ARTICLE   OPEN

Check for updates

Application of single wrist-wearable accelerometry for objective motor diary assessment in fluctuating Parkinson's disease

Matthias Löhle[1,2 ✉], Jonathan Timpka [ID][3,4], Alexander Bremer[1], Hamid Khodakarami[5], Florin Gandor [ID][6,7], Malcom Horne [ID][8,9], Georg Ebersbach[6], Per Odin[3,4] and Alexander Storch [ID][1,2 ✉]

Advanced Parkinson's disease (PD) is characterized by motor fluctuations including unpredictable oscillations remarkably impairing quality of life. Effective management and development of novel therapies for these response fluctuations largely depend on clinical rating instruments such as the widely-used PD home diary, which are associated with biases and errors. Recent advancements in digital health technologies provide user-friendly wearables that can be tailored for continuous monitoring of motor fluctuations. Their criterion validity under real-world conditions using clinical examination as the gold standard remains to be determined. We prospectively examined this validity of a wearable accelerometer-based digital Parkinson's Motor Diary (adPMD) using the Parkinson's Kinetigraph (PKG®) in an alternative application by converting its continuous data into one of the three motor categories of the PD home diary (Off, On and Dyskinetic state). Sixty-three out of 91 eligible participants with fluctuating PD (46% men, average age 66) had predefined sufficient adPMD datasets (>70% of half-hour periods) from 2 consecutive days. 92% of per-protocol assessments were completed. adPMD monitoring of daily times in motor states showed moderate validity for Off and Dyskinetic state (ICC = 0.43–0.51), while inter-rating methods agreements on half-hour-level can be characterized as poor (median Cohen's κ = 0.13–0.21). Individualization of adPMD thresholds for transferring accelerometer data into diary categories improved temporal agreements up to moderate level for Dyskinetic state detection (median Cohen's κ = 0.25–0.41). Here we report that adPMD real-world-monitoring captures daily times in Off and Dyskinetic state in advanced PD with moderate validities, while temporal agreement of adPMD and clinical observer diary data is limited.

INTRODUCTION

Advanced stages of Parkinson's disease (PD) are characterized by the presence of motor complications, which affect about 50% of patients after 5 and 90% after 10 years of disease[1–3]. During the disease course, most patients initially experience progressive shortening of benefit from short-acting dopaminergic drugs such as levodopa[3], referred to as "Wearing-off". The clinical assessment and classification of these transitions between the levodopa response and "Off" state have been comprehensively reviewed[4] and are commonly referred to as "fluctuations". With disease progression, patients tend to present with unpredictable Off periods and dyskinesias[5,6]. Dyskinesias are involuntary movements of head, trunk or extremities that can interfere with patient activity and lower quality of life (QoL)[7]. Unpredictable motor fluctuations are an important aspect for disease management in late stage PD with particularly high impact on health-related quality of life[8,9]. The clinical phenomenology of unpredictable motor fluctuations is multifaceted and includes among others "Delayed On", "Dose failure/No-On" and "Random On-Off"[6,10]. The underlying patho-physiological mechanisms are complex and comprise not only factors affecting the peripheral pharmacokinetics of levodopa but also central (non-dopaminergic) mechanisms (for details, please refer to refs. [6,10]). However, there is still an intensive debate on

their precise phenomenology as well as their response to medication and the resulting treatment strategies[6,11]. Motor fluctuations are accompanied by fluctuations of non-motor symptoms, such as depression, anxiety and pain[12,13]. Since symptom fluctuations significantly impairs QoL, their accurate detection and subsequent treatment provide the opportunity to improve QoL in advanced PD[12,14].

Clinical management and drug development for response fluctuations largely depend on clinical rating instruments such as the PD home diary[15], Unified PD rating scale (UPDRS)[16] or modified Abnormal Involuntary Movement Scale (mAIMS)[17]. Although these assessments are the current reference standard, they have biases and errors, since patients forget to record their motor state, do not adequately recognize the respective motor state, or confuse motor and non-motor symptoms[18–21]. The PD home diary was developed to quantify motor fluctuations. It is frequently used in clinical trials on motor fluctuations and fairly widespread in clinical routine[15,18,22]. For its use, patients are asked to indicate their predominant motor status during half-hour time periods throughout the day using the categories Asleep, Off state, On without dyskinesias (On state), and On with dyskinesia (Dyskinetic state). Recent validation studies however demonstrated that the PD home diary unsatisfactorily reflects actual motor states as compared to simultaneous clinical

[1]Department of Neurology, University Medical Center Rostock, Rostock, Germany. [2]German Center for Neurodegenerative Diseases (DZNE) Rostock-Greifswald, Rostock, Germany. [3]Division of Neurology, Department of Clinical Sciences Lund, Lund University, Lund, Sweden. [4]Department of Neurology, Skåne University Hospital, Lund, Sweden. [5]Global Kinetics Pty Ltd, 31 Queen St., Melbourne, VIC, Australia. [6]Movement Disorders Hospital, Beelitz-Heilstätten, Beelitz, Germany. [7]Department of Neurology, Otto-von-Guericke University, Magdeburg, Germany. [8]Bionics Institute, Melbourne, VIC, Australia. [9]The Department of Medicine, The University of Melbourne, St Vincent's Hospital, Fitzroy, VIC 3010, Australia. ✉email: matthias.loehle@med.uni-rostock.de; alexander.storch@med.uni-rostock.de
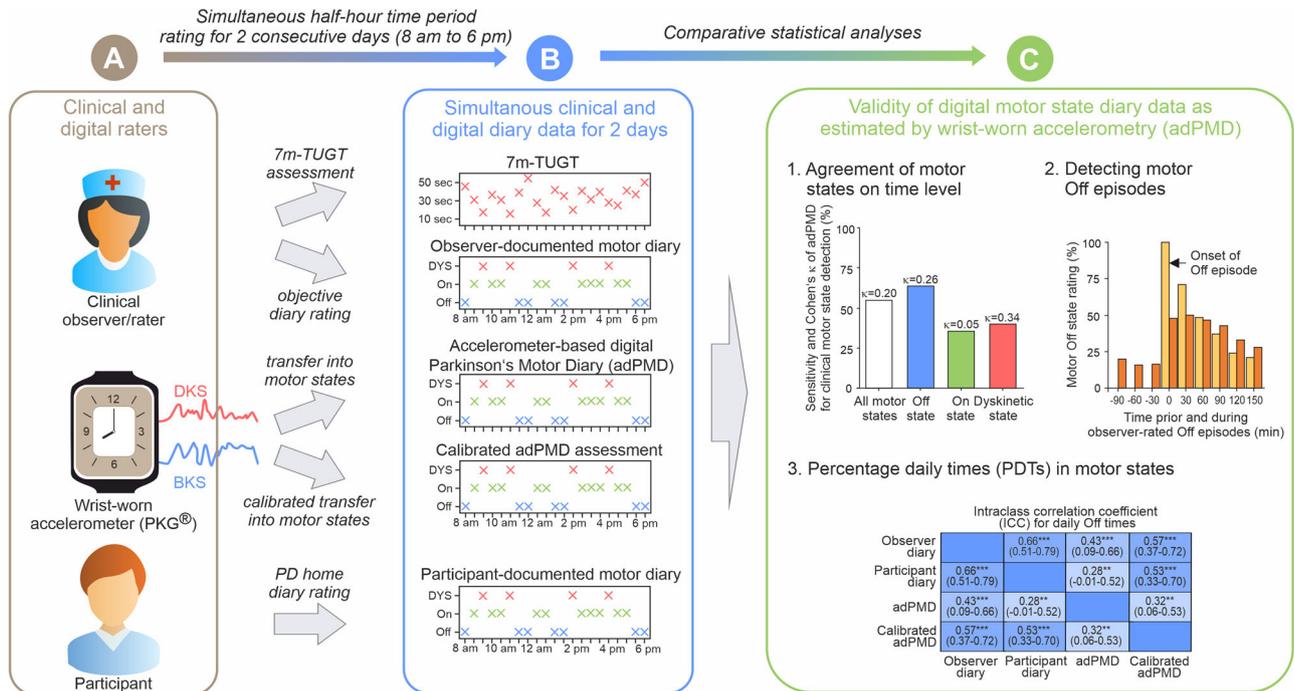
Fig. 1 Graphical synopsis of application study of wrist-wearable accelerometry for objective motor diary assessment in fluctuating Parkinson's disease. The present study was designed to validate wrist-wearable accelerometry-based digital Parkinson's Motor Diary (adPMD) using the Parkinson's Kinetigraph (PKG®) as the hardware device for objective motor diary assessment as an alternative application of this digital health technology (DHT). **A** Motor function was simultaneously assessed by fluctuating Parkinson's disease (PD) patients using the PD home diary data of motor states (Off state, On state dyskinetic state) by the adPMD as well as by clinical observers and participants for each half-hour time period from 8 a.m. until 6 p.m. on 2 consecutive days. In addition, the 7 meter Timed-Up-and-Go test (7 m-TUGT) was performed simultaneously for each time period. **B** The quantitative accelerometer scores for bradykinesia (BKS) and dyskinesia (DKS) were transferred via median scores for each half-hour time period into motor state data similar to the diaries (Off state, On state and Dyskinetic state). **C** Comparative statistical analyses using the clinical rater diary data as the gold-standard outside criterion were applied to estimate the validity of adPMD as a clinically used wrist-worn accelerometry in PD for (1) temporal agreement of motor state ratings for the half-hour time periods, (2) detection of Off episodes following an On phase and (3) assessment of percentage daily times in the various motor states. Ancillary analyses used the PD home diary data as documented by study participants as an alternative outside criterion.

observation[19,23,24]. Consequently, there is a strong need for continuous and objective monitoring of motor function in PD.

Recent advancements in digital health technologies (DHTs) provide user-friendly wearables including smartwatch-based applications using inertial measurement unit sensors (e.g. accelerometers, gyroscopes) combined with sufficient battery life that can be tailored for continuous monitoring of motor symptom severity and thus presumably motor fluctuations[25,26]. Commercially available wearable devices such as STAT-ON (Sense4Care, Barcelona, Spain)[27,28], Parkinson KinetiGraph® (PKG®; Global Kinetics Corporation Ltd., Melbourne, Australia)[29–34], and smartwatch applications such as StrivePD (Rune Labs Inc., San Francisco, USA), and Verily smartwatch (Verily Life Science, San Francisco, CA)[35] have various characteristics (e.g. single versus multiple sensors) and deliverables to assess motor symptoms and their fluctuations in PD, but their application in clinical trials as well as their implementation in routine care is still limited[36,37]. However, the feasibility of using DHTs for measuring motor function/fluctuations has been demonstrated and first clinical evidence is provided that continuous, objective monitoring of motor symptoms using wearable biosensors may enhance clinical decision-making and outcome in PD patients[34,38].

Continuous monitoring of motor symptoms in PD in an uncontrolled real-world environment without interrupting daily activities can be confounded by voluntary inactivity and (over)-activity (e.g. exercise) as well as tremor. Most DHTs overcome this problem by increasing the sample size by recording for several days and by reducing times of inactivity (evening) combined with subsequent averaging/smoothing[29,35,39,40]. It therefore remains largely unknown whether DHT data reflect the motor state at a given time

point as an important information to manage unpredictable motor fluctuations and day-to-day variations in motor function. Available studies compared DHT outcomes with PD home diary data and demonstrated fair to good validity for daily times in the various motor states (mostly expressed as percentage daily times, PDT), but limited temporal agreement at a given time point (accuracy: 53%)[27,31]. Validity data of DHTs to continuously and temporarily monitor motor states similar to the PD home diary in an uncontrolled environment with clinical assessment as the outside criterion are not available.

We here report on the VALIDATE-PD study[23,24], which was designed to apply wrist-wearable accelerometry for objective motor diary assessment in fluctuating PD in a routine clinical environment (see Fig. 1 for study synopsis). Our main aim is to establish whether a wearable accelerometer-based digital Parkinson's Motor Diary (adPMD) using the Parkinson's Kinetigraph (PKG®) device can detect the clinical state at a point in time using simultaneous half-hourly performed clinical examinations by experienced raters with the PD home diary as the main comparator. Importantly, the PKG® system was originally designed to use averaging of daily traces to aid routine clinical care and has no commercially available format of providing either two minutes or half-hour data nor have there ever been any claims for it to function in this time scale. We here use the PKG® accelerometer system in an alternative manner and format by converting its continuous data into one of the three motor categories of the PD home diary resulting in adPMD ratings[15,22,31,41]. We then estimate the validity of the adPMD data with respect to three different diary outcomes: (1) Detecting the diary motor states at a given time point as routinely used clinical parameters to detect unpredictable

Off state and Dyskinetic state periods. (2) Detecting Off episodes (Off state episode following motor On state) as another important aspect in routine care for adaption of medication[42]. (3) PDT of the motor states as a frequently used read-out in clinical trials[18,43–48].

RESULTS

Demographic and clinical data

In three centers in Germany and Sweden, we screened 96 participants for eligibility of whom 91 (95%) were successfully included into the VALIDATE-PD study according to inclusion and exclusion criteria. 63 participants (69%) had 2 days of calibrated adPMD datasets and were included in the final study analyses (see Participants & Methods section). Demographic and clinical characteristics of study cohort are displayed in Table 1. We did not observe relevant differences between the cohorts from the two countries. Clinical scores and antiparkinsonian medication were representative for an advanced PD cohort. A glossary of adPMD terms and clinical fluctuation parameters are displayed in Table 2.

Table 1. Demographic and clinical characteristics of adPMD study cohorts.

	Total cohort (n = 63)	German subcohort (n = 40)	Swedish subcohort (n = 23)	P value (cohorts)
Male/Female, n (%)	29 (46%)/34 (54%)	19 (48%)/21 (52%)	10 (44%)/13 (56%)	0.758[§]
Age, Median (IQR) in years	66 (60–73)	64 (57–69)	70 (60–76)	0.038[$]
Disease duration, Median (IQR) in years	9 (6–14)	10 (8–14)	7 (5–14)	0.159[$]
Symptom duration, Median (IQR) in years	11 (8–15)	11 (9–16)	8 (7–14)	0.076[$]
Duration of fluctuations, Median (IQR) in months	54 (34–99)	70 (38–106)	60 (33–86)	0.607[$]
Hypokinetic fluctuations	60 (35–103)	70 (38–106)	48 (20–86)	0.068[$]
Hyperkinetic fluctuations	36 (24–57)	39 (25–60)	35 (20–64)	0.453[$]
Clinical Phenotype				0.614[§]
Tremor Dominant, n (%)	11 (18%)	7 (18%)	4 (18%)	
Axial Dominant, n (%)	0 (0%)	0 (0%)	0 (0%)	
Appendicular Dominant, n (%)	7 (11%)	3 (8%)	4 (18%)	
Rigor Dominant, n (%)	1 (2%)	1 (3%)	0 (0%)	
Postural Instability and Gait Difficulty, n (%)	44 (70%)	30 (75%)	14 (64%)	
Reported motor complications during structured interview				
Nighttime off, n (%)	54 (86%)	34 (85%)	20 (87%)	1.000[§]
Wearing-off, n (%)	56 (89%)	37 (93%)	19 (83%)	0.247[§]
Delayed on, n (%)	35 (56%)	31 (78%)	4 (17%)	<0.001[§]
On-off phenomenon, n (%)	39 (62%)	26 (65%)	13 (59%)	0.645[§]
Peak-dose dyskinesia, n (%)	47 (75%)	31 (78%)	16 (70%)	0.486[§]
Biphasic dyskinesia, n (%)	12 (19%)	7 (18%)	5 (23%)	0.740[§]
Off-dose dystonia, n (%)	32 (51%)	22 (55%)	10 (48%)	0.583[§]
PD medications				
Levodopa, n (%)	63 (100%)	40 (100%)	23 (100%)	-
Dopamine agonizts, n (%)	43 (68%)	24 (60%)	19 (83%)	0.063[§]
MAO B inhibitors, n (%)	43 (68%)	28 (70%)	15 (65%)	0.695[§]
COMT inhibitors, n (%)	50 (79%)	36 (90%)	14 (61%)	0.009[§]
Amantadine, n (%)	25 (40%)	20 (50%)	5 (22%)	0.027[§]
Levodopa dose (mg per day), Median (IQR)	550 (450–725)	506 (425–700)	550 (500–875)	0.054[$]
Total levodopa equivalent dose (mg per day), Median (IQR), calculated according to[68]	1199 (976–1505)	1331 (1030–1635)	1112 (814–1263)	0.003[$]
Clinical scales				
MDS-PDRS Total score On state, Median (IQR)	63 (45–82)	64 (52–86)	53 (39–58)	0.007[$]
Part I	11 (8–16)	13 (8–14)	13 (7–16)	0.133[$]
Part II	14 (9–19)	16 (10–21)	11 (7–17)	0.039[$]
Part III	28 (18–40)	28 (21–41)	25 (16–31)	0.149[$]
Part IV	8 (6–10)	9 (7–10)	6 (4–8)	0.001[$]
Hoehn & Yahr stage, Median (IQR)	4 (3–5)	4 (3–4)	4 (3–5)	0.676[+]
Montreal Cognative Assessment score, Median (IQR)	27 (25–28)	27 (26–28)	26 (24–28)	0.111[$]
Beck'sches Depression Inventory score, Median (IQR), assessed in German cohort only	10 (4–16)	10 (4–16)		

[§]P values are from χ^2 test or fisher exact test as appropriate.
[$]P values are from Mann-Withney U test.
[+]P values from Jonckheere Terpstra test.

Table 2. Glossary of PKG® and wrist-wearable accelerometer-based digital Parkinson's Motor Diary (adPMD) terms and clinical parameters.

Parameter	Abbreviation	Explanation	Ref.
PKG® quantitative scores per time periods			
Median bradykinesia score	mBKS	30 min median of bradykinesia score <80 (inactivity/non-wrist wear excluded) for diary time period	[29]
Timed bradykinesia score	mBKS$_{timed\ (28)}$	28 min median bradykinesia score with inactivity/non-wrist wear excluded from recordings exactly 28 min around clinical testing (seven 2-min epochs prior and seven 2-min-epochs after clinical testing), German cohort only	
Timed bradykinesia score	mBKS$_{timed\ (16)}$	16 min median bradykinesia score with inactivity/non-wrist wear excluded from recordings exactly 16 min (eight 2-min epochs) prior to but excluding clinical testing, German cohort only	
Dyskinesia score	mDKS	30 min median of dyskinesia score with activity excluded for diary time period	[29]
Timed dyskinesia score	mDKS$_{timed\ (28)}$	28 min median dyskinesia score with activity excluded from recordings exactly 28 min around clinical testing (seven 2-min epochs prior and seven 2-min epochs after clinical testing), German cohort only	
Wrist-wearable accelerometer-based digital Parkinson's motor diary (uncalibrated/calibrated adPMD) data per half-hour periods			
Asleep	-	Asleep as predicted by adPMD through thresholding of BKS (uncalibrated data) or by adPMD through per patient calibration of threshold of 30 min median BKS using day 0 of patient diary (calibrated data)	[31]
Off state	Off state	Motor Off state as predicted by adPMD through thresholding of 30 min median BKS (uncalibrated data) or by adPMD through per patient calibration of threshold of 30 min median bradykinesia scores using day 0 of patient diary (calibrated data)	[31]
On without dyskinesia state	On state	Motor On without dyskinesia state as predicted by adPMD through thresholding of 30 min median BKS and DKS (uncalibrated data) or by adPMD through per patient calibration of threshold of 30 min median BKS and DKS using day 0 of patient diary (calibrated)	[31]
On with dyskinesia state	Dyskinetic state	Motor On with dyskinesia state as predicted by adPMD through thresholding of 30 min median DKS (uncalibrated data) or by adPMD through per patient calibration of threshold of 30 min median dyskinesia score using day 0 of patient diary (calibrated data)	[31]
Observer or participant diary data per half-hour periods			
Asleep	-	Asleep as rated by clinical observer (experienced movement disorder specialist) or study participant	[24]
Off state	Off state	Motor Off state as rated by clinical observer (experienced movement disorder specialist) or study participant	[24]
On without dyskinesia state	On state	Motor On without dyskinesia state as rated by clinical observer (experienced movement disorder specialist) or study participant	[24]
On with dyskinesia state	Dyskinetic state	Motor On with dyskinesia state as rated by clinical observer (experienced movement disorder specialist) or study participant	[24]
Off episode	Off episode	Off state episode of at least 0.5 h duration following a motor On state period of at least 1.5 h	[42]
Off time	Off time	Daily time in Off state (between 8 a.m. and 6 p.m.), expressed in percentage of total half-hours with diary data (PDT)	
On without dyskinesia time	On time	Daily time in On state (between 8 a.m. and 6 p.m.), expressed in percentage of total half-hours with diary data (PDT)	
On with dyskinesia time	Dyskinetic time	Daily time in Dyskinetic state (between 8 a.m. and 6 p.m.), expressed in percentage of total half-hours with diary data (PDT)	
Other clinical parameters			
7 m-Timed Up and Go Test	7m-TUGT	Modified 7 m-TUGT test results as a quantitative clinical measure of bradykinesia	[58]

Temporal agreement of adPMD motor state classification with clinical diary data

To allow for direct comparisons of adPMD and frequent clinical motor assessment data, quantitative PKG® scores were transferred into motor states as decribed previously[31] with the routinely used consistent threshold (75th percentile) in 30 min time periods resulting in adPMD ratings. In total, 2384 (90.1% of all periods or 97.5% of all adPMD scorings) half-hour PKG® quantitative scores were classifiable into motor states with 6.0% classified as Asleep, 40.9% as motor Off state, 32.7% as On state and 20.4% as Dyskinetic state.

To compare adPMD data with clinical observer motor diary data, we next estimated the sensitivities of adPMD for the detection of respective motor states as rated by clinical observers on the group level (Fig. 2a): adPMD ratings correctly recognized 55% of all motor states, but 64% of observed Off states, only 36% of On and 40% of Dyskinetic state simultaneous to the observer. Together with the specificities of adPMD ratings, these figures translated into balanced accuracies (avarage between sensitivy and specificity) between 52% for On state and 65% for Off and Dyskinetic state (Table 3, Supplementary Table 1). Corresponding Cohen's κ values ranged from 0.05 for On state to 0.34 for Dyskinetic state. The adPMD rated On state in 21% of the intervals with observed Off state (Fig. 2b). Even more strikingly, the adPMD decided Off in 45% of those intervals in which the observer had actually noted On state.

In ancillary analyses, we calculated various validity measures of the adPMD ratings for the detection of diary motor state data from study participants (PD Home diary data). The adPMD correctly recognized 40% of all motor states, but 57% of patient-rated Off states, 34% of On states, and 34% of Dyskinetic state simultaneous to the participant's rating in their PD home diaries (Fig. 2c, see

a

b

- ■ adPMD Off state
- ■ adPMD On state
- ■ adPMD Dyskinetic state
- □ adPMD Asleep

c

d

- ■ adPMD Off state
- ■ adPMD On state
- ■ adPMD Dyskinetic state
- □ adPMD Asleep

e

f
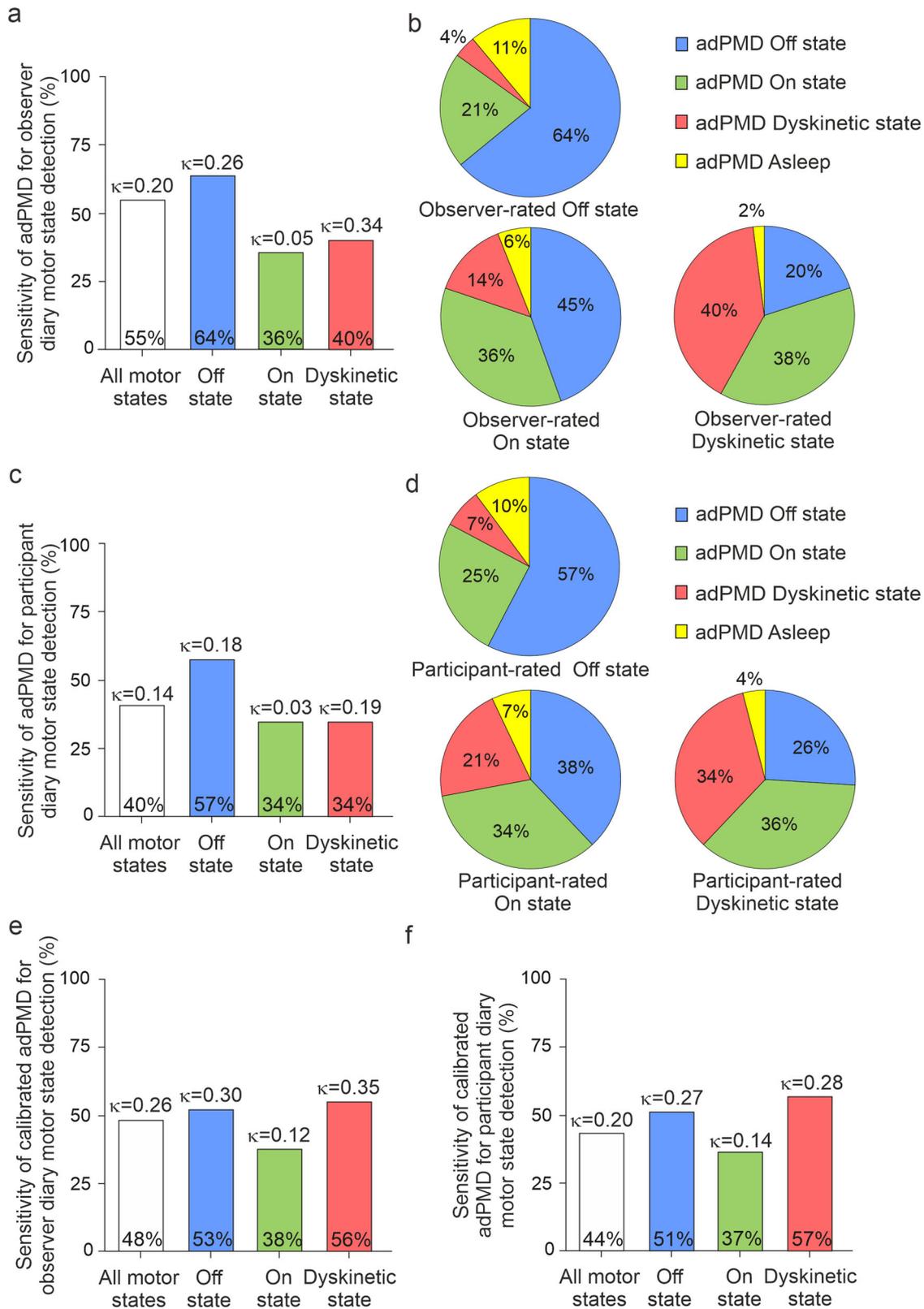
Supplementary Table 2 for other validity measures). The adPMD rated On state in 25% of the intervals with participant-rated Off state (Fig. 2d). Moreover, the adPMD decided Off or On state in 38% and 34%, respectively, of those intervals in which the participants had actually noted On state.

Comparisons of true ratings (true positives+true negatives) by adPMD between clinical observer and participant diary data revealed significantly more true ratings of adPMD for detecting observer as compared to participant ratings when analyzing all motor states or only Off state ($P < 0.001$; McNemar test), but no

Fig. 2 Temporal agreement of uncalibrated and calibrated wearable accelerometer-based digital Parkinson's Motor Diary (adPMD) ratings and observer- and participant-documented diary motor states on the half-hour time period level. a, c Sensitivities expressed in percent and Cohen's κ values of uncalibrated adPMD ratings for the detection all motor states (white color), Off state (blue color), On state (green color) and Dyskinetic state (red color) as rated by clinical observers (**a**) and by participants (**c**) using the PD Home diary categories. Preferred choices in observer-documented diaries (**b**) and participant diaries (**d**) in the respective adPMD motor states. Figures are based on 2339 (88.4% of all periods) simultaneous half-hourly adPMD and clinical observer ratings (**a, b**) and 2338 (88.4% of all time periods) simultaneous adPMD and participant ratings of 63 participants (**c, d**). Sensitivities in percent and Cohen's κ values of calibrated adPMD ratings for the detection all motor states, Off state, On state and Dyskinetic state as rated by clinical observers (**e**) and participants (**f**) using the PD home diary (1877 [70.9%] and 1913 [72.3%] half-hour periods with simultaneous observer or participant ratings of 63 participants; for preferred choices, refer to Supplementary Figs. 4a, b). For additional test performance measures, refer to Table 3, Supplementary Tables 1 and 2.

relevant differences for On and Dyskinetic states ($P \geq 0.05$; McNemar tests). Similar results were obtained for true positive ratings (sensitivity) and true negative ratings (specificity; data not shown).

To address potential patient-inherent factors, we calibrated adPMD data to model the first day of diary entries by each participant (PD Home diary data) for Off and Dyskinetic state. We thus used algorithms classifying the continuous accelerometer data into motor states for every half-hour time period between 8 a.m. and 6 p.m. using scores and diary ratings from day 0 as decribed earlier[31]. Test performance measures including balanced accuracies and Cohen's κ values were similar to those estimated by using uncalibrated adPMD data (Fig. 2e, f; Table 3; Supplementary Fig. 1, Supplementary Tables 1,2).

Validation of adPMD classifications on participant level
Validity of adPMD data when compared to observer diary data on the participant level were estimated using the sensitivity, specificity, balanced accuracy and Cohen's κ for each participant (Fig. 3, Table 3). Taking the observer diary as gold standard criterion, median sensitivities of adPMD data were 41% of all observer-rated motor states, 67% of Off states, but only 27% of On state and 23% of Dyskinetic state (Fig. 3a). The corresponding median balanced accuracies ranged from 52% to 58% and Cohen's κ values from 0.04 to 0.21 (Fig. 3b, Table 3). Importantly, only the sensitivities for Dyskinetic state showed a moderate correlation with the respective PDTs ("number of positive hits"), but all other validity measures—particularly all Cohen's κ values —did not correlate with the magnitude of test outcomes (PDTs) making a relevant proportional bias unlikely ($P \geq 0.05$ from Pearson's correlation test; Fig. 3b, c, e, f). In addition, the validity measures were not associated with potential confounding demographic or clinical parameters ($P \geq 0.05$ from Mann-Whitney U test or Pearson's correlation test as appropriate; see Table 1 for covariable list, data not shown). Calibration of adPMD data did not relevantly change validity measures of adPMD for the detection of clinical observer diary motor states (see Supplementary Fig. 2a, b. Table 3). Similar results were obtained for the test performance measures including sensitivities and Cohen's κ for each participant of adPMD data with participant-rated diary data (Supplementary Fig. 2c–f).

Agreement of adPMD classification with clinical ratings during Off episodes
Rather than compare adPMD and diary scores from half-hour periods, the timing of adPMD scores was synchronized with Off episodes[42]. Off episodes were defined as an Off state period of at least 0.5 h duration following a motor On state period of at least 1.5 h (Supplementary Fig. 3). We did not observe relevant changes of median bradykinesia scores (mBKS) at the beginning of the Off episodes with an increase of 12% (IQR: −9 to 53%) for observer-reported Off episodes ($P = 0.615$; Mann-Whitney U test). Consistently, there was no change in 7m-TUGT as a quantitative clinical bradykinesia outcome measure (17% [8–35%]; $P = 0.693$; Mann-Whitney U-test). There was, however, a moderate correlation

between the change in mBKS with the change of 7m-TUGT results at the start of observer Off episodes ($r = 0.364$, $P = 0.007$; Pearson's correlation test).

We next analyzed the temporal agreement of adPMD ratings and clinical diary ratings with respect to the time course of Off episodes (Fig. 3g, h). Sensitivities during the first 2.5 h of the Off episodes ranged from 31% to 72% for adPMD data with no major differences over the time course of Off episodes. Corresponding balanced accuracy values ranged from 56% to 79% and Cohen's κ values from −0.15 to 0.18. Calibration of adPMD data to participants' individual thresholds did not lead to relevant differences with sensitivities of 43–60%, balanced accuracies of 59–78%, and Cohen's κ values of −0.22 to 0.25 (Supplementary Fig. 4; for data on participant-documented diaries, refer to Supplementary Results).

Correlation of PDTs between adPMD data and clinical ratings
Daily times spent in the three different motor states (expressed here as percentage daily times; PDTs) calculated from PD home diary have been repeatedly used as the primary outcome measures in clinical trials on motor fluctuations in advanced PD[18].

On the group level, we detected similar PDTs for motor On state when comparing adPMD and observer-rated diary data with medians of 33% and 31% daily On time for observer and adPMD data (Fig. 4), but different PDTs for Off and Dyskinetic state with medians of 21% and 46% Off time and 34% and 8% Dyskinetic time for observer and adPMD data. Calibration of adPMD data resulted in an increase in Dyskinetic time at cost of Off time leading to an overall better association of adPMD data with observer diary data on the group level (Fig. 4).

Analyses on the participant level using Pearson correlation analyses of individual PDTs revealed moderate correlations of Off time between adPMD and observer diary data with moderate proportional bias in Bland–Altman plot, but only a weak correlation with MDS-UPDRS percentage Off time (Fig. 4b–d). No relevant correlations of adPMD and clinical diary data were detected for On times (Fig. 4f–h). For Dyskinetic time, we detected a strong correlation between adPMD and observer diary data with no proportional bias, but only a weak correlation between calibrated adPMD data and clinical observer data (Fig. 4j–l). Multivariate regression analyses controlling for the candidate co-variates age, sex, symptom duration, fluctuation duration, disease severity (MDS-UPDRS III motor score), MDS-UPDRS part IV as a measure of motor complication, BDI and MoCA largely confirmed these correlation test results (Supplementary Table 3). Validity analyses using ICC calculation revealed moderate validity for adPMD ratings for Off time and Dyskinetic time when correlated with observer diary data (Fig. 5). Calibration of adPMD data did not lead to relevant changes for most measures, but did reduce the ICC value for Dyskinetic time.

Factors that influence validity of adPMD motor state assessment
To determine factors influencing the transfer of continuous PKG® quantitative scores into ordinal adPMD ratings and thus the

Table 3. Validity parameters of the adPMD according to the various cut-offs as determined by ROC plot analyses for transferring PKG® scores into adPMD states.

	ROC-AUC (95%CI)	adPMD test validity measures on group level				adPMD test validity measures on participant level, median (95%CI)				Percentage daily times, median (95%CI) ICC
		Sensitivity, recall (%)	Specificity (%)	Balanced accuracy (%)	Cohen's κ	Sensitivity, recall (%)	Specificity (%)	Balanced accuracy (%)	Cohen's κ	
Discrimination of Off versus Non-Off state from simultaneous clinical observer ratings										
mBKS (standard cut-off)	0.740 (0.718–0.762)***	64%	67%	65%	0.26	58% (0–100%)	67% (11–100%)	58% (39–83%)	0.13 (−0.15–0.61)	0.43 (0.09–0.66)+++
mBKS (calibrated cut-off)	0.867 (0.854–0.880)***	53%	79%	66%	0.30	47% (0–100%)	83% (31–100%)##	56% (43–83%)	0.20 (−0.16–0.64)	0.57 (0.37–0.72)+++
mBKS (individualized cut-offs, all participants)[a]	0.715 (0.691–0.738)***	62%	74%	68%	0.31	67% (25–100%)	72% (35–100%)	67% (41–90%)##,§§	0.25 (−0.01–0.68)##,§	0.57 (0.38–0.82)+++
mBKS (individualized cut-offs, sig. ROC only)[a,b]	0.822 (0.791–0.853)***	70%	80%	75%	0.48	77% (34–100%)	80% (58–100%)	81% (59–88%)	0.45 (0.23–0.75)	0.74 (0.47–0.89)+++
Discrimination of Dyskinetic state versus Non-dyskinetic state from simultaneous clinical observer ratings										
mDKS (standard cut-off)	0.774 (0.754–0.793)***	40%	90%	65%	0.34	26% (0–88%)	95% (31–100%)	55% (45–83%)	0.21 (−0.13–0.79)	0.51 (0.18–0.71)+++
mDKS (calibrated cut-off)	0.842 (0.828–0.857)***	56%	78%	67%	0.35	57% (0–100%)#	84% (0–100%)###	58% (46–92%)	0.35 (−0.05–0.83)	0.20 (−0.05–0.42)
mDKS (individualized cut-offs, all participants)[a]	0.748 (0.727–0.770)***	75%	69%	72%	0.41	75% (30–100%)##,§	75% (25–100%)###	71% (49–94%)###,§§§	0.41 (−0.03–0.88)#,§	0.34 (0.04–0.59)+
mDKS (individualized cut-offs, sig. ROC only)[a,b]	0.811 (0.785–0.837)***	74%	83%	79%	0.57	81% (46–100%)	81% (67–100%)	80% (68–96%)	0.54 (0.27–0.93)	0.49 (0.10–0.76)+

***Represents $P < 0.001$ from ROC-AUC calculation.

+Represents $P < 0.05$ and +++$P < 0.001$ from intraclass correlation coefficient (ICC) calculation.

#Represents $P < 0.05$, ##$P < 0.01$, ###$P < 0.001$ when compared to respective result with adPMD standard cut-off, §$P < 0.05$, §§$P < 0.01$, §§§$P < 0.001$ when compared to results with calibrated adPMD cut-offs from Friedman tests with post-hoc Wilcoxon test with Bonferroni adjustment for α inflation.

[a]Individual cut-offs were estimated by ROC analysis for each participant for the discrimination between observer Off state and Non-Off state by mBKS and between observer-documented Dyskinetic and Non-Dyskinetic state by mDKS.

[b]Only participants with significant discrimination in ROC analyses (AUC significantly larger than 0.5) between observer Off state and Non-Off state by mBKS ($n = 20$) and between observer-documented Dyskinetic and Non-Dyskinetic state by mDKS ($n = 29$) were used for these analyses (not included in comparative statistics).
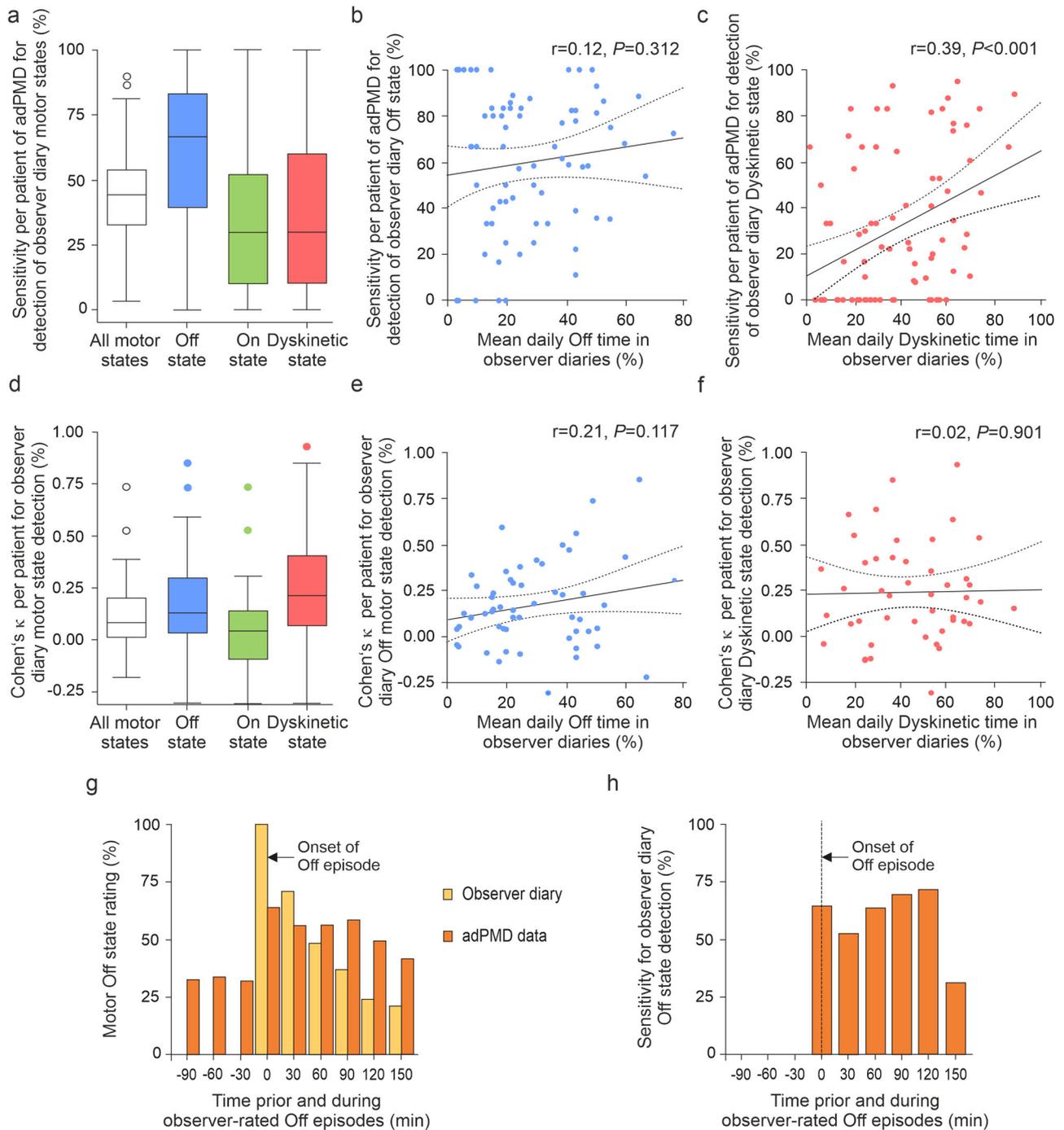
Fig. 3 Temporal agreement between wearable accelerometer-based digital Parkinson's Motor Diary (adPMD) ratings and observer-documented motor states on the participant level and during the time course of Off episodes. Sensitivities (**a**–**c**) and Cohen's κ values (**d**–**f**) of uncalibrated adPMD data for the detection observer-documented motor states on the participant level. Sensitivities expressed in percent (**a**) and Cohen's κ values (**d**) for all motor states (white color), Off state (blue color), On state (green color) and Dyskinetic state (red color) based on simultaneous half-hourly adPMD and clinical observer ratings in 63 participants. 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, and outliers scoring more than 1.5 × IQR but at most 3 × IQR outside the quartiles. Correlation analyses of sensitivities and percentage daily times for Off state (**b**) and Dyskinetic state (**c**). Correlation analyses of Cohen's κ values and percentage daily times for Off state (**e**) and Dyskinetic state (**f**). Numbers in right corner of diagrams (**b**, **c**, **e**, **f**) represent Pearson correlation coefficients and *P* values. For additional test performance measures, refer to Table 3. **g**, **h** Temporal agreement between adPMD and observer-documented motor states during the time course of Off episodes. Off episodes were defined as a minimum of 30 min Off preceded by at least 90 min with On time as judged by the clinical observers. **g** Proportions of Off responses on simultaneous observer ratings (yellow color) and adPMD motor states (brown color) synchronized to the onset of observer-rated Off episodes. **h** Sensitivities of uncalibrated adPMD ratings for the detection of observer Off ratings synchronized to the onset of the observer-documented Off episodes. Values are provided as numbers from 84 Off episodes in 51 participants.
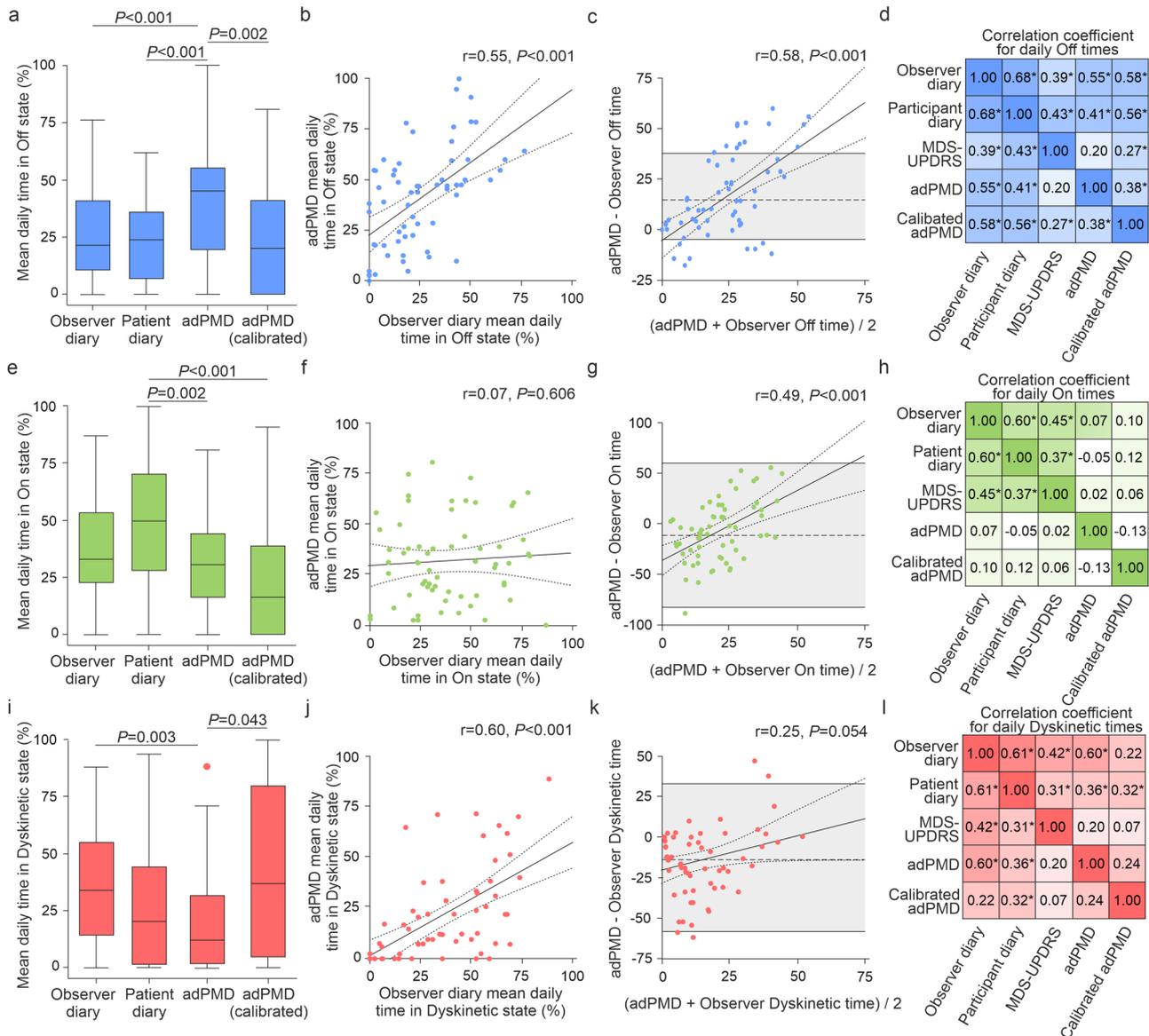
Fig. 4 **Proportions for times spent of motor states as assessed by observer diaries, participant diaries and wearable accelerometer-based digital Parkinson's Motor Diary (adPMD) ratings.** Distribution of daily Off times (**a**), On times (**e**) and Dyskinetic times (**i**) based on the simultaneous, half-hourly performed clinical diary/adPMD ratings from 63 participants from 2 consecutive days (8 am to 6 pm). 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, and outliers scoring more than $1.5 \times IQR$ but at most $3 \times IQR$ outside the quartiles. Displayed P values are from from Friedman tests with *post-hoc* Wilcoxon signed-rank tests with Bonferroni correction for multiple comparisons. Correlation analyses of daily Off times displaying exemplarily the scatter plot of calibrated adPMD and observer Off times and the corresponding Bland-Altman plot (**b**, **c**) and the corellation matrix (**d**). Correlation analyses of daily On times displaying exemplarily the scatter plot of calibrated adPMD and observer On times and the corresponding Bland-Altman plot (**f**, **g**) and the correlation matrix (**h**). **j**–**l** Correlation analyses of daily Dyskinetic times displaying exemplarily the scatter plot of calibrated adPMD and observer Dyskinetic times and the corresponding Bland-Altman plot (**b**, **c**) and the correlation matrix (**d**). Solid lines in scatter plots represent the regression line with 95%CI (dotted lines) and the gray area in (**c**, **g**, **k**) the limit of argreement (mean ± 1.96 SD). Numbers in right corner of scatter plots represent Pearson correlation coefficients and P values. In correlation matrices, * represent $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$ from Pearson's correlation tests.

validity of adPMD motor state assessment, we first determined the influence of thresholding on the validity of temporal agreement of adPMD ratings. Receiver operating curve (ROC) analyses revealed significant sensitivity-specificity relationships for both mBKS and 7m-TUGT results for detecting observer-documented Off state and median dyskinesia scores (mDKS) for the detection of Dyskinetic state (Fig. 6a–d, for precision-recall curve [PRC] plot analyses, refer to Supplementary Text, Supplementary Table 1). The optimal cut-off value (at maximal Youden index) for standard mBKS to discrimate Off versus Non-Off states was 25 units, leading to a maximal Cohen's κ of 0.303. For Dyskinetic state detection, the optimal cut-off of standard mDKS scoring was 2 units with a maximal Cohen's κ of 0.409.

We then determined the individual optimal cut-off values for mBKS and mDKS for the detection of the respective motor states for each participant by using ROC analyses. Discrimination (AUC significantly larger than 0.5) between observer Off state and Non-Off states by mBKS succeeded in 20 out of 56 participants with Off

a

Intraclass correlation coefficient (ICC) for daily Off times

	Observer diary	Participant diary	adPMD	Calibrated adPMD
Observer diary		0.66*** (0.51–0.79)	0.43*** (0.09–0.66)	0.57*** (0.37–0.72)
Participant diary	0.66*** (0.51–0.79)		0.28** (−0.01–0.52)	0.53*** (0.33–0.70)
adPMD	0.43*** (0.09–0.66)	0.28** (−0.01–0.52)		0.32** (0.06–0.53)
Calibrated adPMD	0.57*** (0.37–0.72)	0.53*** (0.33–0.70)	0.32** (0.06–0.53)	

b

Intraclass correlation coefficient (ICC) for daily On times

	Observer diary	Participant diary	adPMD	Calibrated adPMD
Observer diary		0.55*** (0.31–0.71)	0.05 (−0.18–0.30)	0.09 (−0.13–0.31)
Participant diary	0.55*** (0.31–0.71)		−0.04 (−0.23–0.17)	0.09 (−0.10–0.29)
adPMD	0.05 (−0.18–0.30)	−0.04 (−0.23–0.17)		−0.12 (−0.35–0.13)
Calibrated adPMD	0.09 (−0.13–0.31)	0.09 (−0.10–0.29)	−0.12 (−0.35–0.13)	

c

Intraclass correlation coefficient (ICC) for daily Dyskinetic times

	Observer diary	Participant diary	adPMD	Calibrated adPMD
Observer diary		0.59*** (0.39–0.73)	0.51*** (0.18–0.71)	0.20 (−0.05–0.42)
Participant diary	0.59*** (0.39–0.73)		0.35** (0.12–0.54)	0.27** (0.04–0.48)
adPMD	0.51*** (0.18–0.71)	0.35** (0.12–0.54)		0.17 (−0.05–0.38)
Calibrated adPMD	0.20 (−0.05–0.42)	0.27** (0.04–0.48)	0.17 (−0.05–0.38)	

Fig. 5 Reliability (intraclass correlation coeﬁcient; ICC) of motor state times as assessed by observer diaries, participant diaries and wearable accelerometer-based digital Parkinson's Motor Diary (adPMD) ratings. ICC values and their 95%CIs of proportions of Off times (**a**), On times (**b**) and Dyskinetic times (**c**). ICC estimates and their 95%CIs were calculated based on single-rating, absolute-agreement, 2-way mixed-effects models with two rating instruments across 63 participants. ** represents $P < 0.01$ and ***$P < 0.001$ from ICC analyses.

state periods (36%), and between observer-documented Dyskinetic and Non-Dyskinetic states by mDKS in 29 out of 56 participants with Dyskinetic state periods (52%; $P = 0.063$; χ^2 test; Fig. 6e–g, for PRC plot analyses, refer to Supplementary Text, Supplementary Fig. 5). By searching for potential factors predicting significant discrimination of Off and Dyskinetic state by the respective PKG® scores, we did not detect any significant association of major demographic, clinical and observer diary parameters with individual thresholding data (see Supplementary Results, Supplementary Fig. 6a, b for details).

The ranges of individual cut-offs were wide for both mBKS and mDKS (Fig. 6h) with a median Youden index of 0.44 (95%CI: 0.07–0.81) for mBKS, and 0.49 (95%CI: 0.09–0.92) for mDKS. Using these individual cut-offs did—in general—not change the test performance for detecting Off state using mBKS or Dyskinetic state using mDKS (Table 3; Supplementary Fig. 6c, d). However, analyzing only individuals with relevant discrimination in ROC analyses led to a maximal Cohen's κ of 0.484 for mBKS and Off detection and a maximal Cohen's κ of 0.574 for mDKS and Dyskinetic state detection (Table 3; Supplementary Fig. 6e, f).

Using the individualized cut-offs, the adPMD test performance measures on the group level did not change for Off state detection as compared to standard adPMD cut-offs, but raised for Dyskinetic state detection with Cohen's κ value indicating moderate validity (Table 3). On the participant level, the median balanced accuracy and Cohen's κ for Off and Dyskinetic state were significanty increased, for the Dyskinetic state to the moderate validity level (Fig. 6i–l; Table 3). Selecting participants with significant ROC analyses results, test performance measures showed moderate validity for both Off and Dyskinetic state detection (Table 3).

Validity analyses of PDTs from adPMD ratings with individualized thresholds using ICC calculations revealed no relevant changes when compared to standard cut-off adPMD assessments with moderate validity for individualized adPMD data for Off times (ICC = 0.57 [95%CI: 0.38–0.82]) and Dyskinetic time estimation (ICC = 0.34 [95%CI: 0.04–0.59]; Table 3).

We then analyzed the importance of the timing of PKG® recordings with respect to the clinical rating time for the validity of the temporal agreement and PDTs of the adPMD ratings. Using exactly timed PKG® recordings for diary classification with time frames of the 28 min around the clinical ratings (mBKS/DKS$_{timed\ (28)}$) or the 16 min directly prior clinical rating (mBKS$_{timed\ (16)}$) did not change test performances for the temporal agreement of both Off state and Dyskinetic state detection (Fig. 6a, b; Supplementry Fig. 7). For data on timing for the detection of participant-rated motor state, please refer to Supplementary Fig. 8.

DISCUSSION

The primary finding of this prospective observational multi-center cohort study in 63 participants with advanced PD is that the validity of wrist-wearable accelerometer-based digital Parkinson's Motor Diary (adPMD) real-world monitoring of PDTs is moderate for Off time and Dyskinetic time when compared to simultaneous clinical rating (ICC = 0.43–0.51), while temporal agreement at a given time point on the half-hour time level is poor (median Cohen's κ = 0.13–0.21). The accuracy between adPMD and observer were higher than those between adPMD and participant ratings when analyzing all motor states. Synchronizing adPMD data collection to the time of diary entries and using an index day to calibrate the adPMD threshold to patient diary entries did not increase the concordance between patient diary entries and adPMD monitoring. Calibration of adPMD data to clinician ratings improved their concordance of Dyskinetic state detection to the moderate level (Cohen's κ = 0.41).

DHTs have been particularly developed because the features of PD that respond to dopaminergic stimulation fluctuate due to the shorter benefit from levodopa dosing in advanced disease[37]. Consequently, a single assessment fails to capture the presence, timing or severity of these fluctuations and so continuous assessment is required. While patient diaries were developed to address this, there have been misgivings about their accuracy[19,23,24]. For example, in the recent studies, subjects who collected diaries while under video or face-to-face clinical observation failed to report dyskinesia observed by the clinician in 34–64% of instances[19,23,24]. However, diaries, which classify motor function into three motor states[15,22,27,31,39–41] have influenced the way these motor states are considered: The patient initially determines the level of bradykinesia that marks the transition from On to Off state. However, there is likely to be variation during the course of the day as to where that point of transition is perceived as well as variation between subjects. To compare the diary to an objective measurement, including UPDRS III, 7m-TUGT or DHTs, requires a point on that objective measurement to be used as the transition point. Although the observers and participants were intensively trained in diary rating, the application of the criteria for each motor state has inherent
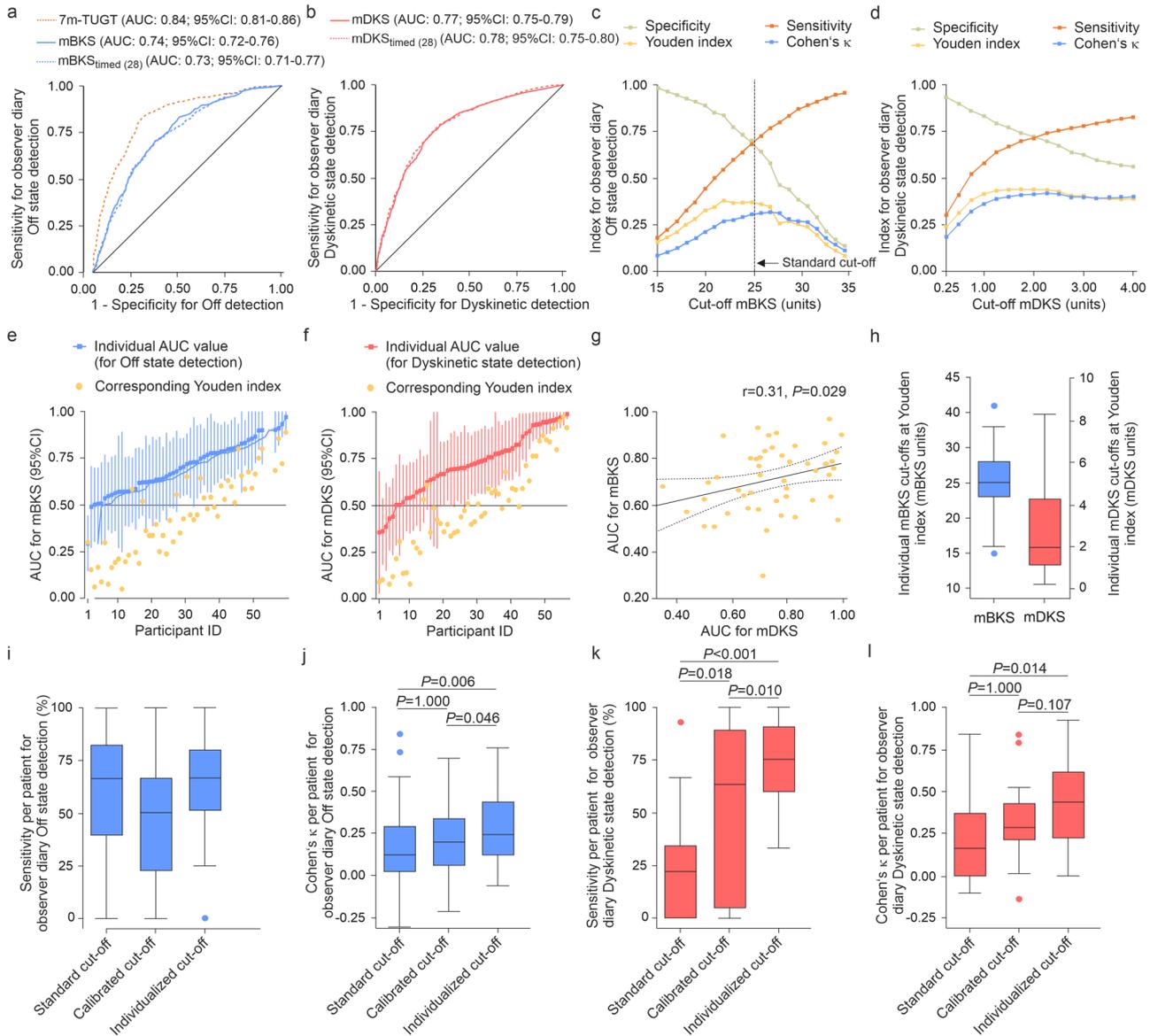
Fig. 6 Test performance of PKG® quatitative scores for the detection of observer-documented Off and Dyskinetic state. a–c Test performance of mBKS transferred into motor diary Off state. **a** Receiver operating curve (ROC) analyses displaying the sensitivity and specificity of the 7m-TUGT results, mBKS and exactly timed mBKS recordings for diary classification with time frames of the 28 min (seven 2-min epochs before and 7 epochs after clinical rating) around the clinical ratings (mBKS$_{timed (28)}$) for the detection of observer-documented Off state and (**b**) the individual AUC values and Youden values of mBKS for the detection the detection of observer-documented Off state. **c** Test performance indices (sensitivity, specificity, Youden index, Cohen's κ for the detection of observer-documented Off state by mBKS cut-offs. Note the standard cut-off value of 25 units used for clinical routine use. **d–f** Test performance of mDKS tranferred into motor diary Dyskinetic state. **d** ROC analyses displaying the sensitivity and specificity of mDKS and exactly timed mDKS recordings for diary classification with time frames of the 28 min (seven 2-minutes epoches before and 7 epoches after clinical rating) around the clinical ratings (mDKS$_{timed (28)}$) for the detection of observer-documented Dyskinetic state and (**b**) the individual AUC values and Youden values of mDKS for the detection the detection of observer-documented Dyskinetic state. **c** Test performance indices (sensitivity, specificity, Youden index, Cohen's κ for the detection of observer-documented Dyskinetic state by mDKS cut-offs. **g** Correlation of individual AUC values of mBKS and mDKS for the detection of the respective motor state. Numbers in right corner of scatter plot represent Pearson correlation coefficient and P value. **h** Distribution of individual mBKS and mDKS cut-offs values for detection of the respective motor state as rated by the clinical observer estimated by ROC analyses normalized to the 50th percentile of normal subjects (mBKS: 25 units; mDKS: 4 units). **i–l** Temporal agreement between wrist-wearable accelerometer-based digital Parkinson's Motor Diary (adPMD) ratings ratings and observer-documented motor states on the participant level using the various cut-offs. Sensitivites expressed in precent (**i, k**) and Cohen's κ values (**j, l**) of adPMD ratings for the detection of observer-documented Off state and Dyskinetic state on the participant level. Displayed P-values are from from Friedmann tests with post-hoc Wilcoxon signed-rank tests with Bonferroni correction for multiple comparisons. Boxplots (**h–l**) 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, and outliers scoring more than 1.5 × IQR but at most 3 × IQR outside the quartiles. Data are based on 2389 simultaneous half-hourly performed observer diary ratings, respectively with simultaneous PKG® recordings from 63 participants from 2 consecutive days (8 a.m. to 6 p.m.). 7m-TUGT and timed PKG® recordings were only performed in the German cohort ($n = 40$, 1361 half-hourly assessments).

inconsistencies that introduce variability into an individual rater's scores and between raters. Furthermore, there is no objective threshold to aid the rater in determining when a switch between motor state categories has occured. Clinical measures of bradykinesia/dyskinesia, such as UPDRS[16] or mAIMS[17], which closely correlate with PKG® scores[29,38–40], are more quantitative and would permit a threshold. However, we chose not to use them because they would consume to much of each 30-min interval and may potentially interfere with PKG® recordings.

We used the PKG® device[29–34] in an alternative application format by converting its continuous data into one of the three motor categories of the PD home diary on a half-hour basis[15,22,31,41] (resulting in adPMD ratings) and analyze its validity using half-hourly clinical ratings as the main validation criterion. We not only analyzed the PDTs as most previous studies[27,39,40], but we were also interested in the temporal agreement of DHT data and clinical ratings at a given time point and during motor Off episodes to detect unpredictable day-to-day differences[49]. Our study settings were as uncontrolled as possible when testing the motor function each half-hour during the waking day, but the frequent ratings and manual diary entries might have nevertheless interfered with motor function. We demonstrated moderate validity for Off time and Dyskinetic time, but only limited correlation/validity for the other measures of motor fluctuations. In particular, temporal agreements between adPMD assessments and simultaneous observer ratings are limited, which can lead to unpredictable Off or Dyskinetic episodes being overlooked in clinical routine.

While our main outside validation criterion assessed by clinical raters overcome the numerous problems with patient diaries including individual selecting their own (possibly varying) threshold, interference with non-motor aspects, diary fatigue and others (see detailed discussion in Ossig and colleagues[31]), clinical examinations are also associated with limitations and inter-rater agreement data are lacking. Even though the clinical raters took global bradykinesia, tremor, dyskinesia and gait function into account, they were not required to score according to specific thresholds or cut-off scores such as to score Off state with reference to a specific UPDRS III score (noting though there was insufficient time for a formal UPDRS III assessment). Similarly, there was no decision about what number of seconds performing a 7m-TUGT constituted Off state. However, we validated observer diary ratings against the results of a simultaneously adjudicated 7m-TUGT[23]: Observer clinical diaries showed better separation of 7m-TUGT results between Off and On states and a lower variability during On without dyskinesia as compared to PD home diary ratings arguing for a higher consistency and reliability of observer ratings. Consequently, we attempted to calibrate the scoring of clinical observers and diaries against the PKG® standard threshold. Our initial analyses thus compared uncalibrated adPMD data with observer and participant diary data and thereby confirmed PKG® standard threshold or target values[29], but with rather weak discriminatory power. This poor overall temporal agreement between adPMD data and observer ratings prompted us to test various strategies to optimize the cut-offs: We repeated all analyses with adPMD data calibrated to the patient's individual thresholds[31] without relevant changes in validity measures. We next tested the effects of individualizing cut-offs according to the clinician ratings as another approach for calibrating adPMD data by using ROC analyses. The first important observation was that only in 36% of participants the detection of Off states and in 56% the detection of Dyskinetic states are successful. However, the use of individual cut-off values in the whole cohort leads to an increase of the validity of adPMD data, which is even further increased by selecting participants with significant ROC results. The wide ranges of cut-off values in combination with the wide ranges of corresponding Youden indices for both Off state and Dyskinetic state detection further demonstrate the importance of

an individualized transfer of DHT data into motor states. Unfortunately, we did not find clinical predictors to identify the individuals with significant motor state detection by the DHT. Since we did not predefine the analytic strategy using individualized cut-offs in our study protocol, the data on thresholding of PKG® data transfer into adPMD motor states needs to be confirmed in a future prospective trial.

To assess whether DHT provides an accurate assessment of the motor states at a particular point in time, the digital monitoring must temporally coincide with the clinical assessment. This forces constraints on digital monitoring that do not easily fit with the design of the PKG®. As the PKG® produces a score every 2 min, notionally that 2-min score could be used to compare with an assessment done contemporaneously. Moreover, best results are obtained from that point in the day being represented by a smoothed moving average (median) over at least 30 min (15 × 2-min epochs) during usual activity which is then averaged again over the same period from 6 days. An important reason for longer periods of monitoring and the use of medians is that it reduces the likelihood of atypical physical activity or inactivity biasing the score. When only 30-minutes time periods are used, the benefit of this averaging is lost using the method adopted in the present study.

The short time period of 30 min applied as the clinical reference period requires exact timing of simultaneous DAT recordings and clinical ratings. We thus examined—in addition to standard 30-min intervals used in the total cohort—two different exactly timed periods for assessing adPMD data excluding the clinical testing time period, each differing in their relation to the time of the clinical rating: One 28 min period exactly around clinical testing, and the second one was a 16-min period directly prior to clinical ratings, which was thought to more likely depict the motor state present when the observer made their clinical ratings. In keeping with the design of the PKG® algorithms, the median of the 2-min epochs were used to produce mBKS and mDKS scores. The correlation between the adPMD data from the two periods and standard recordings for both Off state and Dyskinetic state was similar.

Our validation data are in good agreement with previous reports using passive uncontrolled assessment of motor function by DHTs in comparison to clinical rating scales, such as the MDS-UPDRS (present data and[27,50]). However, these scales do not necessarily reflect motor states at a given time point, but more closely the PDTs or even an average over several days. The higher correlations or validity measures in many previous studies using various wearables including STAT-ON, PKG® and the Verily smartwatch with partly excellent validity indices[29,35,39,40,51], are most likely related to notable differences in the study designs: They used more controlled settings or even a virtual exam for their assessments, in some cases combined with averaging the DHT signals over several hours or even days to reduce the effects of interference of the DHT signal with voluntary motor activity/inactivity[29,35,39,40]. As a limitation of our study, we did not record (voluntary) mobility of the participants during PKG® recordings to enable the investigation of this interference. Indeed, most current DHT systems use as a key idea a statistical approach of measuring multiple times. In this view, the central tendency becomes the likely state of bradykinesia or dyskinesia of the sample, and the variation determines the confidence of that measure. The systems thus measure under conditions of lower variation (controlled conditions) or increase the sample size by recording for several days and between 09:00 and 18:00 (to reduce times of inactivity). This general approach leads not only to a close correlation of mBKS or other DHT measures of bradykinesia and conventional clinical rating scales such as the UPDRS part III motor score but also to a clear response to levodopa medication[29,38–40]. Unfortunately, we did not assess the time points of medication intake to assist the interpretation of our data. The available data also imply

that even though the PKG® is placed on only one wrist it is able to estimate motor dysfunction in PD when used in its standard format[29–34]. However, it cannot be entirely excluded that a single wrist-worn device might not be able to capture full anatomical distribution of symptoms emerging during transitions between motor states and thus needed to detect the various motor states by the alternative adPMD approach. Interference of the adPMD dyskinesia rating with tremor is unlikely in our cohort, because less than 2% of participants showed mild or moderate tremor of the upper extremities. Moreover, DHTs including the PKG® reliably detect tremor and differentiate tremor from dyskinesia[50,52,53]. Together, our data suggest that the combination of frequent virtual examination (active assessment under controlled conditions) and passive monitoring of motor function might potentially help to increase validity of DHTs in detecting motor fluctuations in remote real-world scenarios.

In conclusion, the adPMD using the PKG® device captures daily times in Off and Dyskinetic state assessed using the PD home diary in fluctuating PD with moderate validities as determined by ICC analyses. Limited validity of adPMD data with respect to temporal agreement of motor state ratings at a given time point limits its use for detecting unpredictable Off episodes. The loss of benefit by not smoothing/averaging the DHT signal seems an important factor for this limitation. Our data suggest that the combination of frequent virtual examination (active assessment under controlled conditions) and passive monitoring of motor function might potentially help to increase validity of DHTs in detecting motor fluctuations in remote real-world scenarios. Future studies in larger patient cohorts are warranted to confirm our data on thresholding for PKG® data transfer into motor state diary data and to evidence the adPMD using the PKG® device as a suitable trial endpoint.

PARTICIPANTS AND METHODS
Study protocol approvals and participant consents
This study reports the prospective, observational VALIDATE-PD cohort study[23,24] on alternative application of the wrist-wearable accelerometer Parkinon KinetiGraph® (PKG®; Global Kinetics, Melbourne, Australia) for objective motor diary assessment (wrist-wearable accelerometer-based digital Parkinson's Motor Diary, adPMD). The study was conducted at three hospital centers in Germany and Sweden (University Medicine Rostock, Germany; Movement Disorder Clinic Beelitz-Heilstätten, Germany; Neurology Research Unit, Skåne University Hospital, Lund, Sweden) as described previously[23,24]. The study was approved by the institutional review boards of all participating centers (ethic committee registry numbers A 2017-0115 for Rostock, AS 84(bB)/2018 for Beelitz-Heilstätten and the Regional Ethics Review Board, Lund, Sweden (2017/936). All participants gave written informed consent to participate in the study and were advised both orally and in writing of the nature of the study.

Participants
Participants were eligible for the study if they were over 30 years old, had been diagnosed with PD according to the United Kingdom PD Society Brain Bank criteria, suffered from motor fluctuations observed by the treating physician and/or documented on part 4 of the Movement Disorder Society-revised Unified Parkinson's Disease Rating Scale (MDS-UPDRS) and were able to provide written informed consent.

Exclusion criteria comprised the existence of clinical signs for secondary or atypical parkinsonian syndromes, inability to complete questionnaires and/or patient diaries, lack of cooperation during the study procedures, presence of dementia (defined as scores on the Montreal Cognitive assessment [MoCA] < 21)[54] and/or relevant psychotic symptoms, ongoing treatment with advanced/invasive therapies (deep brain stimulation, subcutaneous apomorphine and levodopa-carbidopa intestinal gel).

We screened 96 participants for eligibility of whom 91 were successfully included into the present study. Four participants were excluded because of MoCA scores below 21 points at screening ($n = 2$) or were not able to sufficiently adhere to diary assessments ($n = 2$). One participant declined further participation due to undisclosed reasons after the screening visit. Of these 91 participants, 63 participants (69%) had 2 days of calibrated adPMD datasets and were included in the final study analyses (for 5 participants, no recordings at day 0 were availble, for 14 participants only 1 day of PKG® recordings were available and for 9 participants less then 30 half-hours (<71%) of observer readings or PKG® recording periods were available). According to MDS-UPDRS part IV, participants reported a median of 18% (IQR: 7–28%) time in Off state, 65% (50–82%) in On state and 13% (6–29%) in Dyskinetic state. Although tremor was not an exclusion criterion, only 1.6% of participants showed mild or moderate resting/postural/action tremor of the upper extremities (MDS-UPDRS tremor items 3.15–3.17).

Clinical assessments
We assessed basic demographic and clinical data including PD medication, clinical phenotype, type of motor complication, Hoehn-Yahr score[55], MDS-UPDRS[56], cognitive screening with MoCA[54] and Beck's Depression Inventory version 2 (BDI-II)[57]. After inclusion into the study, all participants received detailed instructions on the PD home diary[15,22,41]. Participants of the German centers additionally watched a training video explaining all functional states with particular focus on the difference between tremor and dyskinesia[41]. Participants were then asked to indicate their predominant status during half-hour time periods for 3 days using the categories Asleep, Off (herein called Off state), On without dyskinesia (On state), On with non-troublesome dyskinesia, and On with troublesome dyskinesia using the PD home diary)[15]. While diary data for both participants and the observer was initially collected using these categories[15], we eventually combined the categories "On with non-troublesome dyskinesia" and "On with troublesome dyskinesia" into the category "On with dyskinesia" for analysis (Dyskinetic state), since the distinction between non-troublesome and troublesome dyskinesia could only be made by participants, but neither by observers nor the adPMD. In the present study, we only used PD home diary data from the initial day (day 0) for adPMD calibration (see below). On the following 2 days (days 1 and 2), participants were observed by experienced PD raters (A.B., J.T. and acknowledged personnel), who had been trained to identify motor complications in advanced PD patients in the participating hospitals and acquired MDS certification as qualified UPDRS raters prior to study start. The observers/clinical raters independently evaluated motor states half-hourly throughout daytime (8 a.m. through max. 6 pm) based on clinical observations taking global bradykinesia, tremor, dyskinesia and gait function into account. In the German cohort, 7-meter version of the Timed-Up-and-Go test (7m-TUGT) times[58] were recorded but not considered for observer ratings. The exact time points of the clinical rating (synchronized time with the PKG® device) was documented in the German cohort. The 7m-TUGT was canceled at 1 min for patient safety and to not interfere too much with the real-world monitoring environment (total clinical rating and diary assessment time not exceeding 1:30 to 2:00 min). The observer diary ratings were used as the outside criterion for all validity analyses and PD home diary data were used for ancillary analyses (see Table 2 for glossary of clinical ratings).

adPMD assessments

We used the PKG® system as a wrist-worn accelerometer-based device, which was worn at the wrist from day 0 until day 2 on the side more severely affected by PD. The primary outcome measures of the PKG® are the continuous quantitative bradykinesia scores (BKS) and dyskinesia scores (DKS) for 2-minute time periods arising from the analyzing technique of the spectral power of the low frequencies of accelerometer data as developed by Griffith et al.[29]. For the alternative application of the adPMD, we calculated the 30 min median BKS with times of inactivity/non-wrist wear excluded (mBKS) and the 30 min median DKS with times of activity excluded (mDKS; see Table 2 for definitions of PKG® and adPMD measures). The algorithms used to transfer these quantitative measures into motor states were already introduced by ref. [31]. These algorithms convert each half-hour of continuous quantitative mBKS/mDKS into one of the five categories (asleep, motor Off state, On state or On state with dyskinesia [Dyskinetic state], not wearing PKG® device) as used in the observer diary and the PD home diary[23,31]. In brief, the 75th percentiles of BKS and DKS of normal subjects were used from data of the study by Griffith and colleagues[29] and the probability that more BKS or DKS was greater than the 75th percentile of controls in each half-hour period was estimated using the χ^2 test. There are fifteen 2 min epochs in a half-hour and in normal subjects, one in four will be over the 75th percentile of controls. Thus in patients, if 15 or more BKS or DKS are greater than 75th percentile of controls in the hour then the χ^2 test provides a P value < 0.05. These data are herein referred to as uncalibrated adPMD motor state data. As well as correlating adPMD data with the consistent threshold, we additionally calibrated adPMD data according to Ossig and colleagues by modeling the thresholds used by the individual participant[31]. To achieve this, uncalibrated adPMD data were tuned to model day 0 of diary entries by each patient (referred herein as adPMD [calibrated] data) for establishing individual thresholds for Off and Dyskinetic state. Calibration from that first day (day 0) was then applied to the subsequent 2 days. adPMD analyzers were blinded to clinical data and observer/patient diary entries (H.K., N.S., M.H.).

For further analyses, we used exactly timed PKG® recordings from the German cohort for diary classification using the adPMD method with time frames from the 28 min around the clinical ratings (seven 2-min epochs before and seven epochs after clinical rating, separated by one 2-min epoch during which the clinical testing took place; mBKS/DKS$_{timed\ (28)}$) or the 16 min directly prior to clinical ratings (eight 2-min epochs before rating; mBKS$_{timed\ (16)}$).

Adherence to adPMD and clinical diary assessments

All VALIDATE-PD study participants who had 2 days of calibrated adPMD datasets with at least 30 half-hour time periods (>70% of all half-hour periods between 8:00 and 18:00) with simultaneous adPMD and observer diary data were included in final study analyses leading to analyzable datasets from a total of 63 participants (German cohort: n = 40).

In total, we detected 2435 half-hour periods during the waking day with observer diary entries (92.0% of all half-hour periods between 8 a.m. and 6 p.m.). 32 time periods (1.2% of all time periods) rated as Asleep by the observers were excluded from further analyses. From participant diaries, we analyzed 2443 (92.3%) half-hour periods with 40 (1.6%) half-hour periods being excluded due to the rating Asleep. In the German cohort, we generated 1613 half-hour periods (95.4% of all periods) of timed clinical and adPMD assessments and further performed 1566 half-hourly 7m-TUGTs (92.6% of all half-hour periods) used as a clinical measure of bradykinesia[59]. We further detected 2444 (92.4%) half-hour periods with adPMD scores, and 2384 (90.1% of all periods or 97.5% of all PKG® scorings) half-hour adPMD scores were classifiable into motor states with 144 periods (6.0%) classified

as Asleep, 974 (40.9%) as motor Off state, 780 (32.7%) as On state and 486 (20.4%) as Dyskinetic state. Calibration was possible in 1944 half-hour periods (73.5% of all periods or 79.5% of all adPMD scorings) with 115 periods (5.9%) classified as Asleep, 566 (29.1%) as motor Off state, 600 (30.9%) as On state, and 663 (34.1%) as Dyskinetic state.

These numbers translated into 2339 half-hour periods (88.4%) with complete simultaneous ratings of motor states in observer diaries and uncalibrated adPMD data, and 1877 (70.9%) half-hour period with simultaneous observer-ratings and calibrated adPMD data. For participant diaries, we analyzed 2338 half-hour periods (88.4%) with complete simultaneous ratings of motor states, and 1913 (72.3%) half-hour period with simultaneous participant-ratings and calibrated adPMD data. Simultaneous timed assessments and 7m-TUGT were available for 1361 (81.0%) half-hour periods.

For analyses on the participant level, we used PDTs of the three different motor states (time in motor state per total time from 8 a.m. to 6 p.m.) of 2 consecutive days from 63 participants with sufficient datasets from two days with a median of 42 (IQR: 34–42) half-hour datasets or 100% (81–100%) of total datasets for clinical ratings and 40 (IQR: 34–42) half-hour datasets or 95% (81–100%) of total datasets for adPMD recordings. Separate analyses of single day values revealed similar results (data not shown).

Data analyses and statistics

Statistical analyses were performed using IBM SPSS Statistics software version 27 and Excel (IBM Corporation, New York, USA). Values are provided as numbers (percentages) or median (interquartile range, IQR), as appropriate. 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, and outliers scoring more than $1.5 \times IQR$ but at most $3 \times IQR$ outside the quartiles. Pairwise exclusion was used for missing values. As most of the data were not normally distributed, we chose non-parametric tests for statistical comparisons such as Kruskal-Wallis tests with Dunn-Bonferroni post-hoc tests, corrected for multiple comparisons, or Friedman tests with post-hoc Wilcoxon signed-rank tests with Bonferroni correction for multiple comparisons. P values < 0.05 (two-tailed) were considered to be statistically significant.

Test performance of adPMD data for the detection of diary motor states

Statistical analysis of the three motor states (motor Off, On or On with dyskinesia) from clinical observer or PD home the diary data was then confined to all half-hour periods between 8 a.m. and 6 p.m., for which simultaneous ratings from adPMD and observers had been recorded. Assessments on the half-hour time level used the following preformance measures from the 2×2 contingency tables: Sensitivity (or recall) expressed in percent, specificity in percent, accuracy in percent[60], balanced accuracy as the average of sensitivity and specificity in percent[61], precision in percent, Cohen's κ[62], Matthews correlation coefficient (MCC or Φ coefficient)[60,62,63] and F1-score as the harmonic mean of sensitivity and precision[60]. Imbalance of outcome classes was estimated as the imbalance ratio (IR) defined as the number of majority class samples (negative ratings) divided by the number of minority class samples (positive ratings) and values of 5 or higher were considered as relevant (moderate to severe) imbalance[64]. Pearson's correlation test and intraclass correlation coefficient (ICC) estimation were used for correlations of daily times spent in the various motor states on the participant level. Pearson's correlation coefficient $|r| < 0.3$ was considered a weak, $|r| = 0.3–0.59$ a moderate and $|r| \geq 0.6$ a strong 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[65], we interpret κ values or ICC < 0.40 as poor, κ/ICC = 0.40–0.59 as moderate, κ/ICC = 0.60–0.74 as good and κ/ICC = 0.75–1.00 as excellent validity.

Data balancing

Although diary outcome data were not or only mildly imbalanced with IR values between 1.56 and 2.94 for clinical observer diary data and 1.01 and 3.45 for PD home diary data[64], we addressed the challenge of class imbalance by balancing the data using random undersampling the majority class, random oversampling the minority class and combined under-/oversampling of clinical observer and PD home diary data[64,66]. Analyses of balanced datasets revealed no relevant influence of data imbalance in the original datasets with stable values for major performance measures such as sensitivity, specificity, balanced accuracy, Cohen's κ and MCC (for details, refer to Supplementary Results).

Association of diary data with demographic and clinical factors

Pearson's correlation test and multiple linear regression modeling was used to test the association of demographic and clinical candidate factors with adPMD quantitative scores and motor state results as well as with the agreement rates and Cohen' κ values for adPMD motor state results with observer-rated diary entries. We included the variables age, sex, symptom duration, MDS-UPDRS part III motor score as a measure of disease severity, motor fluctuation duration, MDS-UPDRS part IV as a quantitative measure of motor fluctuations, BDI and MoCA as independent covariates into the models using a step-wise selection with $P < 0.05$ for adding and $P < 0.10$ for removing variables. Multicollinearity of candidate variables were excluded by Pearson correlation test ($|r| < 0.5$). Results were confirmed by hierarchical multiple linear regression models.

Temporal connection between Off episodes

For analyzing the temporal connection between Off episodes as rated by the clinical observers, Off episodes (defined herein as an motor Off state of at least 30 min duration following a motor On period of at least 90 min) from all participants were synchronized by summation of all Off state periods as rated by observers using the first 30 min of the motor Off period as the trigger event (start of Off episode). The Off state ratings from the adPMD were then cross-classified by putting them into 2×2 contingency tables for each 30-min motor Off state interval. All diary sets were included for analysis. Individual time periods were excluded from analysis if there was no response or more than one response on either diary.

Association of quantitative PKG® scores with diary data

For analyses of quantitative PKG® scores[29], we initially compared mBKS and mDKS with the corresponding diary data and 7m-TUGT results (used as a quantitative measure of bradykinesia[58]) on the half-hour time period level. Consistently, we detected significantly higher mBKS and higher 7m-TUGT times during Off state periods as compared to On state and Dyskinetic state periods as assessed by the clinical observers, while mDKS were lower during Off state periods compared to On state and Dyskinetic state periods as assessed by clinical observer diaries ($P < 0.001$, Kruskal-Wallis tests). Further analyses of quantitative PKG® scores are reported in Supplementary Results.

Receiver operating curve (ROC) and precision-recall curve (PRC) analyses of PKG® scores

ROC analyses were used to display the correlation between specificity and sensitivity, and PRC plotting sensitivity/recall versus precision to display the fraction of true positives among positive detections of the quantitative PKG® scores for detection of motor states (mBKS for Off states and mDKS for Dyskinetic states). The area under the curve (AUC) with 95% confidence intervals (95%CI) was used to compare the various models (AUCs with 95%CIs of PRC plots were calculated using the trapezoid rule with the assumption of normal data distribution with the SPSS data output). In ROC analyses, a relevant discrimination of motor states by PKG® scores was defined as an ROC-AUC significantly larger than 0.5. The maximal Youden index (J = Sensitivity + Specificity – 1) was used to estimate optimal cut-off values. PRC plots with their AUCs using as an overall performance measure express the susceptibility of the classifiers to imbalance of the data and was thus applied as an accurate comparison of the classifier performance between datasets with difference imbalance ratios[67]. The results of PRC plot analyses are displayed in Supplementary Results.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

REFERENCES

1. Bjornestad, A. et al. Risk and course of motor complications in a population-based incident Parkinson's disease cohort. *Parkinsonism Relat. Disord.* **22**, 48–53 (2016).
2. Garcia-Ruiz, P. J., Del Val, J., Fernandez, I. M. & Herranz, A. What factors influence motor complications in Parkinson disease? A 10-year prospective study. *Clin. Neuropharmacol.* **35**, 1–5 (2012).
3. Ahlskog, J. E. & Muenter, M. D. Frequency of levodopa-related dyskinesias and motor fluctuations as estimated from the cumulative literature. *Mov. Disord.* **16**, 448–458 (2001).
4. Chou, K. L. et al. The spectrum of "off" in Parkinson's disease: What have we learned over 40 years? *Parkinsonism Relat. Disord.* **51**, 9–16 (2018).
5. Jankovic, J. Motor fluctuations and dyskinesias in Parkinson's disease: clinical manifestations. *Mov. Disord.* **20**, S11–S16 (2005).
6. Rota, S. et al. Why do 'OFF' periods still occur during continuous drug delivery in Parkinson's disease? *Transl. Neurodegener.* **11**, 43 (2022).
7. Pechevis, M. et al. Effects of dyskinesias in Parkinson's disease on quality of life and health-related costs: a prospective European study. *Eur. J. Neurol.* **12**, 956–963 (2005).
8. Haahr, A., Kirkevold, M., Hall, E. O. & Ostergaard, K. Living with advanced Parkinson's disease: a constant struggle with unpredictability. *J. Adv. Nurs.* **67**, 408–417 (2011).
9. Rahman, S., Griffin, H. J., Quinn, N. P. & Jahanshahi, M. Quality of life in Parkinson's disease: the relative importance of the symptoms. *Mov. Disord.* **23**, 1428–1434 (2008).
10. Ray Chaudhuri, K., Poewe, W. & Brooks, D. Motor and nonmotor complications of levodopa: phenomenology, risk factors, and imaging features. *Mov. Disord.* **33**, 909–919 (2018).
11. Boyle, A. & Ondo, W. Role of apomorphine in the treatment of Parkinson's disease. *CNS Drugs* **29**, 83–89 (2015).
12. Storch, A. et al. Nonmotor fluctuations in Parkinson disease: severity and correlation with motor complications. *Neurology* **80**, 800–809 (2013).

13. Martinez-Fernandez, R., Schmitt, E., Martinez-Martin, P. & Krack, P. The hidden sister of motor fluctuations in Parkinson's disease: a review on nonmotor fluctuations. *Mov. Disord.* **31**, 1080–1094 (2016).

14. Dodel, R. C., Berger, K. & Oertel, W. H. Health-related quality of life and healthcare utilisation in patients with Parkinson's disease: impact of motor fluctuations and dyskinesias. *Pharmacoeconomics* **19**, 1013–1038 (2001).

15. Hauser, R. A. et al. A home diary to assess functional status in patients with Parkinson's disease with motor fluctuations and dyskinesia. *Clin. Neuropharmacol.* **23**, 75–81 (2000).

16. Fahn, S., Elton, R. L. & Committee, U. D. *Recent Developments in Parkinson's Disease*. Vol. 2 (eds S. Fahn, C. D. Marsden, & D. B. Calne) 153–163, 293–304 (Macmillan Healthcare Information, 1987).

17. May, P. R., Lee, M. A. & Bacon, R. C. Quantitative assessment of neuroleptic-induced extrapyramidal symptoms: clinical and nonclinical approaches. *Clin. Neuropharmacol.* **6**, S35–S51 (1983).

18. Papapetropoulos, S. S. Patient diaries as a clinical endpoint in Parkinson's disease clinical trials. *CNS Neurosci. Ther.* **18**, 380–387 (2012).

19. Erb, M. K. et al. mHealth and wearable technology should replace motor diaries to track motor fluctuations in Parkinson's disease. *NPJ Digit. Med.* **3**, 6 (2020).

20. Stacy, M. et al. Identification of motor and nonmotor wearing-off in Parkinson's disease: comparison of a patient questionnaire versus a clinician assessment. *Mov. Disord.* **20**, 726–733 (2005).

21. Goetz, C. G. et al. Evaluating Parkinson's disease patients at home: utility of self-videotaping for objective motor, dyskinesia, and ON-OFF assessments. *Mov. Disord.* **23**, 1479–1482 (2008).

22. Hauser, R. A., Deckers, F. & Lehert, P. Parkinson's disease home diary: further validation and implications for clinical trials. *Mov. Disord.* **19**, 1409–1413 (2004).

23. Lohle, M. et al. Validation of the PD home diary for assessment of motor fluctuations in advanced Parkinson's disease. *NPJ Parkinsons Dis.* **8**, 69 (2022).

24. Timpka, J. et al. Objective observer vs. patient motor state assessments using the PD home diary in advanced Parkinson's disease. *Front. Neurol.* **13**, 935664 (2022).

25. Dietrichs, E. & Odin, P. Algorithms for the treatment of motor problems in Parkinson's disease. *Acta Neurol. Scand.* **136**, 378–385 (2017).

26. Jimenez-Shahed, J. A review of current and novel levodopa formulations for the treatment of Parkinson's disease. *Ther. Deliv.* **7**, 179–191 (2016).

27. Perez-Lopez, C. 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. *Front. Neurol.* **13**, 835249 (2022).

28. Rodriguez-Martin, D. et al. A new paradigm in Parkinson's disease evaluation with wearable medical devices: a review of STAT-ON(TM). *Front. Neurol.* **13**, 912343 (2022).

29. Griffiths, R. I. et al. Automated assessment of bradykinesia and dyskinesia in Parkinson's disease. *J. Parkinsons Dis.* **2**, 47–55 (2012).

30. Horne, M. K., McGregor, S. & Bergquist, F. An objective fluctuation score for Parkinson's disease. *PLoS One* **10**, e0124522 (2015).

31. Ossig, C. et al. Correlation of quantitative motor state assessment using a kinetograph and patient diaries in advanced PD: data from an observational study. *PLoS One* **11**, e0161559 (2016).

32. Farzanehfar, P. et al. Objective measurement in routine care of people with Parkinson's disease improves outcomes. *NPJ Parkinsons Dis.* **4**, 10 (2018).

33. Farzanehfar, P., Woodrow, H. & Horne, M. Assessment of Wearing Off in Parkinson's disease using objective measurement. *J. Neurol.* **268**, 914–922 (2021).

34. Woodrow, H., Horne, M. K., Fernando, C. V. & Kotschet, K. E., Treat to Target Study, G. A blinded, controlled trial of objective measurement in Parkinson's disease. *NPJ Parkinsons Dis.* **6**, 35 (2020).

35. Burq, M. et al. Virtual exam for Parkinson's disease enables frequent and reliable remote measurements of motor function. *NPJ Digit. Med.* **5**, 65 (2022).

36. Espay, A. J. et al. A roadmap for implementation of patient-centered digital outcome measures in Parkinson's disease obtained using mobile health technologies. *Mov. Disord.* **34**, 657–663 (2019).

37. Bloem, B. R., Post, E. & Hall, D. A. An apple a day to keep the Parkinson's disease doctor away? *Ann. Neurol.*, https://doi.org/10.1002/ana.26612 (2023).

38. Isaacson, S. H. et al. Effect of using a wearable device on clinical decision-making and motor symptoms in patients with Parkinson's disease starting transdermal rotigotine patch: a pilot study. *Parkinsonism Relat. Disord.* **64**, 132–137 (2019).

39. Perez-Lopez, C. et al. Assessing motor fluctuations in parkinson's disease patients based on a single inertial sensor. *Sensors* **16**, https://doi.org/10.3390/s16122132 (2016).

40. Rodriguez-Molinero, A. et al. A kinematic sensor and algorithm to detect motor fluctuations in parkinson disease: validation study under real conditions of use. *JMIR Rehabil. Assist. Technol.* **5**, e8 (2018).

41. Hauser, R. A. et al. Patient evaluation of a home diary to assess duration and severity of dyskinesia in Parkinson disease. *Clin. Neuropharmacol.* **29**, 322–330 (2006).

42. Ossig, C. et al. Timing and kinetics of nonmotor fluctuations in advanced Parkinson's disease. *J. Parkinsons Dis.* **7**, 325–330 (2017).

43. Ferreira, J. J. et al. Opicapone as an adjunct to levodopa in patients with Parkinson's disease and end-of-dose motor fluctuations: a randomised, double-blind, controlled trial. *Lancet Neurol.* **15**, 154–165 (2016).

44. Parkinson Study Group. A randomized placebo-controlled trial of rasagiline in levodopa-treated patients with Parkinson disease and motor fluctuations: the PRESTO study. *Arch. Neurol.* **62**, 241–248 (2005).

45. Rascol, O. et al. Rasagiline as an adjunct to levodopa in patients with Parkinson's disease and motor fluctuations (LARGO, Lasting effect in Adjunct therapy with Rasagiline Given Once daily, study): a randomised, double-blind, parallel-group trial. *Lancet* **365**, 947–954 (2005).

46. Schapira, A. H. et al. Assessment of safety and efficacy of safinamide as a levodopa adjunct in patients with parkinson disease and motor fluctuations: a randomized clinical trial. *JAMA Neurol.* **74**, 216–224 (2017).

47. Katzenschlager, R. et al. Apomorphine subcutaneous infusion in patients with Parkinson's disease with persistent motor fluctuations (TOLEDO): a multicentre, double-blind, randomised, placebo-controlled trial. *Lancet Neurol.* **17**, 749–759 (2018).

48. Olanow, C. W. et al. Continuous intrajejunal infusion of levodopa-carbidopa intestinal gel for patients with advanced Parkinson's disease: a randomised, controlled, double-blind, double-dummy study. *Lancet Neurol.* **13**, 141–149 (2014).

49. Olanow, C. W., Poewe, W., Rascol, O. & Stocchi, F. On-demand therapy for OFF episodes in Parkinson's disease. *Mov. Disord.* **36**, 2244–2253 (2021).

50. Lipsmeier, F. et al. Reliability and validity of the Roche PD Mobile Application for remote monitoring of early Parkinson's disease. *Sci. Rep.* **12**, 12081 (2022).

51. Pulliam, C. L. et al. Continuous assessment of levodopa response in Parkinson's disease using wearable motion sensors. *IEEE Trans. Biomed. Eng.* **65**, 159–164 (2018).

52. Braybrook, M. et al. An ambulatory tremor score for Parkinson's disease. *J. Parkinsons Dis.* **6**, 723–731 (2016).

53. Powers, R. et al. Smartwatch inertial sensors continuously monitor real-world motor fluctuations in Parkinson's disease. *Sci. Transl. Med.* **13**, https://doi.org/10.1126/scitranslmed.abd7865 (2021).

54. Dalrymple-Alford, J. C. et al. The MoCA: well-suited screen for cognitive impairment in Parkinson disease. *Neurology* **75**, 1717–1725 (2010).

55. Hoehn, M. M. & Yahr, M. D. Parkinsonism: onset, progression and mortality. *Neurology* **17**, 427–442 (1967).

56. Goetz, C. G. et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov. Disord.* **23**, 2129–2170 (2008).

57. Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. & Erbaugh, J. An inventory for measuring depression. *Arch. Gen. Psychiatry* **4**, 561–571 (1961).

58. Podsiadlo, D. & Richardson, S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *J. Am. Geriatr. Soc.* **39**, 142–148 (1991).

59. Morris, S., Morris, M. E. & Iansek, R. Reliability of measurements obtained with the Timed "Up & Go" test in people with Parkinson disease. *Phys. Ther.* **81**, 810–818 (2001).

60. Chicco, D. & Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genom.* **21**, 6 (2020).

61. Wei, Q. & Dunbrack, R. L. Jr The role of balanced training and testing data sets for binary classifiers in bioinformatics. *PLoS One* **8**, e67863 (2013).

62. Chicco, D., Warrens, M. J. & Jurman, G. The Matthews Correlation Coefficient (MCC) is more informative than Cohen's Kappa and Brier Score in binary classification assessment. *IEEE Access.* **9**, 78368–78381 (2021).

63. Boughorbel, S., Jarray, F. & El-Anbari, M. Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. *PLoS One* **12**, e0177678 (2017).

64. Kumar, V. et al. Addressing binary classification over class imbalanced clinical datasets using computationally intelligent techniques. *Healthcare* **10**, https://doi.org/10.3390/healthcare10071293 (2022).

65. Cicchetti, D. V. Multiple comparison methods: establishing guidelines for their valid application in neuropsychological research. *J. Clin. Exp. Neuropsychol.* **16**, 155–161 (1994).

66. Lemaitre, A., Nogueira, F. & Aridas, C. K. Imbalanced-learn: a Python toolbox to tackle the curse of imbalanced datasets in machine learning. *J. Machine Learn. Res.* **18**, 1–5 (2017).

67. Saito, T. & Rehmsmeier, M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* **10**, e0118432 (2015).

68. Tomlinson, C. L. et al. Systematic review of levodopa dose equivalency reporting in Parkinson's disease. *Mov. Disord.* **25**, 2649–2653 (2010).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

Conceptualization: M.L., J.T., A.B., G.E., P.O. A.S.,. Funding acquisition: P.O., A.S. Organization of research: M.L., J.T., A.B., A.S., P.O. Execution of research/clinical assessments: M.L., J.T., A.B., F.G., G.E. Data curation: M.L., H.K., M.H., A.S. Blinded assessment of digital data: H.K., M.H. Statistical methodology: M.L., H.K., M.H., G.E., P.O., A.S. Execution of statistical analyses: M.L., P.O., A.S. Writing—original draft: A.S. Visualization: A.S. Manuscript review & editing: M.L., J.T., H.K., A.B., F.G., M.H., G.E., P.O.

FUNDING

COMPETING INTERESTS

M.L. has received honoraria for presentations from Novartis Pharma. Jonathan Timpka has received funding from the Swedish National Government and County Councils through the ALF agreement, the Elsa Schmitz Foundation, and the Swedish Parkinson Foundation. He has received compensation for consultancies from TransPerfect and the Swedish National Board of Health and Welfare, as well as royalties from UNI-MED Verlag. A.B. has nothing to disclose. H.K. was employee of Global Kinetics, Melbourne, Australia. F.G. has received honoraria from AbbVie, BIAL, Merz, and STADA outside the submitted work. M.H. has been a consultant to Global Kinetics (manufacturer of the PKG®), since 2017. He is a founder of Global Kinetics and an inventor of the PKG®. Prior to 2017 he was a member of Global Kinetics and received shares in Global Kinetics as reimbursement. G.E. has received honoraria for advisory Boards, consultancy and presentations from AbbVie Pharma, BIAL, Biogen GmbH, Desitin Pharma, STADA Pharma, Neuroderm Inc., Licher GmbH, UCB Pharma, and Zambon Pharma. He has received royalties from Kohlhammer Verlag and Thieme Verlag. P.O. 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. A.S. has received funding from the Deutsche Forschungsgemeinschaft (German Research Association) and the Helmholtz-Association outside the present study. He has received honoraria for presentations/advisory boards/consultations from Global Kinetics Corporation (manufacturer of the PKG®), Desitin, Lobsor Pharmaceuticals, STADA, Bial, RG Gesellschaft, Zambon, NovoNordisk, and AbbVie outside the present study. He has received royalties from Kohlhammer Verlag and Elsevier Press. He serves as an editorial board member of Stem Cells International.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41746-023-00937-1.

Correspondence and requests for materials should be addressed to Matthias Löhle or Alexander Storch.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.