Identification of clinical and genetic predictors of Parkinson’s disease progression via Bayesian machine learning

Jeanne C. Latourelle¹, Michael T. Beste¹, Tiffany C. Hadzi¹, Robert E. Miller¹, Jacob N. Oppenheim¹, Matthew P. Valko¹, Diane M. Wuest¹, Iya G. Khalil¹, Boris Hayete¹, Charles S. Venuto²

¹ GNS Healthcare, Cambridge, MA ² Center for Human Experimental Therapeutics and the Department of Neurology, University of Rochester, Rochester, NY

Introduction

The clinical progression of Parkinson’s disease (PD) is highly heterogeneous across patients. Identifying features predictive of the rate of disease progression can:
- provide insight into the mechanisms of disease process
- inform clinical trial enrollment
- aid clinical disease management

The aim of this study was to develop data-driven models of PD progression, separately for both motor and cognitive symptoms.

Our novel machine learning platform allows the identification of an optimal ensemble of multivariate predictors from a complex data set including a variety of clinical, genetic, molecular and imaging data.

Methods

Source Data and Study Population

**Discovery Set:** Parkinson’s Progression Markers Initiative (PPMI) database (www.ppmi-info.org)³. Participants with 2+ years of follow-up available as of 12/28/15 were included
- 317 untreated PD patients identified within two years of diagnosis
- 118 age- and sex- matched healthy controls

**Validation Set:** 317 independent de novo PD subjects followed 7+ years from the Longitudinal and Biomarker Study in PD (LABS-PD)²

Modeling Approach and REFS™ Analytical Platform

Rate of clinical progression of two clinical domains, Motor and Cognitive, were estimated using linear mixed effects models of subject-specific annualized rate of change of the appropriate clinical assessment
- Motor: Movement Disorders Society Unified Parkinson’s Disease Rating Scale (MDS-UPDRS) Parts II and III
- Cognitive: Montreal Cognitive Assessment (MoCA)

Potential predictors included medical evaluations (N=18), neurological imaging (N=8), genotyping (N=17,456), and CSF biomarkers (N=70).

GNS Healthcare’s proprietary machine learning platform, Reverse Engineering and Forward Simulation (REFSTM)⁴, was used to build prediction models. Selection of a single model underestimates prediction error, thus REFS learns an ensemble of the most probable models (N=128) given the data (Figure 1).

- Ensemble constructed via Monte Carlo sampling of the posterior model landscape.
- Model additions/subtractions scored based on a maximum entropy structural prior with complexity also penalized by the Bayesian Information Criterion⁵.
- Linear, additive, quadratic, and cubic terms allowed in order to accommodate non-linear effects and sub-populations.
- Confidence of a given relationship X→Y determined by frequency among ensemble.

Validation and Extensions of Prediction Models

Both ensembles combined both novel and established markers of disease progression.

Motor Progression Rates (MDS-UPDRS Parts II & III units/year):
- 80 unique predictors, 11 with >5% frequency
- Baseline motor score, PD status, SWEDD status, PD med use, and sex selected in over 90% of models

Cognitive Progression Rates (MoCA units/year):
- 205 unique predictors, 21 with >5% frequency
- Baseline age, MoCA score, CSF t-tau/Aβ1-42 ratio, PD med use, sex, African ancestry, motor score, PD status, and education selected in over 90% of models

Genetic variants were among the selected features, but at lower frequencies, and often in interactions. The most common genetic predictor of motor progression, a novel interaction between rs9298897 (intronic LINGO2) and rs17710829 (2q14.1) was replicated in LABS-PD (Figure 3).

Conclusions

This study highlights the utility of ensemble prediction models to capture the complex interplay of clinical, genetic, and molecular profiles in disease progression.

- REFS identified Bayesian models that combined established and novel patient factors to predict progression for both motor and cognitive deficits.
- Models allow early detection of patients most likely to have rapid disease progression (Figure 2), enabling more effective trial recruitment and clinical disease management.
- Able to identify and replicate a novel genetic interaction (Figure 3), providing potential mechanistic insight into the disease process.

Acknowledgments

This work was supported by grants from the Michael J. Fox Foundation for Parkinson’s Research and National Institute of Neurological Disorders and Stroke (1R20NS092529-01).

PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson’s Research and funding partners, including Abbvie, Avid Radiopharmaceuticals, Biogen, Bristol-Myers Squibb, Covance, GE Healthcare, Genentech, GlaxoSmithKline, Lilly, Lundbeck, Merck, MesoScaleDiscovery, Pfizer, Piramal, Roche, Servier, UCB, and Golub Capital.

References