HHS Public Access

Author manuscript

*IEEE Trans Neural Syst Rehabil Eng*. Author manuscript; available in PMC 2016 June 06.

Published in final edited form as:

*IEEE Trans Neural Syst Rehabil Eng*. 2009 December ; 17(6): 568–575. doi:10.1109/TNSRE.2009.2034461.

# Application of Modified Regression Techniques to a Quantitative Assessment for the Motor Signs of Parkinson's Disease

**Bambi R. Brewer**,

Department of Rehabilitation Science and Technology, University of Pittsburgh, Pittsburgh, PA 15260 USA

**Sujata Pradhan**,

Physical Therapy Department, University of Pittsburgh, Pittsburgh, PA 15260 USA. She is now with the Department of Rehabilitation Medicine, University of Washington, Seattle, WA 98195 USA

**George Carvell**, and

Physical Therapy Department, University of Pittsburgh, Pittsburgh, PA 15260 USA

**Anthony Delitto**

Physical Therapy Department, University of Pittsburgh, Pittsburgh, PA 15260 USA

Bambi R. Brewer: bbrewer@pitt.edu; Sujata Pradhan: sujatap@u.washington.edu; George Carvell: gcarvell@pitt.edu; Anthony Delitto: delitto@pitt.edu

## Abstract

Effective clinical trials for neuroprotective interventions for Parkinson's disease (PD) require a way to quantify an individual's motor symptoms and analyze the change in these symptoms over time. Clinical scales provide a global picture of function but cannot precisely measure specific aspects of motor control. We have used commercially available sensors to create a protocol called Advanced Sensing for Assessment of Parkinson's disease (ASAP) to obtain a quantitative and reliable measure of motor impairment in early to moderate PD. The ASAP protocol measures grip force as an individual tracks a sinusoidal or pseudorandom target force under three conditions of increasing cognitive load. Thirty individuals with PD have completed the ASAP protocol. The ASAP data for 26 of these individuals were summarized in terms of 36 variables, and modified regression techniques were used to predict an individual's score on the Unified Parkinson Disease Rating Scale based on ASAP data. We observed a mean prediction error of approximately 3.5 UPDRS points, and the predicted score accounted for approximately 76% of the variability of the UPDRS. These results demonstrate that the ASAP protocol can measure differences for individuals who are clinically different. This indicates that the ASAP protocol may be able to measure changes with time in the motor signs of an individual with PD.

## Introduction

TREATMENTS such as medication and exercise [1] are being developed in an attempt to slow or halt the progress of Parkinson's disease (PD); such interventions are termed neuroprotective. However, the evaluation of potential neuroprotective interventions for PD requires sensitive measures to quantify the progress of the disease. A precise measure of disease progression shortens the time period needed to determine whether a given

therapeutic intervention is slowing or stopping the disease process. Various types of PET and SPECT imaging that measure the deterioration of the nigrostriatal dopaminergic system have been widely researched as potential diagnostic and progression biomarkers for PD [2], [3], [4]. However, while the nigrostriatal damage measured by imaging methods is often correlated with functional decline, this correlation predicts less than half the variability observed in motor impairment as measured by the motor subscale of the Unified Parkinson Disease Rating Scale (UPDRS), the standard clinical assessment for PD [5]. The correlation between imaging and function is particularly poor early in the disease process [6]. The disconnect between damage in the brain and clinical functional impairment is also emphasized by the fact that more than 60% of the neurons in the basal ganglia are destroyed before clinical signs of PD are observed [2]. This indicates that the functional effects of neuroprotective intervention cannot be measured by current imaging techniques.

The Unified Parkinson Disease Rating Scale (UPDRS) is the scale most often used to assess the clinical symptoms of PD, including motor symptoms. This scale contains 42 questions and has a maximum score of 147 (a higher score indicates more severe PD) [7]. The UPDRS is divided into four sections: mentation, behavior, and mood; activities of daily living (ADL); motor examination; and complications of therapy. For individuals with early PD, the test–retest reliability of the UPDRS is 0.92 and the reliabilities of the ADL and motor subscales are 0.85 and 0.90, respectively [8]. The UPDRS is a useful tool to obtain a global picture of an individual's Parkinsonian symptoms, but it is primarily valuable after an individual demonstrates clear signs of PD. Early in the disease process, the utility of the UPDRS is limited by a floor effect. In addition, three of the sections of the UPDRS rely on patient self-report, which may limit the objectivity of this assessment. The Grooved Pegboard Test is an alternative clinical measure that can also be used for individuals with and without PD [9]. The Grooved Pegboard Test measures the time required to rotate and insert 25 pegs into holes of various orientations. This test is a complex motor task involving spatial perception and precision grip, and a correlation with PET imaging has been measured for the less affected hand of individuals with PD [10]. However, the results of the test provide no indication as to why a given individual performed well or poorly.

The goal of our work is to develop an objective, quantitative clinical assessment to measure the progression of early to moderate PD. We have created a protocol that uses sensing technology to obtain a quantitative measure of motor impairment in early to moderate PD. We incorporate the performance of multiple, simultaneous tasks in order to magnify the motor deficits of PD [11], [12], [13]. We use the abbreviation ASAP (Advanced Sensing for Assessment of Parkinson's disease) to collectively refer to our experimental environment and the assessment protocol that individuals complete in this environment. Our previous work has demonstrated significant differences in performance on the ASAP protocol between individuals with PD and age-matched controls [11]; in particular, the performance of individuals with PD deteriorated more than that of age-matched controls when the subjects performed a simultaneous cognitive task in addition to the required motor task.

The ASAP protocol collects a large amount of data for each individual; to be interpreted easily, this data must be combined to generate a single "score" for the assessment. In this paper, we describe the use of three modified regression techniques to combine the data

collected by the ASAP protocol into a single ASAP score. We show that the ASAP score can predict an individual's UPDRS score, and we examine those parts of the ASAP protocol that contribute most to this prediction. While the UPDRS is not a perfect assessment, it is a widely used and accepted measure of the signs and symptoms of PD. Comparison with the UPDRS establishes that the ASAP protocol can measure different scores for individuals in different clinical stages of PD. This indicates that the ASAP protocol is measuring the construct it is designed to measure, namely the signs of Parkinson's disease, and it indicates that the ASAP protocol may be able to measure motor changes over time for individuals with PD.

## Methods

### A. Subjects

Thirty individuals with PD participated in this experiment. All subjects remained off medication for 12 h before the testing session. All subjects had ratings between I and III on the Hoehn-Yahr scale when off medication for 12 h (median H-Y score of 2, mean H-Y score of 2.03) [14]. Each subject chose whether he or she preferred to be tested at the University of Pittsburgh or at home. The inclusion/exclusion criteria for this experiment are presented in Table I. All procedures were approved by the Institutional Review Board of the University of Pittsburgh, and all subjects provided informed consent. Of the 30 individuals who completed the experiment, four individuals with UPDRS scores greater than 40 were excluded from the regression analyses because the data were too sparse in this region to allow effective prediction. Thus, 26 individuals were included in the analyses reported here.

### B. Experimental Protocol

The experimental environment for this experiment is shown in Fig. 1(a). Two Nano17 6-axis force/torque sensors from ATI Industrial Automation were mounted on a custom-made, portable platform that was clamped to a table. Each sensor measured six axes of force and torque with resolutions of 0.003 N and 0.01 N-mm, respectively. Subjects used the index finger and thumb to exert force on the sensors, and the mean of the forces exerted on the two sensors was displayed on a computer screen [Fig. 1(b)]. Subjects attempted to modulate the exerted force in order to track a target waveform. Two target waveforms were used, a sine wave with a period of 7.5 s and pseudorandom waveform. The target wave scrolled continuously across the screen so that at each time point, the current target force and the current force exerted by the thumb and index finger were displayed in the center of the computer screen. In addition, the subject was shown a 12.5 s history of his or her performance relative to the target wave. In the case of the sine target, the subject was also shown the upcoming 12.5 s of the target wave. No information about the upcoming wave was given for the pseudorandom target.

The subject tracked each waveform for three minutes while force and torque data were recorded at 100 Hz. During the first minute, the subject tracked the target force with no external cognitive load. During the second minute, the subject tracked the target force while simultaneously counting down from 100 by 1. During the third minute, the subject tracked the target force while simultaneously counting down from 100 by 3. Each subject tracked

each target waveform (sine or pseudorandom) with both the right and left hands. Thus, each subject completed a total of four trials. Before completing these trials, each subject was given 2 min of training with the system using a target waveform that included periods of constant and ramping force. During each testing session, the physical therapist conducting the session also administered the UPDRS. The UPDRS was administered before the ASAP protocol.

### C. Data Analysis

To examine whether the data measured during our testing protocol could be used to predict an individual's UPDRS score, we computed three summary variables for each trial and cognitive load condition. The first summary variable was the tremor integral; we quantified the tremor exhibited by an individual by computing the power spectral density and calculating the area under this curve between 2 and 8 Hz. This range includes the 4–5 Hz window that typically bounds Parkinsonian tremor [15]. After computing the tremor integral, the data was filtered with a low-pass second-order dual-pass Butterworth filter with a cutoff frequency of 2 Hz. The filtered data was used to calculate the remaining summary variables. The second summary variable was the root-mean-square error (RMSE) between the target wave and the subject's force response. The third summary variable was the lag between the target waveform and the subject's force response, which was calculated as the time interval that maximized the cross-covariance between the two waves. The lag was bounded at 2 s for the sine target and 5 s for the pseudorandom target. If the covariance was less than 0.35, the lag was automatically set to the maximum value. These heuristic values were chosen as the maximum lag and minimum covariance at which any relationship was visually apparent between the target waveform and the subject response.

Each variable was calculated for each cognitive load condition for each trial. For each variable, the value of that variable in minute 1 (no external cognitive load) was subtracted from the values for minute 2 (counting down by 1) and minute 3 (counting down by 3). Thus, we obtained a total of 36 predictor variables for each individual (2 hands ×2 waveforms ×3 cognitive load conditions ×3 summary variables). Many of our participants had asymmetrical motor symptoms of PD. For this reason, we divided the data collected for both hands based on the side of better/worse performance, rather than based on the right/left side. The side of better performance was defined as the side with the lower RMSE error averaged over the sine and pseudorandom trials.

Because many of our variables were correlated with one another, standard least-squares linear regression was a poor choice for examining the relationship between the predictor variables and the subject's UPDRS score. One alternative to traditional linear regression is to reduce the dimensionality of the data using principal component analysis. Principal component analysis represents a large number of correlated variables using a smaller number of uncorrelated variables that account for most of the variance in the original data set [16]. We performed principal component analysis on our data set, extracting all components corresponding to an eigenvalue greater than 1 (Kaiser's criterion [17]). We then used the extracted components as the predictors in standard linear regression. We estimated the prediction error using leave-one-out cross-validation [16]. This means that data from a

single individual was removed from the data set, and the regression coefficients relating the extracted principal components to the UPDRS score were computed using the remaining data. These coefficients were then used to estimate the UPDRS score of the removed individual, and this estimate was compared to his or her actual UPDRS score. This process was repeated for each subject, and we computed the mean absolute prediction error.

As alternatives to standard regression using the principal components, we considered ridge regression and lasso regression, two modifications of standard least-squares regression that were designed to be used with correlated predictor variables. Like standard regression, each of these methods calculates a set of coefficients that corresponds to the linear combination of predictor variables that best predicts the outcome variable (the UPDRS score). However, ridge and lasso regression differ from standard regression in the details of the optimization that determines these coefficients. Ridge and lasso regression were used with all 36 original predictor variables, not the extracted principal components. Ridge and lasso regression were not used in combination with principal component analysis because the principal components are uncorrelated; since ridge and lasso regressions were designed to reduce the problem of correlated predictors, the use of either in combination with principal component analysis is somewhat redundant.

Ridge regression finds the coefficients $\hat{\beta}^{\text{ridge}}$ that minimize the quantity $\|y - \hat{\beta}^{\text{ridge}} X\|^2 + \lambda \|\hat{\beta}^{\text{ridge}}\|^2$ where $X$ is an $n \times p$ matrix corresponding to $p$ predictors measured for each of $n$ data points and $y$ is a vector of length $n$ containing the known outcome for each point in $X$ [16], [18]. $X$ and $y$ are known as the training set for the ridge regression. Because the training data is used to determine the coefficients, the success of the regression must be determined based on an independent test set. Ridge regression relies upon a parameter $\lambda$ that is supplied by the experimenter. This parameter acts to shrink the regression coefficients by imposing a penalty for large coefficients; this shrinkage reduces the amount of variance in the coefficients at the cost of an increase in bias. A large value for $\lambda$ increases the amount of shrinkage but also increases the amount of bias in the coefficients [18].

We estimated the prediction error of ridge regression for our data using leave-one-out cross-validation [16]. Each individual in turn was considered as the test set, with the remaining 25 individuals serving as the training set. Within the training set, we then chose a value for $\lambda$ using five-fold cross-validation [16]. We considered $\lambda = 0, 0.001, 0.002, \ldots, 1$ [18]. The training set was partitioned into five subsets. A subset was removed from the training set and ridge regression (Matlab “ridge” function) was used with the remaining data to find, for every value of $\lambda$, the linear combination of predictor variables that best fit the UPDRS scores. The resulting coefficients were then used to estimate the UPDRS scores of the individuals in the removed subset, and the error in these predictions was computed as a function of $\lambda$. This process was repeated for each of the five subsets of the training set, and the mean absolute prediction error was computed as a function of $\lambda$. Because a larger value of $\lambda$ increases the bias of ridge regression [18], we chose the smallest value of $\lambda$ such that the prediction error was within one standard error of the minimum mean prediction error. After choosing a value for $\lambda$, we used this value and the entire training set to generate the coefficients $\hat{\beta}^{\text{ridge}}$. These coefficients were then used to predict the UPDRS score for the individual in the test set. A coefficient with a 95% confidence interval that did not include 0

was considered to be significantly different from 0, which indicates that the predictor variable corresponding to that coefficient has a meaningful effect on prediction of the UPDRS score. The coefficients $\hat{\beta^{ridge}}$ were computed after standardizing each predictor variable to have a mean of 0 and a standard deviation of 1; thus, the values of the ridge regression coefficients can be used to assess the contribution of each variable to the prediction.

Lasso regression is a regression technique that computes the coefficients $\hat{\beta^{lasso}}$ minimize the quantity $\|y - \hat{\beta^{lasso}} X\|^2$ subject to $\sum_{j=1}^{p} |\beta_j| \leq s$, where the $\beta_j$ are elements of $\hat{\beta^{lasso}}$. Lasso regression requires a parameter $s$ that is between 0 and $s_0$, where $s_0$ is the sum of the absolute values of the coefficients found using standard linear regression. Lasso regression has the effect of setting equal to 0 the coefficients of variables with the least predictive value. Lasso regression is similar to ridge regression in that it acts to shrink the regression coefficients; the difference between the two methods is in the form of the optimization function. We used a set of Matlab functions implementing lasso regression that were written by Mark Schmidt.[1] We estimated the prediction error of lasso regression for our data using leave-one-out cross-validation, as described above. We chose $s$ using five-fold cross-validation with the training set; the values of $s$ considered were $s = 0.05s_0, 0.1s_0, 0.15s_0, \ldots, s_0$. We then used the chosen value of $s$ and the entire training set to calculate $\hat{\beta^{lasso}}$. These coefficients were used to predict the UPDRS score of the individual not included in the training set. A coefficient with a 95% confidence interval that did not include 0 was considered to be significantly different from 0. Because the coefficients $\hat{\beta^{lasso}}$ are for the unstandardized data, the magnitudes of the coefficients do not necessarily indicate the relative importance of the predictors.

For comparison, we also computed the mean prediction error of standard linear regression for the entire data set using leave-one-out cross-validation.

## Results

Principal component analysis with the Kaiser criterion resulted in nine extracted components that were used as the predictor variables for standard regression. The predicted and actual UPDRS scores for every subject in the data set are shown in Fig. 2(a). The predicted UPDRS score was significantly correlated with the actual UPDRS score ($p = 0.004$) with $R = 0.54$. The mean absolute prediction error was 7.06 ± 1.37 UPDRS points (mean ± standard error).

We consider next the results for ridge regression. Fig. 3(a) shows an example of the average prediction error for the training set as a function of $\lambda$. The value of $\lambda$ selected by five-fold cross-validation was 0.001 for this example, though the curve is relatively flat except for a sudden increase when $\lambda = 0$. The mean value of $\lambda$ selected by five-fold cross-validation was0.0036 ± 0.0022. The mean absolute prediction error of ridge regression for the entire data set was 3.58 ± 0.69 UPDRS points [Fig. 2(b)]. The correlation between the predicted and actual UPDRS scores was highly significant ($p < 0.001$) with $R = 0.87$.

Twenty-two variables were found to have coefficients more than two standard deviations from zero; details of the variables corresponding to the 10 largest coefficients are found in Table II.

For lasso regression, Fig. 3(b) shows an example of the mean prediction error for the training set as a function of $s$. For this example, the chosen value of $s$ was $0.4s_0$. The mean value of $s$ for our data set was $0.31s_0 \pm 0.020s_0$. For lasso regression, the mean absolute prediction error for the entire data set was 4.57 ± 0.84 UPDRS points. The predicted UPDRS score as a function of actual UPDRS score is shown in Fig. 2(c) for the entire data set. The correlation between the actual and predicted scores was significant ($p < 0.001$)) with $R = 0.78$. For lasso regression, six variables had coefficients that were more than two standard deviations from zero. These variables and their corresponding coefficients are detailed in Table III.

For comparison, Fig. 2(d) shows the UPDRS score predicted by standard linear least-squares regression as a function of actual UPDRS score. The mean prediction error was 167.1 ± 85.4 UPDRS points. The predicted and actual UPDRS scores were not significantly correlated ($R = 0.23$, $p = 0.26$).

## Discussion

The application of high-precision sensors to the assessment of PD can enable investigators to quantify minute changes in the motor signs of this disease. Such a quantitative measure, in combination with brain imaging techniques, will be extremely valuable for clinical trials of potential neuroprotective interventions for PD. In addition, a quantitative measure of motor signs would enable physicians to more easily optimize the medication regime for a specific patient. Such a quantitative measure is not intended as a replacement for the UPDRS; the UPDRS is an inexpensive, well-studied, and easily administered test of the symptoms of PD. Our goal is the supplement the UPDRS with a precise, objective assessment to quantify fine motor signs of PD.

A variety of systems have been investigated for the assessment of motor symptoms of PD. For example, Cleveland Medical Devices, Inc., has developed the Kinesia system, a wireless wearable system that measures Parkinsonian tremor [19]; they measured good correlation between tremor measurements made by their system and clinician ratings of tremor. Montgomery *et al.* [20], [21] used a manipulandum and an LED display to measure wrist flexion as a part of a larger diagnostic test battery, but the specificity and sensitivity of this subtest were only 38% and 67.5%, respectively, for a group of individuals with probable PD. Other proposed assessment tools utilize target tracking, an established paradigm for measuring differences between individuals with and without PD [22], [23]. For example, Allen *et al.* [24] used a joystick and steering wheel designed for video games to measure the ability of individuals to track pseudorandom or sinusoidal waveforms; they measured a significant between-group difference for individuals with and without PD. Digitizing tablets have also been used to quantify performance by individuals with PD while tracing a target spiral [25] or other waveforms [23]. Saunders–Pullman *et al.* [26] showed that several

variables derived from spiral analysis are significantly correlated with UPDRS score and several subscales of the UPDRS for individuals with early PD.

Though a variety of technologies have been explored for PD assessment, few have been used to measure early motor changes in individuals with PD, which is the goal of our research. In addition, our work utilizes a simultaneous task paradigm; performance of a simultaneous cognitive task magnifies the motor deficits of individuals with PD [11], [12], [13]. We believe this will make it easier for us to quantify early deficits and changes in early symptoms. In the work presented here, we compare the ASAP protocol to the UPDRS, the standard clinical scale for PD. We found that there exists a linear combination of ASAP variables that has a strong correlation with the UPDRS score. The UPDRS scores predicted by ridge regression account for 76% of the variance in the actual UPDRS scores. This is somewhat surprising, given that the UPDRS measures a wide range of symptoms of PD while our test focuses only on fine motor control and the change in fine motor control due to cognitive load. However, the ASAP protocol does measure how the major motor signs of PD, particularly bradykinesia and tremor, affect fine motor control. The effects of these signs on fine motor control are likely to be highly correlated with their effects on other motor activities examined by the UPDRS, such as gait and speech. In addition, many signs and symptoms of the disease will be correlated with one another because they are all related to disease progression. This may explain why the ASAP score can explain more than 75% of the variance in the UPDRS even though its focus is fine motor control.

These results show that the ASAP protocol does yield different scores for individuals with divergent clinical scores and indicates that the ASAP protocol may be useful for measuring the progression of an individual's Parkinsonian symptoms. The fact that the ASAP protocol can be highly correlated with the UPDRS also begins to establish the construct validity of the ASAP protocol by demonstrating that this assessment does measure the construct it is designed to measure, namely the motor signs of Parkinson's disease. We anticipate that the ASAP protocol will have performance superior to the UPDRS in some areas, specifically in the quantification of fine motor control in the early stages of the disease process. However, we feel that the comparison of the ASAP protocol to established clinical measures is an important step in the creation of this novel assessment. Construct validity is an essential step in the formation of traditional clinical assessments, and this step should not be neglected for assessments utilizing technology.

The application of technology to assessment of disease often results in large quantities of correlated data. In this work, the sensors used to quantify deficits in fine motor control enabled us to collect a great deal of data, approximately 18000 measurements of force per trial. We summarized this data in terms of three variables with each variable calculated for each hand, waveform, and cognitive load condition for every subject. However, even after this data reduction, we were left with 36 correlated variables; less dramatic data reduction could lead to hundreds of variables. Clinicians cannot sort through large numbers of variables in order to track patient performance and to compare individuals. To interpret the information in a clinically meaningful way, the data must be synthesized into a single "score" that summarizes an individual's performance on the assessment. The correlations between variables and the number of variables relative to the number of subjects necessitate

the use of modified regression techniques to performance this synthesis. In this work, we utilized standard regression after principal component analysis, ridge regression, and lasso regression to create a linear combination of variables that predicts an individual's UPDRS score. Principal component analysis has been used in previous applications of technology to assessment of PD [27]. Ridge regression has been applied to regression problems with correlated predictors for decades [28]; it has been shown to perform well for automatic ICD-9 coding of patient records [29] and predictions from gene expression data [30]. Lasso regression is a newer technique [16]that has also been used with success with gene expression data [30]. These techniques were chosen because they are popular techniques that are intuitively similar to standard linear regression and have been applied with success to comparable problems.

For our data, ridge regression performed best, with a prediction error of about 3.5 UPDRS points. Lasso regression had a prediction error of approximately 4.5 UPDRS points. These two methods are conceptually similar, in that both act to shrink the regression coefficients. Standard regression after principal component analysis performed worst with a prediction error of about 7 UPDRS points. However, all results were far superior to the enormous prediction error for standard linear regression; this demonstrates the inadequacy of standard regression for data sets with highly correlated predictors and the importance of applying appropriate data mining techniques to data-intensive projects such as this one. Such approaches will only become more important as we look at additional variables or examine each variable over smaller time intervals.

Variables identified as important predictors consisted primarily of tremor and lag variables for ridge regression. These correspond to the tremor and bradykinesia symptoms assessed by the UPDRS. Lasso regression identified tremor and RMSE variables as most important. RMSE is dominated by the degree to which individuals undershoot the target wave; it is similar to the micrographia (abnormally small handwriting) observed in individuals with PD. Important predictors were primarily variables from the more affected arm for ridge regression and the less affected arm for lasso regression. The reason for the differences in the variables chosen by ridge and lasso regression is unknown. However, because the prediction error of ridge regression was lower, we have greater confidence in the variables selected by ridge regression. Most of the important predictors for both methods were based on data taken as the user performed a motor and a cognitive task simultaneously. This is an interesting contrast to the results of Cordell *et al.* [27], who found that measures of performance in driving (an activity inherently involving simultaneous tasks) showed little correlation with standard clinical measures for PD, including the UPDRS. This may be because Cordell used principal component analysis (without regression) to combine multiple driving variables into a single score. This is an equally valid approach, but we chose to use a regression technique because we were most interested here in determining whether it was possible to accurately predict UPDRS score using ASAP data.

One limitation of this work is our relatively small sample size (data from 26 individuals used in this analysis). This sample size was appropriate given the preliminary nature of this study, but limits the ability of our results to generalize to the larger population of individuals with early PD. Relative to our sample size, we considered a large number of predictor variables;

this may lead to overfitting of the regression function to the training data. This problem is mitigated somewhat by using principal component analysis (with Kaiser's criterion) or lasso regression, which reduce the number of predictors used. Similarly, ridge regression assigns very small coefficients to many variables, thus reducing their influence. In order to investigate the possibility of overfitting, we utilized leave-one-out cross-validation, in which the regression function is tested using an independent test set. Based on our results, overfitting does not seem to negatively affect our analysis, but this possibility remains a major limitation of this work. We plan to address this concern in our future work by recruiting larger samples and through the investigation of other methods of feature selection to reduce the number of predictor variables considered.

It is important to note that our assessment will not measure all aspects of PD, particularly nonmotor manifestations of the disease. We anticipate that our assessment will be used in conjunction with imaging and other clinical evaluations. However, this work shows that ASAP data can be combined to produce a score that shows a strong correlation with the clinical UPDRS score. This begins the process of validating our assessment and shows that we can measure different ASAP scores for individuals who are clinically different. This indicates that we may be able to track important clinical changes using our test. However, before using our assessment to evaluate the success of neuroprotective interventions for PD, we must establish that this test is reliable and can measure longitudinal change. To this end, the test–retest reliability of the ASAP protocol is currently being measured, and we plan to follow individuals over time to examine the progression of PD using the ASAP protocol.

## Acknowledgments

This work was supported in part by the National Academies Keck Futures Initiative and in part by the Foundation of Physical Therapy Promotion of Doctoral Studies II Scholarship 2007.

## References

1. Crizzle AM, Newhouse IJ. Is physical exercise beneficial for persons with Parkinson\'s disease? Clin J Sport Med. 2006; 16:422–5. [PubMed: 17016120]
2. Becker G, Muller A, Braune S, Buttner T, Benecke R, Greulich W, Klein W, Mark G, Rieke J, Thumler R. Early diagnosis of Parkinson\'s disease. J Neurol. 2002; 249:III/40–8.
3. Brooks DJ, Frey KA, Marek KL, Oakes D, Paty D, Prentice R, Shults CW, Stoessl AJ. Assessment of neuroimaging techniques as biomarkers of the progression of Parkinson\'s disease. Exp Neurol. 2003; 184:S68–79. [PubMed: 14597329]
4. Siderowf A, Stern MB. Preclinical diagnosis of Parkinson\'s disease: Are we there yet? Curr Neurol Neurosci Rep. 2006; 6:295–301. [PubMed: 16822349]
5. Pirker W. Correlation of dopamine transporter imaging with Parkinsonian motor handicap: How close is it? Mov Disord. 2003; 18:S43–51. [PubMed: 14531046]
6. Bohnen NI, Albin RL, Koeppe RA, Wernette KA, Kilbourn MR, Minoshima S, Frey KA. Positron emission tomography of monoaminergic vesicular binding in aging and Parkinson disease. J Cereb Blood Flow Metab. 2006; 26:1198–212. [PubMed: 16421508]
7. Hoehn; Yahr. Staging of Parkinson\'s disease. Unified Parkinson Disease Rating Scale (UPDRS), and Schwab and England Activities of Daily Living [online]. Available: http://www.parkinson.org
8. Siderowf A, McDermott M, Kieburtz K, Blindauer K, Plumb S, Shoulson I. Test-retest reliability of the unified Parkinson\'s disease rating scale in patients with early Parkinson\'s disease: Results from a multicenter clinical trial. Mov Disord. 2002; 17:758–63. [PubMed: 12210871]

9. Mitrushina, MN.; Boone, KB.; D'Elia, LF. Handbook of Normative Data for Neuropsychological Assessment. Oxford Univ. Press; 1999.
10. Bohnen NI, Kuwabara H, Constantine GM, Mathis CA, Moore RY. Grooved pegboard test as a biomarker of nigrostriatal denervation in Parkinson\'s disease. Neurosci Lett. 2007; 424:185–9. [PubMed: 17714864]
11. Pradhan S, Brewer BR, Carvell GE, Sparto PJ, Delitto A, Matsuoka Y. Effects of a secondary cognitive task on assessment of fine motor control using force tracking in individuals with Parkinson\'s disease. J Neurologic Phys Therapy.
12. Brown RG, Marsden CD. Dual task performance and processing resources in normal subjects and patients with Parkinson\'s disease. Brain. 1991; 114:215–31. [PubMed: 1998883]
13. Dalrymple-Alford JC, Kalders AS, Jones RD, Watson RW. A central executive deficit in patients with Parkinson\'s disease. J Neurol Neurosurg Psychiatry. 1994; 57:360–7. [PubMed: 8158188]
14. Hoehn MM, Yahr MD. Parkinsonism: Onset, progression and mortality. Neurology. 1967; 17:427–42. [PubMed: 6067254]
15. Kandel, ER.; Schwartz, JH.; Jessell, TM. Principles of Neural Science. McGraw-Hill; 2000. p. 853-867.
16. Hastie, T.; Tibshirani, R.; Friedman, JH. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer; 2001.
17. Field, A. Discovering Statistics Using SPSS. Sage; 2005.
18. Marquardt DW, Snee RD. Ridge regression in practice. Amer Stat. 1975; 29:3–20.
19. Giuffrida JP, Riley DE, Maddux BN, Heldman DA. Clinically deployable Kinesia TM technology for automated tremor assessment. Mov Disord. 2009; 24(5):723–730. [PubMed: 19133661]
20. Montgomery EB Jr, Koller WC, LaMantia TJ, Newman MC, Swanson-Hyland E, Kaszniak AW, Lyons K. Early detection of probable idiopathic Parkinson\'s disease: I. Development of a diagnostic test battery. Mov Disord. 2000; 15:467–73. [PubMed: 10830410]
21. Montgomery EB Jr, Lyons K, Koller WC. Early detection of probable idiopathic Parkinson\'s disease: II. A prospective application of a diagnostic test battery. Mov Disord. 2000; 15:474–8. [PubMed: 10830411]
22. Carey JR, Deskin KA, Josephson KT, Wichmann RL. Sex differences in tracking performance in patients with Parkinson\'s disease. Arch Phys Med Rehabil. 2002; 83:972–7. [PubMed: 12098158]
23. Hocherman S, Giladi N. Visuomotor control abnormalities in patients with unilateral Parkinsonism. Neurology. 1998; 50:1648–54. [PubMed: 9633706]
24. Allen DP, Playfer JR, Aly NM, Duffey P, Heald A, Smith SL, Halliday DM. On the use of low-cost computer peripherals for the assessment of motor dysfunction in Parkinson\'s disease quantification of bradykinesia using target tracking tasks. IEEE Trans Neural Syst Rehabil Eng. 2007; 15(1):286–94. [PubMed: 17601199]
25. Rudzinska M, Izworski A, Banaszkiewicz K, Bukowczan S, Marona M, Szczudlik A. Quantitative tremor measurement with the computerized analysis of spiral drawing. Neurol Neurochir Pol. 2007; 41:510–6. [PubMed: 18224573]
26. Saunders-Pullman R, Derby C, Stanley K, Floyd A, Bressman S, Lipton RB, Deligtisch A, Severt L, Yu Q, Kurtis M, Pullman SL. Validity of spiral analysis in early Parkinson\'s disease. Mov Disord. 2008; 23:531–7. [PubMed: 18074362]
27. Cordell R, Lee HC, Granger A, Vieira B, Lee AH. Driving assessment in Parkinson\'s disease-A novel predictor of performance? Mov Disord. 2008; 23(9):1217–1222. [PubMed: 18528878]
28. Jain RK. Ridge regression and its application to medical data. Comput Biomed Res. 1985; 18:363–8. [PubMed: 4042638]
29. Xu J, Yu S, Bi J, Lita LV, Niculescu RS, Rao RB. Automatic medical coding of patient records via weighted ridge regression. Proc 6th Int Conf Mach Learn Appl. 2007:260–265.
30. Bovelstad HM, Nygard S, Storvold HL, Aldrin M, Borgan O, Frigessi A, Lingjaerde OC. Predicting survival from microarray data A comparative study. Bioinformatics. 2007; 23:2080–7. [PubMed: 17553857]

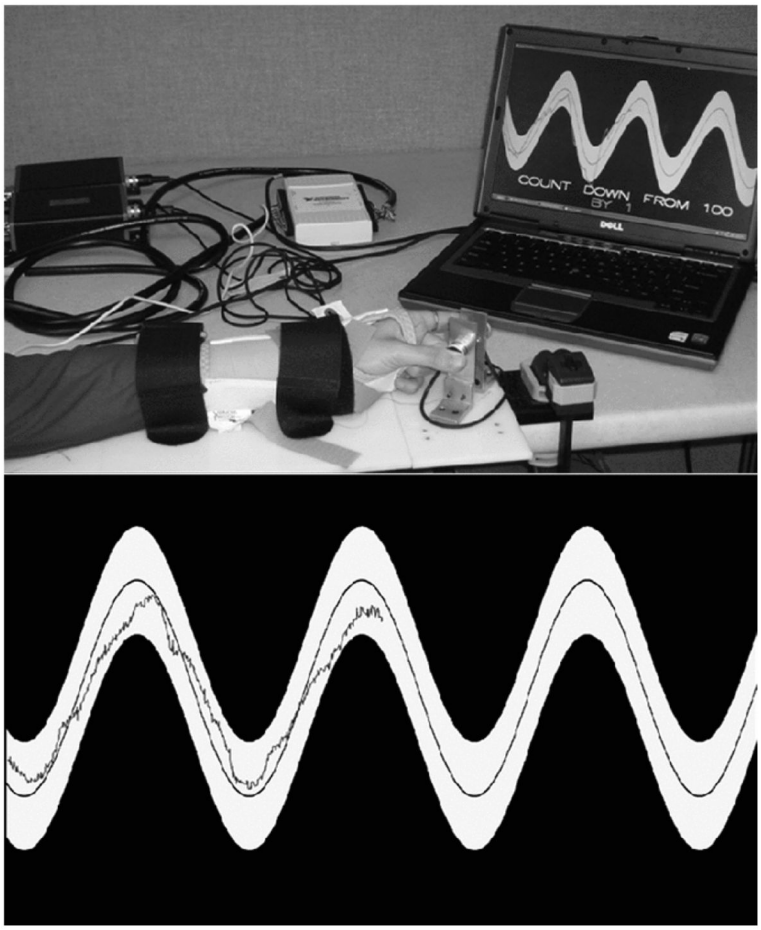


**Fig. 1.**

Experimental setup. The individual used the index finger and thumb to isometrically exert force on the force/torque sensors. He or she modulated the force exerted in order to track a target wave shown on the computer screen. The target wave scrolled continuously across the screen. The current subject force and target force were shown at the center of the screen. A 12.5 s history of the subject's performance relative to the target was also displayed. For the sine target, the upcoming 12.5 s of the target was also displayed.

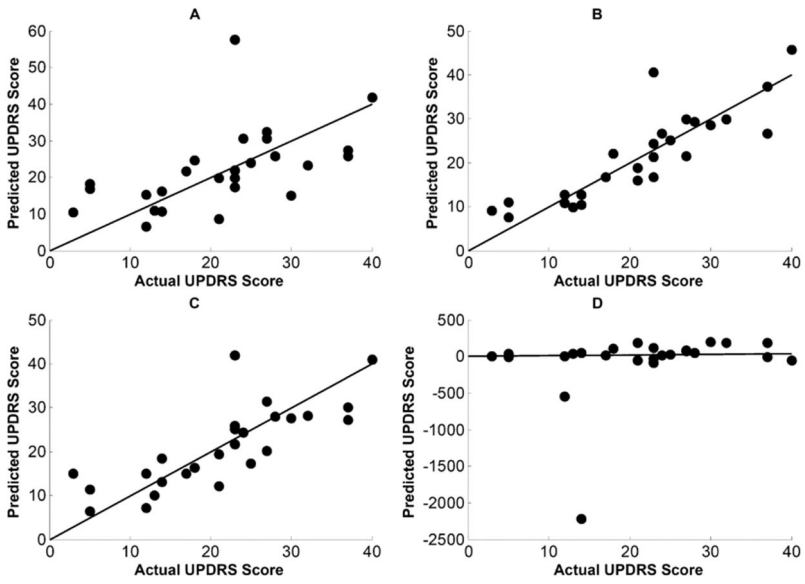


**Fig. 2.**

(a) Standard regression after principal component analysis. The UPDRS score predicted by standard regression after principal component analysis is plotted as a function of the actual UPDRS score for each individual with PD. The line has a slope of 1; if standard regression after principal component analysis perfectly predicted actual UPDRS score, all points would lay directly on this line. The predicted UPDRS score was significantly correlated with the actual UPDRS score ($p = 0.004$) and the mean absolute prediction error was 7.06 ± 1.37 UPDRS points (mean ± standard error). (b) Ridge regression. The predicted UPDRS score was significantly correlated with the actual UPDRS score ($p < 0.001$) and the mean absolute prediction error was 3.58 ± 0.69 UPDRS points. (c) Lasso regression. The predicted UPDRS score was significantly correlated with the actual UPDRS score ($p < 0.001$) and the prediction error was 4.57 ± 0.84 UPDRS points. (d) Standard least -squares regression. The predicted UPDRS was not correlated with the actual UPDRS score ($p = 0.26$), and the prediction error was 167.1 ± 85.4 UPDRS points. A. Regression after PCA. B. Ridge regression. C. Lasso regression. D. Standard regression.

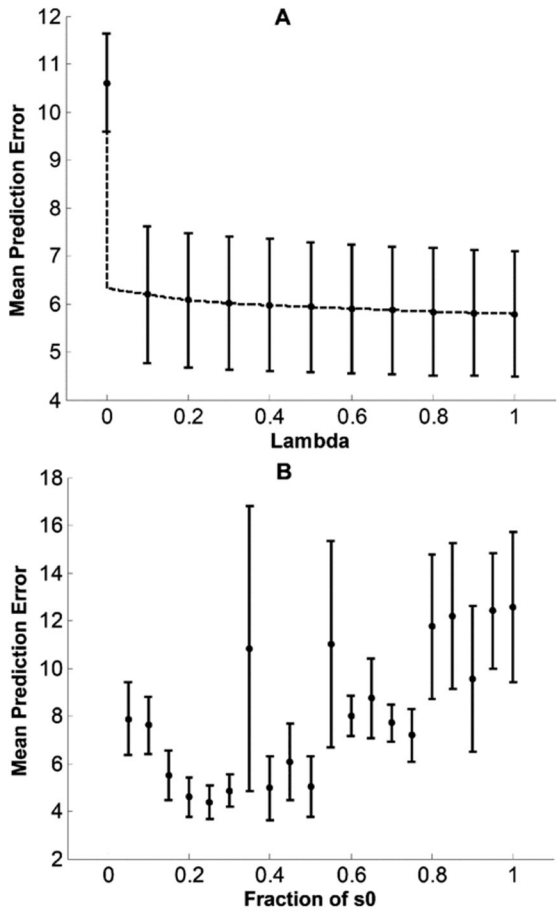


**Fig. 3.**

(a) Ridge regression. An example of the mean prediction error for the training set as a function of $\lambda$. The standard error is shown for selected points. This curve was computed using five-fold cross-validation over the training set. The ridge parameter $\lambda$ was selected as the smallest $\lambda$ such that the prediction error was within one standard error of the minimum prediction error. (b) Lasso regression. An example of the mean prediction error for the training set as a function of $s$. The standard error is shown for each point. This curve was computed using five-fold cross-validation over the training set. The lasso parameter $s$ was selected as the largest $s$ such that the prediction error was within one standard error of the minimum prediction error. A. Ridge regression. B. Lasso regression.

## Table 1

| Inclusion Criteria | Exclusion Criteria |
|---|---|
| 1. Previously established diagnosis of Parkinson's disease (noted by a physician) and a reported history of symptoms of slowed movement, tremor, or difficulty initiating movement for a minimum of one year and/or Hoehn-Yahr scale rating documented by the participant's physician. | 1. Restriction of movement in the upper extremities. |
| 2. 18 years of age or older. | 2. Sensory loss in the hand as determined by superficial sensory testing. |
| 3. Score of 27 or greater on the Mini Mental State Exam. | 3. Loss of vibration in the hand. |
| 4. No history of concurrent CNS disease. | 4. Subject unable to stay off medications 12 hours prior to the appointment. |
| | 5. Inability to come to the testing site accompanied by a friend or family member or unwilling to be tested at their home. |
| | 6. Presence of dyskinesia. |

**Table 2**

Twenty-two variables had ridge regression coefficients that were more than 2 standard deviations from zero. This table gives the details of the ten variables with the largest coefficients. The mean and standard deviation for each coefficient are given.

| Variable | Trial | Affected Arm | Time | Coefficient |
|---|---|---|---|---|
| Tremor integral | Sine | More | Min 1 | 3.30 ± 0.50 |
| Lag | Sine | More | Min 1 | 3.16 ± 0.62 |
| Tremor integral | Sine | More | Min 2 – Min 1 | 2.38 ± 0.63 |
| Lag | Sine | More | Min 2 – Min 1 | 2.81 ± 0.68 |
| Tremor integral | Pseudorandom | More | Min 2 – Min 1 | 5.00 ± 0.44 |
| RMSE | Pseudorandom | More | Min 2 – Min 1 | −2.43 ± 0.24 |
| Lag | Sine | More | Min 3 – Min 1 | −2.71 ± 0.48 |
| Lag | Pseudorandom | More | Min 3 – Min 1 | 2.41 ± 0.46 |
| Lag | Pseudorandom | Less | Min 2 – Min 1 | −3.98 ± 0.29 |
| Tremor integral | Pseudorandom | Less | Min 3 – Min 1 | −5.51 ± 0.35 |

**Table 3**

This table gives the details of the six variables whose coefficients for lasso regression were more than 2 standard deviations from zero. The mean and standard deviation for each coefficient are given.

| Variable | Trial | Affected Arm | Time | Coefficient |
|---|---|---|---|---|
| Tremor integral | Sine | More | Min 2 – Min 1 | 2.20 ± 0.85 |
| RMSE | Pseudorandom | More | Min 2 – Min 1 | 2.48 ± 0.17 |
| RMSE | Sine | Less | Min 1 | −0.27 ± 0.12 |
| Tremor integral | Sine | Less | Min 3 – Min 1 | −1.77 ± 0.28 |
| RMSE | Sine | Less | Min 3 – Min 1 | −0.33 ± 0.16 |
| RMSE | Pseudorandom | Less | Min 3 – Min 1 | −1.35 ± 0.22 |