Supplementary Online Content

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

eMethods. A comprehensive discussion of the learning algorithm used to construct the mobile Parkinson disease score (mPDS).
eEquation. Linear disease severity score objective.
eTable 1. Video instructions of the five activities used in this study.
eTable 2. The top fifteen features determined by ranking mPDS' absolute feature weights.
eFigure. Pictures of the HopkinsPD smartphone application used in the study.

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods: A comprehensive discussion of the learning algorithm used to construct the mobile Parkinson disease score (mPDS)

One general kind of approach for creating a severity score algorithm is based on *supervised learning*: here, experts evaluate the participants at multiple time points to provide the clinical, “gold-standard” score at each time point (e.g., MDS-UPDRS score). Based on these evaluations, a regression function is estimated that maps features (algorithms such as sensor data variability, complexity and summarized frequency information) derived from the smartphone sensor data collected during the smartphone activities into a continuous or discrete-valued score. The key challenge of using such an approach is that it relies heavily on obtaining a large number of gold-standard clinical evaluations, which are very expensive and time-consuming to collect.

Instead, we used a rank based machine learning algorithm—*disease severity score learning* (DSSL)[1]—to create the mobile Parkinson disease score (mPDS). In order to estimate a score from feature data, DSSL uses *weak supervision*[2] where the resulting labels may have an associated error rate. For example, to estimate mPDS parameters, DSSL exploits example pairs of times that are rank ordered in severity such that the severity of symptoms at time t_i is less than that at time t_j. Using the data collected in this study, such example pairs were easily obtained: for an individual responding to medication, the severity of symptoms at a time right before medication administration is assumed to be higher than that an hour after taking their medications.

Given many such pairs, DSSL estimates a score by optimizing the objective shown in **Equation 1**. Here, $\boldsymbol{x}$ represents a feature vector derived from the sensor data recorded during activities collected using HopkinsPD at a given time. A total of 435 features were computed from the five smartphone-enabled test activities. For example, we computed 126 features from the gait and balance tests each to capture changes in body motion, including the mean, median, standard deviation, range, entropy, and dominant frequency from the tri-axial acceleration time-series. 151 features were computed from the tapping test screen touch events, to quantify attributes such as finger tapping speed (e.g., total number of taps within a given period of time), precision of tapping (e.g., range of tap positions normalized by smartphone screen size), and rhythm and inter-tap interval. Detailed descriptions of the features were previously published elsewhere.[3] Each i, j is a numerical index associated with two distinct timestamps, at times t_i and t_j, at which activities were conducted. Each p, q represents two distinct patient indices. The vector $\boldsymbol{w}$ is a vector of weights estimated by DSSL. To compute the mPDS on a new patient at a given time t given a recording of their activities at that time and the resulting feature vector $\boldsymbol{x}$ computed from the sensor data collected during these activities, the linear projections $\boldsymbol{w} \cdot \boldsymbol{x}$ are computed. These linear projections are raw and unscaled. To ease interpretability in a clinical setting, the mPDS is scaled between 0 and 100, where values close to 0 reflect low severity while those close to 100 reflect high severity.

The set O is the set of all available pairs of tuples $(<\boldsymbol{x}_i^p, t_i^p>, <\boldsymbol{x}_j^q, t_j^q>)$ that are ordered by severity; from the development cohort, such pairs are computed automatically based on the activities performed at times right before medication administration and those from the hour after. Severity is assumed to be lower post medication administration. In the second term in Equation 1, L_h is the Huber loss function. This second term in the objective encourages DSSL to estimate a score that satisfies the severity ordering prescribed by the tuples in set O. We had a total of 3074 such pairs available in the development cohort.

The set S, denoted by pairs of tuples $(<\boldsymbol{x}_i^p, t_i^p>, <\boldsymbol{x}^p_{i+1}, t^p_{i+1}>)$, are obtained based on tests taken at consecutive times within a few hours of each other but without medication administration during the interim period. The third term in Equation 1 encourages temporal smoothness for the pairs specified in set S. The coefficients O and S are DSSL regularization parameters and control the relative degree of emphasis on the smoothness between consecutive pairs in the third term of the objective versus maximizing the difference in severity for pairs specified in the second term. These were set using 10-fold cross-validation on the development cohort.

References:

1. Kirill D, Saria S. Learning (predictive) risk scores in the presence of censoring due to interventions. *Machine Learning.* 2016;102(3):323-348.
2. Hernández-González J, Inza I, Lozano JA. Weak supervision and other non-standard classification problems: A taxonomy. *Pattern Recognition Letters.* 2016;69(Supplement C):49-55.
3. 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.* 2016;1601.00960, 2016.

eEquation 1: Linear disease severity score objective

$$\begin{aligned}\min_{\mathbf{w}} \quad & \frac{1}{2}\|\mathbf{w}\|^2 \\ & + \frac{\lambda_O}{|O|} \sum_{(<\mathbf{x}_i^p, t_i^p>, <\mathbf{x}_j^q, t_j^q>) \in O} L_h(1 - \mathbf{w}^\mathsf{T}(\mathbf{x}_i^p - \mathbf{x}_j^q)) \\ & \quad + \frac{\lambda_S}{|S|} \sum_{(<\mathbf{x}_i^p, t_i^p>, <\mathbf{x}_{i+1}^p, t_{i+1}^p>) \in S} \left[\frac{\mathbf{w}^\mathsf{T}(\mathbf{x}_{i+1}^p - \mathbf{x}_i^p)}{t_{i+1}^p - t_i^p} \right]^2\end{aligned}$$

eTable 1: Video instructions of the five activities used in this study.
(1) voice test to say “aaah” for twenty seconds; (2) finger tapping test to alternately tap one’s index and middle finger in a regular rhythm for twenty seconds; (3) reaction time test to press and hold an on-screen button as soon as it appeared and to release it as soon as it disappeared; (4) gait test to walk twenty yards forward, turn around, and return to the starting position; and (5) balance test to stand upright unaided for twenty seconds.

Activity	URL
Voice test	https://youtu.be/BeUFgljiuJI
Balance test	https://youtu.be/xwJsLGdlhsE
Gait test	https://youtu.be/2BidIYn1Nrg
Tapping test	https://youtu.be/tJLqvKHn2XQ
Reaction test	https://youtu.be/Brz2yZp_O7M

eTable 2. The top fifteen features determined by ranking mPDS' absolute feature weights.

Smartphone test protocol	Feature description
Finger tapping	Mean vertical tapping position scaled according to smartphone screen size
Balance	Mean acceleration in the direction of motion when the individual is walking
Gait	Entropy of the acceleration in the direction of motion when the individual is walking
Finger tapping	Mean vertical tapping position on the left button scaled according to smartphone screen size
Finger tapping	Mean square energy of the vertical tapping position scaled according to smartphone screen size
Balance	Mean acceleration in the direction of the gravitational acceleration vector
Gait	Entropy of the acceleration in the side direction (perpendicular to the walking direction)
Finger tapping	Mean horizontal tapping position on the left button scaled according to smartphone screen size
Finger tapping	Mean squared energy of the vertical tapping position on the right button scaled according to smartphone screen size
Gait	Entropy of acceleration in the direction of the gravitational acceleration vector
Finger tapping	Mean horizontal tapping position on the left button scaled according to smartphone screen size
Finger tapping	Median vertical tapping position on the left button scaled according to smartphone screen size
Finger tapping	Mean squared energy of the vertical tapping position on the left button scaled according to smartphone screen size
Balance	Entropy of acceleration in the inclination direction in the spherical coordinate system
Voice	Mean voice amplitude over all 0.5 second frames with voiced signal

eFigure 1: Pictures of the HopkinsPD smartphone application used in the study

a. Gait test	b. Tapping test	c. Voice test