Digital Mobility Measures: A Window into RealWorld Severity and Progression of Parkinson's Disease

RESEARCH ARTICLE

Digital Mobility Measures: A Window into Real-World Severity and Progression of Parkinson's Disease

Anat Mirelman, PhD,[1,2*] Jana Volkov, MSc,[1] Amit Salomon, MSc,[1] Eran Gazit, MSc,[1] Alice Nieuwboer, PhD,[3] Lynn Rochester, PhD,[4,5] Silvia Del Din, PhD,[4,5] Laura Avanzino, MD, PhD,[6,7] Elisa Pelosin, PhD,[6,8] Bastiaan R. Bloem, MD,[9] Ugo Della Croce, PhD,[10] Andrea Cereatti, PhD,[11] Avner Thaler, MD, PhD,[1,2] Daniel Roggen, PhD,[12] Claudia Mazza, PhD,[12] Julia Shirvan, MD,[12] Jesse M. Cedarbaum, MD,[13,14] Nir Giladi, MD,[1,2] and Jeffrey M. Hausdorff, PhD[1,2,15,16]

[1]*Laboratory for Early Markers of Neurodegeneration (LEMON), Center for the Study of Movement, Cognition and Mobility, Neurological Institute, Tel Aviv Medical Center, Tel Aviv, Israel*
[2]*Faculty of Medicine and Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel*
[3]*Department of Rehabilitation Science, KU Leuven, Neuromotor Rehabilitation Research Group, Leuven, Belgium*
[4]*Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle, United Kingdom*
[5]*National Institute for Health and Care Research (NIHR) Newcastle Biomedical Research Centre (BRC), Newcastle University and The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, United Kingdom*
[6]*Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics and Maternal Child Health (DINOGMI), University of Genoa, Genoa, Italy*
[7]*Department of Experimental Medicine, Section of Human Physiology, University of Genoa, Genoa, Italy*
[8]*IRCCS Policlinico San Martino Teaching Hospital, Genoa, Italy*
[9]*Department of Neurology, Radboud University Medical Center, Donders Institute for Brain, Cognition and Behavior, Nijmegen, The Netherlands*
[10]*Department of Biomedical Sciences, University of Sassari, Sassari, Italy*
[11]*Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy*
[12]*Biogen, Cambridge, Massachusetts, USA*
[13]*Coeruleus Clinical Sciences, Woodbridge, Connecticut, USA*
[14]*Yale University School of Medicine, New Haven, Connecticut, USA*
[15]*Department of Physical Therapy, Tel Aviv University, Tel Aviv, Israel*
[16]*Department of Orthopedic Surgery, Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois, USA*

***Correspondence to:** Dr. Anat Mirelman, Laboratory for Early Markers of Neurodegeneration (LEMON), Center for the Study of Movement, Cognition and Mobility, Neurological Institute, Tel Aviv Medical Center, 6 Weizmann Street, Tel Aviv 64239, Israel; E-mail: anatmi@tlvmc.gov.il

Relevant conflicts of interest/financial disclosures: None of the authors report conflicts of interest pertaining to this manuscript. Financial disclosures are provided: A.M. and J.M.H. serve as consultants for The Michael J. Fox Foundation for Parkinson's Research and F. Hoffmann-La Roche AG (JH) in areas relating to the use of wearable sensors to assess mobility in Parkinson's disease (PD). S.D.D. reports consultancy activity for Hoffmann-La Roche Ltd. relating to the use of wearable sensors to assess mobility in PD, outside of this study. B.R.B. has received research grants from Verily Life Sciences and UCB for the development of smartwatch-based digital measures of PD. J.S., D.R., and C.M. are Biogen employees and might hold shares of the company. J.M.C. serves as acting chief medical officer and holds an equity position in VanquaBio.

Funding agencies: This project was funded by the European Commission (FP7 project V-TIME-278169); The Michael J. Fox Foundation for Parkinson's Research (DeFOG, tDCS); the USA MED Research Acquisition Activity; Department of Defense Awarding Office, Award Number W81XWH2010468 (DOD); Israel Science Foundation (ONPar); and Biogen Inc. (BeaT-PD). This work was also supported by the NIHR/Wellcome Trust Clinical Research Facility (CRF) infrastructure at the Newcastle upon Tyne Hospitals NHS Foundation Trust. A.M., A.N., L.R., S.D.D., E.G., A.C., and J.M.H. were supported, in part, by Mobilise-D. Mobilise-D has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (under grant agreement no.: 820820). S.D.D. and L.R. were also supported by the Innovative Medicines Initiative 2 Joint Undertaking (IMI2 JU) project IDEA-FAST—Grant Agreement 853981. This joint undertaking received support from the European Union's Horizon 2020 Research and Innovation Programme and the EFPIA. S.D.D. and L.R. were also supported by the National Institute for Health Research (NIHR) Newcastle Biomedical Research Centre (BRC) based at the Newcastle upon Tyne Hospital NHS Foundation Trust, Newcastle University and the Cumbria, Northumberland and Tyne and Wear (CNTW) NHS Foundation Trust. S.D.D. and L.R. were supported by the NIHR/Wellcome Trust Clinical Research Facility (CRF) infrastructure at the Newcastle upon Tyne Hospitals NHS Foundation Trust. S.D.D. was supported by the UK Research and Innovation (UKRI) Engineering and Physical Sciences Research Council (EPSRC)-National Institute for Health and Care Research (NIHR) (grant ref.: EP/W031590/1). All opinions are those of the authors and not the funders. All funding agencies were not involved in design, acquisition, interpretation or the findings, or the writing of this manuscript.

Received: 5 September 2023; **Revised:** 20 November 2023; **Accepted:** 27 November 2023

Published online 27 December 2023 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mds.29689

ABSTRACT: Background: Real-world monitoring using wearable sensors has enormous potential for assessing disease severity and symptoms among persons with Parkinson's disease (PD). Many distinct features can be extracted, reflecting multiple mobility domains. However, it is unclear which digital measures are related to PD severity and are sensitive to disease progression.

Objectives: The aim was to identify real-world mobility measures that reflect PD severity and show discriminant ability and sensitivity to disease progression, compared to the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) scale.

Methods: Multicenter real-world continuous (24/7) digital mobility data from 587 persons with PD and 68 matched healthy controls were collected using an accelerometer adhered to the lower back. Machine learning feature selection and regression algorithms evaluated associations of the digital measures using the MDS-UPDRS (I–III). Binary logistic regression assessed discriminatory value using controls, and longitudinal observational data from a subgroup ($n = 33$) evaluated sensitivity to change over time.

Results: Digital measures were only moderately correlated with the MDS-UPDRS (part II-$r = 0.60$ and parts I and III-$r = 0.50$). Most associated measures reflected activity quantity and distribution patterns. A model with 14 digital measures accurately distinguished recently diagnosed persons with PD from healthy controls (81.1%, area under the curve: 0.87); digital measures showed larger effect sizes (Cohen's d: [0.19–0.66]), for change over time than any of the MDS-UPDRS parts (Cohen's d: [0.04–0.12]).

Conclusions: Real-world mobility measures are moderately associated with clinical assessments, suggesting that they capture different aspects of motor capacity and function. Digital mobility measures are sensitive to early-stage disease and to disease progression, to a larger degree than conventional clinical assessments, demonstrating their utility, primarily for clinical trials but ultimately also for clinical care.

Key Words: digital mobility measures; Parkinson's disease; wearable sensors; disease progression

Recent years have seen an increase in the use of digital technology in the health arena. Dedicated wearable devices and more ubiquitous smartwatches and smartphones can generate objective, ecologically relevant, information-rich continuous data streams that can provide insights into patient-relevant disease symptoms, unencumbered by recall bias.[1] In addition, digital measures can assess real-world behavior and provide means to detect intermittent, episodic events and disease-related signs that cannot otherwise be objectively assessed in a clinic visit (eg, sleep efficiency),[2,3] reflecting tremendous potential.

The most widely deployed use of digital technologies is in Parkinson's disease (PD).[4] This is perhaps not surprising, as most digital technologies include accelerometers and gyroscopes, enabling the direct assessment of movement and motor function, which are prominent disease signs in PD. However, the early adoption of technology in the field of PD also relates to the limitations of the current clinical assessment.[5] The traditional clinical endpoint of the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS)[6,7] is widely used and has reliable clinometric characteristics,[8] but its utility is limited by the need for specific training and expertise, infrequent assessments conducted in the clinic, within-rater and across-center variability,[2,9] and intra- and inter-day symptom fluctuations.[10] Moreover, the MDS-UPDRS has low sensitivity to change over time, especially in early PD.[11]

Digital technologies were initially applied to mimic conventional PD assessments, like the MDS-UPDRS, Part III.[12,13] However, recently, regulatory bodies indicated that this approach is not satisfactory, as only digitizing an existing clinical assessment will not add to the understanding of meaningful aspects relevant to patients' ability to function in daily life.[9] Moreover, such an approach overlooks the opportunity presented by digital technology to capture patient-relevant symptoms and behaviors that are not necessarily a mirror image of the MDS-UPDRS. For example, everyday mobility or movement during sleep can be captured in a continuous and passive manner, without the engagement of the patient, potentially reflecting a more holistic assessment.[14] However, despite emerging interest, these technologies have yet to be transitioned from exploratory outcomes to well-accepted, widely used measures for clinical practice and clinical trials.[15,16]

Perhaps the most significant challenge relates to the need to establish validity compared to a "gold standard." Comparison to established performance metrics becomes extremely challenging when using novel digital mobility measures, where gold standards do not exist yet.[1] The International Classification of Functioning Disability and Health model suggests that assessments performed in the clinic, which reflect functional

capacity, and assessments performed during daily activities, which are more indicative of the performance of the individual, differ.[17,18] In this regard, the relationship between the traditional assessment of PD symptoms with the MDS-UPDRS and measures obtained using digital technology during habitual daily living should not be expected to be strong. To date, there is limited evidence on these relationships from real-world studies.[19,20] Importantly, the association between digital measures and patient-related outcomes (eg, parts I and II of the MDS-UPDRS) is unclear. Finally, many distinct measures can be extracted from mobility sensors that reflect multiple mobility domains (eg, physical activity, gait, and nocturnal movement).[2,21,22] However, it is not yet established which features are most related to PD severity, which are sensitive to the presence of the earliest phases of the disease, and how they change over time. Such information is essential to move the field forward toward clinical and regulatory acceptance of digital measures, their implementation as endpoints in clinical trials, and incorporation into day-to-day clinical practice.

In the current work, we addressed these gaps by (1) identifying real-world, digital mobility measures that are most reflective of PD severity as measured using the MDS-UPDRS; (2) evaluating the association between digital mobility measures and the three components of the MDS-UPDRS clinical rating scale; (3) assessing the sensitivity of the digital mobility measures to the presence of early disease, as compared to healthy controls (HC); and (4) evaluating whether digital measures can capture disease progression.

Patients and Methods

Participants

We leveraged data collected using wearable sensors from 587 PD patients and 68 HCs. The data were obtained from six studies at four different clinical sites: V-TIME[23] (n = 132), Beat-PD[2] (n = 213 PD patients and 68 HCs), ONPar (n = 72),[20] tDCS[24] (N = 58), DOD (N = 56),[25] and DeFOG (N = 56).[26] Participants were included if they were diagnosed with PD by a movement disorders specialist, using the MDS clinical diagnostic criteria for PD[27]; were mobile without physical assistance from another person; and were willing to wear a small light-weight device for 1 week. Participants were excluded if they had other neurological, orthopedic, or psychiatric disorders. Data from HCs were included based on age and gender matched to the early PD subgroup (disease duration <2 years) with similar exclusion criteria. All studies were approved by local ethics committees, and participants provided informed written consent before participation.

Study Procedures

Participants underwent a neurological evaluation using the MDS-UPDRS[8] during an in-clinic morning visit. Patients using levodopa (L-dopa) were assessed in the *on* medication state, ~1 hour after medication. Participants were fitted with a small body-fixed, waterproof device (Axivity Ltd., York, UK; either AX3 or AX6 models, size: ~23.0 × 32.5 × 7.6 mm; weight: ~11 g; sampling rate: 100 Hz) secured to their lower back (lumbar vertebrae 4–5) with medical-grade tape for 7 days.[28] Accelerometer data were passively collected continuously (24/7) and saved on the device. Then, the participants returned the device to the clinical site via courier or self-addressed envelopes.

Daily-Living Feature Extraction

Daily-living measures were classified into three domains: nocturnal behavior (NB, n = 24 features), activity quantity and distribution patterns (AQDP, n = 106 features), and gait quality (GQ, n = 301). For NB, the night period was segmented based on lumbar angle estimation.[2] This enabled the calculation of features describing sleep quality, quantity, and movement patterns such as the number of trunk rotations per night, trunk rotation smoothness in different axes, and SleepEfficiency.[2]

Physical AQDPs were quantified based on the signal vector magnitude[20]; a general measure of overall physical activity and intensity, defined as moderate-to-vigorous physical activity and low-intensity physical activity[29]; and daily amount of activity. Mobility was further segmented into types of activities (ie, walking and lying[30]) and their distributions throughout and across the day.[22,29]

GQ measures were extracted from continuous segmented walking periods of ≥30 seconds. Spectral analysis of these bouts produced frequency domain measures, for example, amplitude, dispersion, and smoothness.[31,32] Step velocity; length and duration; and their variance, regularity, and symmetry were extracted.[28,32] Mean, standard deviation (SD), 10th (ie, typically the "worst" value) and 90th (ie, typically the "best" value) percentiles, skewness, and kurtosis were also included. Altogether, 431 measures collected passively in the real world were included in the analysis (Supplementary Material S1).

Preprocessing, Categorization, and Feature Selection

Values were z scored, and missing values were imputed using the K-nearest neighbor method. Outliers were considered according to Tukey's rule. Samples were removed if over 15% of their measures were outliers (Supplementary Material section 1).

To identify which real-world mobility measures are most reflective of PD severity as assessed using the MDS-

UPDRS, we used mutual information,[33] a nonparametric supervised estimation of the relation between the labels (MDS-UPDRS parts) and digital mobility measures. To minimize random effects, this process was repeated in a K = 100-fold validation with a random partition seed for each MDS-UPDRS part resulting in a cumulative score. See also Supplementary Material.

Regression Model Training and Selection

Several machine learning models evaluated the association between digital mobility measures and the MDS-UPDRS. Data were split into a training set (80%) and a test set (20%). Model selection and training were performed using a fivefold cross-validation process with Pearson's R^2 as the performance metric. Hyperparameter tuning was done using Grid Search (Supplementary Material Table S2a). The process of model selection was repeated again using five-fold nested cross-validation to evaluate selection stability and uncertainty. Mean absolute error and Pearson's R^2 assessed correlations between the metrics and MDS-UPDRS parts. The model was blind to the test set during model evaluation and training. Model intrinsic feature importance attributes were also explored (Supplementary Material Table S2c).

Sensitivity to Disease and Disease Progression

Forward selection binary logistic regression models explored accuracy in discriminating between recently diagnosed patients with PD (disease duration ≤2 years and Hoehn & Yahr stage ≤2) and age- and sex-matched controls. Receiver operating characteristic curves were created for each model.

Sensitivity to change over time was evaluated using data from a subgroup of participants (BeaT-PD, n = 33) who were assessed ∼12 months apart, over 4 years. For this analysis, we used data collected at baseline versus year 1 (visit 1) and also compared visit 3 (month ∼36) and visit 4 (month ∼48) to show change within 1 year as the disease advances and not progression over time in this small sample. Differences between time points were assessed for the 40 most salient features, as identified earlier. Effect sizes (Cohen's d) were calculated for measures showing significant change over time.

We further explored the effect sizes of features that were not correlated with the MDS-UPDRS but showed significant change over time. This analysis investigated whether digital measures could be uniquely sensitive to progression, reflecting a different construct than the clinical assessment. Their inclusion was also explored in the models comparing recently diagnosed patients with controls. In exploratory analysis, we estimated the sample size needed for clinical trials from the mean ± SD of the measure with the highest effect size.

Results

The PD cohort included a diverse representative patient population; mostly men (60.6%) between 36 and 86 years, with disease duration between 0.5 and 37 years since diagnosis (Table 1). Most patients had mild to moderate disease severity (∼73% Hoehn & Yahr stages I and II); 85.3% were on antiparkinsonian medications, and 28.1% showed cognitive impairments (Montreal Cognitive Assessment <24). Patient characteristics were similar between sites for each corresponding study ($P > 0.132$) (Supplementary Material Table S1D).

A total of 614 of 655 (94%) of the study participants completed 7 days of collection, and 98% completed ≥3 days. Mean gait speed of the PD cohort, as measured in the clinic, was 1.06 ± 0.60 m/s. The typical (median across all bouts) gait speed as measured in the real world was 0.99 ± 0.16 m/s; real-world best (90th) and worst (10th) percentiles were 1.18 ± 0.25 and 0.82 ± 16 m/s, respectively.

Features Associated with the MDS-UPDRS Scores

Because saturation in scoring was observed, 40 features were used. Figure 1 summarizes the 15 (from 40) measures that were most strongly associated with MDS-UPDRS, parts I, II, and III (for a full list, see Supplementary Material S2–S4). The dominant categories in the top 15 features for MDS-UPDRS-I were NB and GQ. For MDS-UPDRS-II, it was GQ, and for MDS-UPDRS-III, AQDP was the most prominent domain.

Machine Learning Regression Models for Estimating Disease Symptoms

The most frequently chosen and accurate models using cross-validation for MDS-UPDRS I and II were Random Forest with an average Pearson's correlation coefficient of r = 0.50 for part I and r = 0.60 for part II. Bayesian Ridge was the most accurate model for MDS-UPDRS-III (r = 0.50). The mean absolute error ranged between 4 and 5 points for MDS-UPDRS, parts I and II, and ∼8 points for MDS-UPDRS, part III (Table 2; Supplementary Material Table S2.d). Models using only age and gender revealed much lower scores (Supplementary Material Table S2b). Nested cross-validation analysis yielded similar results (Supplementary Material Table S2e).

Sensitivity to Disease and Disease Progression

Characteristics of 64 recently diagnosed persons with PD and 68 HCs included in this analysis are presented in Table 1. A simple model containing age and sex differentiated between HCs and PD patients with an accuracy of 58.6%, an area under the curve (AUC) of 0.56, and Nagelkerke R^2 of 0.03 (Fig. 2, model 1). Seven of

TABLE 1 *Demographics and characteristics of all participants, divided by research question*

Characteristic	Association with disease severity	Sensitivity to disease		Sensitivity to change over time*
	All PD participants (N = 587)	Recently diagnosed PD (N = 64)	Healthy controls (N = 68)	Longitudinal cohort (N = 33)
Age (y)	67.99 ± 8.46 (36–86)	62.16 ± 9.89 (36–79)	58.57 ± 8.63 (46–79)	62.72 ± 8.66 (41–78)
Sex (% M, n M/F)	60.6%, 356/231	61%, 39/25	60.2%, 41/27	63.6%, 21/12
Disease duration (y)	6.44 ± 5.74 (0.5–37)	1.08 ± 0.46 (0.5–2.0)	–	2.57 ± 2.05 (0.5–7.0)
Hoehn & Yahr scale (n, %)				
Stage 1	85, 14.6%	51, 79.69%	–	13, 39.39%
Stage 2	342 58.6%	13, 20.31%		20, 60.61%
Stage 3	155, 25.8%			
Stage 4	5, <1%			
MDS-UPDRS, Part I	9.18 ± 6.37 (0–37)	3.23 ± 2.65 (0–11)	0	5.15 ± 3.65 (0–15)
MDS-UPDRS, Part II	12.63 ± 8.43 (0–44)	4.25 ± 3.12 (0–14)	0	6.87 ± 5.00 (0–23)
MDS-UPDRS, Part III	26.54 ± 13.01 (3–84)	11.68 ± 4.77 (3–24)	1.22 ± 1.83 (0–9)	17.87 ± 9.04 (4–39)
LEDD (mg)	558.57 ± 456.57 (0–3600)	153.07 ± 160.43 (0–665)	–	358.53 ± 388.88 (0–1370)
Use of PD medication	96.3%	84.5%	–	90%
Type of medication (%[a])			–	
Levodopa treatment	64.6%	15.6%	–	25%
Dopamine agonists	28.5%	12.5%	–	39%
MAO-B inhibitors	56.6%	77.5%	–	91%
Amantadine	21.3%	9.3%	–	25%
Other[b]	2%	1.6%	–	–
Gait speed (m/s)	1.06 ± 0.26 (0.31–1.70)	1.09 ± 0.22 (0.42–1.65)	1.08 ± 0.21 (0.58–1.74)	1.08 ± 0.12 (0.60–1.61)
MoCA score	24.44 ± 3.61 (17–30)	25.38 ± 2.79 (18–30)	26.41 ± 2.87 (19–30)	26.00 ± 3.38 (19–30)

Data from 22 participants in ~65 parameters were considered outliers and were removed from the analysis. Hoehn & Yahr stages were determined while patients were *on* medication.

Abbreviations: LEDD, levodopa-equivalent daily dose; MAO-B, Monoamine oxidase B; MDS-UPDRS, Movement Disorder Society-Unified Parkinson's Disease Rating Scale; MoCA, Montreal Cognitive Assessment; PD, Parkinson's disease.

*Characteristics at baseline.

[a]Percentage from those taking medication for PD.

[b]Includes Catechol-O-methyltransferase (COMT), anticholinergics, and apomorphine; gait speed presented was measured in a laboratory setting.

Top 15 features selected by ML methods describing best MDS-UPDRS Part I,II and III		
MDS-UPDRS I	**MDS-UPDRS II**	**MDS-UPDRS III**
NumberOfRotations	NumberOfRotations	StepLength
LeastActive5h	StepLength	NonGaitActiveTime
TotalLogAcc 02TO04Oclock	RotationSpectralArcAP	NumberOfRotations
SampleEntropyV Prc10	GaitSpeed	RelativeAmplitude
WakeTimeNight	SpectralArcV	TimeOfRotation
WalkBoutDistance Prc90	FreqWidthAP	RotationSpectralArcML
FreqSlopeV Prc10	GaitSpeed Prc90	MeanSVM
Age	Age	SumSVMperWearTime
StepLength	StepLengthCV Prc10	DailyActivityRatio
PercentWakeNight	TimeOfRotation	CircadianAmplitude
SleepEfficiency	RotationSpectralArcML	SumSVM
RotationSparcAP	WalkBoutDistance	Age
NumberWalkBoutsNight	AmplitudeV Prc10	ModerateToVigorousActivity
SampleEntropyAP	LieToUprightTime	GaitSpeed Prc90

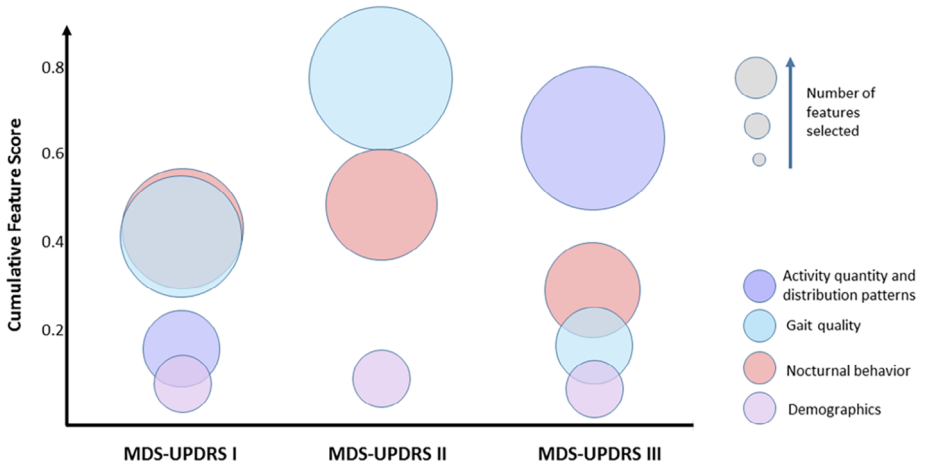

FIG. 1. Selected features and category domains. Bubbles represent the features grouped by domain for each of the first three parts of the MDS-UPDRS (Movement Disorder Society-Unified Parkinson's Disease Rating Scale). The size of the bubble represents the number of features. AP, anteroposterior axis; AQDP, activity quantity and distribution patterns; CV, coefficient of variance; Freq, frequency; GQ, gait quality; ML, mediolateral axis; NB, nocturnal behavior; Prc, percentile; SVM, signal vector magnitude; TLA, total log activity; V, vertical axis. Explanations and details on each of the features are presented in the Supplementary Material.

the 40 MDS-UPDRS-III-associated measures were sufficient to successfully differentiate between the groups with an accuracy of 71.2%, an AUC of 0.79, and Nagelkerke R^2 of 0.34 (Fig. 2, model 2). Model 3 includes seven measures from our exploratory analysis (see later and Supplementary Material S7) showing an accuracy of 65.90%, an AUC of 0.70, and Nagelkerke R^2 of 0.12. Combining measures from models 2 and 3 (14 features) improved the accuracy to 81.10%, the AUC to 0.87, and Nagelkerke R^2 to 0.51 (model 4). Using only the four highest-contributing measures in this model (see Supplementary Material 7) resulted in an accuracy of 77.30%, an AUC of 0.83, and Nagelkerke R^2 of 0.42 (model 5).

Longitudinal data from 33 patients with PD (see Table 1) were used to assess sensitivity to progression

TABLE 2 *ML regression model selected in the majority of repetitions*

MDS-UPDRS part	Chosen model	*R*	*P*-value	Mean absolute error
MDS-UPDRS-I	Random Forest	0.50 ± 0.020	<0.001	4.09 ± 0.02
MDS-UPDRS-II	Random Forest	0.60 ± 0.005	<0.001	4.92 ± 0.04
MDS-UPDRS-III	Bayesian Ridge	0.50 ± 0.000	<0.001	8.15 ± 0.04

Results reflect an average of model performance on the test set. Model selection and training were performed using a fivefold cross-validation process with Pearson's R^2 as the performance metric.
Abbreviations: ML, machine learning; MDS-UPDRS, Movement Disorder Society-Unified Parkinson's Disease Rating Scale.

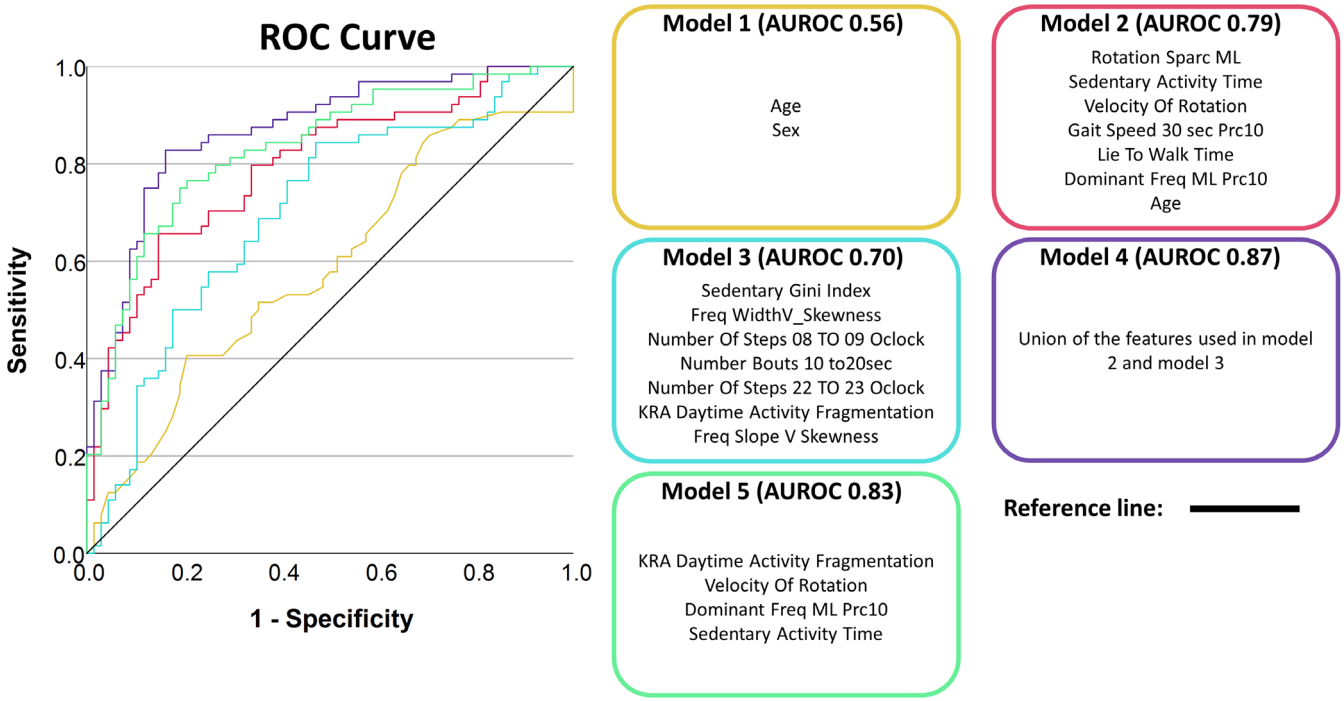

FIG. 2. Models differentiating between recently diagnosed patients with PD (Parkinson's disease) and healthy controls (HC). Model 1 included only age and sex (accuracy: 58.6%, AUC [area under the curve]: 0.56, Nagelkerke $R^2 = 0.03$). Model 2 included 7 of the 40 features associated with MDS-UPDRS-III ([Movement Disorder Society-Unified Parkinson's Disease Rating Scale], accuracy: 71.2%, AUC: 0.79, Nagelkerke $R^2 = 0.34$); model 3: the seven features that were found to be sensitive to change in our exploratory analysis (accuracy: 65.90%, AUC: 0.70, Nagelkerke $R^2 = 0.12$); model 4: the combined 14 features from models 2 and 3 (accuracy: 81.10%, AUC: 0.87, Nagelkerke $R^2 = 0.51$); model 5: the four highest-contributing features in model 3 (accuracy: 77.30%, AUC: 0.83, Nagelkerke $R^2 = 0.42$). The simple model 1 significantly differed from models 2, 4, and 5 ($P < 0.0001$), whereas a trend was observed between models 1 and 3 ($P = 0.077$). Nagelkerke R^2 value was calculated to determine the goodness of fit of the logistic regression models. Higher values indicate a stronger fit (Nagelkerke N.J.D., Biometrika 78: 691–692, 1991).

over time. Mean change in MDS-UPDRS-III was 0.39 ± 5.37, with 25% of patients receiving L-dopa at baseline and 50% at visit 1 (Supplementary Material S9). Mean change in MDS-UPDRS-III between visits 3 and 4 was 3.70 ± 9.02 (Supplementary Material S9). Figure 2 shows that for all three parts of the MDS-UPDRS, several digital mobility features showed greater effect sizes than that of the corresponding MDS-UPDRS score. The measures with the highest sensitivity to change (between baseline and visit 1) relating to MDS-UPDRS-I were MaxSedentaryBoutDuration (effect size = 0.33), a measure reflecting the amount of sedentary behavior, and NumberOfRotations (effect size = 0.19), reflecting nocturnal movement. For features related to MDS-UPDRS-II, only NumberOfRotations significantly worsened over time (5.95 ± 2.59 at baseline vs. 4.14 ± 1.86 visit 1), reflecting an effect size of 0.19 versus 0.10 for the clinical scale. For MDS-UPDRS-III, the most sensitive feature was NumberBoutsUnder60sec, a measure quantifying walking capacity (effect size = 0.34 vs. 0.04 for MDS-UPDRS-III).

In general, the effect sizes became larger 2.34 ± 0.60 years later, for both the digital measures and all parts of the MDS-UPDRS, yet remained higher for digital measures (Fig. 3). With disease progression, the ratio of different digital mobility domains with the highest effect sizes changed. In early years, activity quantity and nocturnal measures showed significant sensitivity to change, whereas GQ measures did not. This changed as the disease progressed, with GQ measures becoming more prominent (Fig. 3; Supplementary Material 10a–f).

In the exploratory analysis, seven measures that were not correlated with the MDS-UPDRS showed significant change between baseline and visit 1. These seven measures were related to within bout regularity of movement across the week and activity at specific times of the day, with effect sizes ranging between 0.52 and 0.36 reflecting even higher sensitivity compared to the clinical scale (for a full list of features, see Supplementary Material S7).

As NumberOfRotations was associated with all three parts of the MDS-UPDRS and showed greater effect sizes than the clinical scale, we estimated the sample size needed to show significant change over time. At 80% power, the digital measure required 83%, 82%, and 95% fewer participants than parts I, II, and III, respectively, at baseline versus visit 1 and 98%, 81%, and 9.8% fewer at visit 3 versus 4. (Supplementary Material Table S11).

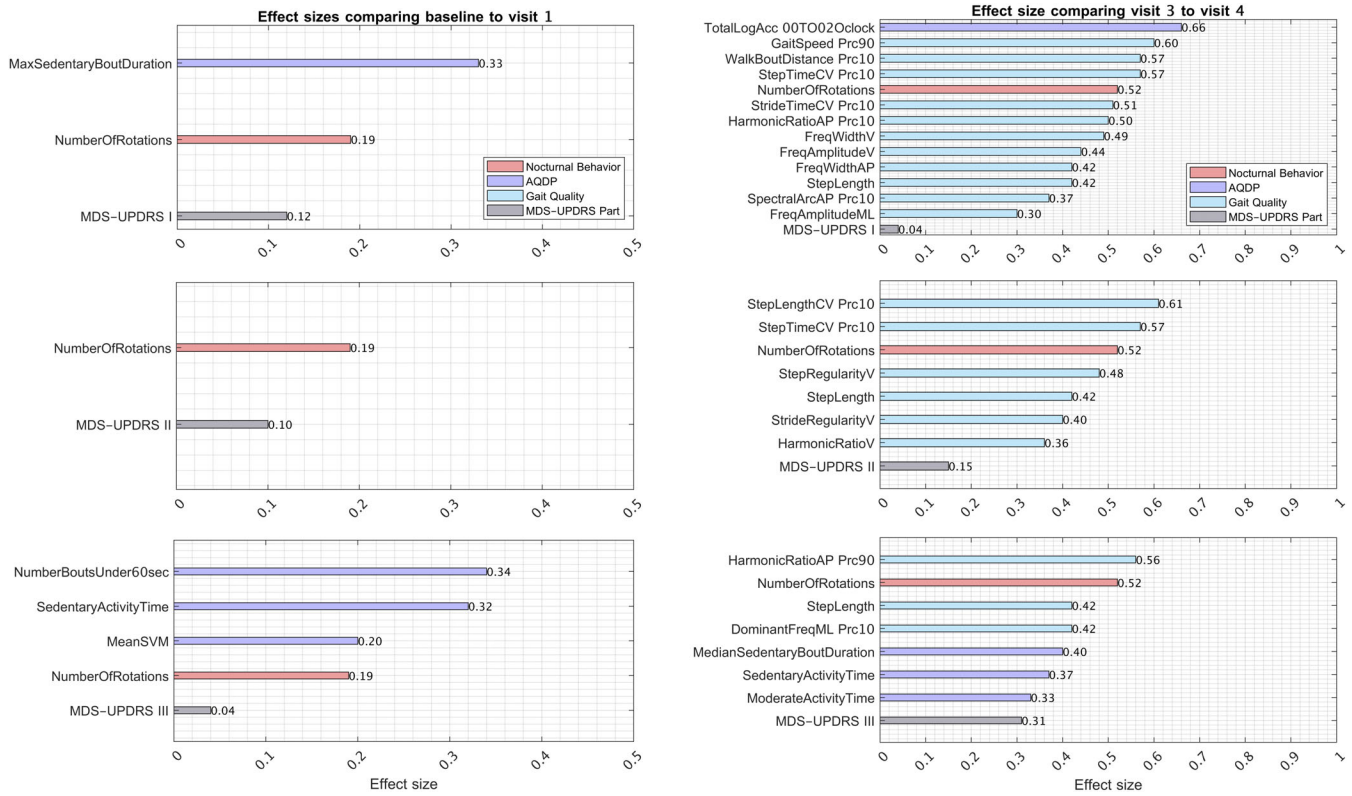

FIG. 3. Effect sizes of features significantly sensitive to change over 1 year. Colors represent domains according to the domain categories; AQDP, activity quantity and distribution patterns; GQ, gait quality; NB, nocturnal behavior. Left: the magnitude of change between baseline and visit 1; right: the magnitude of change between visits 3 and 4 (2.34 ± 0.60 years later, reflecting a progression in the disease). At baseline, the mean disease duration was 2.57 ± 2.05. The mean change in MDS-UPDRS-III (Movement Disorder Society-Unified Parkinson's Disease Rating Scale) was 0.39 ± 5.37 points between baseline and visit 1 and 3.70 ± 9.02 between visits 3 and 4. When adjusting for disease duration, the measures and the ranking of the effect sizes were similar to those presented here, with a higher magnitude due to the reduced standard deviation.

Discussion

In this multicenter study, we utilized multiple machine learning techniques to evaluate the association between an array of real-world digital mobility measures and disease severity, as measured by the widely used PD clinical rating scales, in a large cohort of patients with PD. The analyses produced several important insights: (1) digital mobility measures were only moderately associated with the MDS-UPDRS; (2) different digital mobility measures were associated with each of the three parts of the MDS-UPDRS; (3) digital mobility measures were sensitive to early-stage disease, accurately discriminating recently diagnosed patients with PD from HCs; and (4) digital mobility measures were more sensitive to change over time than each part of the clinical rating scale.

Many studies have evaluated the use of digital technologies for PD symptom monitoring,[12,14,34] generating some evidence for the feasibility and advantages of objective and more frequent monitoring of disease burden with potential as clinical decision-support tools.[19,35-37] Similar to previous studies,[20,38] our results show only moderate associations between digital mobility outcomes and the current gold standard clinical scale. This suggests that the metrics evaluate different constructs; whereas MDS-UPDRS assesses disease signs and symptoms, digital measures assess function, inherently encompassing disease signs but in a behavioral context.[18] Previous work has shown that during daily activities, gait speed of PD patients at home can decrease by 30% compared to that measured in the clinic.[20,38] Movements during daily life are typically self-initiated, embedded in a rich behavioral context, whereas laboratory and clinic-based assessments are usually initiated by an external signal or demand and are executed in isolation.[39] The context of the environment is also different, with the real world imposing greater environmental complexity, whereas performance during in-clinic assessments can be greatly affected by factors such as anxiety or stress, which may worsen or paradoxically improve certain signs.[5,40] As a result, variability in function could be large despite the method of assessment.[12] Moreover, in-clinic assessment is often conducted during *on* medication. In contrast, real-world digital measures, due to the wider time frame of collection, also include motor fluctuations, capturing *on/off* and in between. In this regard, the

clinical and digital measures are not directly comparable, representing "apples and oranges," each important to fully understand and quantify disease severity. Perhaps it is time to consider digital mobility measures as outcomes in their own regard, validating them against meaningfulness to daily living, sensitivity to change over time, and their potential response to therapeutic interventions.

In recent work, patients with PD shared their perspective that slowness of movement was among their most bothersome and important symptoms.[40] Patients reported that PD symptoms had the greatest impact on sleep, job functioning, exercise, communication, relationships, and sense of being,[41] aspects of daily living that can be partially reflected in real-world function collected via digital measures.[42] Indeed our findings show that the measures most strongly associated with disease severity were not related to quality but rather to function coming from domains of sleep (NB) and activity quantity and distribution. The number of rotations at night is of particular interest. It was associated with all three parts of the MDS-UPDRS, contributed to the discrimination model between recently diagnosed patients and HCs, and was sensitive to change over time. Several studies have shown the utility of such a measure,[2,3] which highlights the benefit of passive digital mobility assessment as such measures cannot be captured using conventional testing. Interestingly, a greater number of measures reflecting GQ were more associated with MDS-UPDRS, part II, than with part III. This perhaps relates to the constructs of the clinical scales; part III is heavily focused on tremor and rigidity, whereas part II evaluates difficulties in daily living, including transfers and dressing. It may be helpful to consider that the number of features in the GQ domain is much larger than that in the NB domain and the "curse of dimensionality." Nonetheless, if this imbalance was the only determining factor, one would expect to see that many gait features would have been selected and that a few NBs would have been selected. The results suggest, however, that the feature selection process successfully identified the most relevant features and was not unduly influenced by the number of features in each domain.

Regulators in Europe have recently accepted stride velocity 95th centile[43] in Duchene muscular dystrophy and moderate to vigorous physical activity in idiopathic pulmonary fibrosis as endpoints for clinical trials.[44] In the present study, related digital mobility features were observed as relevant. For example, walking bout distance 90th percentile, gait speed 90th percentile, and step length coefficient of variance 10th percentile, measures that are derived from the distribution of an individual's walking quality across the week and reflect an individual's best gait performance, were among the most prominent features associated with the MDS-UPDRS, demonstrating that such measures are also valuable as capacity measures in PD. In addition, moderate to vigorous activity was one of the top-ranked measures associated with MDS-UPDRS, part III. One can argue that the 90th percentile and vigorous activity reflect capacity, and are more related to the clinical scale, whereas typical activity may reflect function or performance. Measures of reduced function, such as sedentary time and walking bout distance 10th percentile, were also found to be sensitive to change over time. These capacity and function measures were obtained passively in the home, unobtrusively capturing the best and worst performances of the person. Understanding the contribution of these measures in PD has great value and potential as future endpoints. The ongoing Mobilise-D project aims to obtain regulatory qualification for digital mobility outcomes in PD and is well positioned to build on the present work for that important objective.[45]

For use as endpoints in clinical trials, digital mobility measures should meet multiple requirements, such as sensitivity to disease severity and sensitivity to change. Our findings show that digital measures accurately differentiate between recently diagnosed patients with PD and HCs, reflecting the potential use as a screening tool and perhaps an opportunity to also use digital mobility measures in the prodromal stage. Our longitudinal data showed that digital mobility measures are more sensitive to change than the gold standard clinical scale (Fig. 3). In mild to moderate disease, activity and NB measures were most prominent, perhaps reflecting more burden of nonmotor symptoms (ie, problems with sleep and nocturia). In contrast, as the disease progressed, motor symptoms from the GQ domain were more prominent and notably were more sensitive to progression than the clinical scale.

Recently, Brzezicki et al compared standard clinical rating scales and kinematic features of upper-extremity bradykinesia and gait as assessed in a clinic using wearable devices in untreated, recently diagnosed patients with PD assessed quarterly for 2 years.[46] The study showed that commencing antiparkinsonian medication led to masking of progression signals in both clinical and digital measures.[46] Similarly, we also observed only minimal change in MDS-UPDRS in the early years, with 50% of our cohort on antiparkinsonian medication at follow-up. Contrarily, several digital mobility measures still showed significant change. The discrepancy between our findings and those of Brzezicki et al may be attributed to the assessment protocol, conducted in the real-world versus a clinic-based assessment. It may be interesting in the future to combine clinic-based and real-world-based measures to evaluate their value. The minor change in the MDS-UPDRS-III in medicated patients over 1 year that we observed is similar to that found in the Parkinson's Progression

Markers Initiative[47,48] and presents a challenge for clinical trials. The ability to detect small changes with digital mobility measures within 1 year may directly lead to a reduction in sample size needed and thus reduce time and cost of clinical trials.

The present analyses have several limitations. The wearable device used was placed on the lower back, limiting our ability to also assess arm swing, tremor, or dyskinesia, which are known to be impaired in different stages of PD and may impact mobility. In the future, it might be interesting to evaluate the benefit obtained by instrumenting both the wrist and lower back, although this may come with the trade-off of lower compliance and poorer patient acceptance. Our longitudinal cohort was relatively small. Future work such as in the Personalized Parkinson Project,[49] Mobilise-D project,[45] and WatchPD[50] will further explore these findings and the prognostic value of digital mobility measures. We did not assess patients *off* medication or compare naive patients to those *on* medications or assess treatment effects. These and other important questions should be addressed in ongoing[42,45,50] and future studies. Nonetheless, this is one of the largest studies investigating digital mobility technology in PD. Our findings demonstrate the value of quantifying multiple domains of mobility and reveal the ability of digital mobility measures to detect and track meaningful, relevant, sensitive changes that can be used in both clinical trials and clinical care.

Acknowledgments: We thank the participants of the studies for their time and effort and thank Batsheva Cohen, Liat Yahimovich, and Ora Assias for their valuable assistance in recruitment and data collection.

Data Availability Statement

The data supporting the findings presented are available from the corresponding author upon reasonable request. For additional details see SM. The code and algorithms are available in: https://github.com/Jana-art/Axivity_Analysis_Daily_Living.

References

1. Houts CR, Patrick-Lake B, Clay I, Wirth RJ. The path forward for digital measures: suppressing the desire to compare apples and pineapples. Digit Biomark 2020;4:3–12.
2. Mirelman A, Hillel I, Rochester L, et al. Tossing and turning in bed: nocturnal movements in Parkinson's disease. Mov Disord 2020;35:959–968.
3. Sringean J, Taechalertpaisarn P, Thanawattano C, Bhidayasiri R. How well do Parkinson's disease patients turn in bed? Quantitative analysis of nocturnal hypokinesia using multisite wearable inertial sensors. Parkinsonism Relat Disord 2016;23:10–16.
4. Masanneck L, Gieseler P, Gordon WJ, Meuth SG, Stern AD. Evidence from ClinicalTrials.gov on the growth of digital health technologies in neurology trials. NPJ Digit Med 2023;6:23.
5. Bloem BR, Post E, Hall DA. An apple a day to keep the Parkinson's disease doctor away? Ann Neurol 2023;93:681–685.
6. Landers M, Dorsey R, Saria S. Digital endpoints: definition, benefits, and current barriers in accelerating development and adoption. Digit Biomark 2021;5:216–223.
7. Stephenson D, Badawy R, Mathur S, Tome M, Rochester L. Digital progression biomarkers as novel endpoints in clinical trials: a multistakeholder perspective. J Parkinsons Dis 2021;11:S103–S109.
8. Goetz CG, Tilley BC, Shaftman SR, 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 2008;23:2129–2170.
9. Food and Drug Administration. FDA response to Verily's Virtual exam; 2022.
10. Dorsey ER, Venuto C, Venkataraman V, Harris DA, Kieburtz K. Novel methods and technologies for 21st-century clinical trials: a review. JAMA Neurol 2015;72:582–588.
11. Tosin MHS, Simuni T, Stebbins GT, Cedarbaum JM. Tracking emergence of new motor and non-motor symptoms using the MDS-UPDRS: a novel outcome measure for early Parkinson's disease? J Parkinsons Dis 2022;12:1345–1351.
12. Burq M, Rainaldi E, Ho KC, et al. Virtual exam for Parkinson's disease enables frequent and reliable remote measurements of motor function. NPJ Digit Med 2022;5:65.
13. Zhan A, Mohan S, Tarolli C, et al. Using smartphones and machine learning to quantify Parkinson disease severity: the Mobile Parkinson disease score. JAMA Neurol 2018;75:876–880.
14. Adams JL, Dinesh K, Snyder CW, et al. A real-world study of wearable sensors in Parkinson's disease. NPJ Parkinsons Dis 2021;7:106.
15. Espay AJ, Hausdorff JM, Sanchez-Ferro A, et al. A roadmap for implementation of patient-centered digital outcome measures in Parkinson's disease obtained using mobile health technologies. Mov Disord 2019;34:657–663.
16. Rochester L, Mazza C, Mueller A, et al. A roadmap to inform development, validation and approval of digital mobility outcomes: the mobilise-D approach. Digit Biomark 2020;4:13–27.
17. Viceconti M, Hernandez PS, Dartee W, et al. Toward a regulatory qualification of real-world mobility performance biomarkers in Parkinson's patients using digital mobility outcomes. Sensors (Basel) 2020;20(20):5920.
18. World Health Organization. Towards a common language for functioning, disability and health ICF; 1149 2002.
19. Del DS, Kirk C, Yarnall AJ, Rochester L, Hausdorff JM. Body-worn sensors for remote monitoring of Parkinson's disease motor symptoms: vision, state of the art, and challenges ahead. J Parkinsons Dis 2021;11(s1):S35–S47.
20. Galperin I, Hillel I, Del DS, et al. Associations between daily-living physical activity and laboratory-based assessments of motor severity in patients with falls and Parkinson's disease. Parkinsonism Relat Disord 2019;62:85–90.
21. Mantri S, Wood S, Duda JE, Morley JF. Comparing self-reported and objective monitoring of physical activity in Parkinson disease. Parkinsonism Relat Disord 2019;67:56–59.
22. Salomon A, Galperin I, Buzaglo D, et al. Fragmentation, circadian amplitude, and fractal pattern of daily-living physical activity in people with multiple sclerosis: is there relevant information beyond the total amount of physical activity? Mult Scler Relat Disord 2022;68:104108.
23. Mirelman A, Rochester L, Maidan I, et al. Addition of a non-immersive virtual reality component to treadmill training to reduce fall risk in older adults (V-TIME): a randomised controlled trial. Lancet 2016;388:1170–1182.
24. Manor B, Dagan M, Herman T, et al. Multitarget transcranial electrical stimulation for freezing of gait: a randomized controlled trial. Mov Disord 2021;36:2693–2698.
25. Oz S, Dagay A, Katzav S, et al. Monitoring sleep stages with a soft electrode array: comparison against vPSG and home-based detection of REM sleep without atonia. J Sleep Res 2023;32(5):e13909.
26. Zoetewei D, Herman T, Brozgol M, et al. Protocol for the DeFOG trial: a randomized controlled trial on the effects of smartphone-based, on-demand cueing for freezing of gait in Parkinson's disease. Contemp Clin Trials Commun 2021;24:100817.

27. Postuma RB, Berg D, Stern M, et al. MDS clinical diagnostic criteria for Parkinson's disease. Mov Disord 2015;30:1591–1601.
28. Del DS, Galna B, Godfrey A, et al. Analysis of free-living gait in older adults with and without Parkinson's disease and with and without a history of falls: identifying generic and disease-specific characteristics. J Gerontol A Biol Sci Med Sci 2019;74:500–506.
29. Leroux A, Xu S, Kundu P, et al. Quantifying the predictive performance of objectively measured physical activity on mortality in the UK biobank. J Gerontol A Biol Sci Med Sci 2021;76:1486–1494.
30. Weiss A, Herman T, Giladi N, Hausdorff JM. Objective assessment of fall risk in Parkinson's disease using a body-fixed sensor worn for 3 days. PloS One 2014;9:e96675.
31. Balasubramanian S, Melendez-Calderon A, Burdet E. A robust and sensitive metric for quantifying movement smoothness. IEEE Trans Biomed Eng 2012;59:2126–2136.
32. Weiss A, Sharifi S, Plotnik M, van Vugt JP, Giladi N, Hausdorff JM. Toward automated, at-home assessment of mobility among patients with Parkinson disease, using a body-worn accelerometer. Neurorehabil Neural Repair 2011;25:810–818.
33. Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 2005;27:1226–1238.
34. Daneault JF, Vergara-Diaz G, Parisi F, et al. Accelerometer data collected with a minimum set of wearable sensors from subjects with Parkinson's disease. Sci Data 2021;8:48.
35. Debelle H, Packer E, Beales E, et al. Feasibility and usability of a digital health technology system to monitor mobility and assess medication adherence in mild-to-moderate Parkinson's disease. Front Neurol 2023;14:1111260.
36. Omberg L, Chaibub NE, Perumal TM, et al. Remote smartphone monitoring of Parkinson's disease and individual response to therapy. Nat Biotechnol 2022;40:480–487.
37. Sundgren M, Andreasson M, Svenningsson P, Noori RM, Johansson A. Does information from the Parkinson KinetiGraph (PKG) influence the Neurologist's treatment decisions?-an observational study in routine clinical Care of People with Parkinson's disease. J Pers Med 2021;11(6):519.
38. Toosizadeh N, Mohler J, Lei H, Parvaneh S, Sherman S, Najafi B. Motor performance assessment in Parkinson's disease: association between objective in-clinic, objective in-home, and subjective/semi-objective measures. PloS One 2015;10:e0124763.
39. Giannouli E, Bock O, Mellone S, Zijlstra W. Mobility in old age: capacity is not performance. Biomed Res Int 2016;2016:3261567.
40. Port RJ, Rumsby M, Brown G, Harrison IF, Amjad A, Bale CJ. People with Parkinson's disease: what symptoms do they Most want to improve and how does this change with disease duration? J Parkinsons Dis 2021;11:715–724.
41. Mammen JR, Speck RM, Stebbins GT, et al. Relative meaningfulness and impacts of symptoms in people with early-stage Parkinson's disease. J Parkinsons Dis 2023;13:619–632.
42. Mammen JR, Speck RM, Stebbins GM, et al. Mapping relevance of digital measures to meaningful symptoms and impacts in early Parkinson's disease. J Parkinsons Dis 2023;13:589–607.
43. Servais L, Yen K, Guridi M, Lukawy J, Vissiere D, Strijbos P. Stride velocity 95th centile: insights into gaining regulatory qualification of the first wearable-derived digital endpoint for use in Duchenne muscular dystrophy trials. J Neuromuscul Dis 2022;9:335–346.
44. Inacio P. Bellerophon, FDA agree on design of phase 3 INOpulse trial. 1-3-2020.
45. Mikolaizak AS, Rochester L, Maetzler W, et al. Connecting real-world digital mobility assessment to clinical outcomes for regulatory and clinical endorsement-the mobilise-D study protocol. PloS One 2022;17:e0269615.
46. Brzezicki MA, Conway N, Sotirakis C, FitzGerald JJ, Antoniades CA. Antiparkinsonian medication masks motor signal progression in de novo patients. Heliyon 2023;9:e16415.
47. Sample size estimation for clinical trials in de novo Parkinson's disease (PD): results from the Parkinson's progression markers initiative (PPMI) study. Mov Disord Soc conf; 2017;32(suppl 2). https://www.mdsabstracts.org/abstract/sample-size-estimation-for-clinical-trials-in-de-novo-parkinsons-disease-pd-results-from-the-parkinsons-progression-markers-initiative-ppmi-study/
48. Simuni T, Siderowf A, Lasch S, et al. Longitudinal change of clinical and biological measures in early Parkinson's disease: Parkinson's progression markers initiative cohort. Mov Disord 2018;33:771–782.
49. Bloem BR, Marks WJ Jr, Silva de Lima AL, et al. The personalized Parkinson project: examining disease progression through broad biomarkers in early Parkinson's disease. BMC Neurol 2019;19:160.
50. Adams JL, Kangarloo T, Tracey B, et al. Using a smartwatch and smartphone to assess early Parkinson's disease in the WATCH-PD study. NPJ Parkinsons Dis 2023;9:64.

Supporting Data

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.