Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Noah Adler Date

Assessing Motor Function in Parkinson's Disease using a Web-based, Computerized and User-friendly Tool

By
Noah Adler
Master of Science
Mathematics and Computer Science

_____
Eugene Agichtein, Ph.D.
Advisor

_____
Jorge Juncos, M.D.
Committee Member

_____
Li Xiong, Ph.D.
Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D. Dean of the James T. Laney School of Graduate Studies

_____ Date

Assessing Motor Function in Parkinson's Disease using a Web-based, Computerized and User-friendly Tool

By
Noah Adler

Advisor:
Eugene Agichtein, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science
in Mathematics and Computer Science
2014

Abstract

Assessing Motor Function in Parkinson's disease using a Web-based, Computerized and User-friendly Tool

By Noah Adler

Parkinson's disease (PD) is a neurodegenerative disease resulting in motor- and movement-related impairments. A clinical diagnosis of Parkinson's disease requires clinically detectable motor symptoms, which do not occur until six to eight years after the nigral neurons in the brain begin to degenerate. By detecting PD at an earlier stage, patients can begin therapy sooner, and consequently receive better treatment and care. Therefore, in order to detect motor defects prior to clinical detection, we developed a web-based, user-friendly computer task called Predictive Movement and Trajectory Tracking (PMATT). This task was administered to 23 PD patients and 14 normal controls while recording computer cursor movements. Using machine learning techniques, we calculated fifteen significant motor-related behavioral metrics which strongly distinguish the two groups of patients. By implementing a J48 classifier with these behavioral metrics, over 97% of subjects were correctly classified with an AUC of 0.992. From these results, we conclude that PMATT may be a helpful tool in screening for PD. Since it is easily scalable and automated for individual use, PMATT can be effortlessly administered to the general population. Furthermore, its use in research may help provide insights into the development of motor impairment in pre-clinical PD and help track symptom progression with a higher precision than is currently possible.

Assessing Motor Function in Parkinson's Disease using a Web-based, Computerized and User-friendly Tool

By
Noah Adler

Advisor:
Eugene Agichtein, Ph.D.

A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science
in Mathematics and Computer Science
2014

List of Tables:

List of Figures:

Introduction:

There are more than 4.1 million people worldwide living with Parkinson's disease (PD). (Dorsey et al.) Michael J. Fox, a high profile actor and major advocate, who himself has PD, teaches us that "acceptance [of PD] doesn't mean resignation—it means understanding that something is what it is and that there's got to be a way through it."(Rader) With this in mind, many patients with PD, specialists, and researchers are continuously in search of better treatments for this disabling condition. However, research into therapy, especially neuro-protective therapy that is designed to slow symptom progression, is currently limited by our inability to detect the illness until mid-stages of the disease process. Current therapy only addresses symptoms and does not address the disease process itself. We know that the motor symptoms used to make the diagnosis of PD appear ≥6-8 years after the start of the morbid pathology of the illness. (Schapira) In quantitative terms, symptoms of PD appear only after 60% of striatal dopamine is depleted and mechanisms used to compensate for this loss have been exhausted. (Becker et al.)

Since the brain's compensatory mechanisms triggered by the disease process allow the patient to remain without clinically relevant motor symptoms during this prodromal phase, scientists have begun to identify various non-motor processes early in the course of PD that may serve as biomarkers to help identify those at risk, but still in early stages (pre-motor phase) of PD, that could be helped with neuro-protective therapy. It is the hope of current research strategies to identify cost-effective biomarkers that could help identify those at risk. Thus far, these consist of non-motor signs such as anosmia and RBD (Rapid Eye Movement Behavioral Disorder), which appear several years before the onset of motor symptoms. These biomarkers, though interesting, represent a departure of the motor core

of this movement disorder. As such, and given the frequency of these symptoms in the elderly, they lack in the sensitivity and specificity that is expected of a good biomarker. In order to have sufficient confidence in making the diagnosis of "prodromal PD" to initiate potentially life-long therapy, investigators need a biomarker closer to the core of this disease, that is, a 'motor biomarker' that antecedes the onset of motor symptoms. Such a *motor biomarker* would be ideally suited to help the development of neuro-protective therapies during the prodromal phase of PD.

Today, with the widespread use of computers and the internet, researchers have developed methods and tasks for detection of neurologic dysfunction that could be administered over the internet. Such methodologies would be ideally suited to conduct the large epidemiologic studies needed to set the stage for the neuro-protective trials above. With this in mind, our laboratory developed VPCW (Web-based Visual Paired Comparison), a task that can identify amnestic mild cognitive impairment (aMCI), a common precursor of Alzheimer's disease, which is another important neurodegenerative condition. The VPCW task contains two parts, a simple target tracking task and an image viewing task. In order to replace and simulate the tracking of eye movements in a previous version, the visual paired comparison (VPC) task, VPCW records and analyzes the user's mouse movements. This modification allows the original laboratory-based task to be administered over the web. The computer-based version is simple enough that it can be administered to any subject, regardless of the level of computer experience. Because of these unique characteristics, it provides methodological advantages in the detection of and in the study of aMCI progression. (Agichtein et al.)

The goal of this Parkinson's-focused study was to develop a web-based task sensitive to the fine motor deficits in PD using technology that could be implemented online using the toolbox provided by the VPCW study. While aMCI impairs hippocampal-based memory and learning, in PD a different region of brain is affected, the basal ganglia. Unlike the memory impairment of aMCI, the first symptoms of PD are motor defects linked to dopamine depletion in the basal ganglia. Because the image viewing task of VPCW mostly detects memory impairment, we developed the Predictive Movement and Trajectory Tracking (PMATT) task based on the other part of VPCW, the target tracking task, which allows us to focus directly on easily reproducible motor defects in PD. Traditional motor scales used in PD focus on gross motor movement that are functionally impairing such as tremor and bradykinesia. Yet, we know that fine motor skills, though not as impairing, appear sooner than the gross motor changes that we rely upon to make the diagnosis. The classic example of this is handwriting, which gradually becomes smaller and more cramped in the years approaching the diagnosis of PD. (Fahn) Like handwriting, the target tracking task of VPCW allows the recording and assessing of fine motor skills by asking subjects to move the cursor to randomly appearing stationary targets every five seconds. Since this task is accomplished by tracking mouse movements, and this capability is retained in PMATT, we are able to administer the task over the internet. The task is also simple and easy to explain to the subjects (see methods section).

Correspondingly, studies by Rascol et al. and others found abnormal eye movements in subjects with early PD. This supports the use of tasks that directly, or as in our case indirectly, use eye tracking as a reliable method to detect early PD. As shown in Figure 1,

normal controls performed smooth, steady eye movements when following circular targets along a sinusoidal path. On the other hand, PD subjects were unable to follow the targets readily, due to a tendency for their eye movements to be jerky and shaky. (Rascol et al.) The previously mentioned VPCW study proved that tracking of computer mouse movements is an effective surrogate for the direct tracking of eye movements using infrared devices during cognitive testing. Thus, building on the findings of Rascol et al., the PMATT task implements moving targets using movements of increasing complexity analogous to a dose response: linear, diagonal, and sinusoidal.

Other studies in early PD subjects have shown that these subjects perform normally on simple motor tasks, since the mechanism compensating for lower dopamine levels in the striatum is functioning properly. However, their performance deteriorates as the complexity of the task increases, thus readily separating their dose-response curve from that of controls. In one such study by Stern et al., both PD and normal control (NC) participants traced segments of patterns as shown in Figure 2. Results showed that PD subjects made more errors and their errors were more significant compared to those in controls. Moreover, as the complexity of the tasks increased, relative performance of PD subjects declined further. Stern et al. concluded from their results that "perceptual motor impairment in Parkinson's disease is a form of intellectual impairment associated with higher-order motor control of sequential and predictive voluntary movements." Based on these findings, we chose to add layers of complexity to PMATT by progressively reducing target size. As PMATT focuses on these same motor control movements, we expect it will be very sensitive and accurate at detecting *perceptual motor impairment*, and thus, defects in fine motor control in PD. If true, the PMATT task and these findings

could contribute to the study of pre-motor phases of PD and help with the design of future neuro-protective trials in PD.

Current methods for detection and diagnosis of Parkinson's:

A common problem in PD research is that a clinical diagnosis cannot be established until the neurobiology of the illness is well established, as noted above, >6 years into the course of its morbid pathology. Becker et al. highlight this limitation of the study of early PD in describing how 60% the nigrostriatal neurons of the substantia nigra (SN) are lost before a neurologists can establish the diagnosis of PD. In order to open the possibility of introducing neuro-protective therapies before the 'late' onset of classic motor symptoms, we need tools to detect PD motor dysfunction during its pre-clinical stages.

A seemingly logical first step at trying to make an earlier diagnosis would be to ask a target population if they have relevant symptoms. Dahodwala et al. reviews how a screening questionnaire based on awareness of motor symptoms called the PDSQ (Parkinson's Disease Screening Questionnaire) has been used to detect mild parkinsonian signs in the general population. With it, the highest sensitivity and specificity found were 59% and 63% respectively (Dahodwala et al.). Other screening questionnaires tested so far have shown comparably low sensitivity and specificity, creating a need for alternative strategies. This has led to the focus on parkinsonian non-motor symptoms, which have been shown to antedate the onset of clinically important motor symptoms.

Non-motor symptoms of PD include anosmia and RBD. (Schapira) Early studies have shown that "thirty-eight percent (11/29) [of subjects with RBD] were eventually diagnosed as having a parkinsonian disorder (presumably Parkinson's disease) at a mean interval of 3.7 ± 1.4 (SD) years after the diagnosis of RBD+, and at a mean interval of

12.7 ± 7.3 years after the onset of RBD." (Schenck et al.) However, the low percentage of subjects with RBD who developed PD disqualified RBD by itself as a sufficiently sensitive discriminative tool for detecting prodromal PD.

Similarly, smell identification tests seem to do well at identifying subjects with *early PD*. Picillo et al. evaluated the presence of anosmia in early PD subjects and controls using the University of Pennsylvania Smell Identification Test (UPSIT). Applying logistic regression to the results, "84.5% of subjects were correctly classified into control or PD group[s]. Reliability measures (95% CI) were: specificity 88.2% and sensitivity 82%." (Picillo et al.) The UPSIT thus seems to be a substantial improvement over the motor screening questionnaires. However, its positive predictive value in prodromal stages of PD when applied to a general asymptomatic population remains to be established. The test itself is not practical or cost effective in large populations ($25-30/test + time and personnel to analyze). It also not amenable to, and thus lacks the features and advantages of, a computer based task. For example, computer based tasks can provide strong data gathering and analytic capacity for large field studies. The PMATT task has a web-based recording system to track mouse movement data. This data is more objective than smell identification and is much easier to collect and interpret. Other benefits provided by a computer- based task include services like explanatory visual aids and game-like interfaces to deliver the detection tool in a user-friendly manner. For these reasons, PMATT may provide an efficient and convenient tool for prodromal detection of PD and the monitoring of symptoms during its pre-clinical and clinical phases.

Hypothesis/Expected Results:

Utilizing the PMATT task, this pilot study aims to extract behavioral metrics that correlate with the subject's fine motor skills. These metrics (further explained in methods) are expected to separate normal age-matched control subjects from a cross section of PD subjects and patients with early PD. We expect to increase the discriminative ability of this task by using a dose response approach in presenting increasingly complex movements during task administration. The ultimate aim is to create a tool that will allow us to test the motor system in PD before clinical motor symptoms are apparent. It may also allow us to detect small, clinically unapparent changes in motor control that may be important in detecting neuro-protective effects of drugs or therapies.

Methods:

Study population:

We compared two groups of volunteers: 23 with Parkinson's disease (PD) diagnosed according to standard criteria (Hughes et al.) and 14 normal controls (NC). We recruited PD subjects through the Movement Disorders Clinic at Emory University with the assistance of Dr. Jorge Juncos and his colleagues. The PD volunteers represented a cross section of early to mid-stage levels of disability as classified by the modified Rankin Scale (mRS; see below). They were all optimally treated for their motor and non-motor symptoms. We recruited NC participants from the community, many of which were the spouses of the PD subjects. Exclusion criteria included known neurologic condition or motor disorder (except for PD). Normal controls could not be on any medication known to alter dopaminergic transmission. All subjects completed the informed consent process in accordance with the Institutional Review Board at Emory University.

There were 8 men and 6 women in the control group and 13 men and 10 women in the PD group. Table 1 shows there was no significant difference in age between groups (NC mean ± SD = 71.3 ± 8.1 and PD = 67.7 ± 7.3; $p > .05$) and no significant difference in levels of education (NC mean ± SD = 17.57 ± .79 and PD = 17.74 ± 1.02; $p > .05$). There was no significant group difference in general cognition as measured by the Montreal Cognitive Assessment (MoCA) test as shown in Table 1. All participants scored > 22/30 on MoCA with some individuals having mild cognitive impairment (i.e., scores between 22 and 26) (Aggarwal et al). The disability levels of PD subjects was established by Dr. Juncos and measured using the mRS. (Simuni et al.) The distribution was: Stage 1 (unilateral symptoms with no balance problems) = 5; Stage 2 (bilateral symptoms with no balance problems) = 8; Stage 3 (balance problems but ambulatory and independent) = 10. The self-administered PDQ-39 questionnaire was used to assess quality of life. (Tan et al.)

Clinical Features:

Cognition: Aggarwal et al. validate the Montreal Cognitive Assessment (MoCA) as an adequate screening tool to evaluate gross cognitive abilities in the general population. It has also been used extensively and validated in PD. (Aggarwal et al.) The investigator (Noah Adler) administered the assessment to all study participants.

Quality of Life: The 39-item, self-administered Parkinson's Disease Questionnaire (PDQ-39, Appendix 1) is validated by Tan et al. as a tool for assessing health-related quality of life (HRQoL) in PD patients. It contains thirty-nine questions divided into eight unique domains of PD: mobility, activities of daily living (ADLs), emotional wellbeing, stigma,

social support, cognition, communication, and bodily discomfort. All PD volunteers completed this questionnaire.

Depression: The Patient Health Questionnaire-2 (PHQ-2) contains two multiple choice questions that Kroenke et al. validate for evaluating depression. Kroenke et al. describe the tool itself:

> "The stem question is, 'Over the last 2 weeks, how often have you been bothered by any of the following problems?' The 2 items are 'little interest or pleasure in doing things' and 'feeling down, depressed, or hopeless.' For each item, the response options are 'not at all,' 'several days,' 'more than half the days,' and 'nearly everyday' scored as 0, 1, 2, and 3, respectively."

Every subject in this study answered both of the PHQ-2 questions.

Data Security:

In order to securely store personal healthcare data, we used REDCap, which allowed us "to build and manage online surveys and databases quickly and securely." (Harris et al.)

Task Overview:

Each participant was seated in front of a 15" laptop computer in a brightly lit room with no distractions. The PMATT task was explained prior to its administration. This included a detailed demonstration video and specific instructions. The test itself was conducted under the supervision of the investigator (Noah Adler) to establish that the subject was following the instructions correctly.

Participants performed the task in full screen mode to avoid distractions, but a smaller rectangle (1024 x 768 pixels) of the screen was used for the task itself. Targets emerged

in 3 different sizes: large (radius 45 pixels), medium (radius 36 pixels), and small (radius 26 pixels).

The PMATT Task:

This task requires the user to move the computer mouse such that the cursor remains on the circular white targets, and only on the circular white targets on the screen. While the mouse is on the circle, the circle turns green to indicate the correct placement of the mouse. The targets travel in pre-defined patterns designed to create a *dose response of movement complexity*. To challenge attention and capacity for response inhibition, different visuo-spatial *'distractors'* appear on the screen during the task. These are the patterns:

- *Stationary*: Circular, white targets appear in random locations on the screen and remain on the screen for 5 seconds. Six targets of each size appear: first the large targets, then the medium targets, and finally the small targets. In addition to the circular white targets, blue and white square distractors randomly appear, interspersed throughout the pattern. A total of 18 stationary targets emerge: 9 without distractors, 6 with blue distractors (2 large, 2 medium, 2 small), and 3 with white distractors (1 large, 1 medium, 1 small).
- *Long Stationary*: Two long stationary targets appear for 15 seconds each.
- *Linear motion*: Targets move horizontally across the screen from left to right for 11 seconds each at a speed of 93 pixels/second (shown in Figure 3). These targets emerge at random heights. The user must follow two targets of each size, totaling 6 linear trajectories.

- *Diagonal motion*: Targets move from the top left to the bottom right or bottom left to top right for 13 seconds at a speed of 98.5 pixels/second (shown in Figure 4). The user is expected to follow one medium target and one small target along each of these paths, totaling 4 diagonal trajectories.
- *Sinusoidal motion*: Targets travel in a sinusoidal pattern horizontally across the screen. The sinusoidal pattern has amplitude of 60 pixels and completes a total of 4 full sine periods (a total of 8π radians) as it travels across the screen from left to right. These targets move at three speeds: slow (16 seconds), moderate (12 seconds), and fast (8 seconds). A total of nine targets of this pattern appear: one slow target of each size, then one moderate speed target of each size, and finally one fast target of each size. Thus, a total of 9 targets must be tracked on the sinusoidal trajectory.

Data processing:

We record mouse movements throughout the task to track timestamps, the cursor's x and y coordinates, and the target's x and y coordinates. For consistency in analysis, these data points were interpolated every 20 milliseconds using a nearest-neighbor interpolation method.

Behavioral Metrics:

We calculated numerous behavioral metrics as seen in Appendix 2. After some preliminary analysis (further discussed in the results and discussion sections), we determined that the three primary behavioral metrics for the PMATT task are:

1) TimeOnTarget
2) NumberOfEntriesIntoTarget

3) MotorPersistence

TimeOnTarget is the number of milliseconds in which the cursor remains on the target, which is an indirect measure of "higher-order motor control of sequential and predictive voluntary movement," as described in Stern et al. *NumberOfEntriesIntoTarget* is the number of times the cursor goes from outside the area of the target into the area enclosed by the target, which is an indirect measure of tremor. *MotorPersistence* monitors the ability to keep the cursor on target throughout the task, not just for a particular portion of the task. To do this, using time intervals of 300 milliseconds, we count how many times the cursor is on the target. We used time intervals of 300 milliseconds because it provided better indicators of PD than similar time intervals. Thus, MotorPersistence determines the participant's ability to remain on the target throughout the task and can detect minute changes and fluctuations that TimeOnTarget does not accurately detect.

We subcategorized these three metrics by motion pattern (e.g. stationary, linear, and sinusoidal) and then further split by target size. This allowed the detection of minuscule differences in the subject's actions.

Classification models:

In order to discriminate if a subject's behavioral metrics came from the PD or NC group, we use classification models. The first constructed models used Weka's NBTree (naïve bayes tree) classifier with single behavioral metrics to determine the assignment of group label for the subject. The NBTree classifier finds the optimal threshold to distinguish PD and NC subjects and uses this threshold as the basis for the classification model. Recognizing that additional information provides a better model, we used Weka's J48 classifier with a confidence factor of 0.25 and a minimum number of objects in each leaf

of five to simultaneously calculate the optimal thresholds of multiple, jointly used behavioral metrics. Therefore, the resulting model is defined by a decision tree, splitting at each of the thresholds that have been optimized for joint use among the selected behavioral metrics. (Bhargava et al.)

Linear Regression:

Linear regression models, built using python's scikit library, helped to calculate the PD subjects' level of impairment. To do this, we used the lasso method with all fifteen behavioral metrics as inputs. (Efron et al.)

Evaluating Models:

The most common method of evaluating a model involves testing it on the same data that it used to train the model. This is advantageous when using small datasets because the model can be built utilizing all the data. Additionally, each model's performance metrics are easily calculated, allowing for a simple, straightforward comparison between different models. On the other hand, this method of evaluation does not account for the unforeseen variability in new data. Thus, in order to strengthen our analysis, we also look at five-fold cross validation.

Five-fold cross validation is often used to evaluate a model's expected performance on new data. It involves splitting the dataset, causing a smaller sample of 80% of the data to be used in the model creation. Thus, the held out 20% can be treated like new data. By splitting the model five times and averaging the results, any errors caused by only using a smaller portion of the data for training the model will be minimized. By testing the models on new data, cross validation determines the predicted performance of these models when used on a larger sample. Since the subjects in this study are controls and

patients with PD at early to mid-stages of illness, they closely resemble our eventual target population. Therefore, the cross validation performance metrics used here are a statistically sound approximation of the anticipated PMATT task results if were to use it in screening field studies.

Performance Metrics:

Performance metrics evaluate the strength of a model. For classification models, we report accuracy, AUC, precision, recall, and F-measure. Accuracy is the number of subjects correctly classified by the model. AUC is the Area under the ROC (Receiver Operating Characteristic) Curve. Precision is calculated as true positive results divided by the total number of positive results. Recall is calculated as true positive results divided by the total number of positive subjects. F-measure is the harmonic mean of precision and recall, thus providing a single value that evaluates these two metrics together.

For linear regression models, we report R-squared and RMSE metrics. R-squared is the coefficient of determination representing how well the model performs when compared with a model determined by the mean value of the data. RMSE stands for root mean squared error. All these performance metrics are commonly used and well validated; therefore, we use these metrics to evaluate the results of this study.

Results:

Overview:

One of the main goals of the PMATT study is to provide a tool to detect early defects and small changes in PD motor control that simultaneously consider high (e.g., response inhibition) and low order (e.g., movement time) variables with high accuracy. To achieve this, we generated numerous behavioral metrics associated with motor function and,

based on analysis, chose the metrics most suitable for further analysis. Then, using the selected behavioral metrics, we built classification models to screen for PD. To accomplish this, we first built models using each individual behavioral metric and analyzed its contribution to discriminative ability of the model. After realizing that a better model would provide an improved classification method, we used a J48 classifier to select optimal feature thresholds simultaneously. This classifier resulted in a model dependent on two behavioral metrics: MotorPersistenceDiagonalSmall and NumberOfEntriesIntoTargetSinusoidalSmall. Using these models, our data analysis suggests that PMATT can successfully detect PD in most subjects.

The long term goal of this project is to examine these behavioral metrics, in order to study the progression of pre-clinical and clinical motor defects in PD. To keep the project anchored on clinically relevant motor function, we first showed that PMATT's behavioral metrics correlate strongly with mRS, PDQ-39 scores, and depression scores, which can indirectly affect motor function. At the same time, we showed that our behavioral metrics do not correlate with age or cognitive ability. Therefore, we conclude that PMATT is linked to motor signs and symptoms of PD and not linked to other common elderly ailments. We used this information and linear regression models to optimize the use of the behavioral metrics to fit mRS, PDQ-39 scores and depression scores. This provided insight into the advantages and limitations of PMATT by using the behavioral metrics, which do not fit models regressing to emotional well-being or disease associated stigma, to successfully fit models for regressing motor skills.

Significant Behavioral Metrics:

Many of the selected individual behavioral metrics correspond with fine motor skills. To increase the power of our analysis, we used a two-tailed two-sample unequal variance t-test between the PD and NC groups to determine the optimal metrics. As seen in Table 2, TimeOnTarget, NumberOfEntriesIntoTarget, and MotorPersistence represent the behavioral metrics with significance (p) values less than 0.0001. To further optimize these metrics, each one is separated into diagonal and sinusoidal target motions, and then even further into the different target sizes. Thus, we determined that these fifteen behavioral metrics are optimal for additional analysis. Of these fifteen metrics, three are not significant ($p > .05$) when comparing the PD and NC groups: NumberOfEntriesIntoTargetDiagonalMedium, NumberOfEntriesIntoTargetSinusiudalBig, and NumberOfEntriesIntoTargetSinusiudalMedium.

Table 2 also shows that TimeOnTarget, NumberOfEntriesIntoTarget, and MotorPersistence remain significant ($p < .05$) even when comparing NC to the smaller subset of early PD subjects (mRS <= 2; n = 13). Due to the smaller dataset, the p-values are higher. In contrast, when comparing early PD with moderate PD (mRS = 3; n = 10), only 2 behavioral metrics show significance ($p < .05$): TimeOnTargetSinusoidalMedium and MotorPersistenceSinusoidalMedium.

Classification Results:

The best performing models using individual behavioral metrics for classifying the PD and NC groups are MotorPersistenceDiagonalSmall and NumberOfEntriesIntoTargetSinusoidalSmall. As seen in Table 3, model 1 classifies PD subjects as those with MotorPersistenceDiagonalSmall less than 190, while model 2

classifies PD subjects as those with NumberOfEntriesIntoTargetSinusoidalSmall less than 17.33. Table 4 displays the performance measures of these comparable models. When model 1 is tested on the training data, it correctly classified 32 out of 37 of the subjects with an AUC of 0.821 and an F-measure of 0.857. These performance measures are very similar to model 2, which correctly classified 31 out of 37 of the subjects with an AUC of 0.842 and an F-measure of 0.840. Additionally, the table shows that, when tested using five-fold cross validation, model 1 correctly classified 31 out of 37 of the subjects with an AUC of 0.744 and an F-measure of 0.831, while model 2 correctly classified 30 out of 37 of the subjects with an AUC of 0.750 and an F-measure of 0.809. From these results, we conclude that these models perform strongly compared to the current methods of screening, such as questionnaires (sensitivity of 59% and specificity of 63%) or olfactory testing (accuracy of 84.5%). Can a better model be built?

To build a better model, we implemented weka's J48 decision tree classifier in order to utilize multiple behavioral metrics simultaneously. This model took all the features as inputs and determined two jointly assessed behavioral metrics to provide an exceptional model for classifying subjects as PD or NC: NumberOfEntriesIntoTargetSinusoidalSmall and MotorPersistenceDiagonalSmall. As seen in Table 3, the J48 model classifies PD subjects as those with NumberOfEntriesIntoTargetSinusoidalSmall less than 17.33 **and** MotorPersistenceDiagonalSmall less than 215. All other subjects are classified as NC. When tested on training data, the model correctly classified 36 out of 37 of the subjects with an AUC of 0.992 and an F-measure of 0.973. Furthermore, when using five-fold cross validation, 32 out of the 37 subjects were correctly classified with an AUC of 0.882

and an F-measure of 0.867. These results indicate that the joint model detects PD with extraordinary accuracy.

Correlation:

Table 5 depicts the correlation coefficients between the modified Rankin Scale (mRS), age, MoCA, depression scores, PDQ and the fifteen behavioral metrics mentioned above. These provide insights into many important aspects of PD presumed relevant to fine and gross motor control. As expected, the mRS and PDQ are highly correlated to each other. Additionally, the mRS, depression scores, and PDQ all have high correlations (> .4) with many of the behavioral metrics. In contrast, age and MoCA scores are not correlated with any other fields. From this we conclude that PMATT is a strong tool for calculating behavioral motor metrics.

Linear Regression Models:

In an attempt to gain knowledge and understanding about how these models inform motor and cognitive defects in PD, we built linear regression models. Table 6 portrays depression, PDQ, and mRS fields and their R-squared and RMSE values when tested on the training data and on five-fold cross validation. As evident from the R-squared values, these models perform strongly when tested on the training data. However, when using five-fold cross validation, only the models for the depression scores, PDQ mobility and ADLs (activities of daily living), and mRS fit well. In contrast, those fields with negative R-squared values, meaning they perform worse than using the average, include PDQ emotional wellbeing, stigma, and total.

Discussion:

Significant Behavioral Metrics:

In this analysis, many motor metrics of PMATT showed significant ($p<.05$) differences between the NC group and our cross section of PD patients. Some of these metrics include: movement time, reaction time, maximum velocity, time on target, number of entries into the target, and motor persistence. In the past, the use of these metrics to distinguish a subject with early PD and age-matched controls has been difficult because of their lack of sensitivity and their huge variance in aging populations plagued by cognitive changes, tremors unrelated to PD (e.g., drug induced or essential tremor), and non-specific slowing due to arthritis and other factors. For example, Jahanshahi et al. reviewed many studies that tried to use reaction time to distinguish PD and controls. They concluded that reaction time (RT) appears to be "a slowness in response initiation in Parkinson's disease, which is a stage of processing common to all RT conditions," (Jahanshahi et al.) and this commonly seen in the elderly population. Other studies proved the inadequacy of RT as a metric to discern early PD from normal aging, again due to lack of sensitivity. Accordingly, these behavioral and motor metrics have fallen into disfavor in recent years for use in these investigations.

To address these historical, but we think remediable, limitations of these metrics, we chose to focus on a subset of these metrics that are more likely to provide precision, sensitivity, and simplicity in the task of separating early PD motor dysfunction from normal ranges of functioning. After reviewing available data and studying the likely reasons for previous failures, we chose parameters that are closely linked to the fundamental motor defects in PD. We then selected behavioral metrics for their high precision ($p<.0001$). In the process, we found that simple linear measures at low speeds using a standard target size have been previously incapable of discriminating PD subjects

and controls. Thus, as part of the PMATT task, we opted to administer the targets at different speeds, using both linear and non-linear movements, and targeting different size objects.

Using this strategy we found that the most discriminant metrics were *TimeOnTarget*, *NumberOfEntriesIntoTarget*, and *MotorPersistence*. As anticipated, the hardest motions for the subjects were the diagonal and sinusoidal ones. As such, these patterns were also the most likely to discriminate between PD and NC groups. Additionally, the higher level of difficulty created by the progressive reduction in target size led to an increase in the significance level for many of the metrics. This early finding informed our subsequent analysis using the fifteen most sensitive behavioral metrics.

Three behavioral metrics were insignificant ($p > .05$):

NumberOfEntriesIntoTargetDiagonalMedium,

NumberOfEntriesIntoTargetSinusoidalBig, and

NumberOfEntriesIntoTargetSinusoidalMedium. The NumberOfEntriesIntoTarget behavioral metrics is an indirect measure of tremor; thus, we determine that these larger target sizes are incapable of detecting minor tremors. This is likely caused by two reasons. First, the target size is larger than the tremor amplitude, so the tremor does not cause the cursor to leave the target. Secondly, following the trajectory of the larger targets is simpler, so it does not cause a dose response of movement complexity. To account for this issue in future studies, target sizes must be optimized to account for minor tremors and to cause a dose response of movement complexity.

Following this analysis, we compared the NC group with the early PD group (mRS <= 2). This allowed us to determine the effectiveness of the behavioral metrics at distinguishing

those at early stages of the disease. Not surprisingly, all the significant metrics for distinguishing NC from PD are also significant for distinguishing NC from early PD. This validates the ongoing investigation of PMATT as a tool that could serve as a *'motor-biomarker'* of the pre-clinical phase of PD.

Only two metrics remained significant when comparing early vs moderate PD: TimeOnTargetSinusoidalMedium and MotorPersistenceSinusoidalMedium. As a possible explanation, both metrics are heavily dependent on remaining on the target, making the size of the target critical. Since subjects with moderate PD have more severe motor dysfunction, they were unable to remain on the medium target, but those with early PD, and consequently only minor motor dysfunction, were capable of tracking the target successfully on the medium target size. As seen in Table 2, the smaller the target, the more significant the behavioral metric is at discriminating between NC and early PD. Thus, on the small target, the early PD subjects likely experience similar motor dysfunction as moderate PD subjects, causing a lack of significance at distinguishing early PD from moderate PD. Further work is needed to confirm this explanation.

Classification:

We built individual behavioral metric models for classifying participants into PD and NC groups. The top performing single metric model, model 1, uses MotorPersistenceDiagonalSmall, classifying 86.49% of subjects correctly on training data. Model 2 is the second best single metric model using NumberOfEntriesIntoTargetSinusoidalSmall to classify 83.78% of subjects correctly on training data. Both these models accurately classify over 81% of subjects in cross validation.

However, we aimed to build a better model by simultaneously using all the behavioral metrics employing a J48 classifier. This classifier utilized a double metric classification model that performs significantly better than the single metric models, correctly classifying 97.30% on training data. Thus, the J48 model validates our hypothesis that PMATT can distinguish between NC and a cross section of PD subjects. By correctly classifying 86.49% of subjects in cross validation, we validate utilizing PMATT's behavioral metrics to classify new subjects.

Due to the small size of our subject pool, implementing a model using three or more features would cause substantial over fitting. Thus, by increasing the size of the subject pool, we would be able to utilize multi-dimensional models, which would strengthen the detection abilities of this tool.

Correlations:

Table 5 shows the correlation between the behavioral metrics discussed above and the many clinical features of the Parkinson's patients. By gaining understanding about the correlation between these different attributes, we hope to refine the behavioral metrics that PMATT produces and determine their clinical significance. For instance, we found that unlike the effects of aging on motion and cognition, our experimental metrics did not correlate with age or MoCA scores. Accordingly, the PMATT behavioral metrics, the mRS, the depression index, and the PDQ score were all independent of age and cognition.

In contrast, the high correlation between the PMATT behavioral metrics and the clinically important indices of mRS, depression, and PDQ suggest that PMATT is anchored on clinically relevant measures of PD and may serve to detect small changes in

clinical symptoms during neuro-protective therapies that we have failed to detect using conventional clinical measures like the United Parkinson Disease Rating Scale (UPDRS). As a result of this, we conclude that PMATT's behavioral metrics are sensitive to PD-specific change and not a function of common elderly ailments such as aging or cognitive decline.

Linear Regression Models:

One of the main goals of this ongoing work is to use the PMATT task to learn about motor symptoms related to early PD. Since the linear regression models built using PMATT's behavioral metrics fit the PDQ-39 mobility scores and mRS scores well, PMATT scores may help as surrogates to track these areas of dysfunction in PD by possibly adding precision without sacrificing simplicity to the process.

We also attempted to fit PMATT behavioral metrics to non-motor domains of the PDQ-39, such as emotional wellbeing and stigma. These linear regression models show an unequivocal lack of correlation in these domains, with these models performing worse than a model that just returns the average of the input data. One potential explanation is that PMATT was designed to assess motor function and thus appears insensitive to other forms of disability in PD. Given the multiple domains of the PDQ-39, it is not surprising to find that the linear regression models did not correlate with the total score of the PDQ-39.

Future Work:

How will PMATT inform the study of PD?

Discriminant analysis: Using a cross sectional sample of patients with mild to moderate PD we have shown that PMATT can distinguish PD from NC subjects with high

accuracy. Even when we limit our analysis to those with the earlier stages of PD, PMATT can still distinguish both groups with a robust statistical significance. PMATT could thus be developed as a tool to screen elderly populations for PD. Current detection tools require big, "clinically relevant" changes in motor function for detection of PD. These changes may be "too much to ask for" when using drugs that address the lack of dopamine production in the brain instead of addressing specific PD motor symptoms like bradykinesia and tremor.

Better understanding motor function: Additionally, while the linear regression models may not accurately fit all aspects of PD, PMATT provides a useful tool to better understand motor (e.g., movement execution) and cognitive (e.g., motor planning) aspects of motor control in PD. With this information, we hope to detect defects in motor performance before they become clinically apparent. This would have important implications for neuro-protective therapy. In the ideal world this information will be necessary to study early intervention in asymptomatic individuals and to help track the results of early interventions in minimally affected individuals. The presumption is that it would be easier to slow or reverse these changes with pharmacologic interventions in PD when the changes are minor. Other potential uses of this technology include the use of PMATT as a surrogate to index and track an individual's disability and its correlation with his/her ability to perform activities of daily living like driving and using tools at home. These reports could also be used to predict the level of assistance a patient is likely to need at home.

Beyond screening: Since PMATT can be implemented and administered over the web, it can easily be accessed and completed by anyone in the world. Therefore, it can be used as

a screening tool in large scale projects in Parkinson's disease and other neurodegenerative conditions. Potential users for this include researchers, clinicians, and individuals. Additional uses include long term tracking of the clinical effects of drugs during development or post-marketing. An example of this would be the detection of small, otherwise clinically unimportant changes in motor function that can track the effects of neuro-protective agents. An additional, and potentially important, use would be as a biomarker to pre-clinical (not yet apparent on clinical exam) motor involvement in individuals at risk of developing PD. Based on current research, these individuals include those with genetic forms of PD, those with RBD and those with anosmia. (Mollenhauer)

Potential research directions:

This study provides new insights into the usage of mathematical models that can improve the use of cognitive and motor indices of dysfunction in PD. By modeling and analyzing various parameters simultaneously, we achieved significant improvements in detection capabilities of PMATT's calculated behavioral metrics. More complex models would likely increase the power of the PMATT tool, but since this is a pilot study aiming for simplicity, we kept our models simple and straightforward. We also expect that simultaneous modeling of various symptoms commonly seen in PD, such as motor symptoms, RBD, and anosmia, would create a more advanced tool that encompasses all facets of the disease, thus providing a multifaceted tool for detection.

In the future, to further PMATT as a potential pre-symptomatic tool to detect 'pre-motor' PD, we would want to validate the discriminant ability of PMATT in patients with very early and untreated disease and compare them to a larger sample of controls, controlling for age, education, cognitive function, medical conditions and non-PD medications. Once

we can establish and validate detection thresholds for individual PMATT parameters, we would be ready to begin detecting 'incipient PD' in populations at risk. Currently, it is very difficult to make presumptive diagnosis of 'incipient PD' in patients with RBD or anosmia in the absence of motor signs since the discriminant ability of these non-motor signs independent of motor indices remains poor. Once established as a pre-clinical index of PD, PMATT could then be used as a practical tracking tool during the pre-clinical course of the illness.

Potential Limitations:

Test subjects may find the PMATT task to be boring, in which case attention to the testing may be compromised during task administration. To curtail this, we attempted to keep the task as short as possible (total run time of eight minutes). Additionally, to ensure the subjects remained motivated and attentive, their performance was constantly monitored by the study investigator (Noah Adler). After careful observation, we confirmed that the task was suitably completed by all the subjects. Since we have now validated the PMATT task, it can be made more interesting by adding elaborate attention and 'gaming style' graphics and sound effects.

Another limitation is that the elderly subjects may find the computer game to be daunting and intimidating due to little or no computer experience. Many attributes of the task aid in resolving these issues. First of all, before starting the task, demonstration videos are shown to explain the task to the user. Secondly, written and verbal instructions assist the user throughout. Lastly, the task was kept simple and straightforward. With these aids employed, all subjects in this study were capable of participating and completing PMATT (monitored by Noah Adler).

A third limitation is that this study only confirms PMATT as a screening tool. The hypothesized diagnosis must be confirmed by a physician. Since this is just a pilot study, concluding that this task can be used as a screening tool is a very significant step. The next significant step for us is extending this tool to an early detection tool. In order to do this, a much larger, longitudinal dataset must be gathered including a large number of untreated, early-stage PD subjects. Additionally, the task must be evaluated when compared to supplementary PD scales such as UPDRS, which was not implemented at this stage due to time and personnel constraints.

Lastly, when the PMATT task is implemented in practice, we expect to find a learning effect in subjects that participate multiple times. Lumos Labs, also known as Lumosity, performed studies on the cognitive effects of training on attention and working memory. They compared a group of trained subjects against a control group to find that "the trained group improved significantly more than the control group on untrained measures of visual attention and working memory." (Hardy et al.) Similar to these results, if subjects were to use PMATT regularly, for example to track the effects of a new drug on motor function, they would likely begin to experience a training effect that would result in improved performance unrelated to the drug itself. This effect could be diminished by providing additional versions of the task and implementing various randomization algorithms. Furthermore, by modeling this training effect, we aim to detect previously unnoticed variations in the learning abilities of PD patients.

Conclusion:

The data presented here supports the view that the PMATT may be an effective tool to detect early, and possibly pre-motor stages, of PD. If true, PMATT could be simplified to

serve as a practical and cost effective tool to screen large populations for early signs of Parkinsonism. Such a tool has a significant potential in facilitating current and future research in neuro-protective and other therapies in PD. With further refinements, it could aid physicians in monitoring the effectiveness of treatment or produce summary reports on a patient's ability to perform activities of daily living. Similar examples elaborated on throughout this paper show that PMATT can provide an additional tool to investigate symptomatic and neuro-protective treatments in PD.

Figures and Tables:

Figure 1

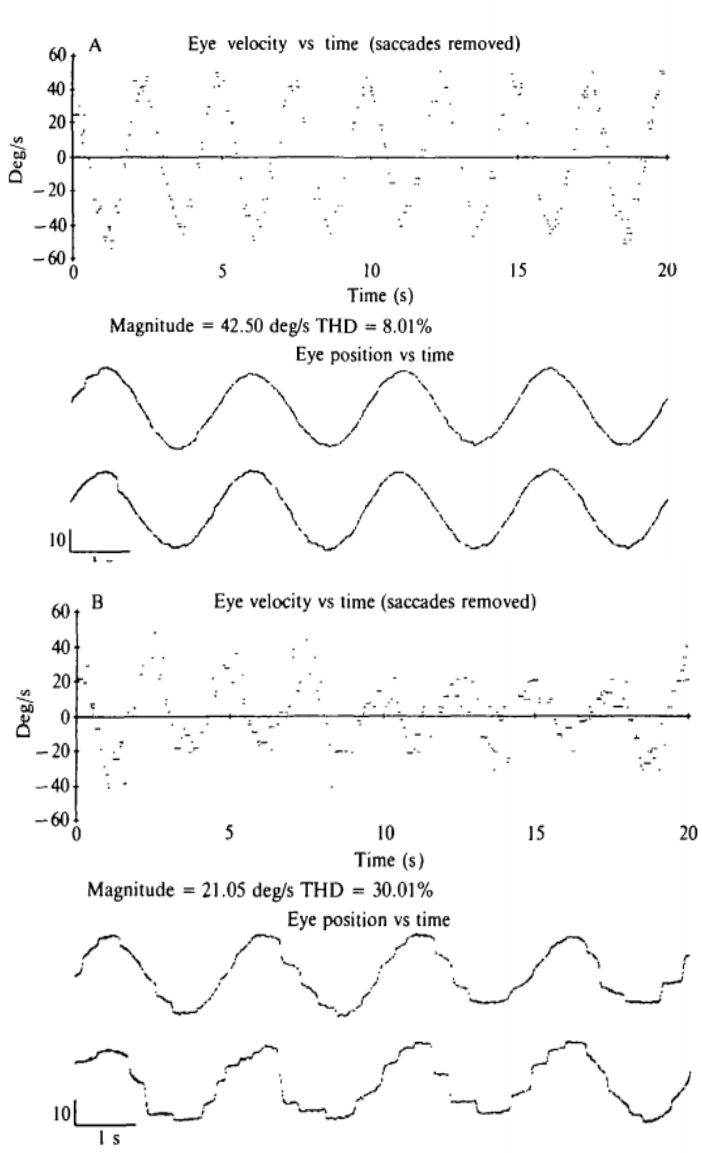


FIG. 2. Examples of smooth pursuit tests (40 deg/s, 0.4 Hz). A, *normal control*. An analysed data graph displays eye velocity versus time. The statistical results for magnitude (= peak eye velocity) and THD are displayed below this graph and above that for eye position versus time. B, *parkinsonian patient* (stage IV of Hoehn and Yahr). Note that the smooth pursuit eye velocity is markedly decreased. This explains the low value for peak eye velocity at the frequency of the stimulus (magnitude) and the high THD which illustrate the patient's difficulty in matching his eye movement correctly with that of the target. This difficulty makes him to catch up the target with small saccades, explaining the saccadic or cogwheel aspect of the pursuit observed on the graph displaying eye position versus time.

Rascol, O., M. Clanet, J. L. Montastruc, M. Simonetta, M. J. Soulier-Esteve, B. Doyon, and A. Rascol. "Abnormal Ocular Movements In Parkinson's Disease." Brain 112.5 (1989): 1193-214. Web.

Figure 2

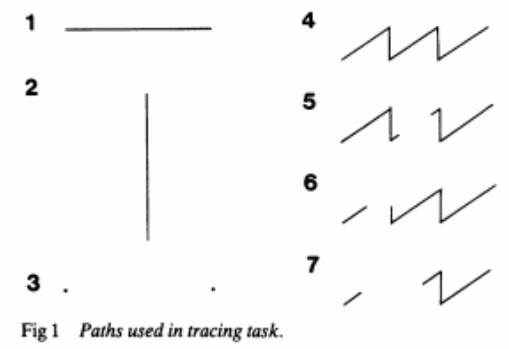
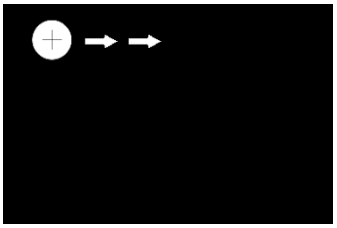


Fig 1 *Paths used in tracing task.*

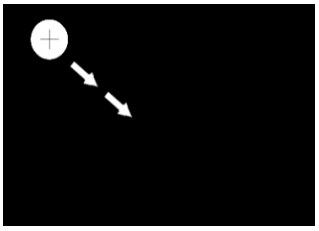
Stern, Yaakov, Richard Mayeux, Jeffrey Rosen, and Joyce Ilson. "Perceptual Motor Dysfunction in Parkinson's Disease: A Deficit in Sequential and Predictive Voluntary Movement." Journal of Neurology, Neurosurgery & Psychiatry 46.2 (1983): 145-51. Web.

Figure 3



Target traveling in a linear pattern

Figure 4



Target traveling in diagonal pattern

Table 1

	NC (N=14)	PD (N=23)			PD vs. NC
		mRS: 1 (N=5)	mRS: 2 (N=8)	mRS: 3 (N=10)	
Age	71.33(2.15)	67.50 (1.77)	64.58(2.23)	70.50(2.63)	*NS*
Years of Education	17.57(0.79)	17.20(1.20)	17.50(0.98)	18.20(0.96)	*NS*
Montreal Cognitive Assessment	27.71(0.46)	27.60(0.40)	26.50(0.63)	26.20(0.65)	*NS*

Subjects demographics and neuropsychological assessment scores. Legend: NC – normal control, PD – Parkinson's disease, mRS – modified Rankin Score. Columns 2-5 are the scores obtained by the groups for the descriptors in Column 1. Column 5 is the p value for a two-tailed two-sample unequal variance t-test. Significance (p) values are reported when below 0.05, otherwise reported as *NS* (not significant).

Table 2

Behavioral Metric	NC(n=14)	PD(n=23)	NC vs. PD	NC vs. Early PD	Early PD vs. Moderate PD
TimeOnTarget					
Diagonal movements					
Medium targets	318.46(13.85)	211.91(18.03)	4.12E-05	0.002242	NS
Small targets	247.82(12.61)	156.35(14.05)	2.69E-05	0.001205	NS
Sinusoidal movements					
Big targets	459.21(15.73)	335.86(25.36)	0.000223	0.012246	NS
Medium targets	400.71(18.55)	268.74(24.11)	0.000116	0.021961	0.015968
Small targets	278.79(22.31)	175.26(19.86)	0.001595	0.040168	NS
NumberOfEntriesIntoTarget					
Diagonal movements					
Medium targets	7.46(0.85)	9.09(0.75)	NS	NS	NS
Small targets	13.11(0.66)	10.52(0.73)	0.012812	0.048825	NS
Sinusoidal movements					
Big targets	7.62(1.04)	8.58(0.56)	NS	NS	NS
Medium targets	12.43(1.10)	10.61(0.62)	NS	NS	NS
Small targets	17.88(0.73)	12.28(0.99)	6.05E-05	0.000895	NS
MotorPersistence					
Diagonal movements					
Medium targets	241.07(6.67)	176.30(12.95)	0.000101	0.003550	NS
Small targets	218.93(7.02)	149.13(11.33)	8.66E-06	0.000529	NS
Sinusoidal movements					
Big targets	333.57(7.37)	259.13(16.96)	0.000368	0.012014	NS
Medium targets	309.76(9.39)	222.90(15.99)	4.64E-05	0.008716	0.024067
Small targets	257.38(11.16)	170.87(15.95)	8.60E-05	0.005528	NS

The mean scores and standard errors values for the PMATT behavioral metrics. For each behavioral metric we report the p-value from the two-tailed two-sample unequal variance t-test. Significance (p) values are reported when below 0.05, otherwise reported as *NS* (not significant). Data = mean ± SD.

Table 3

	Single Behavioral Metric Model		Double Behavioral Metric Model
	Model 1	Model 2	J48 Model
MotorPersistence DiagonalSmall	<190	NA	<215
NumberOfEntriesIntoTarget SinusoidalSmall	NA	<17.33	<17.33

The inequalities show the dependency of each model on each behavioral metric that is needed for a subject to be classified as PD. When two behavioral metrics are present, both inequalities must hold true for the subject to be classified as PD by the model.

Table 4

		Accuracy	AUC	Precision	Recall (sensitivity)	F-Measure
Model 1	**Training**	86.49% (32/37)	0.821	0.889	0.865	0.857
	5-fold Cross Validation	83.78% (31/37)	0.744	0.847	0.838	0.831
Model 2	**Training**	83.78% (31/37)	0.842	0.846	0.838	0.840
	5-fold Cross Validation	81.08% (30/37)	0.750	0.809	0.811	0.809
J48 Model	**Training**	97.30% (36/37)	0.992	0.975	0.973	0.973
	5-fold Cross Validation	86.49% (32/37)	0.882	0.880	0.865	0.867

The training data and five-fold cross validation performance metrics for the three classification models. For each model, we report accuracy, AUC, precision, recall, and F-measure.

Table 5

*PD subjects only	mRS	Age	MoCA	Depression	PDQ
Modified Rankin Score	1	0.2290	-0.2909	0.1473	**0.6124**
Age	0.2290	1	0.2075	0.0513	0.0481
MoCA	-0.2909	0.2075	1	-0.0140	-0.1232
Depression	0.1473	0.0513	-0.0140	1	**0.5468**
PDQ	**0.6124**	0.0481	-0.1232	**0.5468**	1
TimeOnTarget					
Diagonal movements					
Medium targets	**-0.4376**	-0.2773	0.0150	-0.3783	**-0.4751**
Small targets	**-0.4128**	-0.2022	0.0150	**-0.4359**	**-0.5147**
Sinusoidal movements					
Big targets	**-0.5011**	-0.2394	0.0655	**-0.4612**	**-0.4950**
Medium targets	**-0.6447**	-0.1142	0.1229	-0.3594	**-0.5412**
Small targets	**-0.5080**	-0.0191	0.0204	**-0.4421**	**-0.5595**
NumberOfEntriesIntoTarget					
Diagonal movements					
Medium targets	-0.2695	-0.1719	0.0483	-0.3249	-0.1437
Small targets	-0.2782	-0.0940	-0.0568	-0.3875	-0.2561
Sinusoidal movements					
Big targets	0.0946	-0.0208	-0.0457	**-0.5037**	-0.0382
Medium targets	0.1467	-0.1376	-0.1021	**-0.4794**	-0.0759
Small targets	**-0.4395**	-0.0667	0.1702	**-0.5296**	**-0.4899**
MotorPersistence					
Diagonal movements					
Medium targets	**-0.4479**	-0.2732	0.0069	**-0.4231**	**-0.4761**
Small targets	**-0.4214**	-0.2197	0.0175	**-0.4523**	**-0.4853**
Sinusoidal movements					
Big targets	**-0.4747**	-0.2618	0.0552	**-0.5166**	**-0.4861**
Medium targets	**-0.6119**	-0.1129	0.0921	**-0.4201**	**-0.5547**
Small targets	**-0.5038**	-0.0256	0.0673	**-0.5126**	**-0.5668**

Pearson correlation values among PD subjects comparing horizontally: modified Rankin Scale, Age, MoCA, Depression, and PDQ to vertically: modified Rankin Scale, Age, MoCA, Depression, PDQ, and PMATT behavioral metrics. Bolded values show strong correlation.

Table 6

	Training data		Cross validation	
	R squared	**RMSE**	**R squared**	**RMSE**
depQ1	0.989	0.944	**0.594**	0.969
depQ2	0.978	0.787	**0.697**	0.834
pdqMobility	0.909	0.710	**0.430**	0.815
pdqADLs	0.985	0.720	**0.758**	0.770
pdqEmotionalWellbeing	0.924	0.863	-8.448	0.957
pdqStigma	1	0.822	-1.944	0.823
pdqTotal	0.982	0.610	-0.386	0.677
mRS	0.908	0.651	**0.283**	0.718

Comparison of linear regression models using the behavioral metrics to fit depression, PDQ, and mRS. For each model we report R-squared and RMSE. Bolded R-squared values show well fit regression models.

References:

- Adler, Charles H., Thomas G. Beach, Joseph G. Hentz, Holly A. Shill, Erika Driver-Dunckley, Marwan N. Sabbagh, Lucia I. Sue, Sandra A. Jacobson, Christine M. Belden, and Brittany N. Dugger. "Low Clinical Diagnostic Accuracy of Early vs Advanced Parkinson Disease." Neurology (2014): 406-12. Web.
- Aggarwal, Arun, Maria G. Bernardi, and Lyn Wright. "Cognitive Screening Tool in Parkinson's Disease: Mini Mental State Examination (MMSE) Versus Montreal Cognitive Assessment (Moca)." Journal of Alzheimer's Disease & Parkinsonism 01.S1 (2012): n. pag. Web.
- Agichtein, Eugene, Dmitry Lagun, Noah Adler, Cecelia Manzanares, Jongho Shin, Allan I. Levey, Elizabeth Buffalo, and Stuart Zola. "An Internet-Based Automated Behavioral Task for Detection of Amnestic Mild Cognitive Impairment." (in submission)
- Becker, Georg, Antje Muller, Stefan Braune, Thomas Buttner, Reiner Benecke, Wolfgang Greulich, Wolfgang Klein, Gunter Mark, Jurgen Rieke, and Reiner Thumler. "Early Diagnosis of Parkinson's Disease."Journal of Neurology 249.0 (2002): 1. Web.
- Bhargava, Neeraj, Gerja Sharma, Ritu Bhargava, and Manish Mathuria. "Decision Tree Analysis on J48 Algorithm for Data Mining."International Journal of Advanced Research In Computer Science and Software Engineering 3.6 (2013): 1114-1119. Web.

- Coeytaux, Alessandra, Keith Wong, Ron Grunstein, and Simon JG Lewis. "REM Sleep Behaviour Disorder More than Just a Parasomnia." Australian Family Physician 42.11 (2013): 785-88. Web.
- Dahodwala, Nabila, Lee Kubersky, and Andrew Siderowf. "Can a Screening Questionnaire Accurately Identify Mild Parkinsonian Signs?" Neuroepidemiology 39.3-4 (2012): 171-75. Web.
- Dorsey, E. R., R. Constantinescu, J. P. Thompson, K. M. Biglan, R. G. Holloway, K. Kieburtz, F. J. Marshall, B. M. Ravina, G. Schifitto, A. Siderowf, and C. M. Tanner. "Projected Number of People with Parkinson Disease in the Most Populous Nations, 2005 through 2030." Neurology 68.5 (2007): 384-86. Web.
- Efron, Bradley, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. "Least Angle Regression." The Annals of Statistics 32.2 (2004): 407-99. Web.
- Fahn, Stanley. "Description of Parkinson's Disease as a Clinical Syndrome." Annals of the New York Academy of Sciences 991.1 (2003): 1-14. Web.
- Hall, Mark, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.
- Hardy, Joseph L., David Drescher, Kunal Sarkar, Gregory Kellett, and Michael Scanlon. "Enhancing Visual Attention and Working Memory with a Web-based Cognitive Training Program." Mensa Research Journal 42.2 (2011): 13-20. Web.
- Harris, Paul A., Robert Taylor, Robert Thielke, Jonathon Payne, Nathaniel Gonzalez, Jose G. Conde, Research electronic data capture (REDCap) - A

metadata-driven methodology and workflow process for providing translational research informatics support, J Biomed Inform. 2009 Apr;42(2):377-81.

- Hughes, Andrew J., Susan E. Daniel, Linda Kilford, and Andrew J. Lees. "Accuracy of Clinical Diagnosis of Idiopathic Parkinson's Disease: A Clinico-pathological Study of 100 Cases." Journal of Neurology, Neurosurgery & Psychiatry 55.3 (1992): 181-84. Web.
- Jahanshahi, Marjan, Richard G. Brown, and David Marsden. "Simple And Choice Reaction Time And The Use Of Advance Information For Motor Preparation In Parkinson's Disease." Brain 115.2 (1992): 539-64. Web.
- Kroenke, Kurt, Robert L. Spitzer, and Janet BW Williams. "The Patient Health Questionnaire-2: Validity of a Two-Item Depression Screener." MEDICAL CARE 41.11 (2003): 1284-292. Web.
- Mollenhauer, Brit, Ellen Trautmann, Friederike Sixel-Doring, Tamara Wicke, Jens Ebentheuer, Martina Schaumburg, Elisabeth Lang, Niels K. Focke, Kishore R. Kumar, Katja Lohmann, Christine Klein, Michael G. Schlossmacher, Ralf Kohnen, Tim Friede, and Claudia Trenkwalder. "Nonmotor and Diagnostic Findings in Subjects with De Novo Parkinson Disease of the DeNoPa Cohort." Neurology 81.14 (2013): 1226-234. Web.
- Pedregosa, F., Varoquaux, G., Gramfort, A. & Michel, V. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830.
- Picillo, Marina, et al. "The Use Of University Of Pennsylvania Smell Identification Test In The Diagnosis Of Parkinson's Disease In Italy."

Neurological Sciences: Official Journal Of The Italian Neurological Society And Of The Italian Society Of Clinical Neurophysiology 35.3 (2014): 379-383. MEDLINE. Web. 26 Oct. 2014.

- Rader, Dotson. "Michael J. Fox at 50: 'I Don't Look at Life as a Battle'" Parade. N.p., 1 Apr. 2012. Web. 24 Oct. 2014.
- Rascol, O., M. Clanet, J. L. Montastruc, M. Simonetta, M. J. Soulier-Esteve, B. Doyon, and A. Rascol. "Abnormal Ocular Movements In Parkinson's Disease." Brain 112.5 (1989): 1193-214. Web.
- Schapira, Anthony H.V. "Neurobiology and Treatment of Parkinson's Disease." Trends in Pharmacological Sciences 30.1 (2009): 41-47. Web.
- Schenck, CH, SR Bundlie, and MW Mahowald. "Delayed Emergence of a Parkinsonian Disorder in 38% of 29 Older Men Initially Diagnosed with Idiopathic Rapid Eye Movement Sleep Behaviour Disorder." Neurology46.2 (1996): 388-93. Web.
- Simuni, Tanya, Sheng T. Luo, Kelvin L. Chou, Hubert Fernandez, Bo He, and Sotirios Parashos. "Rankin Scale as a Potential Measure of Global Disability in Early Parkinson's Disease." Journal of Clinical Neuroscience 20.9 (2013): 1200-203. Web.
- Stern, Yaakov, Richard Mayeux, Jeffrey Rosen, and Joyce Ilson. "Perceptual Motor Dysfunction in Parkinson's Disease: A Deficit in Sequential and Predictive Voluntary Movement." Journal of Neurology, Neurosurgery & Psychiatry 46.2 (1983): 145-51. Web.

- Tan, Louis C.s., Nan Luo, Mohammed Nazri, Shu Chuen Li, and Julian Thumboo. "Validity and Reliability of the PDQ-39 and the PDQ-8 in English-speaking Parkinson's Disease Patients in Singapore." Parkinsonism & Related Disorders 10.8 (2004): 493-99. Web.

Appendix 1:

PDQ-39 item
Mobility
Had difficulty doing the leisure activities you would like to do
Had difficulty looking after your home, e.g. DIY, housework, cooking
Had difficulty carrying bags of shopping
Had problems walking 1 km[b]
Had problems walking 100 m[b]
Had problems getting around the house as easily as you would like
Had problems getting around in public[c]
Needed someone else to accompany you when you went out
Felt frightened or worried about falling in public
Been confined to the house more than you would like
Activities of daily living (ADL)
Had difficulty washing yourself
Had difficulty dressing yourself[c]
Had problems doing up buttons or shoe laces
Had problems writing clearly
Had difficulty cutting up your food
Had difficulty holding a drink without spilling it
Emotional well-being
Felt depressed[c]
Felt isolated and lonely
Felt weepy or tearful
Felt angry or bitter
Felt anxious
Felt worried about your future
Stigma
Felt you had to conceal your Parkinson's from people
Avoided situations which involve eating or drinking in public
Felt embarrassed in public due to having Parkinson's disease[c]
Felt worried about other people's reaction to you
Social support
Had problems with your close personal relationships[c]
Lacked support in the ways you need from your spouse or partner
Lacked support in the ways you need from your family or close friends
Cognitions
Unexpectedly fallen asleep during the day
Had problems with your concentration, e.g. when reading or watching TV[c]
Felt your memory was bad
Had distressing dreams or hallucinations
Communication
Had difficulty with your speech
Felt unable to communicate with people properly[c]
Felt ignored by people
Bodily discomfort
Had painful muscle cramps or spasms[c]
Had aches or pains in your joints or body
Felt unpleasantly hot or cold

Tan, Louis C.s., Nan Luo, Mohammed Nazri, Shu Chuen Li, and Julian Thumboo. "Validity and Reliability of the PDQ-39 and the PDQ-8 in English-speaking Parkinson's Disease Patients in Singapore." Parkinsonism & Related Disorders 10.8 (2004): 493-99. Web.

Appendix 2:

Behavioral metrics:

1. reachTargetTime
2. startMoveTime
3. reachTargetTimeMinusStartMoveTime
4. maxVelocity
5. timeIdle
6. timeInMotion
7. totalDistanceTravelled
8. initialDistanceFromTarget
9. normTotalTravelled
10. TimeOnTarget
11. NumberOfEntriesIntoTarget
12. numberOfExitsFromTarget
13. MotorPersistence

Each of the above behavioral metrics was calculated for the following:

- average for entire task
- average for each pattern (stationary, long stationary, linear, diagonal, and sinusoidal)
- average for each target size (large, medium, and small)
- average for each pattern at each target size (examples: stationary large and linear small)
- average for stationary targets without distractors
- average for stationary targets with distractors
- average for stationary targets with blue distractors
- average for stationary targets with white distractors
- average for each speed on sinusoidal task (fast, moderate, and slow)