

## Research Article

# Agreement between Parkinson Disease Home Diary and Observer Assessments before and after Structured Patient Training

Carin Janz <sup>1,2</sup> Jonathan Timpka <sup>1,2</sup> Matthias Löhle <sup>3,4</sup> Alexander Bremer <sup>3</sup>  
Florin Gandor <sup>5,6</sup> Georg Ebersbach <sup>5</sup> Alexander Storch <sup>3,4</sup> and Per Odin <sup>1,2</sup>

<sup>1</sup>Division of Neurology, Department of Clinical Sciences Lund, Lund University, Lund, Sweden

<sup>2</sup>Department of Neurology, Rehabilitation, Memory and Geriatrics, Skåne University Hospital, Lund, Sweden

<sup>3</sup>Department of Neurology, University of Rostock, Rostock, Germany

<sup>4</sup>German Center for Neurodegenerative Diseases (DZNE) Rostock-Greifswald, Rostock, Germany

<sup>5</sup>Movement Disorders Clinic, Beelitz-Heilstätten, Beelitz, Germany

<sup>6</sup>Department of Neurology, Otto-von-Guericke University, Magdeburg, Germany

Correspondence should be addressed to Carin Janz; [carin.janz@med.lu.se](mailto:carin.janz@med.lu.se)

Received 26 September 2022; Revised 22 May 2023; Accepted 29 May 2023; Published 15 June 2023

Academic Editor: Gessica Sala

Copyright © 2023 Carin Janz et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Objectives.** Recent studies showed only fair agreement between observer and patients' motor state assessments on the Parkinson's disease (PD) home diary (HD). This could possibly be explained by the patients' insufficient knowledge about motor fluctuations. Therefore, the study is aimed at investigating the effect of structured training concerning motor fluctuations on the agreement between observer and HD ratings and daily motor state times. **Methods.** Participants from a previous validation study of the HD were invited back for a study extension. This interventional study consisted of a screening visit including a structured training concerning motor fluctuations and one day of motor ratings onsite during which observer and patient simultaneously and independently evaluated the patient's motor state every half hour. **Results.** Observer and 20 patients completed 316 pairs of motor state assessments. The overall agreement was 68% before training and 76% after training ( $P = .093$ ) and Cohen's  $\kappa$  increased from .438 to .559 ( $P = .320$ ). There was no significant improvement in the correlation/reliability of HD-documented daily motor state time when compared with observer ratings. Moreover, before training, the agreement in observed "on with dyskinesias" was 58%, and after training, it was 80% ( $P = .074$ ). **Conclusion.** Our structured patient training in motor fluctuations did not significantly improve the agreement between observer and HD or the reliability of daily times spent in the different motor states as an aggregate measure of HD in this PD patient group. However, there are indications of an improvement in the participants' ability to detect dyskinesias.

## 1. Introduction

Motor fluctuations affect up to 75% of Parkinson's disease (PD) patients after 4 to 6 years of levodopa treatment [1], and the PD Home Diary (HD) is often used to evaluate the occurrence of motor fluctuations [2]. The HD divides the PD motor symptomatology into four distinct motor states: "off," "on without dyskinesias," "on with nontroublesome dyskinesias," and "on with troublesome dyskinesias" [3]. In a first validation study, Hauser et al. [2] showed that there were positive correlations between both "off" and "on with troublesome dyskinesias" time and patient-rated "bad time,"

as well as between both "on" and "on with nontroublesome dyskinesia" time and patient-rated "good time." A subsequent study that investigated the correlation between HD ratings and the patient's assessment of their motor state on a visual analog scale showed that the predictive reliability was reasonable and that the HD showed a good test-retest reliability [3].

The HD is often used in clinical practice as a complementary source of information on the occurrence of motor fluctuations over time and may as such be used to optimize the pharmacological treatment [4]. Two recent studies based on similar protocols tested the agreement between

simultaneous objective observer and HD ratings in the evaluation of motor states [5, 6]. Both studies showed that the agreement between the observer as the gold standard and the HD assessments was only fair with a lack of temporal agreement. One possible explanation is that the patients did not have sufficient knowledge about motor fluctuations to complete the HD accurately. They may need more information about when motor fluctuations occur, how they look, and to what motor state they typically correspond. In the study by Timpka et al. [6], the patients received brief oral and written instructions on the HD motor states, but no structured training concerning motor fluctuations. In addition, Löhle et al. [5] used an instructional video to enhance the understanding of the different motor states before starting the HD ratings.

Patient education efforts aimed at persons with PD often contain general information about PD and address management of psychosocial aspects of the disease, the importance of activity, and social support [7–9]. In the late 1990s, Goetz et al. [10] showed that the application of a video particularly produced for the education on motor fluctuations significantly improved rater agreement between patient and observer on an on-off diary containing the motor states “on,” “off,” and “on with dyskinesia.” Accordingly, it was suggested that clinical trials assessing motor fluctuations should include this video as part of the training protocol. An observational study that evaluated patient and clinician agreement over diary entries on a four-category on-off diary used the video for training in motor fluctuations before the patient and observer independently completed the diary every 30 minutes for four hours [11]. The study showed a high agreement between observer and patient and that structured training can yield good agreement between patients and clinicians when assessing motor states.

Despite the widespread use of the HD to optimize and evaluate PD treatment [4], validation studies show only fair agreement between observer and HD assessments [5, 6]. Consequently, exploring methods to enhance agreement between patient and observer is important. The aim of this study was to investigate the effect of structured patient training concerning motor fluctuations on the agreement between observer and HD ratings in the evaluation of the PD motor state and daily motor state times.

## 2. Materials and Methods

This interventional study was part of an international collaboration on the evaluation of symptom fluctuations in PD, VALIDATE-PD [5, 6]. Previously, one study was conducted in Sweden and another in Germany, utilizing a similar study design. Both studies investigated the agreement between HD and observer motor state assessments. This extension of the Swedish study added structured training in motor fluctuations before assessments. The study was conducted at the Neurology Research Unit, Skåne University Hospital, Lund, Sweden.

**2.1. Participation Criteria.** Participants were eligible for the study if they had a diagnosis of PD according to the United Kingdom PD Society Brain Banks [12], were over 30 years

old, experienced motor fluctuations according to a neurologist’s assessment or according to the Movement Disorder Society sponsored revision of the Unified Parkinson Disease Rating Scale (MDS UPDRS) part IV, were able to fill in patient diaries, and sign an informed consent.

Exclusion criteria were signs of secondary or atypical Parkinsonian syndromes, signs of dementia (Montreal Cognitive Assessment (MoCA) <21) or psychotic symptoms, inability to complete patient diaries and patient questionnaires, and lack of cooperation during assessments. Furthermore, the presence of other conditions that interfered with the patient’s ability to consent, to participate in the study, or that made it difficult to clinically assess the patient were not allowed.

**2.2. Participant Selection.** Participants from the previous Swedish validation study of the HD were considered for this study extension [6], but they still had to meet the participation criteria. Potential participants received information about the extended study in the mail and were then contacted by phone. Those interested in participating in the study extension were invited for a screening visit that included an evaluation of participation criteria and the signing of an informed consent.

**2.3. Instruments and Assessments.** MoCA was used to screen for cognitive impairment [13]. A lower score indicates more cognitive impairment, and the maximum score is 30 points. The MDS-UPDRS was performed to characterize the included participants [14]. A higher score indicates more PD symptoms, and the maximum score is 260 points. During the training, the patient was shown the “Instructional DVD for Motor Fluctuations Diaries in PD,” endorsed by the MDS [10].

The motor states that the patients and observers could choose between in the HD were “asleep,” “off,” “on without dyskinesia,” “on with nontroublesome dyskinesia” or “on with troublesome dyskinesia” [3]. The latter two categories were replaced by “on with dyskinesias” in the analyses.

**2.4. Study Design.** All patients completed a screening visit including a structured training about motor fluctuations followed by one office-hour day of on-site ratings.

**2.4.1. Screening Visit.** The screening visit included cognitive screening using the MoCA and clinical evaluation with MDS-UPDRS. Information about the participant’s PD diagnosis, motor fluctuations, disease, and demographics were collected.

**2.4.2. Patient Training.** After inclusion in the study, the patient received a 50-minute-long training in motor fluctuations and motor symptoms. The training consisted of written information with definitions of common motor symptoms in PD occur (see supplementary material (available here)), an educational video [10], and a discussion about the patient’s own experience of motor fluctuations. The participants were also instructed in the use of a HD. First, the key nomenclature of motor symptoms was explained and exemplified. Then, the patients were shown an image describing the occurrence of motor fluctuations

in relation to plasma concentrations of levodopa to exemplify why motor fluctuations occur (see supplementary material). The patients were then shown the training video on motor fluctuations and on/off diaries [10]. For patients not fluent in English, the spoken text in the video was translated to Swedish by the rater (CJ). The participants were asked questions during the viewing of the video, for example, regarding which motor state the people in the video were in. The training ended with a discussion about the patient's own motor symptoms and fluctuations.

**2.4.3. Observation Day after Training.** During the full day on site (8.30 AM–4.00 PM), the participants were asked every 30 minutes to walk seven meters to a chair, sit down in the chair, rise, and walk back again. When finished with the task, they were asked to note their motor state in the HD. At the same time, the observer independently evaluated the motor state based on the observations of the patient during preparation for and during the 7-meter walk, taking into account tremor, dyskinesia, bradykinesia, and gait function. The author CJ functioned as a trainer and observer and had not met the participants before the study. CJ, a MD and PhD student, had previous experience with clinical rating of motor fluctuations in PD and had completed the MDS-UPDRS training program prior to the study. Also, CJ had 14 months of experience working exclusively with clinical PD research. The assessments after training were compared with the same patients' assessments before training during the initial VALIDATE-PD study (baseline).

**2.5. Statistical Analysis.** Values are provided as medians (interquartile range (IQR)). Pairwise exclusion was used for missing values. Levodopa equivalent doses (LED) were calculated according to Tomlinson et al. [15]. The agreement between HD and the observer was calculated with percentages and Cohen's kappa ( $\kappa$ ). A weighted kappa ( $\kappa_w$ ) was calculated to take different levels of disagreement into consideration. Agreement was interpreted as slight (<.20), fair (.21-.40), moderate (.41-.60), substantial (.61-.80) or almost perfect agreement (.81-1.00) [16]. The McNemar-Bowker test with a *post hoc* McNemar test with Bonferroni adjustment when appropriate was used to test for symmetry of disagreements between the rating procedures. The Wilcoxon signed rank test was used to assess differences in percentage agreement before and after structured training. HD and observer ratings before structured training were retrieved through post hoc analysis of data in the Timpka et al. [6] study. The delta function in the R package "multi-agree" was used to compare dependent pairwise kappa coefficients through Hotelling's T-square statistics [17]. Pearson's correlation test and intraclass correlation coefficient (ICC) estimation were used for correlations of daily times spent in the various motor states at the participant level. Pearson's correlation coefficient  $|r| < 0.3$  was considered a weak,  $|r| = 0.3 - 0.59$  a moderate and  $|r| < 0.6$  a strong agreement/correlation. ICC estimates and their 95% confidence intervals (95% CIs) were calculated based on single-rating, absolute-agreement, 2-way mixed-effects models with two rating instruments across all participants.

According to the guideline by Cichetti [18], we interpret  $\kappa$  values or ICC <0.40 as poor,  $\kappa$ /ICC = 0.40-0.59 as moderate,  $\kappa$ /ICC = 0.60-0.74 as good, and  $\kappa$ /ICC = 0.75-1.00 as excellent reliability.  $P < .05$  was considered statistically significant. IBM SPSS 27.0, RStudio, and GraphPad Prism were used to perform the statistical analyses and to build graphs.

**2.6. Ethics Review.** The study was approved by the Swedish Ethical Review Authority (Dnr 2022-00550-02) and performed in line with the principles of the Declaration of Helsinki. Written informed consent was obtained from the patients participating in the study.

### 3. Results

**3.1. Demographic and Clinical Data.** Figure 1 illustrates how the patient selection was performed. The 40 participants in the original study were considered for participation [6], and 20 of them could be included in this study extension. Demographic data and clinical characteristics are displayed in Table 1. The median LED was 929 mg (IQR: 751-1154) in the initial study and 1058 mg (IQR: 737-1233) in the extension study.

**3.2. Diary Training on Temporal Agreement of Motor State Rating.** Out of the expected 320 pairs of observer and HD ratings, 316 pairs were completed (99%). Ratings in observer diaries and HDs were distributed between "off," "on without dyskinesias," and "on with dyskinesias" with no differences between the two ratings in the distribution between the different motor states neither before ( $P = .495$ ) nor after training ( $P = .236$ ; both McNemar-Bowker test; Figure 2(a)). Significant differences were noted when comparing the diary assessments before and after structured training for both observer and HD ratings (both  $P < .001$ ;  $\chi^2$  tests). The after-training assessments showed a decrease in "on without dyskinesias" and an increase in "on with dyskinesias" in both assessments, compared to baseline ( $P < .001$ ;  $\chi^2$  tests with Bonferroni adjustment; Figure 2(a)).

Cohen's  $\kappa$  and  $\kappa_w$  for the agreement between observer and HD before and after training are shown in Table 2. The overall  $\kappa$  was .438 before training and .559 after training ( $P = .320$ ).  $\kappa_w$  was .474 before training and .559 after training ( $P = .539$ ). Examination of the agreement for the respective motor states revealed substantial values for "off" both before and after training ( $\kappa = .644$  and  $\kappa = .608$ ;  $P = .625$ ). There was a fair agreement for "on without dyskinesias" ( $\kappa = .390$ ) before training and a moderate agreement after training ( $\kappa = .560$ ;  $P = .191$ ). Values for "on with dyskinesias" also improved from fair ( $\kappa = .388$ ) to moderate after training ( $\kappa = .543$ ;  $P = .299$ ).

Agreement between observers and HD ratings, using the observer as the gold standard, is shown in Figures 2(b) and 2(c). The overall agreement between patient and observer ratings was 68% before training and 76% for the same patients after training ( $P = .093$ ; Wilcoxon signed rank test). The patients did not improve their ability to detect neither "off" ( $P = .340$ ) nor "on without dyskinesias" ( $P = .802$ ), but there was a trend towards improved agreement

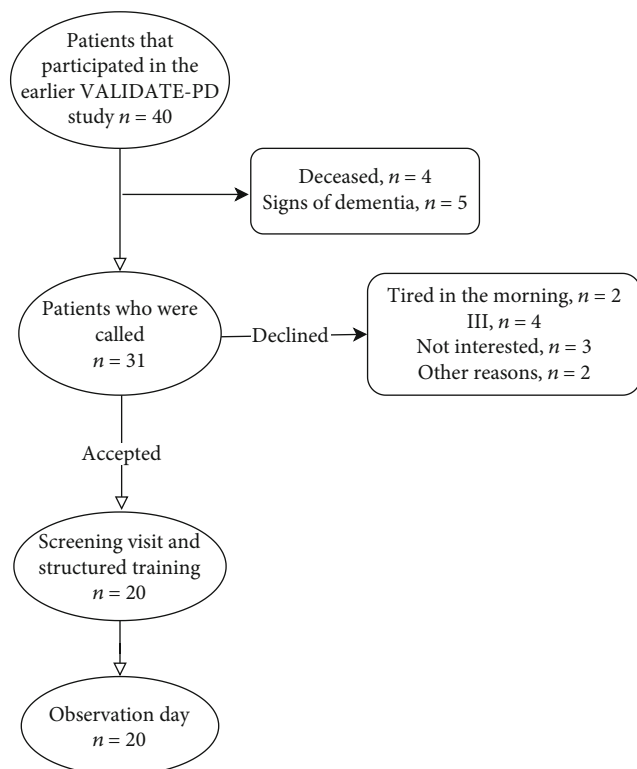


FIGURE 1: Flowchart of patient selection.

concerning “on with dyskinesias”. Before training, 58% of the HD ratings agreed with observed “on with dyskinesias,” and after training, the agreement was 80% ( $P = .074$ ). When participants were in “off” before the training, the most common mistake was to rate themselves as “on without dyskinesias” instead. After training, the most common mistake was to rate themselves as “on with dyskinesias” (Figure 2(c)).

### 3.3. Effects of Diary Training on Daily Motor State Times.

The aggregate measure of daily times spent in the three different motor states is the most frequent read-out in clinical trials on motor fluctuations in advanced PD when using the HD [4]. We thus analyzed the daily percentage times spent in the three different motor states (8.30 AM–4.00 PM) on the participant level from all 20 participants with respect to the diary training session. As shown in Figure 3(a), we detected similar percentage daily times spent in all three motor states when comparing observer diary data and HD with no significant differences between the two diary ratings for all motor states ( $P \geq .05$ , Friedman test with post hoc Wilcoxon Rank test with Bonferroni adjustment). After training, the median number of motor state changes was 2 (IQR = 0.75–5) according to the patient and 3.5 (IQR = 1.75–7) according to the observer. Pearson correlation analyses of the individual times spent in the three different motor states revealed strong correlations of percentage daily times spent in all three motor states before and after diary training (Figures 3(b)–3(d)). Multiple linear regression model analyses revealed that the daily time spent in all three motor states as assessed by the HD was significantly associated with the respective daily times as rated by the observers

TABLE 1: Demographic and clinical characteristics<sup>a</sup>.

Male/female	10 (50%)/10 (50%)
Age, in years	70 (65–77)
Symptom duration, in years	13 (10–16)
Disease duration, in years	11 (9–14)
Months, since initial screening visit <sup>b</sup>	36 (34–39)
Hoehn and Yahr stages	2 (1–3)
LED <sup>c</sup> in initial study	929 (751–1154)
LED in study extension	1058 (737–1233)
MDS-UPDRS <sup>d</sup> total in initial study	38 (26–55)
Part I	7 (4–11)
Part II	8 (5–11)
Part III	18 (13–24)
Part IV	4 (3–8)
MDS-UPDRS total, in study extension	55 (43–78)
Part I	10 (8–13)
Part II	15 (9–19)
Part III	29 (21–40)
Part IV	5 (3–7)
Duration of motor fluctuations, in months	68 (52–111)
Hypokinetic fluctuations	68 (51–101)
Hyperkinetic fluctuations	63 (56–93)
Motor fluctuation symptoms	
Nightly off	12 (60%)
Wearing off	16 (80%)
Delayed on or no on	11 (55%)
On-off phenomena	16 (80%)
Peak dose dyskinesia	12 (60%)
Biphasic dyskinesia	13 (65%)
Off dose dyskinesia	9 (45%)
MoCA <sup>e</sup> total in initial study	26 (25–28)
Cognitive impairment <sup>f</sup> in initial study	
Normal cognition	11 (55%)
Mild cognitive impairment	9 (45%)
Dementia	0 (0%)
MoCA total in study extension	27 (25–28)
Cognitive impairment in study extension	
Normal cognition	11 (55%)
Mild cognitive impairment	9 (45%)
Dementia	0 (0%)

<sup>a</sup>Data from the patients included in both the earlier validate study and this study extension. Presented as median (IQR: interquartile range) or percentages. <sup>b</sup>Time passed between the initial screening visit and the study extension screening visit. <sup>c</sup>LED: levodopa equivalent dose. <sup>d</sup>MDS-UPDRS: Movement Disorder Society sponsored the revision of the Unified Parkinson Disease Rating Scale. <sup>e</sup>MoCA: Montreal Cognitive Assessment. <sup>f</sup>Normal cognition, MoCA >25; mild cognitive impairment; MoCA 21–25; dementia, MoCA <21.

(“off:”  $r^2 = .646$ ,  $\beta = .809$ , and  $P < .001$ ; “on without dyskinesias:”  $r^2 = .660$ ,  $\beta = .804$ , and  $P < .001$ ; and “on with dyskinesias:”  $r^2 = .581$ ,  $\beta = .745$ , and  $P < .001$ ), but not with diary training (“off:”  $\beta = .041$ ,  $P = .680$ ; “on without dyskinesias:”

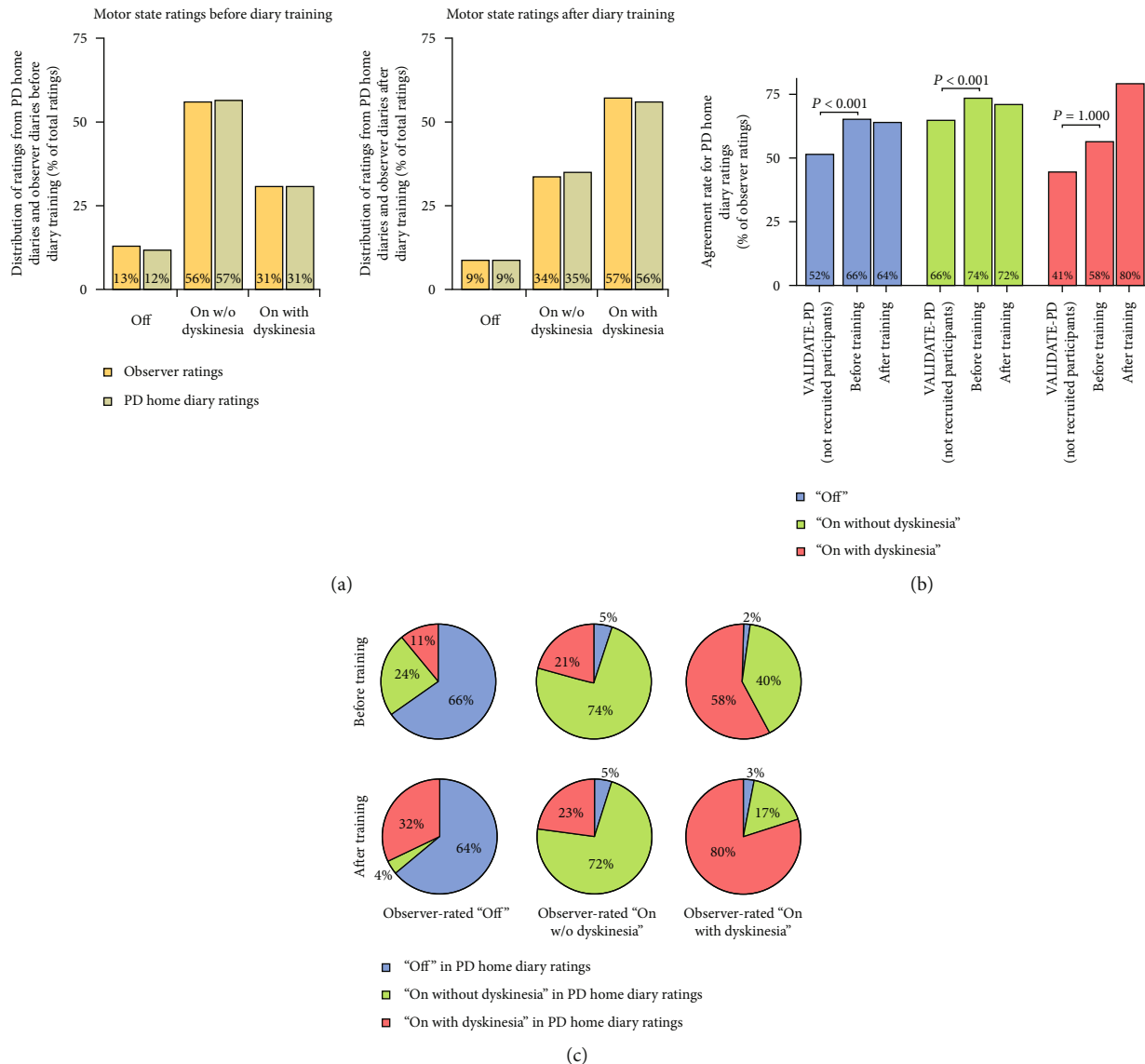


FIGURE 2: Proportion and temporal agreement of motor states assessed by observer diaries and patients Parkinson's disease (PD) home diaries before and after diary training. (a) Proportion of "off," "on without dyskinesia," and "on with dyskinesia" as assessed by the observer and patient diaries before and after diary training. "Before training" is the participants' ratings before structured training about motor fluctuations, and "after training" is the same patients' choices after training ( $n=20$ ). McNemar-Bowker tests revealed no significant results with  $P=.495$  for baseline and  $P=.236$  for after training data. (b) Temporal agreement rate for "off," "on without dyskinesia," and "on with dyskinesia" with the observer ratings as the reference for comparison before and after diary training. Wilcoxon signed rank test tests revealed no significant results when comparing the overall agreement between HD and the observer of HD before and after training assessments ( $P=.093$ ). Agreement was 66% for observed "off" before training and 64% after training ( $P=.340$ ), 74% for observed "on without dyskinesias" before training and 72% after training ( $P=.802$ ), and 58% for observed "on with dyskinesias" before training and 80% after training ( $P=.074$ ). As a comparative group, data from participants of the initial Swedish VALIDATE-PD study cohort, which were not recruited into the study extension, are provided ( $P$  values are from McNemar tests). (c) Participants' choices on the PD home diary in each respective observed motor state before and after diary training. Number of observations in each observed motor state including all patients ( $n=40$ ) in the initial study: "off:"  $n=218$ ; "on without dyskinesias:"  $n=651$ ; and "on with dyskinesias:"  $n=419$ . Numbers of observations in each observed motor state before training (before training): "off:"  $n=84$ ; "on without dyskinesias:"  $n=362$ ; and "on with dyskinesias:"  $n=204$ . Number of observations in each observed motor state after training: "off:"  $n=28$ ; "on without dyskinesias:"  $n=109$ ; and "on with dyskinesias:"  $n=179$ . Note that rounding errors might lead to total percentages above 100%.

$\beta = -.021$ ,  $P = .838$ ; and "on with dyskinesias:"  $\beta = .041$ ,  $P = .725$ ). Consistently, reliability analyses using ICC calculation revealed good to excellent reliability for HD data for all three motor states independent of the diary training (Table 3).

## 4. Discussion

The primary finding of this study was that our structured training program neither significantly improved the overall agreement between observer and HD ratings nor the

TABLE 2: Agreement between HD and observer before and after training, presented as Cohens kappa ( $\kappa$ ) and Cohens weighted kappa ( $\kappa_w$ ).

N = 20	Before training			After training			P value <sup>b</sup>
	Cohen's $\kappa$	95% CI <sup>a</sup>		Cohen's $\kappa$	95% CI		
Overall agreement	.438	.373	.503	.559	.473	.645	.320
Observed "on without dyskinesias"	.390	.319	.461	.560	.464	.656	.191
Observed "off"	.644	.552	.734	.608	.451	.765	.625
Observed "on with dyskinesias"	.388	.312	.464	.543	.447	.637	.299

	Cohen's $\kappa_w$	95% CI		Cohen's $\kappa_w$	95% CI		P value
Overall agreement	.474	.410	.537	.559	.468	.650	.539

<sup>a</sup>CI: confidence interval. <sup>b</sup>Significance levels was calculated on the difference before and with treatment using the delta method in the "multiagree" package in R.

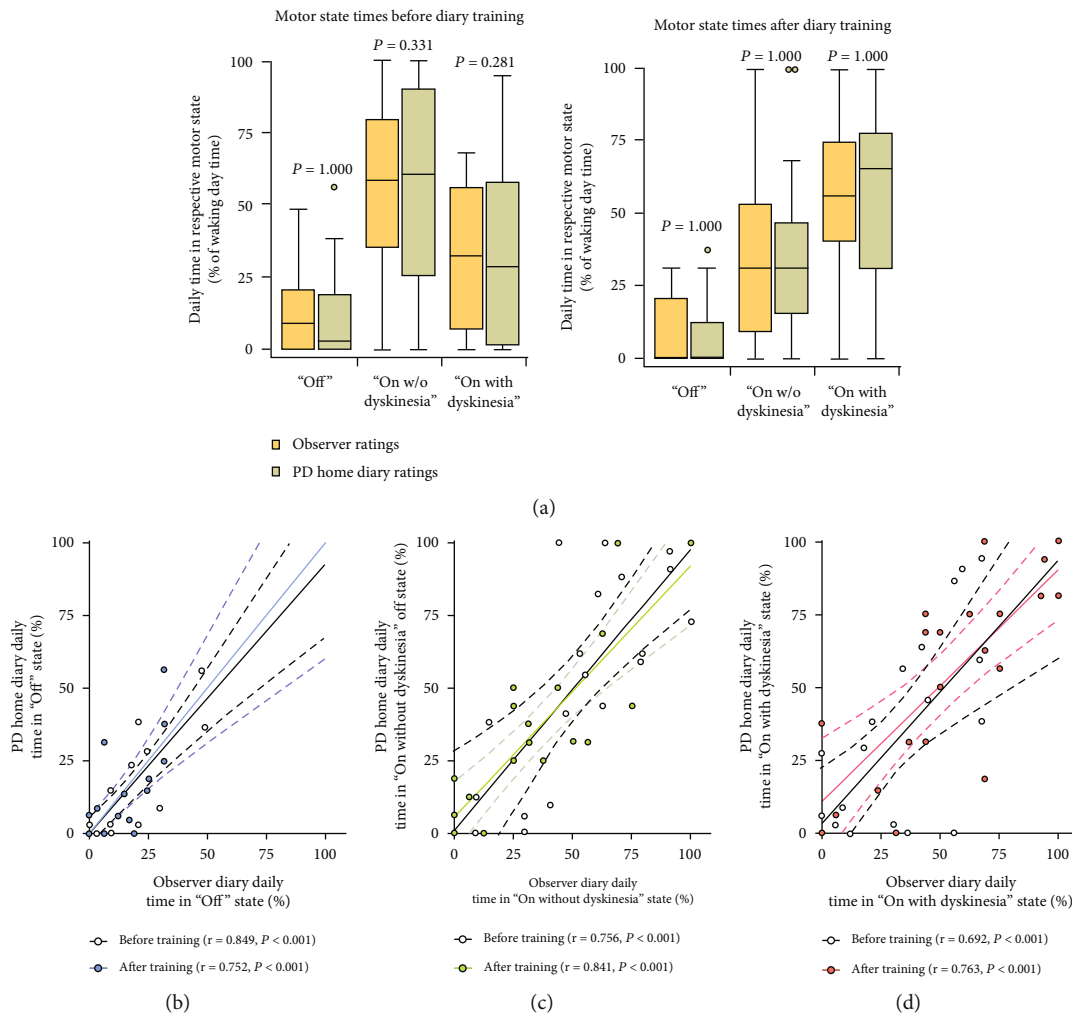


FIGURE 3: Proportions for time spent in motor states assessed by observer diaries and PD home diaries on participant level with respect to diary training. (a) Distribution of daily time proportions of "off," "on without dyskinesia," and "on with dyskinesia" based on the simultaneous, half-hourly performed diary ratings from 20 participants before and after diary training. Boxplots are shown with a central mark at the median, bottom, and top edges of the boxes at 25th and 75th percentiles, respectively, whiskers out to the most extreme points within 1.5 times the interquartile range. Displayed P values are from Friedmann tests with post hoc Wilcoxon signed-rank tests with Bonferroni correction for multiple comparisons. Correlation analyses of mean proportions of "off" (b), "on without dyskinesia" (c), and "on with dyskinesia" (d) with respect to the diary training. Solid lines in represent the regression line with 95% CI (dotted lines). Values in upper right corner are the correlation coefficients and P values from Pearson's correlation tests.

correlation/reliability of HD-documented daily motor state time when compared with objective observer ratings. As a secondary finding, we saw indications that training in motor

fluctuations might improve patients' ability to detect dyskinesias. Before training, 58% of the HD ratings in observed "on with dyskinesias" were in agreement with the

TABLE 3: Reliability of the PD home diary according to intraclass correlation coefficient (ICC) calculation<sup>a</sup>.

HD vs. observer diary	ICC (95% CI)	Value	F-test with true value 0		P value
			df1	df2	
Before training					
Off	0.85 (0.66-0.94)	11.98	19	19	<0.001
On without dyskinesia	0.74 (0.45-0.89)	6.46	19	19	<0.001
On with dyskinesia	0.68 (0.34-0.86)	4.995	19	19	<0.001
After training					
Off	0.73 (0.44-0.89)	6.23	19	19	<0.001
On without dyskinesia	0.85 (0.65-0.94)	11.56	19	19	<0.001
On with dyskinesia	0.77 (0.50-0.90)	7.40	19	19	<0.001

<sup>a</sup>ICC estimates and 95% CIs were calculated based on single-rating, absolute-agreement, 2-way mixed-effects models with two raters (patient diary and observer diary) across 20 participants.

simultaneous observer assessment, compared to 80% after training ( $P = .074$ ).

There are various factors that increase the risk of developing dyskinesias in PD patients, two of which are prolonged disease duration and higher LED dosage [19, 20]. We noted significantly less “on without dyskinesias” and more “on with dyskinesias” in the after-training assessments compared to before training for both observer and HD ratings. Since the previous study (before training) was conducted some years ago, the patients had a longer disease duration in the after-training assessments (median: 7 years before training; 13 years after training) and a higher LED dose (median: 941 before training; 1058 after training) which might explain the observed differences [6].

We did not observe differences in the distribution of the different motor states between HD and observer ratings at baseline and after diary training. This is in contrast to the initial validation study with significantly more “on without dyskinesias” and less “on with dyskinesias” in observer as compared to HD ratings [6]. Consistent with the present study, *post hoc* analysis of the subgroup of participants not included in the extension study showed that this subcohort with significant differences in motor state distribution between observer and HD ratings ( $P < .001$ ) accounts for the differences in the initial Swedish VALIDATE-PD study cohort. The overall agreement between observer and HD ratings in the present extension study could be characterized as moderate both before and after training with no significant increase in Cohen’s  $\kappa$  from  $\kappa = .438$  to  $\kappa = .559$  ( $P = .320$ ). In agreement with the above-mentioned observation on motor state distributions, in ancillary analyses, we detected a significantly higher agreement in the present subcohort as compared to the nonrecruited participant group of the initial VALIDATE-PD study ( $P < .001$ ; McNemar-Bowker-test; Figure 2(b)). The daily times spent in the three different motor states as an aggregate measure calculated from HD showed good to excellent reliability with ICC values ranging from 0.68 for “on with dyskinesias” to 0.85 for “off” before training and 0.73 for “off” to 0.85 for “on without dyskinesias” after training. Structured diary training did not change the already good-to-excellent reliability of these aggregate HD measures at baseline in our cohort. These ICC values are in general higher as reported for the

German and the original Swedish cohort of the VALIDATE-PD study [5, 6] and the ICC values of the HD when compared to wearable sensor data [21].

These data implies that the participants included in the study extension might be those who, even before training, were more effective at assessing their motor state. One explanation might be that these participants had the greatest interest in learning about their disease and that they as a result had some general knowledge about motor fluctuations even before the structured training. Furthermore, 16 out of 20 participants had experience from at least one other prior study at the Research Unit and might potentially be more experienced in the terminology and assessment of motor fluctuations than the average person with PD. It is possible that the training would have had a greater impact if the patients had less experience going into the study. Ancillary analyses showed that there were no significant differences in age, MoCA score, MDS-UPDRS score, disease duration, or duration of motor fluctuations between the participants that were included in the study extension and those that were not (data not shown).

Löhle et al. [5] only found a fair agreement between the HD and observer ratings, even though participants in that cohort had a somewhat more in-depth training in motor fluctuations before the conduct of the study than those in the study by Timpka et al. [6]. Part of the reason for the difference in the level of observer-HD agreement presented by Löhle et al. and that in the present study may still be the level of participant training. It has previously been shown that videos in the patient’s primary language explaining motor fluctuations are valuable training that help patients’ complete diaries correctly [22], which is why we helped patients with real-time translations of the video. Additionally, the training was extended by adding a glossary, a picture explaining motor fluctuations, and a discussion about the patient’s own motor fluctuations.

There was a trend towards improvement in the patient’s ability to determine whether they had dyskinesias or not. Before training, they agreed with the observer in 58% of the time periods and after training in 80% of the time periods ( $P = .074$ ), and the agreement improved from fair to moderate after training ( $P = .299$ ). However, the patients spent significantly more time in “on with dyskinesias” and

less time in “on without dyskinesias” in the observer as compared to HD ratings [6], which might have affected the result. Previous studies have concluded that there is a low awareness of dyskinesias among PD patients and that it appears to be related to metacognitive deficits in the self-monitoring system [23, 24]. According to our study, another explanation could be a fundamental lack of understanding of the concept of “dyskinesias.” During the structured training, several patients said that they finally understood what we meant by “dyskinesias” and that they now realized that they experienced it.

This realization, however, is contradicted by another finding. Although the agreement in observed “off” was virtually unchanged before and after training, more patients rated themselves as “on with dyskinesias” while being “off” after training than before. One explanation might be that the training made the participants more vigilant and aware that something felt wrong, akin to the previously shown correlation between both “off” and “on with troublesome dyskinesia” and “bad time” [2], but that they were not fully capable to determine in what way the body felt different. The difficulties for many patients to differ tremor from dyskinesia are well-known [22]. In addition, even though the percentage that agreed with the observer in observed “on without dyskinesias” after training did not improve, the  $\kappa$  value improved from fair ( $\kappa = .390$ ) to moderate ( $\kappa = .560$ ) ( $P = 191$ ). This indicates that patients enhanced their ability to determine when they were not in “on without dyskinesias” and thus further suggests that patients became more aware of when something felt different in their body after training. Whether the dyskinesias are “troublesome” or not, it is an inherently subjective dichotomization, and the observer is often unsuitable to make that distinction without asking the patient. However, subanalyses (data not shown) were conducted and indicated that structured training in motor fluctuations did not affect the patient’s subjective dichotomization between troublesome and nontroublesome dyskinesia. As noted in the previous study [6], it is worth considering that nonmotor fluctuations could have influenced HD ratings. Specifically, when evaluating the “off” state, it is possible that patients were also taking nonmotor symptoms into account.

This study has several limitations. Firstly, not all initial participants could be included, and only one full day of HD registrations was carried out which resulted in less data than in the initial study. It is unlikely that patient cognition affected the result since the MoCA score did not change from the initial study. However, the patients’ motor deterioration might have impacted the result. Also, the included patients seemed to be unusually experienced in motor fluctuations and clinical studies even before training which might have resulted in a reduced effect of additional training. Furthermore, it would have been better if the observer had been allowed to make a somewhat more thorough examination instead of just passively observing participants’ movements. Dyskinesia can be assessed through observation, but it is possible that mild “off” can be overlooked without, e.g., examining the level of rigidity. However, the full day onsite had to be conducted in the same way as the initial

study to be able to compare the results. A walking test was selected in the initial study due to its ease of administration and its ability to provide insights into dyskinesia and off symptoms, including posture and balance. However, patients’ ability to recognize dyskinesias may vary depending on the context. For example, some patients may be more adept at identifying dyskinesias while engaging in activities like eating or drinking than during walking. Also, just as in the initial study, the observer was not a movement disorder specialist and could thus be considered less accurate than the gold standard. Nevertheless, the same rater performed all ratings in the extension study, which is a strength.

This patient sample is too small to perform further sub-analysis, but follow-up studies that further investigate the impact of motor symptoms, cognitive status, previous experience in clinical studies, and experience in motor fluctuations on the agreement between HD and an observer are justified. The PD patient diaries currently available have several limitations, such as the absence of medication tracking, functional assessments, and registration of nonmotor symptoms [25]. Therefore, technical solutions such as eDiaries and wearable sensors that can provide a broader and more objective picture of the patient’s motor fluctuations are warranted [25, 26].

Together, our structured patient training in motor fluctuations did not significantly improve the overall agreement between observer and HD or the reliability of daily times spent in the different motor states as an aggregate measure of the HD in this group of patients with PD and motor fluctuations. However, there were indications of an improvement in the ability to detect dyskinesias. It is essential to be aware that even after patient training, there is still a lack of agreement between HD ratings and gold-standard observer ratings. The difficulties in collecting reliable data on motor status and motor fluctuations thus remain an obstacle that needs to be addressed, both in practical clinical work and in clinical studies.

## Data Availability

Data is available on request from the authors.

## Ethical Approval

The study was approved by the Swedish Ethical Review Authority (Dnr 2022-00550-02) and performed in line with the principles of the Declaration of Helsinki.

## Consent

Written informed consent was obtained from the patients participating in the study.

## Disclosure

The funders were not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.



## Conflicts of Interest

CJ received funding from the Elsa Schmitz Foundation. JT has received funding from the Swedish National Government and County Councils through the ALF agreement, the Swedish Parkinson Foundation, and the Elsa Schmitz Foundation and has received compensation for consultancies from AbbVie and TransPerfect, as well as royalties from UNI-MED Verlag. AS has received funding from the Deutsche Forschungsgemeinschaft (German Research Association) and the Helmholtz-Association. He has received honoraria for presentations/advisory boards/consultations from Desitin, Global Kinetics, Lobsor Pharmaceuticals, STADA, Bial, RG Gesellschaft, Zambon, NovoNordisk, and AbbVie. He has received royalties from Kohlhammer Verlag and Elsevier Press. He serves as an editorial board member of Stem Cells International. PO has received funding from AbbVie, Lund University Medical Faculty, Multipark, the Swedish Parkinson Foundation, Health Care Region Skåne, and Åhlens Foundation. He has received honoraria for lectures and expert advice from AbbVie, Bial, Britannia, Ever Pharma, Global Kinetics, Lobsor, Nordic Infucare, Stada, and Zambon. He has received royalties from UNI-MED Verlag.

## Authors' Contributions

CJ did the conception, organization, and execution of the research project; designed and executed statistical analysis; and wrote the first draft, review, and critique of the manuscript. JT and PO did the conception and organization of the research project, designed, reviewed, and critiqued of the statistical analysis, and reviewed and critiqued the manuscript. ML and AB did the conception, organization, and execution of the initial research VALIDATE-PD research project; did the design, review, and critique of the statistical analysis; and did the review and critique of the manuscript. FG and GE worked on the review and critique of the statistical analysis and manuscript. AS was assigned to the conception, organization, and execution of the initial research VALIDATE-PD research project; the design, review, and critique of the statistical analysis; the review and critique of the manuscript; and the execution of the statistical analysis.

## Acknowledgments

This study extension received no external funding. The initial study received support from the Global Kinetics Corporation, Melbourne, Australia, and the Skåne University Hospital Foundation and Donations. The authors want to thank Helene Jacobsson at Clinical Studies Sweden. We also want to thank Sofia Christiansson, Jeanette Härnberg, and Monica Scharfenort for contributing to the study. The Restorative Parkinson Unit, led by PO, thanks the Medical Faculty at Lund University, Multipark, the Swedish Parkinson Academy, the Swedish Parkinson Foundation, the Skåne University Hospital Foundation and Donations, and Global Kinetics Corporation for their support. JT thanks the Swedish Parkinson Foundation, the Elsa Schmitz Foundation, and

the Swedish National Government and County Councils for their support through the ALF agreement.

## Supplementary Materials

Glossary and picture used in the structured training. The glossary contains written information with definitions of common motor symptoms in Parkinson's disease (PD). The picture illustrates the occurrence of motor fluctuations in relation to plasma concentrations of levodopa, aiming to provide an example of why motor fluctuations happen. (*Supplementary Materials*)

## References

- [1] N. Tambasco, M. Romoli, and P. Calabresi, "Levodopa in Parkinson's disease: current status and future developments," *Current Neuropharmacology*, vol. 16, no. 8, pp. 1239–1252, 2018.
- [2] R. A. Hauser, J. Friedlander, T. A. Zesiewicz et al., "A home diary to assess functional status in patients with Parkinson's disease with motor fluctuations and dyskinesia," *Clinical Neuropharmacology*, vol. 23, no. 2, pp. 75–81, 2000.
- [3] R. A. Hauser, F. Deckers, and P. Leher, "Parkinson's disease home diary: further validation and implications for clinical trials," *Movement Disorders*, vol. 19, no. 12, pp. 1409–1413, 2004.
- [4] S. S. Papapetropoulos, "Patient diaries as a clinical endpoint in Parkinson's disease clinical trials," *CNS Neuroscience & Therapeutics*, vol. 18, no. 5, pp. 380–387, 2012.
- [5] M. Löhle, A. Bremer, F. Gandor et al., "Validation of the PD home diary for assessment of motor fluctuations in advanced Parkinson's disease," *Parkinson's Disease*, vol. 8, no. 1, 2022.
- [6] J. Timpka, M. Löhle, A. Bremer et al., "Objective observer vs. patient motor state assessments using the PD home diary in advanced Parkinson's disease," *Frontiers in Neurology*, vol. 13, article 935664, 2022.
- [7] G. Simons, S. B. Thompson, and M. C. Smith Pasqualini, "An innovative education programme for people with Parkinson's disease and their carers," *Parkinsonism & Related Disorders*, vol. 12, no. 8, pp. 478–485, 2006.
- [8] C. Derollez, N. Carriere, M. Kyheng et al., "Factors that predict a change in quality of life among Parkinson's disease patients participating in a patient education program," *Revue Neurologique*, vol. 177, no. 9, pp. 1151–1159, 2021.
- [9] H. Sunvisson, S.-L. Ekman, H. Hagberg, and J. Lökk, "An education programme for individuals with Parkinson's disease," *Scandinavian Journal of Caring Sciences*, vol. 15, no. 4, pp. 311–317, 2001.
- [10] C. G. Goetz, G. T. Stebbins, L. M. Blasucci, and M. S. Grobman, "Efficacy of a patient-training videotape on motor fluctuations for on-off diaries in Parkinson's disease," *Movement Disorders*, vol. 12, no. 6, pp. 1039–1041, 1997.
- [11] J. Reimer, M. Grabowski, O. Lindvall, and P. Hagell, "Use and interpretation of on/off diaries in Parkinson's disease," *Journal of Neurology, Neurosurgery, and Psychiatry*, vol. 75, no. 3, pp. 396–400, 2004.
- [12] A. J. Hughes, S. E. Daniel, L. Kilford, and A. J. Lees, "Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases," *Journal of Neurology, Neurosurgery, and Psychiatry*, vol. 55, no. 3, pp. 181–184, 1992.

- [13] Z. S. Nasreddine, N. A. Phillips, V. Bédirian et al., "The Montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment," *Journal of the American Geriatrics Society*, vol. 53, no. 4, pp. 695–699, 2005.
- [14] C. G. Goetz, B. C. Tilley, S. R. Shaftman et al., "Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): scale presentation and clinimetric testing results," *Movement Disorders*, vol. 23, no. 15, pp. 2129–2170, 2008.
- [15] C. L. Tomlinson, R. Stowe, S. Patel, C. Rick, R. Gray, and C. E. Clarke, "Systematic review of levodopa dose equivalency reporting in Parkinson's disease," *Movement Disorders*, vol. 25, no. 15, pp. 2649–2653, 2010.
- [16] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.
- [17] S. Vanbelle, "Comparing dependent kappa coefficients obtained on multilevel data," *Biometrical Journal*, vol. 59, no. 5, pp. 1016–1034, 2017.
- [18] D. V. Cicchetti, "Multiple comparison methods: establishing guidelines for their valid application in neuropsychological research," *Journal of Clinical and Experimental Neuropsychology*, vol. 16, no. 1, pp. 155–161, 1994.
- [19] J. C. Sharma, C. G. Bachmann, and G. Linazasoro, "Classifying risk factors for dyskinesia in Parkinson's disease," *Parkinsonism & Related Disorders*, vol. 16, no. 8, pp. 490–497, 2010.
- [20] C. Warren Olanow, K. Kieburtz, O. Rascol et al., "Factors predictive of the development of levodopa-induced dyskinesia and wearing-off in Parkinson's disease," *Movement Disorders*, vol. 28, no. 8, pp. 1064–1071, 2013.
- [21] C. Perez-Lopez, J. Hernandez-Vara, N. Caballol et al., "Comparison of the results of a Parkinson's Holter monitor with patient diaries, in real conditions of use: a sub-analysis of the MoMoPa-EC clinical trial," *Frontiers in Neurology*, vol. 13, article 835249, 2022.
- [22] R. A. Hauser, H. Russ, D. A. Haeger, M. Bruguere-Fontenille, T. Müller, and G. K. Wenning, "Patient evaluation of a home diary to assess duration and severity of dyskinesia in Parkinson disease," *Clinical Neuropharmacology*, vol. 29, no. 6, pp. 322–330, 2006.
- [23] M. Amanzio, S. Monteverdi, A. Giordano, P. Soliveri, P. Filippi, and G. Geminiani, "Impaired awareness of movement disorders in Parkinson's disease," *Brain and Cognition*, vol. 72, no. 3, pp. 337–346, 2010.
- [24] C. Vitale, M. T. Pellecchia, D. Grossi et al., "Unawareness of dyskinesias in Parkinson's and Huntington's diseases," *Neurological Sciences*, vol. 22, no. 1, pp. 105–106, 2001.
- [25] J. A. Vizcarra, Á. Sánchez-Ferro, W. Maetzler et al., "The Parkinson's disease e-diary: developing a clinical and research tool for the digital age," *Movement Disorders*, vol. 34, no. 5, pp. 676–681, 2019.
- [26] H. Hasan, D. S. Athauda, T. Foltynie, and A. J. Noyce, "Technologies assessing limb bradykinesia in Parkinson's disease," *Journal of Parkinson's Disease*, vol. 7, no. 1, pp. 65–77, 2017.