctuations can on an individual level be unreliable, because when they are asked the patients do not exactly remember how their condition has been. Patient diaries capture fluctuations better, but even these are often not filled out at the correct time [1]. Frequent, time-stamped data capture is necessary to get reliable data and to prevent recall bias [2], and patients' subjective ratings may not correlate to measurements of improvement [3]. This suggests that combining elements of common rating scales with frequent subjective and objective, observer-independent tests before and after a treatment intervention will cover more aspects of outcome than what can be achieved by the established scales alone. In contrast to the standard clinical approaches for assessment and follow-up of PD symptoms, telemedicine methods provide means for remote, long-term, and repeated assessments of symptoms and their fluctuations potentially improving the accessibility and efficiency of care and increasing the patient compliance [4]. These methods can be based on data gathered by wearable sensors [5] [6] and smartphone technology [7]. However, these methods have not yet to any greater extent found their way to routine clinical settings.

A new real-time electronic patient diary with motor tests for telemetering and accurately capturing symptom fluctuations, aimed at providing valid data during fluctuations was evaluated. This test battery consists of repeated measurements of self-assessments of common PD symptoms and objective measures of the motor function (tapping and spiral drawing), for use in patients' home environments. The tests were applied as secondary outcome measures in a prospective 36-month study of levodopa-carbidopa intestinal gel (LCIG) in patients with advanced PD [8].

The present analysis was performed with the following objectives: 1) to determine if PD progression over time can be followed in results of the test battery, 2) to assess sensitivity of the test results in relation to treatment intervention, 3) to assess the relative ability of the test scores and clinical rating scores to detect change, 4) to assess their correlations, and 5) to estimate reproducibility.

II. MATERIALS AND METHODS

A. Patients

Seventy-seven patients with advanced PD were recruited in an open-longitudinal 36-month study at ten hospitals in Sweden and Norway from January 2006 to August 2010 [8]. On

TABLE I
CLINICAL FEATURES AT BASELINE

	LCIG-naïve	LCIG-nonnaïve	Combined
Patients (n, gender)	37 (23m; 14f)	40 (27m; 13f)	77 (50m; 27f)
Age (years)	63 ± 10	65.5 ± 10	65 ± 11
Years with levodopa	10 ± 5	15 ± 8.25	13 ± 7
Years with LCIG		1.96 ± 1.84	
Hoehn and Yahr stage	2.5 ± 1	2.75 ± 2	2.5 ± 1
Total UPDRS	47 ± 18	48 ± 17.5	48 ± 19

Data presented as median ± interquartile range. Abbreviations: m; male, f; female.

inclusion, 40 of them were treated with LCIG (Duodopa, AbbVie) and 37 patients were candidates for switching from conventional oral PD treatment to LCIG [9]. In the latter group, patients were LCIG treatment-naïve at study start. The LCIG dose was adjusted to obtain optimal clinical response for the individual subject by maximizing their functional on time during the day, minimizing their number of off episodes and bradykinesia, and minimizing on time with dyskinesia [8]. Written informed consent was given and the regional ethical review board (in Stockholm, Sweden), after reviewing the study protocol, determined that their approval was not required. Patient characteristics at the baseline are shown in Table I.

B. Telemetry System

Patients performed repeated, time-stamped, and remote assessments of their subjective and objective health indicators using a wireless telemetry test battery implemented on a touch screen handheld device [10]. The test battery consisted of a patient diary section for collecting self-assessments of common PD symptoms, a motor test section for collecting objective measures of the motor function through a set of upper limb motor tests, and a scheduler for restricting the operation to a multitude of predetermined limited time slots. The overall system was implemented in a client-server architecture for collecting and wirelessly transmitting the remote data to a central server for storage and offline processing. On each test occasion, the data are transmitted from the handheld device over a universal mobile telecommunication system to the server where so-called remote device manager (RDM, Nordforce Technology AB) software is executed. RDM is a commercial software platform aimed to provide a high availability data communication link with a very high security level over wireless Internet. Once the data are received at the server side, those are initially stored as separate extensible markup language files which in turn are interpreted by a custom-designed software in .NET. This software parses, processes, and stores both the raw and summarized data into relational database tables. The test battery was implemented on a Qtek 2020i Pocket PC device having a screen size of 60 mm × 80 mm.

C. Telemetry Assessments

Both the LCIG-naïve and LCIG-nonnaïve patients used the test battery four times per day during week-long test periods over the course of the study. In order to enable a reliable assess-

ment during fluctuations, the sample size was determined based on a target variance of an e-diary, as previously reported in the study performed by Nyholm *et al.* [2]. On each test occasion, patients were instructed to place the device on a table, be seated in a chair and use an ergonomic stylus to perform the tests. The LCIG-naïve patients used the test battery at baseline (before LCIG), month 0 (first visit; at least three months after permanent intraduodenal LCIG), and thereafter quarterly for the first year and biannually for the second and third years. The LCIG-nonnaïve patients used the test battery from the first visit, i.e., month 0.

Twelve of the 77 patients did not use the test battery because seven of them were not Swedish-speaking and five were unwilling to use the device or were thought unable to handle it. Out of the remaining 65 patients, 35 were LCIG-nonnaïve and 30 LCIG-naïve. Twenty-seven out of the 37 initially included LCIG-naïve patients started LCIG treatment during the study. In 20 LCIG-naïve patients, assessments with the test battery were available during oral treatment and at least one test period after having started infusion treatment. Three LCIG-naïve patients did not use the test battery at baseline but had at least one test period of assessments thereafter. Hence, $n = 23$ in the LCIG-naïve group. In total, symptom assessments in the full sample (including both LCIG-naïve and nonnaïve patients) were collected during 379 test periods and 10079 test occasions.

A more detailed description of the test battery and its included items can be found elsewhere [10], however an outline is provided here. On each test occasion, patients were first asked to answer the PD-related questions and then to perform the motor tests. There were seven self-assessment questions (q) relating to the previous four hours or this morning, including “ability to walk” (q1), “proportion of time spent in off, on, and dyskinetic” (q2), “off at worst” (q3), “dyskinetic at worst” (q4), “painful cramps” (q5), “satisfied with functioning” (q6), and “momentary motor condition from −3 (very off) to +3 (very dyskinetic)” (q7). The questions were selected based on two previously published e-diaries [2], [11]. The motor tests included different tapping tests and tracing a predrawn Archimedean spiral. The tapping tests consisted of four tests, including “alternate tapping using right hand” (q8), “alternate tapping using left hand” (q9), “tapping with increasing speed using dominant hand” (q10), and “tapping with random chasing using dominant hand” (q11). Patients were asked to perform tapping of square areas (“fields”) as shown on the screen of the device. The fields had a side of approximately 15 mm and at least one button was active. All fields became inactive after 20 s after the first field was pressed. For the spiral drawing test, patients traced a predrawn Archimedean spiral using the dominant hand, and the test was repeated three times per test occasion (q12-q14). An overview of the raw data items is illustrated by Fig. 3. Motor test designs were inspired by other tests and test batteries such as those described by Giovannoni *et al.* [12], Taylor Tavares *et al.* [13], Williams *et al.* [14], and Liu *et al.* [15]. Tests were performed once during four fixed time slots each day during test periods of one week. Data entry was blocked except during the four daily time slots of one hour’s length: 08:00–09:00, 12:00–13:00, 16:00–17:00, and 20:00–21:00.

D. Clinical Assessments

The clinical evaluation included the UPDRS and PDQ-39 scales, performed in afternoons at the start of each test period [8]. LCIG-naïve patients were evaluated at baseline and follow-up test periods, whereas LCIG-nonnaïve patients were evaluated starting from month 0. In total, clinical assessments were done during 369 test periods.

E. Data Processing

Tapping variables of the mean speed (number of taps per 20 s) and accuracy (percentage of correct taps) were calculated and used in subsequent analysis. The spiral drawings were processed by a method including wavelet transforms which generated a “spiral score” representing PD-related drawing impairments [16]. The scorings provided by this method resulted in strong correlations with averaged scorings of two trained neurologists. A “mean spiral score” (MSS) was defined as the mean of the three spiral scores per test occasion.

The week-long time series of the test battery items were processed and summarized into scores representing patients' symptom severities over a test period, using machine learning and statistical methods [17]. Initially, the time series were summarized into statistical features for representing symptom levels and variations followed by using them in data-driven modeling for deriving six conceptual symptom dimensions and an “overall test score” (OTS). There were four subjectively reported dimensions including “walking” (based on q1), “satisfied” (q6), “dyskinesia” (q2- dyskinesic and q4), and “off” (q2-off and q3) and two objectively measured including “tapping” (based on speed and accuracy of q8–q11) and “spiral” (MSS of q12–q14). The OTS was defined to represent the global health condition of the patient during a test period by combining the subjective and objective measures into a composite score, with the aim of providing similar information content as the total UPDRS. The OTS was defined as a linear combination of the first principal components of the six dimensions and their respective coefficients, estimated by the least-squares multiple linear regression model with total UPDRS as a dependent variable. The model was built using a stratified tenfold cross-validation approach by repeating the whole process, training, and testing, ten times with different random samples. The ranges of the OTS and dimensions were scaled from 0 to 1, where 0 represents bad function and 1 good function. After linear regression, the objectively measured dimensions (tapping and spiral) were assigned a higher weight (65%) than subjectively reported dimensions (walking, satisfied, off, and dyskinesia), similarly as the motor section of the UPDRS scale has the highest weight in the total UPDRS.

F. Statistical Methods

Change in the test battery scores over time, that is at baseline and follow-up test periods, was assessed with linear mixed-models [18] using a restricted maximum likelihood estimation method with patient ID as a random effect and test period as a fixed effect of interest. The dependent variables were the test battery scores; the raw test battery items, OTS, and dimensions. To assess the reproducibility of OTS and total UPDRS scores,

correlation coefficients between their scores during “month 3” and “month 6” test periods were calculated. The difference in time between these test periods was 3 months, which is short enough to not be much influenced by disease progression. The LCIG treatment had at this point been used for at least some months in all subjects so treatment settings were in general kept stable. The Wilcoxon signed-rank test was used to assess the difference in the median percentage absolute error (month 6–month 3) for OTS and total UPDRS. The relative ability to detect change from baseline (oral treatment) to month 0 (LCIG) of total UPDRS and OTS was determined by the effect size correlation (r), representing the magnitude of treatment effect, c.f., e.g., [19]. The r was calculated as the square root of the squared t-test value divided by the sum of squares of t-test value plus the degrees of freedom, using $r = \sqrt{(t^2/(t^2 + df))}$, where t (t-test value) and df (degrees of freedom) were extracted from LME models applied on data consisting of tests collected during baseline and month 0 test periods. The scale which had the highest r value was defined as the most sensitive to treatment response. Correlations between OTS, dimensions, and clinical rating scores were assessed using Spearman's rank correlation coefficients (ρ). Statistical analyses were carried out using the statistical software R.

III. RESULTS

The mean OTS scores at the baseline were 0.47 and 0.51 for LCIG-naïve and LCIG-nonnaïve patients, respectively (see Fig. 1). In LCIG-naïve patients, the mean OTS improved to the first test period on LCIG treatment and this improvement remained statistically significant until month 24. The patient number decreased successively during the study (see legend to Fig. 1).

The maximum improvement, 0.15 units (32%) higher than at the baseline ($p < 0.001$), was seen at month 3. Mean dimension scores of LCIG-naïve patients from baseline to the 36-month follow-up are shown in Fig. 2. The mean scores of all subjectively reported dimensions improved significantly throughout the course of the study, except “walking” at month 36 ($p = 0.41$, $n = 4$). However, there were no significant differences in mean scores of objectively measured dimensions between baseline and other test periods, except improved “tapping” at months 6 and 36, and “spirals” at month 3 ($p < 0.05$).

There were no significant changes in mean OTS scores, dimensions, or raw question items in LCIG-nonnaïve patients over the course of the study, except for worse mean OTS at month 36 ($p < 0.01$, $n = 16$) (see Fig. 1). Reasonably, similar trends were observed in mean scores of raw motor items between LCIG-naïve and LCIG-nonnaïve patients. Correlations between months 3 and 6 were stronger for OTS (0.91) than for total UPDRS (0.79). The median percentage absolute error was 10.5 for OTS and 19.7 for total UPDRS, and the difference between the two errors was significant ($p = 0.034$).

Fig. 3 shows mean scores of the raw test battery items of LCIG-naïve patients. The mean scores of raw question items (q1–q6) improved significantly, except for q5 (perceived painful cramps) at month 24 ($p = 0.18$). All changes were seen at the first LCIG period and were sustained over the study period. In

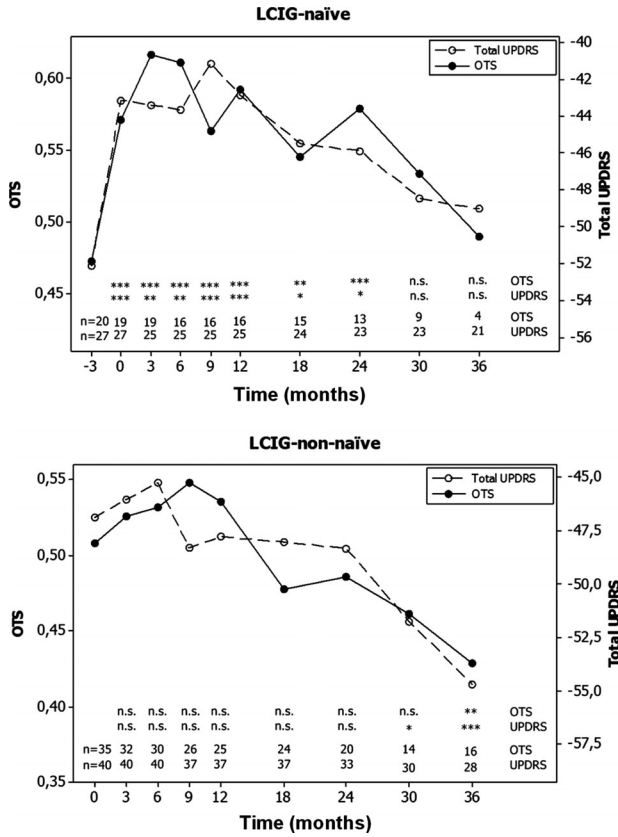


Fig. 1. Trends of mean OTS (solid line) and total UPDRS (dashed line) scores of LCIG-naïve and nonnaïve patients over the study duration. Y-axes: a high score means good function. Note that the total UPDRS axis is reversed. Significant mean differences are shown with respect to initial test periods, month-3 (before LCIG, upper graph) and month 0 (lower graph). * = $p < 0.05$, ** = $p < 0.01$, and *** = $p < 0.001$. Abbreviation: n.s., not significant.

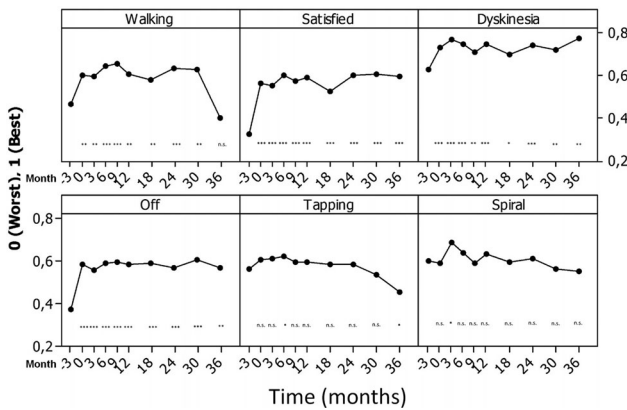


Fig. 2. Trends of mean dimension scores of LCIG-naïve patients over the study period. A high score represents good function. Significance in mean differences are shown with respect to baseline before LCIG (-3). * = $p < 0.05$, ** = $p < 0.01$, and *** = $p < 0.001$. Abbreviation: n.s., not significant. Number of observations in all graphs are the same as in Fig. 1, upper graph, OTS.

general, patients at the baseline had a tendency to self-rate their momentary motor conditions (q7) as being “slightly off” with a mean score of -0.32 on a -3 to $+3$ scale. This question showed little and insignificant changes over time.

Tapping speeds improved starting from the first test period on LCIG treatment and the change remained significant to month

30. However, tapping accuracy was not improved by LCIG and deteriorated during alternate tapping (q8–q9) and random chasing (q11), but accuracy improved during tapping with a system-generated increasing speed (q10). The largest deteriorations in tapping accuracy were observed at month 36. The trend of spiral scores indicates that drawing performance of patients started worsening from month 6. Investigating mean scores of raw test battery items, OTS, and dimensions of LCIG-non-naïve patients asserted that there were generally no prompt changes as in the case of the LCIG-naïve patients.

The r values (effect sizes) for the OTS and total UPDRS were 0.696 and 0.536, respectively, indicating that OTS has a greater effect size than total UPDRS.

The Spearman’s rhos (absolute values) between OTS and clinical ratings in the full sample were as follows: total UPDRS ($\rho = 0.59$), total PDQ-39 ($\rho = 0.59$), and UPDRS II ($\rho = 0.64$) and UPDRS III ($\rho = 0.48$). The ρ between total UPDRS and total PDQ-39 was 0.52. Among subjectively reported dimensions, “Dyskinesia” had the strongest correlation to both total UPDRS and UPDRS III with rhos of 0.41 and 0.29, respectively. “Walking” had the strongest correlations to total PDQ-39 ($\rho = 0.49$) and to UPDRS II ($\rho = 0.45$). Among objectively measured dimensions, “Tapping” had the strongest correlations: total UPDRS ($\rho = 0.53$), total PDQ-39 ($\rho = 0.52$), and UPDRS II ($\rho = 0.58$) and UPDRS III ($\rho = 0.43$).

IV. DISCUSSION

Validated methods to remotely and objectively assess outcomes in subjects with PD over time are needed. Relevant approaches to remote monitoring of PD symptoms include e-diaries [20], wearable inertia sensor systems (e.g., [5] and [6]), various testing tools (e.g., [4]), and video-based monitoring systems (e.g., [21]). The objective methods for quantifying the motor function can potentially complement and enhance the physician and patient perspectives. There has however been a lack of mechanisms that combine these different types of remote information into scores to provide a more holistic presentation of patients’ general health, their symptom fluctuations, and treatment effects. Our study relates to a relatively large number of patients distributed among multiple clinics, reporting during a relatively long time. The main finding of this study is that using this technology it is possible to monitor PD symptoms remotely. Correlations between OTS and clinical rating scales (UPDRS and PDQ-39) were adequate indicating that the test battery contains important elements of the information of these well-established scales. The trends of the test battery scores over time were similar to the trends of conventional clinical rating scores throughout the study (see Fig. 1).

The clinimetric properties (reliability and validity criteria) of OTS and dimensions have been previously evaluated in a separate study with PD patients in different stages [22]. In that study, the internal consistency of symptom dimensions was good, there was adequate test–retest reliability, and OTS median score was different between advanced patients, experiencing on–off fluctuations and less severe, clinically stable patients having an early stage of the disease. These results are now confirmed by the present analysis and it is demonstrated that OTS can have better

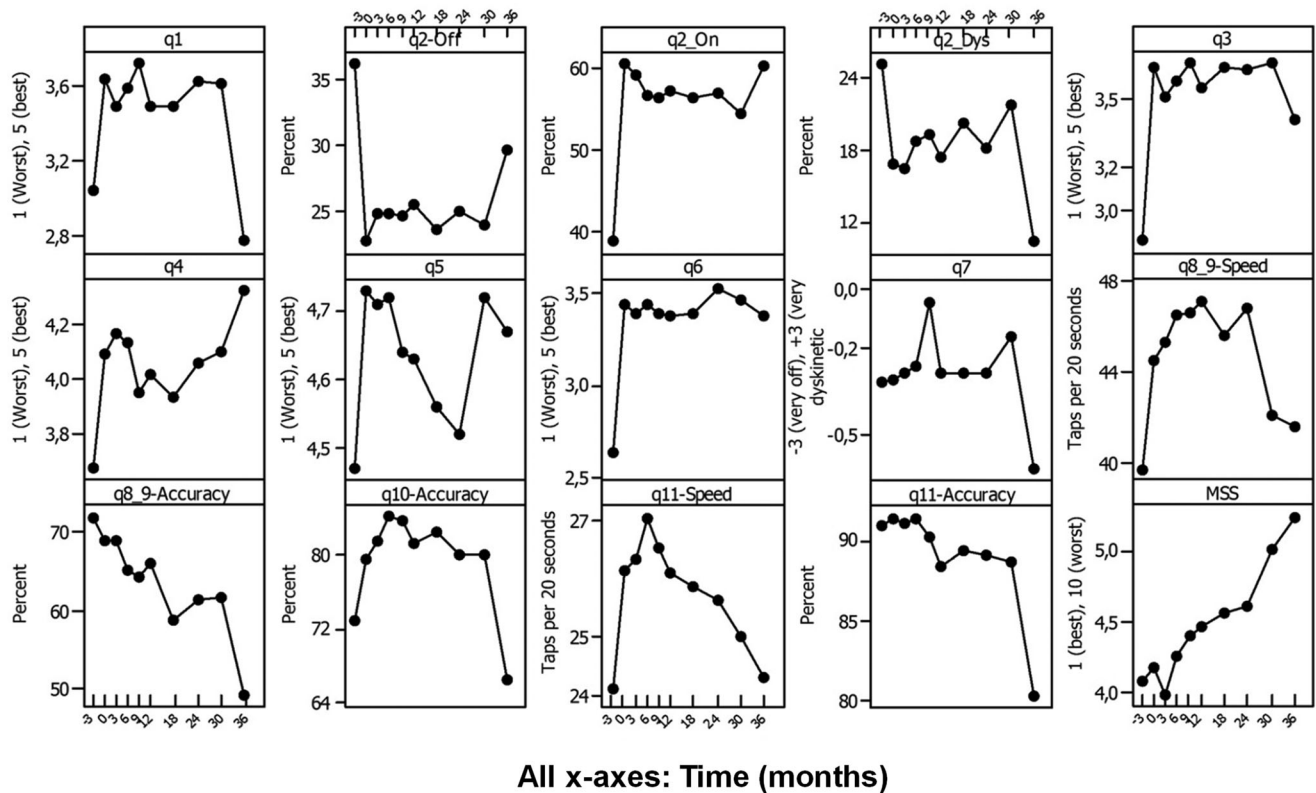


Fig. 3. Trends of mean scores of raw test battery items of LCIG-naïve patients over time. “Ability to walk” (q1), “proportion of time spent in off, on, and dyskinetic” (q2), “off at worst” (q3), “dyskinetic at worst” (q4), “painful cramps” (q5), “satisfied with functioning” (q6), and “momentary motor condition from –3 (very off) to +3 (very dyskinetic)” (q7). “alternate tapping using right hand” (q8), “alternate tapping using left hand” (q9), (q8 and q9 are presented as the average of both hands), “tapping with increasing speed using dominant hand” (q10), and “tapping with random chasing using dominant hand” (q11). MSS is the mean of the three spiral scores of the test occasion. Numbers of occasions: Test period –3 before LCIG ($n = 523$), 0 ($n = 523$), 3 ($n = 487$), 6 ($n = 404$), 9 ($n = 439$), 12 ($n = 372$), 18 ($n = 321$), 24 ($n = 332$), 30 ($n = 238$), 36 ($n = 113$). Note that y axes are presented as truncated/different scales.

reproducibility than total UPDRS. A possible explanation for the good reproducibility of the OTS is that this score is based on about 28 test occasions spread evenly throughout the day during a test week, whereas the UPDRS was based on a single assessment occasion, which is not as suitable when symptoms have a large natural variation.

It is a limitation of this study that the data are unbalanced, i.e., sample sizes for the test periods are not equal. The employment of mixed-effects models however allowed us to use all the data available, account dependencies within and between patients, and model mean scores of test battery items, OTS, and dimensions, with patient ID as random effects. Nevertheless, the low number of reports during month 30 and 36 makes conclusions from those test periods weak. Unbalanced data with home diaries have also been reported in a recent 54-week study with advanced PD patients, performed by Fernandez *et al.* [23]. At week 54, there was a 68% drop in diary responses compared to the baseline. This illustrates difficulties of keeping patients self-reporting in long-term studies. In this study, the difference in the number of test occasions using the test battery from the first test period (baseline or month 0) to month 36 was 78% and 53% at LCIG-naïve and LCIG-nonnaïve patients, respectively. Possible reasons for this high drop-out rate include: the lack of feedback to patients about their test results; absence of immediate and direct benefit to patients’ own health; usage of the test battery was voluntary; fatigue development during the long-term study [24];

the lack of feedback to nurses about their patients’ performance leading to deterioration of their interest toward the test battery; and technical malfunction issues. Patients were however very compliant with using the test battery with median compliance of 93% and user evaluations show acceptable usability [10], [22]. The need for feedback of test results is highlighted for motivating continual data entry in practice but in this study setting, feedback was not allowed.

In both LCIG-naïve and LCIG-nonnaïve patients, there was a tendency of deterioration in scores after the first year with LCIG treatment. This may reflect expected natural progression of PD. In a study performed by Jankovic and Kapadia [25], the average annual decline rate in total UPDRS scores was 1.34 during the “on” state and 1.58 during the “off” state. Patients with an older age at onset had more rapid progression of PD compared to those with a younger age at onset. In an interim report of this study [8], functional improvement with LCIG was determined by the differences between baseline and consequent visits for the total UPDRS scores. The total UPDRS was significantly improved throughout the 12-month study period and the average improvement was 9 units between baseline and month 0, and 9.6 units between baseline and month 12. Using the same methodology, a similar trend was observed in the mean OTS scores during the same study period; 0.1 units (in a 0–1 scale) between baseline and month 0, and 0.12 units between baseline and month 12 (see Fig. 1) [26], demonstrating similar level of

improvement. Along with total UPDRS other items of UPDRS like dyskinesia duration (item 32) and off duration (item 39) also improved significantly throughout the study period.

In LCIG-naïve patients, mean scores of subjectively reported dimensions, including “walking,” “satisfied,” “dyskinesia,” and “off,” improved significantly throughout the study, but the objectively measured dimensions, that is “tapping” and “spiral,” did not improve. These results are in line with those of a three-year follow-up study [27] of 25 advanced PD patients treated with LCIG. Significant improvements with LCIG were seen only in motor complications (UPDRS-IV) and quality of life (PDQ-39). Tapping accuracy scores deteriorated during our study, whereas tapping speed scores improved promptly and remained stable after the initial LCIG treatment period. In contrast to tapping accuracy, the tapping speed differed between intermediate stage patients experiencing on-off fluctuations, and clinically stable patients [28]. This finding indicates an earlier deterioration of the tapping speed than that of tapping accuracy in advancing patients.

The next step in our study is to assess the usefulness of the system in everyday clinical practice. The focus will be on determining the benefits and problems of using this system by healthcare professionals in clinical practice including a cost-benefit analysis.

In conclusion, in this 3-year follow-up study of advanced PD patients treated with LCIG, we found that it is possible to monitor PD progression over time using a telemetry test battery. The significant improvements in the mean OTS scores indicate that the test battery is able to measure a functional improvement with LCIG that is sustained over at least 24 months. When compared to total UPDRS, the OTS was more sensitive and had better reproducibility.

REFERENCES

- [1] A. A. Stone, S. Shiffman, J. E. Schwartz, J. E. Broderick, and M. R. Hufford, “Patient compliance with paper and electronic diaries,” *Control Clin. Trials*, vol. 24, pp. 182–199, 2003.
- [2] D. Nyholm, J. Kowalski, and S. M. Aquilonius, “Wireless real-time electronic data capture for self-assessment of motor function and quality of life in Parkinson’s disease,” *Mov. Disord.*, vol. 19, pp. 446–451, 2004.
- [3] M. B. Davidson, D. J. McGhee, and C. E. Counsell, “Comparison of patient rated treatment response with measured improvement in Parkinson’s disease,” *J. Neurol. Neurosurg. Psychiatry*, vol. 83, pp. 1001–1005, 2012.
- [4] C. G. Goetz, G. T. Stebbins, D. Wolff, W. DeLeeuw, H. Bronte-Stewart, R. Elble, M. Hallett, J. Nutt, L. Ramig, T. Sanger, A. D. Wu, P. H. Kraus, L. M. Blasucci, E. A. Shamim, K. D. Sethi, J. Spielman, K. Kubota, A. S. Grove, E. Dishman, C. B. Taylor, “Testing objective measures of motor impairment in early Parkinson’s disease: Feasibility study of an at-home testing device,” *Mov. Disord.*, vol. 24, pp. 551–556, 2009.
- [5] T. O. Mera, D. A. Heldman, A. J. Espay, M. Payne, and J. P. Giuffrida, “Feasibility of home-based automated Parkinson’s disease motor assessment,” *J. Neurosci. Methods*, vol. 203, pp. 152–156, 2012.
- [6] R. I. Griffiths, K. Kotschet, S. Arfon, Z. M. Xu, W. Johnson, J. Drago, A. Evans, P. Kempster, S. Raghav, M. K. Horne, “Automated assessment of bradykinesia and dyskinesia in Parkinson’s disease,” *J. Parkinson’s Disease*, vol. 2, pp. 47–55, 2012.
- [7] D. Pan, R. Dhall, and D. B. Petitti, “A mobile cloud-based Parkinson’s disease assessment system for home-based monitoring,” *JMIR Mhealth Uhealth*, vol. 26, 2015.
- [8] S. E. Pålhagen, N. Dizdar, T. Hauge, B. Holmberg, R. Jansson, J. Linder, D. Nyholm, O. Sydow, M. Wainwright, H. Widner, A. Johansson, “Interim analysis of long-term intraduodenal levodopa infusion in advanced Parkinson disease,” *Acta Neurol. Scand.*, vol. 126, pp. 29–33, 2012.
- [9] D. Nyholm, “Duodopa treatment for advanced Parkinson’s disease: A review of efficacy and safety,” *Parkinsonism Relat. Disord.*, vol. 18, pp. 916–929, 2012.
- [10] J. Westin, M. Dougherty, D. Nyholm, and T. Groth, “A home environment test battery for status assessment in patients with Parkinson’s disease,” *Comput. Methods Programs Biomed.*, vol. 98, pp. 27–35, 2010.
- [11] D. Nyholm, A. I. Nilsson Remahl, N. Dizdar, R. Constantinescu, B. Holmberg, R. Jansson, S. M. Aquilonius, H. Askmark, “Duodenal levodopa infusion monotherapy vs oral polypharmacy in advanced Parkinson disease,” *Neurology*, vol. 64, pp. 216–223, 2005.
- [12] G. Giovannoni, J. V. Schalkwyk, V. U. Fritz, and A. J. Lees, “Bradykinesia akinesia incoordination test (BRAIN TEST): An objective computerised assessment of upper limb motor function,” *J. Neurol. Neurosurg. Psychiatry*, vol. 67, pp. 624–629, 1999.
- [13] A. L. T. Tavares, G. S. Jefferis, M. Koop, B. C. Hill, T. Hastie, G. Heit, H. M. Bronte-Stewart, “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,” *Mov. Disord.*, vol. 20, pp. 1286–1298, 2005.
- [14] L. M. Williams, E. Simms, C. R. Clark, R. H. Paul, D. Rowe, and E. Gordon, “The test-retest reliability of a standardized neurocognitive and neurophysiological test battery: Neuromarker,” *Int. J. Neurosci.*, vol. 115, pp. 1605–1630, 2005.
- [15] X. Liu, C. B. Carroll, S. Y. Wang, J. Zajicek, and P. G. Bain, “Quantifying drug-induced dyskinesias in the arms using digitized spiral-drawing tasks,” *J. Neurosci. Methods*, vol. 144, pp. 47–52, 2005.
- [16] J. Westin, S. Ghiamati, M. Memedi, D. Nyholm, A. Johansson, M. Dougherty, T. Groth, “A new computer method for assessing drawing impairment in Parkinson’s disease,” *J. Neurosci. Methods*, vol. 190, pp. 143–148, 2010.
- [17] M. Memedi, J. Westin, D. Nyholm, M. Dougherty, and T. Groth, “A web application for follow-up of results from a mobile device test battery for Parkinson’s disease patients,” *Comput. Methods Programs Biomed.*, vol. 104, pp. 219–226, 2011.
- [18] J. Pinheiro and D. Bates, *Mixed-Effects Models in S and S-Plus*. New York, NY, USA: Springer-Verlag, 2000.
- [19] C. G. Goetz, G. T. Stebbins, K. A. Chung, R. A. Hauser, J. M. Miyasaki, A. P. Nicholas, W. Poewe, K. Seppi, O. Rascol, M. A. Stacy, J. G. Nutt, C. M. Tanner, A. Urkowitz, J. A. Jaglin, S. Ge, “Which dyskinesia scale best detects treatment response?” *Mov. Disord.*, vol. 28, pp. 341–346, 2013.
- [20] S. S. Papapetropoulos, “Patient diaries as a clinical endpoint in Parkinson’s disease clinical trials,” *CNS Neurosci. Ther.*, vol. 18, pp. 380–387, 2012.
- [21] F. Marzinik, M. Wahl, C. M. Doletschek, C. Jugel, C. Rewitzer, and F. Klostermann, “Evaluation of a telemedical care programme for patients with Parkinson’s disease,” *J. Telemed. Telecare*, vol. 18, pp. 322–327, 2012.
- [22] J. Westin, M. Schiavella, M. Memedi, D. Nyholm, M. Dougherty, and A. Antonini, “Validation of a home environment test battery for assessments in advanced Parkinson’s disease,” *Neurol. Sci.*, vol. 33, pp. 831–838, 2012.
- [23] H. H. Fernandez, A. Vanaganas, P. Odin, A. J. Espay, R. A. Hauser, D. G. Standaert, K. Chatamra, J. Benesh, Y. Pritchett, S. L. Hass, R. A. Lenz, “Levodopa-carbidopa intestinal gel in advanced Parkinson’s disease open-label study: Interim results,” *Parkinsonism Relat. Disord.*, vol. 19, pp. 339–345, 2013.
- [24] G. Alves, T. Wentzel-Larsen, and J. P. Larsen, “Is fatigue an independent and persistent symptom in patients with Parkinson’s disease?” *Neurology*, vol. 63, pp. 1908–1911, 2004.
- [25] J. Jankovic and A. S. Kapadia, “Functional decline in Parkinson disease,” *Arch. Neurol.*, vol. 58, pp. 1611–1615, 2001.
- [26] J. Westin, D. Nyholm, A. Johansson, M. Memedi, M. Dougherty, T. Groth, S. Pålhagen, “12-month results from a novel test battery used in a duodenal levodopa infusion trial,” *Eur. J. Neurol.*, vol. 17, p. 21, 2010.
- [27] M. Zibetti, A. Merola, V. Ricchi, A. Marchisio, C. A. Artusi, L. Rizzi, E. Montanaro, D. Reggior, C. De Angelis, M. Rizzone, L. Lopiano, “Long-term duodenal levodopa infusion in Parkinson’s disease: A 3-year motor and cognitive follow-up study,” *J. Neurol.*, vol. 260, pp. 105–114, 2013.
- [28] M. Memedi, J. Westin, and D. Nyholm, “Spiral drawing during self-rated dyskinesia is more impaired than during self-rated off,” *Parkinsonism Relat. Disord.*, vol. 19, pp. 553–556, 2013.

Authors’ photographs and biographies not available at the time of publication.