PPMI DATA Challenge 2016

Principal Investigator/Applicant: Brian J Smith, PhD
Institution/Company/University: University of Iowa
Date PPMI dataset downloaded: 8/23/2016
Challenge Question Analyzed:
☑WHAT FACTORS AT BASELINE PREDICT CLINICAL PROGRESSION?

Summary of Results Table

Summary of Results	*We developed models and identified baseline factors to predict clinical progression of Parkinson's Disease (PD) across a range of motor, non-motor, and imaging outcomes; and synthesized the results into a decision support system for clinical trials design.*
Description of Methods Used	*Cross-validation was used to assess the accuracies of a large set of modeling techniques, including traditional regression and machine learning, in predicting PD outcomes. After selecting the best fitting model unique to each outcome, baseline predictors of clinical progression were determined by variable importance.*
Impact of Results	*Our predictive models were integrated into a web application that will allow clinical trialists to determine appropriate sample sizes for their studies based on desired outcomes of interest. The models provide baseline predictions needed to determine eligibility. This development could greatly improve the efficiency and efficacy of PD clinical trials.*

A Decision Support System for Parkinson's Clinical Trials Design

Brian J Smith, Brandon D Butcher, Stephanie J Lussier, Ryan A Peterson

Introduction

Methods are needed to predict patient outcomes in Parkinson's disease (PD). By better understanding outcomes, PD can be treated more effectively and in a patient-specific manner. The PPMI study is an ideal setting in which to apply the latest developments in predictive modeling and machine learning due to the large number of patients and disease-related information collected over time.

In this project, we utilize PPMI data to predict measures of clinical progression from baseline patient factors. Our approach involves the development of a comprehensive suite of R programs to (1) import the data, (2) apply data management techniques to create analytic datasets, (3) develop predictive models, and (4) integrate results into a web application that estimates sample sizes needed for clinical trials performed in patients who are predicted to progress more rapidly.

PPMI Study Data and Management

All available PPMI datasets and study documents were downloaded. The datasets were imported individually into R datasets. Subject characteristics, biospecimens, imaging, medical history, motor and non-motor assessments, and study enrollment datasets were merged by patient number and visit to

create analytic datasets of baseline predictors and measures of clinical progression through year 2 (V01-V06). Analysis focused on enrolled de novo PD subjects. Baseline factors were excluded if there was more than 20% missingness; otherwise, missing values were imputed as described in the Methods section. Subjects were excluded in analyses if the progression measure of interest was missing.

Predictive Modeling

Selection of Outcome Variables

Outcome variables that indicate clinical progression were chosen by review of the 2016 PPMI annual investigators meeting. Indicators of clinical progression were grouped into three domains: (1) Imaging, (2) Non-motor, and (3) Motor. For each outcome, clinical progression was defined as the absolute change from baseline to each of years 1 and 2, the relative baseline change, or the two-year rate of change (slope) estimated using all available interim measurements. In all, over 100 baseline factors were considered in modeling these outcomes.

Models Tested

The following predictive modeling methods were applied to each outcome: stochastic gradient boosting (GBM), lasso and elastic net regularization (ENET)*, AIC-based stepwise variable selection (STEP)*, neural networks (NNET), partial least squares (PLS)*, random forests (RF), and support vector machines with linear or radial basis functions (SVM). Asterisks denote methods that are linear and additive in the predictors. Models were fit with the caret and associated R packages to find the ones with best predictive accuracy as measured by root mean squared error (RMSE).

Each learning method has its own strengths and weaknesses. Common considerations when using machine learning are model interpretability and flexibility. We chose to include a set of methods that spans these two considerations. For instance, ENET and STEP have the forms of linear regression models and were thus chosen for their interpretability. Conversely, GBM, NNET, RF, and SVM are more flexible models that allow for non-linear and/or interaction effects of baseline predictors and were chosen for their potential to capture more complex relationships with PD progression.

Importance of Cross-Validation

The recommendations of Kuhn (2016) were followed to estimate RMSE with 5 repeats of 10-fold cross validation in order to obtain honest, out-of-sample estimates of predictive accuracy. Our approach

ensures realistic estimates of accuracy in predicting new patient outcomes. Conversely, within-sample estimates, obtained using the same data for model fitting and prediction, would be overly optimistic.

Repeated cross-validation was implemented as follows: (1) randomly partition the data into 10 folds, (2) predict the outcomes in each fold with a model fit to the remaining folds, and (3) repeat the partitioning and prediction 5 times. Within each fold, k-Nearest Neighbors was employed to impute missing predictor values before fitting the corresponding model. Fold-specific RMSE and R^2 were computed and averaged over all folds to obtain final estimates of predictive accuracy.

Results

Our modeling methods were applied to each outcome, and the best models selected based on cross-validated RMSE. Since RMSE depends on the scale of an outcome, predictive accuracy is reported here in terms of R^2 due to its standardized scale (0-1) and familiar interpretation. Year 2 outcome results are summarized in the adjacent table.

Motor Outcome	**R^2**	**Nonmotor Outcome**	**R^2**
UPDRS[c,SVM]	5.5	ESS Total[c,ENET]	17.0
UPDRS I[a,ENET]	9.0	GDS Total[c,ENET]	18.6
UPDRS II[c,ENET]	4.6	JLO Total[a,ENET]	21.1
UPDRS III[c,RF]	9.2	MCA Total[a,ENET]	27.5
Imaging Outcome	**R^2**	QUIP Total[c,ENET]	29.3
AI Caudate[c,ENET]	24.3	REM Total[a,ENET]	12.6
AI Putamen[c,RF]	42.4	Scopa Total[c,ENET]	11.7
Mean Putamen[b,RF]	12.3	STAI Total[c,ENET]	23.5
Mean Striatum[b,PLS]	5.1	Total Recall[a,ENET]	28.6

[a] Absolute baseline change at year 2
[b] Relative baseline change at year 2
[c] 2-year rate of change

R^2 estimates for AI Putamen, MCA Total, QUIP Total, and Total Recall are reasonably high for human subjects research, suggesting that baseline predictors can explain an appreciable amount of variation in these outcomes. For other outcomes, R^2 estimates are comparatively low. This could be due to a few reasons. First, the R^2 estimates are legitimate, out-of-sample estimates of predictive accuracy. Within-sample prediction estimates would yield inflated R^2. For example, the within-sample R^2 for Mean Putamen is 84.2 vs the reported out-of-sample estimate of 12.3. Second, we hypothesize that PD medication use is a confounder. Patients with higher UPDRS scores at an interim visit are more likely to be treated with PD medications than those with lower scores, making prediction of the baseline change more difficult. While we could have controlled for PD medication use, it was not a baseline predictor in our cohort of patients who were not taking medication at enrollment. Moreover,

we felt that in order to successfully design clinical trials one should not assume that patients will or will not get PD medication in the future.

Relative variable importance (VI) is summarized in the table below for 4 select models. VI values are scaled so that 100 represents the most important predictor in the model and 0 the least important. The methodology for calculating VI can be found in the caret documentation. Only the top 10 predictors are reported here due to space limitations. The "Model Info" table in the web application described in the next section provides a more detailed list of the relative importance of baseline predictors across all outcomes considered in our project.

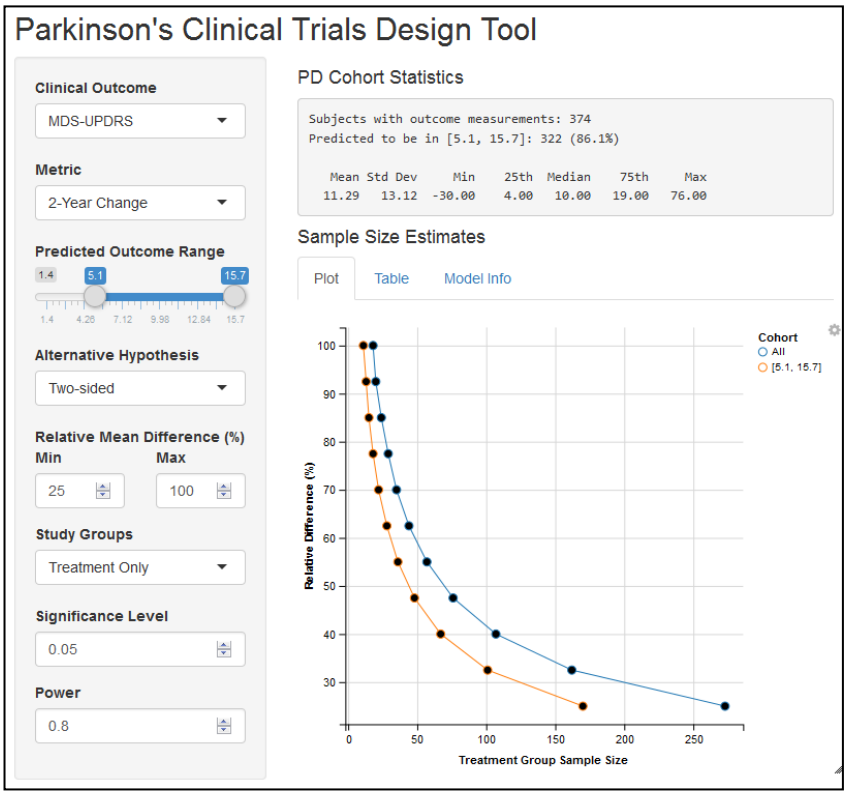
UPDRS I		**MCA Total**		**Mean Putamen**		**GDS Total**	
Predictor	**VI**	**Predictor**	**VI**	**Predictor**	**VI**	**Predictor**	**VI**
UPDRS I	100	MCA Total	100	Mean Putamen	100	GDS Total	100
UPDRS II	48.0	DVS LNS	37.0	Count Density Ratio	74.0	SNP rs17649553	8.2
Baseline Age	36.1	Baseline Age	27.0	Mean Striatum	71.2	SCOPA Total	7.9
SNP rs17649553	24.6	DVS SFTANIM	24.9	Serum Uric Acid	70.4	UPDRS II	7.7
BMI	21.3	DVT Total Recall	24.3	SNP rs76904798	56.5	SNP rs76904798	4.2
Abeta 42	20.3	T-tau/Abeta 42	23.8	Mean Caudate	50.2	SNP rs11724635	3.6
DVT Total Recall	16.6	DVT SDM	20.0	DVS SFTANIM	49.5	WBC	3.3
SNP rs76904798	15.8	GDS Total	18.7	Yrs with PD DX at BL	47.6	JLO Total	2.0
SNP rs823118	14.8	JLO Total	17.3	QUIP Total	47.6	MCA Total	1.5
ApoEe3/e2	14.0	MSE (mseadlg)	13.3	Lymphocytes	46.2	STAI Total	0.7

The most important predictor at baseline tended to be the baseline measurement of that outcome. Age at baseline was also found to be important across a range of outcomes, most notably UPDRS I and MCA Total. An interesting pattern in this dataset is that the multiple sources of data (motor scores, non-motor scores, imaging, and biospecimens) were important predictors for outcomes out of their own domain. For example, SNP rs76904798 was found to be important in UPDRS I, Mean Putamen, and GDS Total.

Web Application

A web application was created with the R shiny package to illustrate how our predictive models might be used to design future clinical trials. It is displayed in the figure on the next page and available at https://bjsmith.shinyapps.io/ppmi2016/. The app estimates sample sizes for user-specified clinical outcomes, ranges of predicted outcome values for which to include subjects, and desired detectable differences in group means (Control vs Treatment). Calculations are based on one and two-sample

standard normal tests of means, with standard deviations computed from corresponding PPMI observed outcomes. The app can be used to assess the decreases in sample sizes that could result if trial inclusion was restricted to ranges of values predicted by our models.



Discussion

The relevance of our project to PD research is that the resulting models could be used to identify patients whose disease is predicted to progress more quickly. Such patients could be prioritized for the development of new treatments and clinical trials. The impact on clinical trials is that smaller sample sizes and resources would be needed to study patients within a predicted range due to less heterogeneity (variability) in their outcomes. To aid in the usability of our models, they were integrated into a web app for clinical trials design. Moreover, our software programs are streamlined to automate the data management, predictive modeling, and web app updating workflow. They are maintained in a git repository to track changes, accommodate multiple contributors, and aid in code sharing. Consequently, our programs can be rerun as the PPMI study data are updated or revised with input from PD researchers.

Reference

Kuhn, M, Johnson, K (2016) *Applied Predictive Modeling,* New York, NY: Springer.