The sound of Parkinson’s disease: a model of audible bradykinesia

Debbie de Graaf (debbie.degraaf@radboudumc.nl)
Radboud University Medical Centre; Donders institute for Brain, Cognition and Behaviour, Department of Neurology, Parkinson Centre Nijmegen (ParC) Nijmegen

Rui Araújo
Department of Neurology, Centro Hospitalar Universitário São Joao, Department of Clinical Neurosciences and Mental Health, Faculty of Medicine, University of Porto, Porto, Portugal

Madou Derksen
Dutch Hospital Data (DHD), Utrecht, Netherlands

Koos Zwinderman
Academic Medical Center, Department of Cardiology, P.O. Box 22660, 1100 DD Amsterdam, The Netherlands

Nienke M de Vries
Radboud university medical center, Donders Institute for Brain, Cognition and Behavior, Department of Neurology, Nijmegen, The Netherlands

Joanna Inthout
Radboud University Medical Center, Department for Health Evidence Nijmegen, the Netherlands

Bastiaan Bloem
bas.bloem@radboudumc.nl  https://orcid.org/0000-0002-6371-3337

Article

Keywords:

Posted Date: September 21st, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3255275/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

Additional Declarations: (Not answered)
The sound of Parkinson’s disease: a model of audible bradykinesia

Debbie de Graaf, MSc¹, Rui Araújo, MD, PhD², Madou Derksen, MSc³, Koos Zwinderman, PhD⁴, Nienke M de Vries, PhD¹, Joanna IntHout, PhD⁵, Bastiaan R Bloem, MD, PhD¹

¹ Radboud University Medical Center; Donders Institute for Brain, Cognition and Behavior; Department of Neurology; Center of Expertise for Parkinson & Movement Disorders; Nijmegen, The Netherlands
² Department of Neurology, Centro Hospitalar Universitário São João, Department of Clinical Neurosciences and Mental Health, Faculty of Medicine, University of Porto, Porto, Portugal
³ Dutch Hospital Data (DHD), Utrecht, Netherlands
⁴ Academic Medical Center, Department of Cardiology, P.O. Box 22660, 1100 DD Amsterdam, The Netherlands
⁵ Radboud University Medical Center, Department for Health Evidence Nijmegen, the Netherlands

Corresponding author
Debbie de Graaf, MSc
Radboud University Medical Centre
Center of Expertise for Parkinson & Movement Disorders, R914
Postbus 9101
6500 HB Nijmegen
The Netherlands
Debbie.degraaf@radboudumc.nl

Conflict of interest: Authors declare no conflicts of interest

Funding statement: The Radboudumc Center of Expertise for Parkinson & Movement Disorders was supported by a center of excellence grant of the Parkinson’s Foundation. This publication is part of the project ROBUST: Trustworthy AI-based Systems for Sustainable Growth with project number KIC3.LTP.20.006, which is (partly) financed by the Dutch Research Council (NWO), Verily Life Sciences LLC, and the Dutch Ministry of Economic Affairs and Climate Policy (EZK) under the program LTP KIC 2020-2023. All content represents the opinion of the authors, which is not necessarily shared or endorsed by their respective employers and/or sponsors. The Personalized Parkinson Project is co-funded by Verily Life Sciences LLC, Radboud university medical center and by the PPP Allowance made available by Health~Holland, Top Sector Life Sciences & Health, to stimulate public-private partnerships.

Word count: 3720
Number of figures: 2
Number of tables: 3
Number of supplementary files: 9
Abstract

Background: Evaluation of bradykinesia is based on four motor tasks from the MDS-UPDRS. Visually scoring these motor tasks is subjective, resulting in significant interrater variability. Recent observations suggest that it may be easier to hear the characteristic features of bradykinesia, such as the decrement in amplitude or force of repetitive movements.

Objectives: To evaluate whether audio signals derived during four MDS-UPDRS tasks can be used to detect and grade bradykinesia, using two machine learning models.

Methods: 54 patients with Parkinson’s disease and 28 healthy controls were filmed while executing the bradykinesia motor tasks. Several features were extracted from the audio signal, including number of taps, speed, amplitude, decrement and freezes. For each motor task, two supervised machine learning models were trained, Logistic Regression (LR) and Support Vector Machine (SVM).

Results: Both classifiers were able to separate patients from controls reasonably well for the leg agility task, area under the receiver operating characteristic curve (AUC): 0.92 (95%CI: 0.78-0.99) for LR and 0.93 (0.81-1.00) for SVM. Also, models were able to differentiate less severe bradykinesia from severe bradykinesia, particularly for the pronation-supination motor task, with AUC: 0.90 (0.62-1.00) for LR and 0.82 (0.45-0.97) for SVM.

Conclusions: This audio-based approach discriminates PD from healthy controls with moderate-high accuracy and separated individuals with less severe bradykinesia from those with severe bradykinesia. Sound analysis may contribute to the identification and monitoring of bradykinesia.
Introduction

Parkinson’s disease (PD) is the fastest growing neurological disorder in the world (1, 2). The phenotype includes both motor and non-motor functions. Bradykinesia is a cardinal disease feature. Established diagnostic criteria define bradykinesia as “slowness to initiate a movement with a progressive reduction in speed and amplitude” (3). Assessment of bradykinesia is paramount for the clinical diagnosis and for monitoring disease progression in patients with PD.

The current gold standard for assessing the severity of PD symptoms and signs is the Movement Disorders Society Unified Parkinson’s Disease Rating Scale (MDS-UPDRS). Evaluating bradykinesia within the MDS-UPDRS subscale III involves a trained clinician observing and visually scoring the patient’s performance during four motor tasks: 1) finger tapping; 2) pronation-supination; 3) toe tapping; and 4) leg agility. Each task is assessed according to speed, amplitude and rhythmicity and is scored from 0 (normal motor activity) to 4 (severe motor impairment) (4). Although training and expertise is required for accurate MDS-UPDRS scoring, it is known that the interpretation of the motor examination is associated with large interrater variability (5), presumably because of the subjective nature of the scoring system, and the relatively crude steps in the five-step ordinal rating.

In an effort to overcome this issue, several automated quantitative tools have been developed to detect and score bradykinesia more objectively (6-11). Although their nature and complexity vary considerably, these approaches rely mostly on traditional aspects of bradykinesia, e.g. typing speed, accuracy and speed variation. None of the previous studies using automated quantitative technologies focused primarily on the audible signs originating from the motor tasks. Also, a decrement in either amplitude, speed, or both, has not always been analysed in prior work, even though this is the characteristic hallmark of parkinsonian bradykinesia.

Previously, a case study of a professional violinist who developed PD ignited the possibility of measuring an audible dimension of bradykinesia (12). This patient reported progressive difficulties playing the violin. Clinical examination revealed bradykinesia and tremor compatible with PD. Sound analysis of his finger tapping and pronation-supination movements showed a clear decrement in tap intensity and frequency of the sequential taps over time, which correlated with the patient’s symptoms and motor examination. These observations motivated the concept of “audible bradykinesia”, the advantage being that this would offer an objective and quantitative measure of bradykinesia.

Here, we aim to create a model that could capture the sound characteristics of the motor tasks of the MDS-UPDRS. We hypothesized that sound analysis, by virtue of its quantitative nature, would add valuable, objective and readily available information to bradykinesia assessments. Specifically, we sought to address this issue by creating machine learning models using sound features extracted from the four traditional bradykinesia motor tasks. Our aim was to create models that could distinguish PD patients from healthy controls, and that would also distinguish patients with less
severe bradykinesia from those with more severe bradykinesia during a phase when PD signs were most pronounced.

Patients and methods

Patient cohort

We included 54 patients with PD and 28 healthy controls. Participants from an ongoing study (13) were asked to participate in this study, their partners were asked to participate as healthy controls. Motor status was filmed using a standardized procedure, in a practically defined OFF state (more than 12 hours after intake of the last dose of dopaminergic medication). Patients were also tested and filmed during a regular ON state, but these data were not used for the present analysis. Exclusion criteria for healthy controls were: 1) age younger than 50 years; and 2) physical limitations that could influence the motor examination. Clinical details of all participants are shown in Table 1. All participants provided informed consent. Study procedures were approved by the Medical Ethical Committee of Nijmegen (CMO: NL59694.091.16).

Bradykinesia assessment (MDS-UPDRS)

Patients performed four bradykinesia tasks of the MDS-UPDRS section III: finger tapping; pronation-supination; toe tapping and; leg agility, while sitting in a non-creaking chair. Performance was videotaped with a standard video camera (Panasonic HD Camcorder HC-V180, 720p/25p) that was placed on a tripod in a standard position in the examination room, at about 1.5 meters distance of the participants. All participants were asked to remove rings, bracelets, watches and any other devices that could produce interfering sounds. Participants were instructed not to speak to avoid interfering with sound collection and to perform all tasks at a steady and fast pace. For finger tapping, participants were instructed to raise their index finger as high as possible and tap the finger on the table. For pronation-supination, participants were instructed to tap the palm and back of their hand repeatedly on the table. For toe tapping, they were asked to raise their toes as high as possible and tap as quickly as possible. For leg agility, they were instructed to raise their legs as high as possible and stomp on the ground. All motor tasks were performed for 30 seconds. Only the most affected side of patients with PD and the dominant side of the healthy controls was used in the analysis. Performance on each video was visually rated by two assessors with extensive expertise in both PD and using the MDS-UPDRS. If disagreement occurred in grading the MDS-UPDRS, a third independent assessor scored the videos whereafter the three assessors reached consensus together.

Pre-processing audio data

The video files were loaded into Matlab 2022a (14), and the audio and sampling frequency of the videos were extracted. The files were filtered for background noise using a 6th order Butterworth high pass filter, and a Fast Fourier Transformation
(FFT) was performed for each motor task separately to find the cut-off frequency (Appendix 1). Lastly, the audio files were smoothed with a moving average window with a filter length of 91. After filtering the audio files for background noise, only the absolute values of the audio datapoints were used for further analysis.

**Feature extraction**

The audio signals from the most affected side of patients (left hand n=25; right hand n=29) and from the dominant side of controls (left hand n=2; right hand n=26) were analyzed. All equations and formulas regarding the audio features are referred to in Appendix 2. All features extracted from the audio signal are listed in Appendix 3. First, a peak-finder algorithm was applied to find peak amplitudes and locations. The threshold to find these peaks was calculated for each audio file separately. These peaks represented the taps, claps and stomps in the video. Afterwards, the following four categories were defined and their features were measured: 1) amplitude, volume which is defined by tap sound intensity; 2) speed, frequency which is defined by the number of taps/second; 3) decrement in amplitude, defined as a progressively lower amplitude throughout the bradykinesia task (see equations 7 and 8 in Appendix 2); 4) halts and freezes, defined by above average inter-tap intervals (see equation 9 in Appendix 2). We expected features related to speed and amplitude to be lower in patients with PD compared to healthy controls, and the decrement to be more pronounced in patients. Additionally, features related to hesitations and freezes were expected to be more prevalent in patients with more severe bradykinesia. In the LR model, when a feature was selected in more than 40% of the 100 repeats, it was considered an important predicting feature for that specific motor task.

**Classification**

After data acquisition was accomplished, pre-processing of the audio data was performed and the features were extracted. For each motor task two specialized predictive classification models with the selected features were created to train two supervised machine learning models, that is, Logistic Regression (LR) and Support Vector Machine (SVM). The LR model used the LASSO feature-selection method to obtain the most significant features, and the SVM model used all features.

The experiments aimed to see whether sound features of the MDS-UPDRS bradykinesia motor tasks distinguished patients with PD from healthy controls. Additionally, we evaluated whether sound features could differentiate less severe bradykinesia from more severe bradykinesia. The classification task was thus composed of two experiments. In experiment one we compared PD patients (labelled 1) with healthy controls (labelled 0). In experiment two we compared PD patients with more severe bradykinesia, MDS-UPDRS 3-4 (labelled 1), with PD patients with less severe bradykinesia, MDS-UPDRS 1-2 (labelled 0). In total, there were 2 experiments x 4 motor tasks x 2 models = 16 models.

**Data analysis**
Hundred times, the dataset was randomly divided into a training and a test set, where the training contained 70% of cases and 70% of controls, while the test set contained the remaining 30% of cases and controls. Each iteration, logistic regression using LASSO for feature selection was conducted on the 12 sound features and age in the training data, with 3 fold cross validation in experiment 1 and with leave-one-out cross validation in experiment 2 (because of less participants in experiment 2), to determine the lambda and corresponding coefficients of the model with minimum deviance, and to calculate the corresponding AUC. The number of times a feature was selected was counted and descriptive statistics of the coefficients of the selected features were provided. On the same dataset, an SVM model with linear kernel was fitted, using all sound features and age. For both LR and SVM, models containing only an intercept and no other variables were fitted to set a reference value for the reliability of the models. The resulting LR and SVM models were applied on the test data, to predict and plot the probability of being a PD patient (experiment 1) or of being a grade 3 or 4 PD patient (experiment 2), and the AUC in the test data was assessed. After training and testing the models 100 times, the mean AUC values with corresponding 95% ranges, reflecting 95% confidence intervals, were calculated to assess classification performance. All the intercept-only models have an AUC of 0.5. Also Brier scores (mean squared errors) were calculated on the test data and presented in histograms, together with the Brier score of the intercept-only model, to assess the accuracy of the predictions; lower Brier scores reflect higher model accuracy (15).

The analyses were conducted with the following packages: R.matlab (16), glmnet (17) for logistic regression, e1071 (18) for SVM, pROC (19) for the AUC, and FSA (20) for summary statistics, in the statistical software R version 4.1.3 (21) with RStudio (22).

**Code availability**
The underlying code for this study [and training/validation datasets] is not publicly available but may be made available to qualified researchers on reasonable request from the corresponding author.

**Results**

Two independent assessors scored the four motor tasks for all 54 participants, leading to 216 scores. In total, the assessors reached agreement for 204 tasks, resulting in an inter-rater reliability score of 94.5%. A third independent assessor scored the 12 remaining tasks where no consensus was reached.

Group size for experiment 1 and 2 is shown in Table 2. Due to poor sound quality, we excluded six videos for the finger tapping task, one video for the pronation-supination task, two videos for the toe tapping task, and one video for the leg agility task.

**Experiment 1: Differentiating patients with PD from healthy controls**
The mean value of each feature and corresponding minimum and maximum values, how many times a particular feature was selected over the 100 repeats and the mean coefficient (\(\beta\)) for each time a feature was selected for each motor task in experiment 1 are given in Appendix 4. The features with the best predictive value differed between the tasks.

The predicted values of the estimated LR and SVM models are shown in Figure 1. The red line is the “best” threshold averaged over 100 repeats. As seen in Figure 1, audible signals from the leg agility motor task differentiated patients from controls for both LR and SVM models. The mean AUC values averaged over the 100 repeats for each motor task and for each model are given in Table 3 and Appendix 6. The leg agility had the highest AUC values in both the LR (0.92 (0.78-0.99)) and SVM (0.93 (0.81-1.00)) model, and also the lowest Brier scores (Appendix 7). The results are clearly eyeballing the graphs in Appendix 7. For this experiment, only the leg agility scores are better than the baseline intercept model. The other tasks perform at chance.

For the leg agility task, the number of freezes (\(F_n\)), the mean change of the tapping amplitude (\(\Delta TA\)) and mean change of the number of taps (\(\Delta Tn\)) were most frequently selected. The variation coefficient in tapping speed (\(VTs\)) was the most selected feature for the toe tapping motor task together with the mean change of the tapping amplitude (\(\Delta TA\)) and the total number of taps (\(Tn\)).

For the finger tapping motor task, the most distinguishing features between patients and controls were the variation coefficient of the tapping speed (\(VTs\)) and the mean change of tapping amplitude (\(\Delta TA\)). For pronation-supination, the average tapping amplitude (\(TA\)), the total number of taps (\(TA\)) and the mean change of maximum tapping amplitude (\(\Delta MTA\)) were the most discriminating features.

Experiment 2: Differentiating patients with less severe bradykinesia from those with more severe bradykinesia.

The mean value of each feature and corresponding minimum and maximum values, how many times a particular feature was selected over the 100 repeats and the mean coefficient (\(\beta\)) for each time a feature was selected for each motor task in experiment 2, are given in Appendix 5.

The predicted values of the estimated LR and SVM models are shown in Figure 2. The mean AUC values averaged over the 100 repeats for each motor task and for each model are given in Table 3 and Appendix 8. The pronation supination task had the highest AUC values in both the LR (0.90 (0.62-1.00)) and the SVM (0.82 (0.45-0.97)) models. For the lower body motor tasks, toe tapping had the highest AUC values for both the LR (0.80 (0.58-0.97)) and SVM models (0.78 (0.57-0.95)). Brier scores are presented in Appendix 9. For this experiment, only the pronation/supination and toe tapping scores are better than the baseline intercept model. The other tasks perform at chance.
For the finger tapping task, the variation coefficient of the tapping speed (VTs) distinguished best between less severe and more severe bradykinetic patients. For the pronation-supination motor task, the standard deviation of tapping interval (TISD) and the variation coefficient of the tapping speed (VTs) were the most distinguishing features.

For the lower limb, regarding the toe tapping task, the average tapping amplitude (TA), the total number of taps (Tn) and the mean change of the number of taps (∆Tn) were the most relevant features. For the leg agility test, the variation coefficient in tapping speed (VTs) and the standard deviation of the tapping interval (TISD) were the most distinguishing features.

Discussion

We present a novel automated and quantitative tool for the objective measurement of bradykinesia based on audio signals. Motivated by previous findings (12), we performed an exploratory study on the concept of “audible bradykinesia”. Patients and controls performed the usual motor tasks for bradykinesia; sound features obtained during these tasks were subsequently extracted. We analyzed variations in tapping speed and intensity, and trained machine learning models to detect and grade bradykinesia. Previous studies mainly focused on speed and latency of the sequential taps (7-11). Our study may allow for a more in-depth, multimodal analysis of bradykinesia by also analyzing tap intensity and tap amplitude based on sound analysis. The elegance of our approach is that we study a commonly used set of clinical tasks that are part of the standard MDS-UPDRS test battery analyzed using very simple equipment. Sound analysis could facilitate its introduction as outcome parameter both in clinical trials and perhaps in daily clinical practice if this is confirmed in further work.

In experiment 1 (patients vs. controls), the leg agility test was the best test to discriminate patients from healthy controls. Of note, both speed and amplitude characteristics were selected as distinguishing features, underlining the relevance and usefulness of “listening” to the sound produced by MDS-UPDRS motor tasks.

In experiment 2 (MDS-UPDRS 1-2 vs. MDS-UPDRS 3-4), pronation-supination and toe tapping performed best in discriminating less severe PD patients from more severe PD patients for the upper body and lower body respectively. In this experiment, speed features were good discriminators. A comparable study used biomechanical features derived from wearable sensor data in different machine learning models (LR, SVM, and a neural network model) (23). This previous study had significant differences in methodology from the present experiment. The raw data used were obtained in PD patients tested in the ON-medication status, whereas we tested patients in a practically defined OFF state; their videos were significantly shorter (10 seconds) than ours (30 seconds); and different motor tasks were employed, such as opening and closing hands, rest tremor and heel tapping. Different methods likely
contribute to different results, and standardization procedures for future studies are recommended.

We also examined the diagnostic value of the various motor tasks. The toe tapping motor task was introduced in the revision of the original MDS-UPDRS scale aiming to capture subtle changes that the original MDS-UPDRS scale did not detect (24). Our results strengthen the thesis that toe tapping may be a particularly useful motor task when assessing bradykinesia severity in the lower limbs.

Our study is not without limitations. First, model performance and accuracy are below optimal values. We recognize that 60-70% of the area under the ROC curve statistics may curtail the application of this method in clinical practice as it stands, but our study provides important proof-of-concept data to a new and hitherto unexplored dimension of bradykinesia. Also, most patients had been diagnosed only 3-5 years ago, which may curtail generalization of the present results to either prodromal stages, when subtle slowing may arise in the final months leading up to the actual diagnosis, or to late-stage PD where repetitive movements may become difficult to perform. Our main reason was to establish the existence of an audible dimension of bradykinesia in persons with established PD according to the MDS-UPDRS clinical criteria. One potentially helpful outcome of the present work is that it may alert clinicians to the fact that bradykinesia can already be heard readily in daily practice, even without equipment. The clinicians in our center have actually begun to listen to bradykinesia, in addition to looking for the visible decrement in the quality of the movements. Perhaps its most immediate application would be in clinical trials, where there is a fairly pressing need for development of new objective outcomes that can measure the response to treatment, or the rate of disease progression (25). Future studies will help clarify the extent to which audible bradykinesia is useful in clinical practice for the prodromal and late phase of PD.

Third, we did not include patients in ON status. The main objective was to evaluate whether audio signals derived during four MDS-UPDRS tasks can be used to detect and grade bradykinesia, using two machine learning models. Hence, we explored the phase when PD signs were most pronounced. We acknowledge that the absence of patients in ON status is a limitation, and further work is necessary to see the effect of medication in the audible aspects of bradykinesia. Additionally, we did not include hand opening and closing motor tasks in our analysis because we anticipated that this test would generate a lower sound intensity compared with the other motor tasks. Future studies could now examine whether other tests for bradykinesia in the hands yield a similarly measurable audible signal. The non-dominant hand was not evaluated for the same reason: we hypothesized it would create a less robust audio signal, lowering the discriminatory capacity of the models. This can be extrapolated from previous work where finger tapping with the dominant finger is faster than the non-dominant finger (26). However, sometimes the non-dominant hand was evaluated if it was the most affected side of the PD patients.

The fact that we applied several strategies to minimize ambient sound could be a confounding factor, as this is not the case in real-life conditions. However, the clinical offices in our center are well isolated soundwise, in order to respect the
privacy of the medical consultation, and we can well imagine that this would be largely similar in most clinical practices around the world. In daily clinical practice, the data collection procedure only required participants not to speak during the motor tasks and to remove any bracelets or rings that could potentially interfere with the sound analysis. In sporadic occasions, we had to substitute a squeaky chair. We believe that, as far as “laboratory” conditions go, it would be difficult to design a more ecological approach than the one we used. We used a video camera in a standardized position in order to minimize confounders; however, it is likely that audio features captured by more “conventional” means such as smartphones or even remote devices could afford the same results. This issue, however, would have to be addressed in future studies.

Overall, the sample size of our study was relatively small, which makes – in combination with a relatively high number of predictors - the risk of overfitting realistic. Therefore, we evaluated the model’s performance on independent test data sets. However, a larger sample may increase the reliability of the models’ performance and the accuracy of the results, and result in more stable sets of selected features. Lastly, model performance and accuracy might be improved by using other audible features; we selected those that seemed most important and best reflected the core aspects of bradykinesia, focusing on speed, amplitude and their variation over time. Given the exploratory nature of the present study, it is unsurprising that different features were selected for each motor task; future studies with larger samples should allow researchers to identify which models and features are the most appropriate to assess bradykinesia.

In conclusion, we were able to identify and grade the severity of bradykinesia using data from sound analysis generated by established MDS-UPDRS motor tasks with moderate to high AUC value. Even though both machine learning models were effective at this, we cannot indicate whether one model is superior to the other. Future research will be necessary in order to effectively build a tool to detect bradykinesia using audio signals, especially in early-stage PD, where some aspects of bradykinesia may be better heard than seen.

Acknowledgements

The Radboudumc Center of Expertise for Parkinson & Movement Disorders was supported by a center of excellence grant of the Parkinson’s Foundation. This publication is part of the project ROBUST: Trustworthy AI-based Systems for Sustainable Growth with project number KICH3.LTP.20.006, which is (partly) financed by the Dutch Research Council (NWO), Verily Life Sciences LLC, and the Dutch Ministry of Economic Affairs and Climate Policy (EZK) under the program LTP KIC 2020-2023. All content represents the opinion of the authors, which is not necessarily shared or endorsed by their respective employers and/or sponsors. The Personalized Parkinson Project is co-funded by Verily Life Sciences LLC, Radboud university medical center and by the PPP Allowance made available by Health~Holland, Top Sector Life Sciences & Health, to stimulate public-private partnerships. Authors thank Veerle van Laarhoven, Janna Maas, Sabine van Zundert and Marjan Meinders for their contributions.
Financial disclosure

DdG, MD, KZ and JiH have no financial disclosures to report. RA has received fees from serving on a scientific advisory board for Bial, and has received speaking fees for speaking at conferences from Roche and Bial. NMdV reports grants from The Netherlands Organisation for Health Research and Development (ZonMw) and The Michael J Fox Foundation. BRB currently serves as Editor in Chief for Journal of Parkinson's disease; serves on the editorial board of Practical Neurology and Digital Biomarkers; has received honoraria from serving on the scientific advisory board for AbbVie, Biogen, and UCB; has received fees for speaking at conferences from AbbVie, Zambon, Roche, GE Healthcare, and Bial; and has received research support from the Netherlands Organization for Scientific Research, The Michael J. Fox Foundation, UCB, AbbVie, the Stichting Parkinson Fonds, the Hersenstichting Nederland, the Parkinson's Foundation, Verily Life Sciences, Horizon 2020, the Topsector Life Sciences and Health, the Gatsby Foundation, and the Parkinson Vereniging.

Competing interest

Data availability
The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Author roles

1. Research project: A. Conception, B. Organization, C. Execution;
3. Manuscript Preparation: A. Writing of the first draft, B. Review and Critique;

DdG: 1B, 1C, 3A
RA: 1A, 1B, 3A, 3B
MD: 1A, 1B, 1C, 2B
KZ: 1A, 1B, 2A, 2B, 2C
NdV: 1B, 2C, 3A, 3B
JiH: 2B, 2C, 3B
BRB: 1A, 1B, 2C, 3A, 3B

References


S+ to analyze and compare ROC curves. BMC Bioinformatics, 12, p. 77. DOI: 10.1186/1471-2105-12-77.


**Tables**

**Table 1:** Participant characteristics of PD patients (n=54) and healthy controls (n=28)

<table>
<thead>
<tr>
<th></th>
<th>PD Patients</th>
<th>Healthy Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex, man (n%)</strong></td>
<td>37 (69)</td>
<td>6 (21)</td>
</tr>
<tr>
<td><strong>Age, years (mean(SD))</strong></td>
<td>63 (10)</td>
<td>65 (7)</td>
</tr>
<tr>
<td><strong>Range:</strong> 33-86</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Disease duration, years</strong></td>
<td>5.1 (1.5)</td>
<td></td>
</tr>
<tr>
<td><strong>Range:</strong> 2-9</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dominant hand (n%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>11 (20)</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Right</td>
<td>43 (80)</td>
<td>26 (93)</td>
</tr>
<tr>
<td><strong>Most affected (MA) side</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>25 (46)</td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>29 (54)</td>
<td></td>
</tr>
<tr>
<td><strong>MDS-UPDRS III OFF MA (n%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finger Tapping</td>
<td>1 (2)</td>
<td></td>
</tr>
<tr>
<td>Pronation Supination</td>
<td>1 (2)</td>
<td></td>
</tr>
<tr>
<td>Toe Tapping</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Leg Agility</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>6 (11)</td>
<td>10 (19)</td>
</tr>
<tr>
<td>1</td>
<td>29 (54)</td>
<td>23 (43)</td>
</tr>
<tr>
<td>2</td>
<td>15 (28)</td>
<td>19 (35)</td>
</tr>
<tr>
<td>3</td>
<td>3 (6)</td>
<td>2 (4)</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 2: Group size (n) for experiment 1 and 2 for each motor task.

<table>
<thead>
<tr>
<th></th>
<th>Experiment 1</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PD Patient</td>
<td>Healthy</td>
<td>Less Severe</td>
<td>Severe</td>
<td></td>
</tr>
<tr>
<td>Finger Tapping</td>
<td>48</td>
<td>28</td>
<td>33</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Pronation Supination</td>
<td>53</td>
<td>28</td>
<td>37</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Toe Tapping</td>
<td>52</td>
<td>28</td>
<td>32</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Leg Agility</td>
<td>53</td>
<td>28</td>
<td>43</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

### Table 3: AUC values with confidence intervals (CI) for each motor task in experiments 1 and 2.

#### Experiment 1

<table>
<thead>
<tr>
<th></th>
<th>LR</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
<td>Test</td>
</tr>
<tr>
<td>Finger Tapping</td>
<td>0.78 (0.50-0.94)</td>
<td>0.64 (0.47-0.82)</td>
</tr>
<tr>
<td>Pronation Supination</td>
<td>0.79 (0.50-0.87)</td>
<td>0.65 (0.49-0.83)</td>
</tr>
<tr>
<td>Toe Tapping</td>
<td>0.78 (0.50-0.90)</td>
<td>0.64 (0.50-0.79)</td>
</tr>
<tr>
<td>Leg Agility</td>
<td>0.94 (0.88-1.00)</td>
<td>0.92 (0.78-0.99)</td>
</tr>
</tbody>
</table>

#### Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>LR</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
<td>Test</td>
</tr>
<tr>
<td>Finger Tapping</td>
<td>0.79 (0.50-0.94)</td>
<td>0.64 (0.47-0.87)</td>
</tr>
<tr>
<td>Pronation Supination</td>
<td>0.95 (0.89-1.00)</td>
<td>0.90 (0.62-1.00)</td>
</tr>
<tr>
<td>Toe Tapping</td>
<td>0.90 (0.81-0.95)</td>
<td>0.80 (0.58-0.97)</td>
</tr>
<tr>
<td>Leg Agility</td>
<td>0.89 (0.50-1.00)</td>
<td>0.77 (0.50-0.99)</td>
</tr>
</tbody>
</table>
Figure 1

Predicted values of the estimated LR and SVM models of experiment 1
Legend: Predicted outcome against true status (patient or control) for the four motor tasks in the LR models (top rows) and the SVM models (bottom rows) of experiment 1. The red line is the “best” threshold averaged over the repeats. Predicted probabilities of being a patient above the red line may be classified as “patient” and below as “controls”. Each dot represents one patient or one control and these individuals occur multiple times depending on their selection into the 100 test sets. The black square represents the predicted probability of being a patient resulting from an intercept-only model.
Figure 2

Predicted values of the estimated LR and SVM models of experiment 2

Legend: Predicted outcome against true status (severe or less severe bradykinesia) for the four motor tasks in the LR models (top rows) and the SVM models (bottom rows) of experiment 2. The red line is the “best” threshold averaged over the repeats. Predicted probabilities of having more severe bradykinesia above the red line may be classified as “severe” and below as “less severe”. Each dot represents one patient and these individuals occur multiple times depending on their selection into the 100 test sets. The black square represents the predicted probability of having more severe bradykinesia resulting from an intercept-only model.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.pdf