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.

JAMA Neurol. doi:10.1001/jamaneurol.2018.0809
Published online March 26, 2018.

© 2018 American Medical Association. All rights reserved.
Current Parkinson disease (PD) measures are subjective and rater-dependent and require in-clinic assessments.\textsuperscript{1,2} As a result, clinical trials using these measures are long, expensive, and can generate false positives or negatives.\textsuperscript{2,3} Many motor symptoms of PD are well-suited to objective measurement by smartphones.\textsuperscript{4-6} Smartphone assessment has been evaluated in PD, but most studies focus on a specific feature (eg, gait), rather than overall symptom burden.\textsuperscript{5,7} 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),\textsuperscript{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.\textsuperscript{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),\textsuperscript{9} the Hoehn and Yahr stage,\textsuperscript{10} and the Timed Up and Go assessment\textsuperscript{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 intercept interval from the finger-tapping activity).\textsuperscript{12} Rather than replicating an existing PD score using regression, we used a rank-based machine-learning algorithm, disease severity score learning (DSSL),\textsuperscript{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\textsuperscript{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 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 capture 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. \textit{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 (\(\rho\)) 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 (\(\alpha = .05\)). 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.

Key Points

\textbf{Question} Can a smartphone be used to quantify Parkinson disease motor symptom severity?

\textbf{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.

\textbf{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.
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 MDS-UPDRS part III–only subscore decreased by a mean (SD) of 10.4 (4.6) in response to dopaminergic therapy.

In addition, 7 off-medication vs on-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. Second, the score provides an objective measure of PD symptom severity, not impacted by interrater variability. Third, the mPDS, unlike current standard measures, objectively weighs activity features. The MDS-UPDRS part III is biased toward tremor-dominant disease, 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, 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. Combining smartphone data with the machine-learning methods outlined here may also provide opportunities for developing objective severity measures in other neurological conditions.
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.
Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Presentation: This paper was presented at the NVIDIA GPU Technology Conference; March 26, 2018; San Jose, California.

REFERENCES