JAMA Neurology | Brief Report

Using Smartphones and Machine Learning to Quantify Parkinson Disease Severity
The Mobile Parkinson Disease Score

Andong Zhan, MS; Srihari Mohan; Christopher Tarolli, MD; Ruth B. Schneider, MD; Jamie L. Adams, MD; Saloni Sharma, MD; Molly J. Elson, BA; Kelsey L. Spear, MPH; Alistair M. Glidden, BS; Max A. Little, PhD; Andreas Terzis, PhD; E. Ray Dorsey, MD; Suchi Saria, PhD

IMPORTANCE Current Parkinson disease (PD) measures are subjective, rater-dependent, and assessed in clinic. Smartphones can measure PD features, yet no smartphone-derived rating score exists to assess motor symptom severity in real-world settings.

OBJECTIVES To develop an objective measure of PD severity and test construct validity by evaluating the ability of the measure to capture intraday symptom fluctuations, correlate with current standard PD outcome measures, and respond to dopaminergic therapy.

DESIGN, SETTING, AND PARTICIPANTS This observational study assessed individuals with PD who remotely completed 5 tasks (voice, finger tapping, gait, balance, and reaction time) on the smartphone application. We used a novel machine-learning–based approach to generate a mobile Parkinson disease score (mPDS) that objectively weighs features derived from each smartphone activity (eg, stride length from the gait activity) and is scaled from 0 to 100 (where higher scores indicate greater severity). Individuals with and without PD additionally completed standard in-person assessments of PD with smartphone assessments during a period of 6 months.

MAIN OUTCOMES AND MEASURES Ability of the mPDS to detect intraday symptom fluctuations, the correlation between the mPDS and standard measures, and the ability of the mPDS to respond to dopaminergic medication.

RESULTS The mPDS was derived from 6148 smartphone activity assessments from 129 individuals (mean [SD] age, 58.7 [8.6] years; 56 [43.4%] women). Gait features contributed most to the total mPDS (33.4%). In addition, 23 individuals with PD (mean [SD] age, 64.6 [11.5] years; 11 [48%] women) and 17 without PD (mean [SD] age 54.2 [16.5] years; 12 [71%] women) completed in-clinic assessments. The mPDS detected symptom fluctuations with a mean (SD) intraday change of 13.9 (10.3) points on a scale of 0 to 100. The measure correlated well with the Movement Disorder Society Unified Parkinson Disease's Rating Scale total (r = 0.81; P < .001) and part III only (r = 0.88; P < .001), the Timed Up and Go assessment (r = 0.72; P = .002), and the Hoehn and Yahr stage (r = 0.91; P < .001). The mPDS improved by a mean (SD) of 16.3 (5.6) points in response to dopaminergic therapy.

CONCLUSIONS AND RELEVANCE Using a novel machine-learning approach, we created and demonstrated construct validity of an objective PD severity score derived from smartphone assessments. This score complements standard PD measures by providing frequent, objective, real-world assessments that could enhance clinical care and evaluation of novel therapeutics.

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Suchi Saria, PhD, Malone Hall 333, Johns Hopkins University, 3400 N Charles St, Baltimore, MD 21218 (ssaria@cs.jhu.edu).

Current Parkinson disease (PD) measures are subjective and rater-dependent and require in-clinic assessments.[1,2] As a result, clinical trials using these measures are long, expensive, and can generate false positives or negatives.[2,3] Many motor symptoms of PD are well-suited to objective measurement by smartphones.[4-6] Smartphone assessment has been evaluated in PD, but most studies focus on a specific feature (eg, gait), rather than overall symptom burden.[6,7] We developed an Android smartphone application (named *HopkinsPD*) that assesses 5 activities (voice, finger tapping, gait, balance, and reaction time; eMethods, eTable 1, and the eFigure in the Supplement),[8] which can be completed as often as desired and includes reporting of medication administration. We created a mobile Parkinson disease score (mPDS) to serve as an objective measure of PD and tested construct validity by evaluating the ability of the mPDS to detect intraday symptom fluctuations, the correlation between this measure and current standard PD measures, and the ability of the mPDS to respond to dopaminergic therapy.

Methods

Study Population

Individuals with PD who owned Android smartphones were invited to download HopkinsPD through the Parkinson Voice Initiative.[8] Data from participants who completed at least 1 complete set of activities before and after their first daily dose of dopaminergic medication (development cohort) were used to develop the mPDS. We also recruited individuals with and without PD to complete smartphone activities alongside current standard assessments (clinic cohort); tests included the Movement Disorder Society Unified Parkinson Disease's Rating Scale (MDS-UPDRS),[9] the Hoehn and Yahr stage,[10] and the Timed Up and Go assessment[11] at baseline, month 3, and month 6.

All study procedures were approved by the University of Rochester research subjects review board. Development cohort participants provided electronic consent for data analysis with application download. The clinic cohort participants provided written informed consent.

Creating the mPDS

Data from the development cohort were processed to extract novel disease features from each of the 5 activities (eg, the intertap interval from the finger-tapping activity).[12] Rather than replicating an existing PD score using regression, we used a rank-based machine-learning algorithm, disease severity score learning (DSSL),[13] to derive an independent measure of PD symptom severity: the mPDS, which is scaled from 0 to 100, with high numbers reflecting greater symptom severity. To weigh unique features, the algorithm exploits weak supervision[14] based on the assumption that symptom severity is higher immediately preceding dopaminergic medication administration compared with a point 1 hour after medication administration. Given many such pairs, DSSL estimates a score by optimizing an objective function to correctly rank as many pairs as possible. Further description of the method

Key Points

Question Can a smartphone be used to quantify Parkinson disease motor symptom severity?

Findings In this study, a machine learning approach was able to generate an objective severity score for Parkinson disease from smartphone sensor data. The score captured intraday symptom fluctuations, correlated strongly with current standard rating scales, and detected response to dopaminergic therapy.

Meaning A smartphone-derived severity score for Parkinson disease is feasible and provides an objective measure of motor symptoms inside and outside the clinic that could be valuable for clinical care and therapeutic development.

can be found in the eMethods and the eEquation in the Supplement. Open-source code for feature extraction and the DSSL learning algorithm was made available at https://github.com/dashan-emr/mpds.

Outcome Measures

We evaluated the ability of the mPDS to capture symptom variability by evaluating the average intraday range in mPDS among home-performed assessments in those with PD in the clinic cohort. Smartphone and current standard assessments completed within 2 hours of each other were used to compare the mPDS with current standard measures in individuals with PD. Pearson correlation was calculated between the mPDS and the MDS-UPDRS total score and part III–only subscore (which examines motor signs of PD), the Timed Up and Go assessment, and the Hoehn and Yahr stage. *P* values associated with the Pearson correlation of 1 rating scale vs another were computed from 2-tailed single-hypothesis tests with the null hypothesis that these correlations are 0. The test statistic was computed by multiplying the estimated correlation (ρ) by the square root of $(N-2)/(1-\rho^2)$ and conforms to a *t* distribution with $n-2$ *df* (where n is the number of cross-sectional points). These *P* values should be interpreted for each test as the probability of an uncorrelated system producing a dataset with a Pearson correlation at least as extreme as the one observed. We evaluated the ability of the mPDS to respond to dopaminergic therapy in the clinic cohort by comparing the mPDS derived during optional, clinic-performed, on-medication vs off-medication evaluations of individuals. A 1-tailed Wilcoxon signed-rank test was used to assess significance (α = .05). Statistical analysis was performed with R, version 3.4.1 (R Project for Statistical Computing) and Python, version 2.7.10 (Python Software Foundation).

Results

A total of 250 individuals with PD downloaded HopkinsPD; 129 (51.6%) fulfilled requirements for the development cohort. An additional 23 individuals with PD and 17 without PD constituted the clinic cohort. Baseline characteristics are shown in **Table 1**. Briefly, participants ranged in age from a mean (SD) of 58.7 (8.6) years in the development cohort to 64.6 (11.5) years

Table 1. Baseline Characteristics

| | No. (%) | | | |
| | Smartphone Users (n = 250) | Development Cohort (n = 129) | Clinic Cohort With Parkinson Disease (n = 23) | Clinic Cohort Without Parkinson Disease (n = 17) |
Characteristic				
Demographic				
Age, y, mean (SD)	57.2 (9.4)	58.7 (8.6)	64.6 (11.5)	54.2 (16.5)
Women	95 (38.0)	55 (42.6)	11 (48)	12 (71)
White race	225 (90.0)	123 (95.3)	22 (96)	16 (94)
Hispanic/Latino	15 (6.0)	9 (7.0)	0 (0)	0 (0)
College graduate	238 (95.2)	121 (93.7)	14 (61)	8 (47)
Using internet or email at home	250 (100)	129 (100)	21 (91)	17 (100)
Clinical				
Time since diagnosis, y, mean (SD)	4.4 (4.9)	4.3 (4.4)	7.0 (4.1)	N/A
Taking levodopa	96	97	90	N/A
MDS-UPDRS total score, mean (SD)	NA	NA	55.0 (26.5)	4.6 (4.6)
MDS-UPDRS III score, mean (SD)	NA	NA	26.9 (11.2)	1.2 (1.7)
Timed Up and Go Test, s, mean (SD)	NA	NA	11.2 (3.3)	8.1 (1.3)
Hoehn and Yahr score, stage, mean (SD)	NA	NA	2.1 (0.7)	0.0 (0.0)

Abbreviations: MDS-UPDRS, Movement Disorder Society Unified Parkinson Disease's Rating Scale; NA, not applicable.

and 54.2 (16.5) years in the clinic cohort with and without PD, respectively; 161 of 169 individuals (95.3%) in the development and clinic cohorts combined were white. Those with PD completed 58 in-clinic assessments (22 [96%] at baseline, 18 [78%] at month 3, and 18 [78%] at month 6); those without PD completed 37 assessments (17 [100%] at baseline, 8 [47%] at month 3, and 12 [71%] at month 6).

Creating the mPDS

During 6 months, development cohort participants performed a mean (SD) of 48 (61) complete activity sets (range, 2-278). A total of 435 unique features were extracted from the 5 smartphone tasks; of these, 8 features from the finger-tapping activity, 3 from the balance activity, 3 from the gait activity, and 1 from the voice activity contributed most toward mPDS generation (eTable 2 in the Supplement). The relative weighting of features in generating the mPDS was gait (33.4%), balance (23.2%), finger tapping (23.0%), voice (17.0%), and reaction time (3.4%). The mean (SD) mPDS (across all assessments) was 30.3 (15.0) in control participants; this was 47% lower than in those with PD (mean [SD] score, 57.5 [16.9]).

Outcomes

During 6 months, clinic cohort participants performed a mean (SD) of 210 (323) complete activity sets (range, 2-996). The mPDS detected a mean (SD) intraday change of 13.9 (10.3) points among those with PD. The **Figure**, A depicts intraday severity fluctuations. The mean (SD) MDS-UPDRS part IV score (which assesses motor complications) was 4.6 (4.3) points. A total of 16 smartphone and standard assessment pairs met criteria for analysis. **Table 2** shows the correlation matrix between the MPDS and standard clinical measures. There was good to excellent correlation between the mPDS and the MDS-UPDRS total ($r = 0.81$, $P < .001$) and part III-only subscore ($r = 0.88$, $P < .001$), the Timed Up and Go test ($r = 0.72$, $P = .002$), and the Hoehn and Yahr stage ($r = 0.91$, $P < .001$). The Figure, B shows the ability of the mPDS to monitor symptom severity more frequently than standard measures.
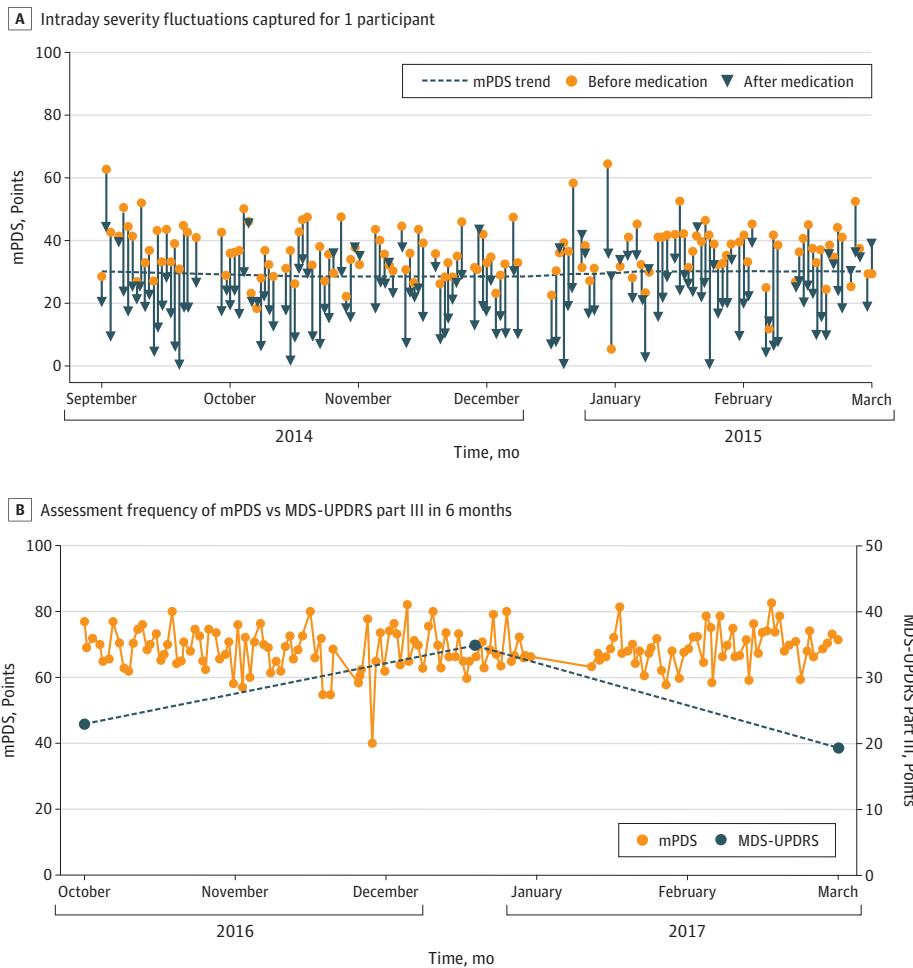
In addition, 7 on-medication vs off-medication pairs of assessments in individuals with PD who were either taking or not taking medication were performed in the clinic cohort. The mPDS decreased by a mean (SD) of 16.3 (5.6) points in response to dopaminergic therapy, with significant Wilcoxon signed rank test (W, 28; $P = .01$). The MDS-UPDRS part III-only subscore decreased by a mean (SD) of 10.4 (4.6) in response to dopaminergic therapy.

Discussion

The mPDS is a novel measure that provides rapid, remote, frequent, and objective assessment of PD symptom severity on widely available smartphones. We demonstrated construct validity by showing that the mPDS can capture intraday fluctuations characteristic of PD, correlate with current standard PD measures, and respond to dopaminergic medication administration.

The mPDS is complementary to current standard PD measures. First, assessments can be performed frequently in real-world settings.[15] Second, the score provides an objective measure of PD symptom severity, not impacted by interrater variability.[16] Third, the mPDS, unlike current standard measures, objectively weighs activity features. The MDS-UPDRS part III is biased toward tremor-predominant disease,[1] with only 5 of 33 items assessing gait or balance. In contrast, 56.6% of mPDS items are derived from gait or balance activities. Finally, unlike current standard measures, which can take years and significant resources to develop,[1] the mPDS was generated quickly from a relatively small number of participants using automated techniques that can account for noise in data collected from multiple smartphone sensors and self-reported medication administration.[17] Combining smartphone data with the machine-learning methods outlined here may also provide opportunities for developing objective severity measures in other neurological conditions.

Figure. Mobile Parkinson Disease Score Assessment During 6 Months

A Intraday severity fluctuations captured for 1 participant

B Assessment frequency of mPDS vs MDS-UPDRS part III in 6 months

A and B, mPDS indicates the mobile Parkinson disease score. B, The total number of tests depicted is 152; MDS-UPDRS indicates the Movement Disorder Society Unified Parkinson's Disease Rating Scale.

Table 2. Correlation Matrix Between the Mobile Parkinson Disease Score (mPDS) and Standard Parkinson Disease Outcome Measures

Test[a]	mPDS, r (P Value)	MDS-UPDRS part III only, r (P Value)	MDS-UPDRS Total, r (P Value)	Timed Up and Go time, r (P Value)	Hoehn and Yahr stage, r
mPDS	1.00				
MDS-UPDRS part III–only subscore	0.88 (<.001)	1.00			
MDS-UPDRS total	0.81 (<.001)	0.82 (<.001)	1.00		
Timed Up and Go assessment	0.72 (.002)	0.74 (.002)	0.27 (.36)	1.00	
Hoehn and Yahr stage	0.91 (<.001)	0.96 (<.001)	0.80 (<.001)	0.70 (.003)	1.00

Abbreviations: mPDS, Mobile Parkinson disease score; MDS-UPDRS, Movement Disorder Society Unified Parkinson Disease's Rating Scale.

[a] These findings are based on 16 cross-sectional points that met the criteria for analysis.

Limitations

This study has several limitations. Participants were generally white, college-educated, people who owned Android smartphones and thus were not representative of the broader PD population. Only 51.6% of those who downloaded the application met criteria for inclusion in the development cohort. Additionally, the clinic cohort included only 7 assessments to evaluate the responsiveness of the mPDS to dopaminergic therapy administration, and only 16 smartphone and in-person assessment pairs met criteria for the correlation analysis. However, to our knowledge, this represents one of the largest longitudinal smartphone assessments of PD.

Conclusions

Further validation of the mPDS in a larger sample with patient-relevant anchors is needed. New iterations of the application for Android and iOS smartphones will expand participation and include additional features and functionality that could provide new insights into PD.

Author Affiliations: Department of Computer Science, Johns Hopkins University, Baltimore, Maryland (Zhan, Mohan, Terzis, Saria); Department of Neurology, University of Rochester Medical Center, Rochester, New York (Tarolli, Schneider, Adams, Dorsey); Center for Health + Technology, University of Rochester Medical Center, Rochester, New York (Tarolli, Adams, Sharma, Elson, Spear, Glidden, Dorsey); Department of Mathematics, Aston University, Birmingham, England (Little); Armstrong Institute for Patient Safety and Quality, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland (Saria); Department of Health Policy and Management, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland (Saria).

REFERENCES

1. Goetz CG, Tilley BC, Shaftman SR, et al; Movement Disorder Society UPDRS Revision Task Force. 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(15): 2129-2170.

2. 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(5):582-588.

3. Beal MF, Oakes D, Shoulson I, et al; Parkinson Study Group QE3 Investigators. A randomized clinical trial of high-dosage coenzyme Q10 in early Parkinson disease: no evidence of benefit. *JAMA Neurol*. 2014;71(5):543-552.

4. Arora S, Venkataraman V, Zhan A, et al. Detecting and monitoring the symptoms of Parkinson's disease using smartphones: a pilot study. *Parkinsonism Relat Disord*. 2015;21(6): 650-653.

5. Espay AJ, Bonato P, Nahab FB, et al; Movement Disorders Society Task Force on Technology. Technology in Parkinson's disease: challenges and opportunities. *Mov Disord*. 2016;31(9):1272-1282.

6. Artusi CA, Mishra M, Latimer P, et al. Integration of technology-based outcome measures in clinical trials of Parkinson and other neurodegenerative diseases. *Parkinsonism Relat Disord*. 2018;46(suppl 1):S53-S56.

7. Ellis RJ, Ng YS, Zhu S, et al. A validated smartphone-based assessment of gait and gait variability in Parkinson's disease. *PLoS One*. 2015;10 (10):e0141694.

8. Little MA. Parkinson's voice initiative. http://www.parkinsonsvoice.org/. Accessed August 30, 2017.

9. Lang AE, Eberly S, Goetz CG, et al. Movement Disorder Society Unified Parkinson Disease Rating Scale experiences in daily living: Longitudinal changes and correlation with other assessments. *Mov Disord*. 2013;28(14):1980-1986.

10. Hoehn MM, Yahr MD. Parkinsonism: onset, progression, and mortality. *Neurology*. 1967;17(5): 427-427.

11. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc*. 1991;39(2):142-148.

12. Zhan A, Little MA, Harris DA, et al. High frequency remote monitoring of Parkinson's disease via smartphone: platform overview and medication response detection; arXiv preprint. https://arxiv.org/abs/1601.00960. Published 2016. Accessed March 9, 2018.

13. Kirill D, Saria S. Learning (predictive) risk scores in the presence of censoring due to interventions. *Mach Learn*. 2016;102(3):323-348.

14. Hernández-González J, Inza I, Lozano JA. Weak supervision and other non-standard classification problems: a taxonomy. *Pattern Recognit Lett*. 2016; 69(suppl C):49-55.

15. Robles-García V, Corral-Bergantiños Y, Espinosa N, et al. Spatiotemporal gait patterns during overt and covert evaluation in patients with Parkinson's disease and healthy subjects: is there a Hawthorne effect? *J Appl Biomech*. 2015;31(3):189-194.

16. Martínez-Martín P, Rodríguez-Blázquez C, Mario Alvarez, et al. Parkinson's disease severity levels and MDS-Unified Parkinson's Disease Rating Scale. *Parkinsonism Relat Disord*. 2015;21(1):50-54.

17. Antonini A, Martinez-Martin P, Chaudhuri RK, et al. Wearing-off scales in Parkinson's disease: critique and recommendations. *Mov Disord*. 2011; 26(12):2169-2175.