

ARTICLE

Application of longitudinal item response theory models to modeling Parkinson's disease progression

Haotian Zou¹  | Varun Aggarwal²  | Glenn T. Stebbins³ | Martijn L. T. M. Müller²  |
Jesse M. Cedarbaum⁴ | Anne Pedata² | Diane Stephenson² | Tanya Simuni⁵ |
Sheng Luo⁶

¹University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

²Critical Path Institute, Tucson, Arizona, USA

³Rush University Medical Center, Chicago, Illinois, USA

⁴Coeruleus Clinical Sciences LLC, Woodbridge, Connecticut, USA

⁵Northwestern University Medical Center, Chicago, Illinois, USA

⁶Duke University, Durham, North Carolina, USA

Correspondence

Sheng Luo, Department of Biostatistics and Bioinformatics, Duke University Medical Center, 2424 Erwin Road, Suite 1102 Hock Plaza, Box 2721, Durham, NC 27710, USA.

Email: sheng.luo@duke.edu

Funding information

Critical Path for Parkinson's Consortium; National Institute on Aging, Grant/Award Number: P30AG028716, P30AG072958 and R01AG064803

Abstract

The Movement Disorder Society revised version of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS) parts 2 and 3 reflect patient-reported functional impact and clinician-reported severity of motor signs of Parkinson's disease (PD), respectively. Total scores are common clinical outcomes but may obscure important time-based changes in items. We aim to analyze longitudinal disease progression based on MDS-UPDRS parts 2 and 3 item-level responses over time and as functions of Hoehn & Yahr (H&Y) stages 1 and 2 for subjects with early PD. The longitudinal item response theory (IRT) modeling is a novel statistical method addressing limitations in traditional linear regression approaches, such as ignoring varying item sensitivities and the sum score balancing out improvements and declines. We utilized a harmonized dataset consisting of six studies with 3573 subjects with early PD and 14,904 visits, and mean follow-up time of 2.5 years (± 1.57). We applied both a unidimensional (each part separately) and multidimensional (both parts combined) longitudinal IRT models. We assessed the progression rates for both parts, anchored to baseline H&Y stages 1 and 2. Both the uni- and multidimensional longitudinal IRT models indicate significant worsening time effects in both parts 2 and 3. Baseline H&Y stage 2 was associated with significantly higher baseline severities, but slower progression rates in both parts, as compared with stage 1. Patients with baseline H&Y stage 1 demonstrated slower progression in part 2 severity compared to part 3, whereas patients with baseline H&Y stage 2 progressed faster in part 2 than part 3. The multidimensional model had a superior fit compared to the unidimensional models and it had excellent model performance.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) parts 2 and 3 reflect patient-reported functional impact and clinician-reported severity of motor signs of Parkinson's disease (PD), respectively.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *CPT: Pharmacometrics & Systems Pharmacology* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

WHAT QUESTION DID THIS STUDY ADDRESS?

We aim to apply longitudinal item response theory (IRT) models to investigate the change of patient-reported impact of motor signs (part 2) and clinician-reported motor signs (part 3) of MDS-UPDRS scores over time and as functions of Hoehn & Yahr (H&Y) stages 1 and 2 for subjects with early PD.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

In both parts 2 and 3, we identified significant worsening time effects and differential baseline H&Y stage 2 effects (as compared with stage 1) on baseline severities and progression rates.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?

Multidimensional longitudinal IRT modeling simultaneously monitors changes in different, but related, outcome domains over time. It provides an operational method to allow testing the efficacy of domain-specific target therapies and to separate possible domain-specific improvements balanced out by domain-specific deteriorations that would be missed by standard statistical methods.

INTRODUCTION

A rich pipeline of therapeutic candidates is being advanced for the treatment of Parkinson's disease (PD) with many targets focused on underlying pathophysiology of the disease.¹ There is growing recognition that early intervention is needed for success in halting or slowing disease progression.² However, current measures are not sensitive to change in early stages of the disease. There is an urgent need to identify clinically meaningful measures of disease progression at the earliest signs of disease onset. Severity and progression of PD can be evaluated using disease-specific clinical rating scales. The International Parkinson and Movement Disorder Society revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS) is a widely used scale for measuring PD features in clinical and research practice.³ MDS-UPDRS consists of 65 items, measured by a 5-point Likert scale (0–4, with higher values denoting an increased severity), in four parts: part 1, Non-Motor Aspects of Experiences of Daily Living (13 items); part 2, Motor Aspects of Experiences of Daily Living (13 items); part 3, Motor Examination (33 items), and part 4, Motor Complications (6 items). Parts 2 and 3 are frequently used as outcomes in cohort studies (e.g., PPMI⁴ and TRACKING⁵) and randomized controlled trials.^{6–9} The current research focuses on MDS-UPDRS parts 2 and 3, which assess patient-reported impact of motor signs of PD and clinician-reported severity of motor signs of PD, respectively. Analyses are carried out in an integrated database of observational and clinical trials targeting early manifest PD.

To analyze data from MDS-UPDRS, the item scores of each part are typically summed to obtain a sum score, which is modeled as a continuous variable using

either mixed models or generalized estimating equations. Reliance on the sum score assumes an equal importance of each item and any sub-scores relative to the sum score, equal discrimination of each item, and equal sensitivity of change for each item. However, such assumptions may not be valid, particularly when disease progression patterns or treatment effects apply to some items of impairment differently from other items.¹⁰

Item response theory (IRT) modeling is a statistical framework that relates an individual's responses to items on a rating scale to the underlying disease severity, considered as the latent variable, termed "theta." Typical IRT analyses usually operate under the assumption of unidimensionality (i.e., all items in part 2 measure a single core attribute of the underlying disease and all items in part 3 measure another single core attribute). However, parts 2 and 3 measure related clinical constructs, obtained through different perspectives (i.e., the patient's experience vs. the clinician's examination). Given the different nature of each part, issues around combining and interpretation arise and it has been recommended to analyze each part as a separate but correlated domain.³ Previous IRT investigations involving PD are contributory to the field, but have the following limitations: (1) used cross-sectional IRT models on longitudinal data^{11,12}; (2) adopted a two-step sequential parameter estimation process that does not estimate all parameters simultaneously¹³; and (3) assumed unidimensionality¹⁴ or modeled the disease progression separately to each of the latent variables but without regard to their internal relationship.^{12,15} In comparison, multidimensional longitudinal IRT models with multiple latent variables offer the unique opportunity of investigating different progression patterns and treatment responses in more than one related, but nevertheless

distinct, domains of diseases.^{16,17} These longitudinal IRT models are novel because they address the limitations of the previous IRT studies and fully utilize the repeated measures from the same subject, simultaneously estimate all unknown parameters, and sufficiently account for the correlation between different theta components.^{16,17}

Among multiple end points defining PD severity, the Hoehn & Yahr (H&Y) scale is a commonly used categorical measure to assess overall PD dysfunction.¹⁸ Specifically, patients with PD in H&Y stages 1 and 2 are early in the disease continuum and able to perform daily activities. These patients are the target population of many clinical studies. The purpose of this study was to apply longitudinal IRT models to investigate change over time and as functions of H&Y stages 1 and 2 at baseline for the multidimensional traits of patient perceived motor impact on function and clinician assessed severity of motor symptoms captured by the parts 2 and 3 MDS-UPDRS scores, respectively, in a harmonized dataset consisting of six studies all focused on early manifest stages of PD.

METHODS

Study population

Our analysis dataset was built based on a harmonized dataset consisting of four observational studies (PPMI,⁴ TRACKING,⁵ ICICLE,¹⁹ and OXFORD²⁰) and two randomized controlled trials (STEADY-PD3²¹ and SURE-PD3²²). Informed consent was obtained from participants of each individual study. The details of these studies can be found in the corresponding literature. The patient level item-level data from individual studies were curated and harmonized according to Clinical Data Interchange Standards Consortium (CDISC) Parkinson's Disease Therapeutic Area User Guide (TAUG-PD). The final harmonized dataset contains a total of 3641 subjects and 18,282 visits, with mean follow-up time of 2.5 years (± 1.57). There were 335 subjects from the STEADY-PD3 study, with 170 patients on isradipine drug and 165 patients on placebo. Among 296 patients from the SURE-PD3 study, there were 147 patients on inosine treatment and 149 patients on placebo. The remaining 3010 subjects were from the four observational studies (PPMI, TRACKING, ICICLE, and OXFORD). As inclusion/exclusion criteria, we selected subjects and visits with available MDS-UPDRS parts 2 and 3 scores and without missing item scores, and H&Y stage 1 or 2 at baseline. To achieve a larger sample size, we included subjects in both the active and placebo groups in both STEADY-PD3 and SURE-PD3 studies, because a sensitivity analysis that excluded subjects in the active group generated very similar results (refer to Section 3

in the Supplemental Material S1). All MDS-UPDRS parts 2 and 3 observations were included in the data analysis. After applying all the inclusion/exclusion criteria, the analysis dataset consists of a total 3573 subjects and 14,904 visits. This study was approved by the Duke University institutional review board (Protocol ID: Pro00107266).

Unidimensional longitudinal item response theory model

We applied unidimensional longitudinal IRT models separately to MDS-UPDRS parts 2 and 3 items. The unidimensional analysis assumed that there was a single latent variable, indicated as theta, in IRT models, representing either the underlying motor disability's impact on daily functions (referred to as part 2 severity) when applied to 13 part 2 items (items 2.1–2.13 with a total score range of 0–52), or the underlying motor impairment through clinical examination (referred to as part 3 severity) when applied to 33 part 3 items (items 3.1–3.18 with a total score range of 0–132). The unidimensional longitudinal IRT model consists of two levels. The first level, a graded-response measurement model (please refer to Figures S1 and S2 and Model 1 in the Supplemental Material S1), quantifies the relationship between the response of each item and theta. The probability of every score in each item was determined by five parameters: the discrimination parameter and the four location parameters. Higher value in the discrimination parameter suggests that this item is powerful for determining the individual's latent severity. The location parameters (also called difficulty parameters) are the probability threshold for transitioning from score 0 to 1 (normal to slight), from 1 to 2 (slight to mild), from 2 to 3 (mild to moderate), and from 3 to 4 (moderate to severe). The second level structural model (Model 2 in the Supplemental Materials S1) regresses parts 2 or 3 severity, indicated as theta, on time in years, H&Y stage, and its interaction with time, with subject-specific random intercepts, and random slopes of time. The random intercepts allow each individual to have a personalized severity at baseline, whereas the random slopes account for individuals' unique progression rates in severity. The correlation between the random intercepts and random slopes was modeled by a correlation coefficient. Because the IRT models are overparameterized, certain constraints need to be set to make the models identifiable (estimable).²³ Specifically, the random intercepts were assumed to have standard normal distribution with mean 0 and variance 1, and the random slopes were assumed to follow a normal distribution with mean 0.

We investigated the progression in parts 2 and 3 severities and the effects of H&Y stage 2 in comparison to

H&Y stage 1. We determined significant effects using 95% credible intervals (95% CIs; the Bayesian equivalence of 95% CIs) and p values, defined as $p = 1 - \Phi(|\text{Mean}| / \text{SD})$, where $\Phi(\cdot)$ denotes the cumulative distribution function of standard normal distribution, mean and SD are obtained from the posterior samples, approximately corresponding to a two-sided p value.

Multidimensional longitudinal item response theory model

The multidimensional longitudinal IRT approach allows for more than one latent variable. We considered the latent variable of disease severity, termed theta, as multivariate with part 2 and part 3 severities as separate but correlated theta components, the former captured by 13 part 2 items and the latter covering 33 part 3 items. The first level graded-response measurement models (please refer to Figure S3 and Models 3 and 4 in the Supplemental Materials S1) are similar to the unidimensional longitudinal IRT model. In the second level, structural models (Models 5 and 6 in the Supplemental Material S1), both the parts 2 and 3 theta components were regressed on time in years, H&Y stage, and its interaction with time, in addition to subject-specific random intercepts and random slopes of time to allow each participant to have his/her baseline severity and rates of progression in both parts 2 and 3 domains. To account for the relationship between the two thetas, the correlation among the random intercepts and random slopes in both domains was represented in a correlation matrix. Because IRT models are overparameterized, additional constraints need to be imposed to make the models identifiable (estimable). Specifically, the random intercepts U_{i0} and U_{i2} (in Models 5 and 6 in the Supplemental Materials S1) were assumed to have a standard normal distribution marginally with mean zero and variance one, fixing the mean and variance of latent overall severity at baseline when time was zero. This assumption ensures that, for subjects with H&Y stage 1 at baseline, the multivariate latent variable thetas had a mean zero at baseline and variance one for both parts 2 and 3.

We applied the multidimensional longitudinal IRT model to investigate the progression in parts 2 and 3 parkinsonian severities and the effects of H&Y stage 2 in comparison to stage 1. Because the latent variables, or theta components, are unitless, the rates of progression and H&Y stage responsivity in parts 2 and 3 domains can be directly compared and tested, thereby providing additional clinically relevant information not available in the unidimensional longitudinal IRT model. In both unidimensional and multidimensional models, we assumed a linear relationship between the latent variable theta and time based on the linear progression patterns displayed in

Figure 1. For model comparison and selection, we used the Deviance Information Criterion (DIC), where lower values indicate better fit to the model.

Model fitting using Bayesian inference

The analyses were conducted using Bayesian inference based on Markov Chain Monte Carlo (MCMC) posterior simulations, implemented in Stan (version 2.28, Data S1 and Data S2)²⁴ via interface in R statistical program (version 4.1.2).²⁵ We used vague (noninformative) prior information on all parameters in the models. The selection of prior distributions and parameters, initial values, and convergence assessment are detailed in our prior work.²⁶ The longitudinal IRT model (or latent variable model) proposed in the prior work²⁶ effectively modeled the latent disease severity and provided accurate personalized predictions of future outcome trajectories and risk of a target event of interest. To facilitate understanding, we also presented p values calculated by $p = 1 - \Phi(|\text{Mean}| / \text{SD})$, where $\Phi(\cdot)$ denotes the cumulative distribution function of standard normal distribution, approximately corresponding to a two-sided p value in a frequentist setting.²⁷

To facilitate clinical interpretation of the regression coefficients, we applied simulation techniques to express them in terms of MDS-UPDRS point scores. These simulation techniques used the posterior samples from Bayesian inference and were associated with increased variance. Therefore, we expressed the primary statistical results in unitless theta values, but provided the simulation MDS-UPDRS point-based results in the Supplemental Material S1.

Diagnostics of model performance

The longitudinal IRT model performance was assessed comprehensively using diagnostics based on simulation and residual assessments. Specifically, based on the estimated parameters and the models, we simulated response of each category of all MDS-UPDRS parts 2 and 3 items across all visits. We compared the observed and simulated proportions via mirror plots and scatter plots. Moreover, we computed the residual of each item (defined as the difference between the observed item value and this item's respective weighted prediction score, which is the sum of each category score; i.e., 0–4, multiplied by this category's simulated probability) and investigated the correlation between residuals. Last, we also generated sum scores of 13 part 2 items and 33 part 3 items from the simulated data and compared them with the observed data.

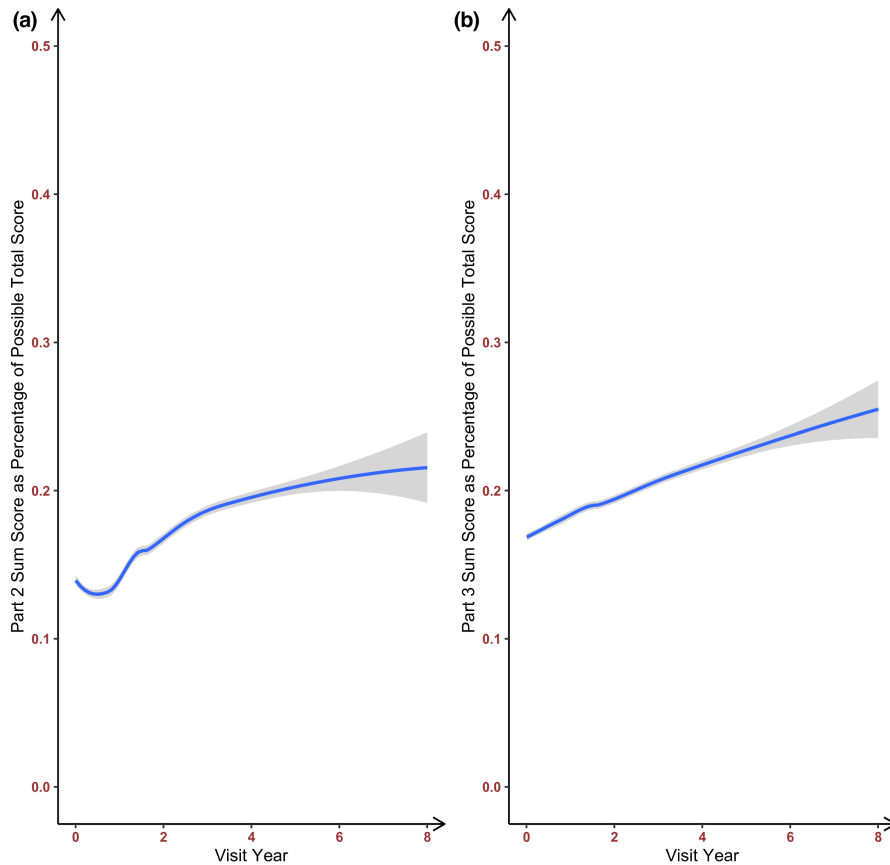


FIGURE 1 LOWESS curves of part 2 (a) and part 3 (b) sum scores based on 3573 subjects and 14,904 visits.

RESULTS

The baseline characteristics of the analysis dataset are displayed in Table 1. At the baseline visit, 1216 (34.0%) subjects had H&Y stage 1, whereas the remaining 2357 (66.0%) subjects had H&Y stage 2. Comparing with subjects with H&Y stage 1, those with H&Y stage 2 at baseline were significantly older, more likely to be male sex, had longer PD duration, and worse motor severity manifested by MDS-UPDRS parts 2 and 3 sum scores. The LOWESS curves of part 2 (Figure 1a) and part 3 (Figure 1b) sum scores show gradual increase in both parts 2 and 3 summed scores suggesting clinical decline in a linear rate.

Model comparison and the selection of final model

The multidimensional IRT model with the interaction term for H&Y stage and time (DIC = 1,223,375) had a superior fit than the model without the interaction term (DIC = 1,223,396), and two unidimensional IRT models together (DIC = 306,318 + 917,334 = 1,223,652). We also attempted to include the exponential progression rates in the multidimensional IRT model but it failed to converge due to its complexity. Hence, we selected the multidimensional IRT model with an interaction term as the final

model and presented the results from this model and its unidimensional counterparts.

Unidimensional longitudinal IRT model results for MDS-UPDRS part 2 items

With the unidimensional longitudinal IRT model using all 13 part 2 items, there was a significant time effect, demonstrating a deterioration in part 2 at the rate of 0.216 theta units per year (95% CI: 0.194, 0.237, $p < 0.001$). As compared with baseline H&Y stage 1, H&Y stage 2 was associated with a significantly greater severity of 0.492 theta units (95% CI: 0.421, 0.569, $p < 0.001$) at baseline in part 2. Baseline H&Y stage 2 was significant in changing the progression rate by -0.034 theta units per year (95% CI: -0.059 , -0.008 , $p = 0.004$), demonstrating a total rate of 0.182 theta units per year (95% CI: 0.167, 0.197, $p < 0.001$) for part 2. Figure 2a displays in theta values the different part 2 severity progression rates of subjects with baseline H&Y stage 1 (red curve) and 2 (blue curve). We found that part 2 severity at baseline and its rate of progression had significant negative correlation (correlation coefficient -0.165 , 95% CI: -0.214 , -0.113 , $p < 0.001$), suggesting that patients with more pronounced part 2 severity at baseline tend to progress slower in patient-reported functional impact of motor symptoms of PD. The regression

TABLE 1 Baseline characteristics of the analysis dataset with 3573 subjects and 14,904 visits.

	Mean	SD	Min	Max	Missing
Age	65.4	9.5	31	89	2
Sex, F:M, <i>N</i> (%)	1272: 2301 (35.6%: 64.4%)				0
Time since Dx (years)	1.1	0.9	0	3.5	1
MDS-UPDRS Part 2 Sum Score	7.9	5.7	0	40	0
MDS-UPDRS Part 3 Sum Score	22.5	10.5	0	85	0
Baseline Hoehn & Yahr					
Stage 1, <i>N</i> (%)	1216 (34.0%)				
Stage 2, <i>N</i> (%)	2357 (66.0%)				
	H&Y Stage 1 (1216, 34.0%), mean (SD)		H&Y Stage 2 (2357, 66.0%), mean (SD)		<i>p</i> Value
Age (years)	63.1 (9.6)		66.5 (9.3)		<0.001
Female (%)	466 (38.3%)		806 (34.2%)		<0.001
PD duration (years)	0.9 (0.8)		1.1 (0.9)		<0.001
MDS-UPDRS Part 2 Sum score	6.2 (4.8)		8.8 (5.9)		<0.001
MDS-UPDRS Part 3 Sum score	15.7 (7.1)		26.0 (10.2)		<0.001

Note: The *p* value is based on a two-group *t*-test with unequal variances.

Abbreviations: Dx, diagnostic; F, female; H&Y, Hoehn & Yahr; M, male; Max, maximum; MDS-UPDRS, Movement Disorder Society-Unified Parkinson's Disease Rating Scale; Min, minimum; PD, Parkinson's disease; SD, standard deviation.

parameters are displayed in [Table 2](#) (upper panel). The location and discrimination parameters are shown in [Table S1](#).

Unidimensional longitudinal IRT model results for MDS-UPDRS part 3 items

With the unidimensional longitudinal IRT model using all 33 part 3 items, there was a significant overall time effect, demonstrating a deterioration in part 3 severity at the rate of 0.305 theta units per year (95% CI: 0.274, 0.334, $p < 0.001$). As compared with baseline H&Y stage 1, H&Y stage 2 was associated with a significantly greater severity of 1.179 theta units (95% CI: 1.094, 1.258, $p < 0.001$) at baseline in part 3. Baseline H&Y stage 2 was significant in changing the progression rate by -0.153 theta units per year (95% CI: -0.185 , -0.118 , $p < 0.001$), demonstrating a total rate of 0.152 theta units per year (95% CI: 0.131, 0.172, $p < 0.001$) for part 3. [Figure 2b](#) displays in theta values the different part 3 severity progression rates of subjects with baseline H&Y stage 1 (red curve) and 2 (blue curve). We found that part 3 severity at baseline and its rate of progression had significant negative correlation (correlation coefficient -0.229 , 95% CI: -0.271 , -0.188 , $p < 0.001$), suggesting that patients with more pronounced part 3 severity at baseline tend to progress slower in clinician-reported severity of motor signs of PD. The regression parameters are shown in

[Table 2](#) (lower panel). The location and discrimination parameters are shown in [Table S2](#).

Multidimensional longitudinal IRT model for both MDS-UPDRS part 2 and part 3 items

We applied the multidimensional longitudinal IRT model to jointly assess the 13 part 2 items and 33 part 3 items. There were significant overall time effects on deterioration for both parts 2 and 3 severities, demonstrating an increase in part 2 severity at the rate of 0.222 theta component values per year (95% CI: 0.201, 0.243, $p < 0.001$) and in part 3 severity at the rate of 0.312 theta component values per year (95% CI: 0.283, 0.343, $p < 0.001$). As compared with baseline H&Y stage 1, H&Y stage 2 was associated with a significantly greater severity of 0.492 (95% CI: 0.418, 0.562, $p < 0.001$) and 1.181 (95% CI: 1.098, 1.267, $p < 0.001$) theta component values for parts 2 and 3 at baseline, respectively. For part 2, baseline H&Y stage 2 was significant in changing the progression rate by -0.039 theta units per year (95% CI: -0.065 , -0.013 , $p = 0.001$), demonstrating a total rate of 0.183 theta units per year (95% CI: 0.169, 0.199, $p < 0.001$). For part 3, baseline H&Y stage 2 was also significant in changing the progression rate by -0.151 theta units per year (95% CI: -0.186 , -0.118 , $p < 0.001$), demonstrating a total rate of 0.161 theta units per year (95% CI: 0.140, 0.182, $p < 0.001$).

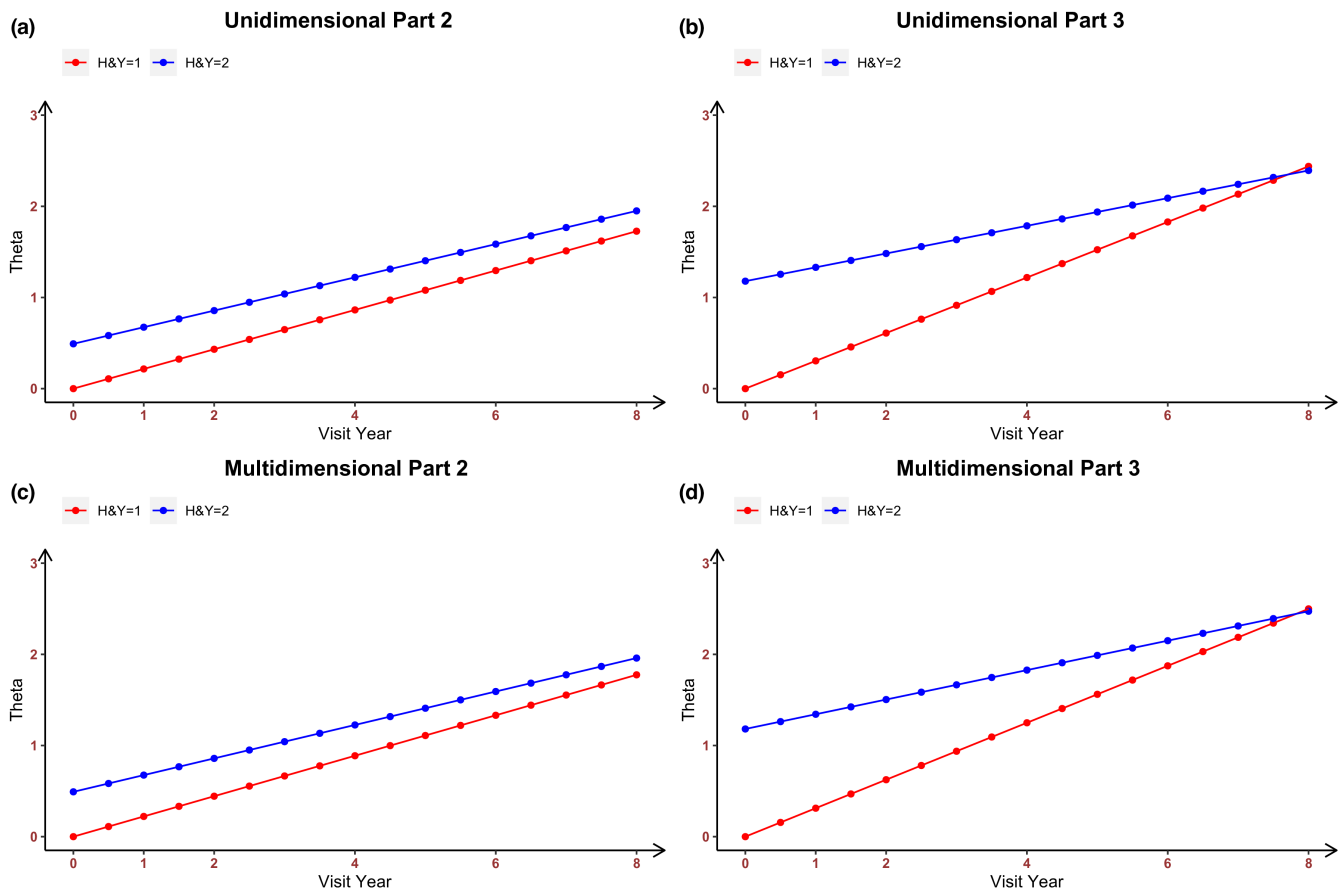


FIGURE 2 The estimated rates of progression among subjects with Hoehn & Yahr stage 1 (red color) and stage 2 (blue color) in part 2 severity (a) and part 3 severity (b) from the unidimensional longitudinal IRT models, and in part 2 severity (c) and part 3 severity (d) from the multidimensional longitudinal IRT models. IRT, item response theory.

TABLE 2 Parameter estimates and 95% CIs from two separate unidimensional longitudinal IRT models for MDS-UPDRS part 2 items (upper panel) and part 3 items (lower panel)

Parameters	Mean	SD	95% CI	p Value	RSE
Part 2 Severity					
β_2 : Time (years)	0.216	0.011	0.194, 0.237	<0.001	5.07
β_1 : H&Y Stage 2	0.492	0.038	0.421, 0.569	<0.001	7.63
β_3 : H&Y Stage 2 * Time	-0.034	0.013	-0.059, -0.008	0.004	37.67
ρ	-0.165	0.026	-0.214, -0.113	<0.001	15.70
Part 3 Severity					
β_2 : Time (years)	0.305	0.015	0.274, 0.334	<0.001	5.01
β_1 : H&Y Stage 2	1.179	0.042	1.094, 1.258	<0.001	3.53
β_3 : H&Y Stage 2 * Time	-0.153	0.017	-0.185, -0.118	<0.001	11.40
ρ	-0.229	0.021	-0.271, -0.188	<0.001	9.27

Note: ρ is the correlation coefficient of random intercepts and random slopes. The p value was calculated by: $p = 1 - \Phi(|\text{Mean}| / \text{SD})$, where $\Phi(\cdot)$ denotes the cumulative distribution function of standard Normal distribution. Refer to Model (2) in the Supplemental Material S1 for the meanings of β parameters. RSE was computed as standard error divided by the absolute value of the estimate and multiplied by 100.

Abbreviations: CI, confidence interval; H&Y, Hoehn & Yahr; IRT, item response theory; MDS-UPDRS, Movement Disorder Society-Unified Parkinson's Disease Rating Scale; RSE, relative standard error; SD, standard deviation.

The rates of increase in parts 2 and 3 severities were significantly different across the two baseline HY stages. Specifically, subjects with baseline H&Y stage 1 had a

slower rate of decline in part 2 than part 3, with a difference of -0.090 theta units per year (95% CI: -0.118 , -0.062 , $p < 0.001$). In contrast, subjects with baseline H&Y

TABLE 3 Parameter estimates and 95% CIs from the multidimensional longitudinal IRT model for part 2 and part 3 items

Parameters	Mean	SE	95% CI	p Value	RSE
Part 2 Severity					
β_2^{P2} : Time	0.222	0.011	0.201, 0.243	<0.001	4.98
β_1^{P2} : H&Y Stage 2	0.492	0.038	0.418, 0.562	<0.001	7.64
β_3^{P2} : H&Y Stage 2 * Time	-0.039	0.013	-0.065, -0.013	0.001	33.50
ρ_{01}	-0.160	0.025	-0.209, -0.111	<0.001	15.93
σ_{u1}	0.267	0.007	0.254, 0.280	NA	2.59
Part 3 Severity					
β_2^{P3} : Time	0.312	0.015	0.283, 0.343	<0.001	4.79
β_1^{P3} : H&Y Stage 2	1.181	0.043	1.098, 1.267	<0.001	3.65
β_3^{P3} : H&Y Stage 2 * Time	-0.151	0.017	-0.186, -0.118	<0.001	11.48
ρ_{23}	-0.215	0.021	-0.257, -0.174	<0.001	9.90
σ_{u3}	0.407	0.009	0.390, 0.425	NA	2.22
ρ_{02}	0.360	0.017	0.326, 0.394	<0.001	4.79
ρ_{03}	0.053	0.022	0.009, 0.097	0.010	42.71
ρ_{12}	-0.098	0.024	-0.144, -0.051	<0.001	24.83
ρ_{13}	0.610	0.020	0.569, 0.648	<0.001	3.31

Note: ρ_{01} is the correlation coefficient of random intercepts and random slopes in part 2 model; ρ_{02} is the correlation coefficient of random intercepts in part 2 model and random intercepts in part 3 model; ρ_{03} is the correlation coefficient of random intercepts in part 2 model and random slopes in part 3 model; ρ_{12} is the correlation coefficient of random slopes in part 2 model and random intercepts in part 3 model; ρ_{13} is the correlation coefficient of random slopes in part 2 model and random slopes in part 3 model; ρ_{23} is the correlation coefficient of random intercepts and random slopes in part 3 model; σ_{u1} is the standard error of the random slopes in part 2 model; σ_{u3} is the standard error of the random slopes in part 3 model. The p value was calculated by: $p = 1 - \Phi(|\text{Mean}| / \text{SE})$, where $\Phi(\cdot)$ denotes the cumulative distribution function of standard normal distribution. The p values for σ_{u1} and σ_{u3} were not computed because the test statistic does not follow normal distribution. Refer to Models (5) and (6) in the Supplemental Material S1 for the meanings of β parameters. RSE was computed as standard error divided by the absolute value of the estimate and multiplied by 100.

Abbreviations: CI, confidence interval; H&Y, Hoehn & Yahr; IRT, item response theory; RSE, relative standard error; SD, standard deviation.

stage 2 had a faster rate of decline in part 2 than part 3, with a difference of 0.022 theta units per year (95% CI: 0.005, 0.041, $p = 0.010$). Figure 2c,d displays the progression rates in theta component values per year for parts 2 and 3 severities, respectively, from the multidimensional longitudinal IRT model. The regression parameters are shown in Table 3. The location and discrimination parameters are shown in Tables S3 (for part 2 items) and S4 (for part 3 items).

The estimated correlation coefficients of random effects in Table 3 indicate negative correlation between baseline severity and progression rate in both parts 2 and 3, as indicated by significant values of ρ_{01} and ρ_{23} . Further highlighting that parts 2 and 3 are two separate but correlated domains, we found that their severities at baseline were positively correlated (with a significant correlation coefficient ρ_{02} of 0.360, 95% CI: 0.326, 0.394, $p < 0.001$), suggesting that subjects with more pronounced part 2 severity scores tend to have worse part 3 severity ratings at baseline, and vice versa. Moreover, the progression rates in both parts 2 and 3 were positively correlated (with a significant correlation coefficient ρ_{13} of 0.610, 95%

CI: 0.569, 0.648, $p < 0.001$), suggesting that subjects who progressed faster in part 2 tend to also progress faster in part 3, and vice versa. The diagnostics of the multidimensional longitudinal IRT model are presented in Section S2 (please refer to the Figures S4a,b, S5a-d, and S6-S9 in the Supplemental Material S1). The multidimensional longitudinal IRT model provides excellent goodness of fit and model performance in the analysis of parts 2 and 3 data. For comparison, the visual predictive checks for the sum scores from two separate unidimensional IRT models are presented in Figure S9 in the Supplemental Material S1.

DISCUSSION

In this study, we applied longitudinal IRT models to investigate the change of patient-reported impact of motor signs (part 2) and clinician-reported motor signs (part 3) of MDS-UPDRS scores over time and as functions of H&Y stages 1 and 2 for subjects with early PD. We have identified significant time effects in increasing severities for both parts 2 and 3. Baseline H&Y stage 2 was associated

with significantly higher baseline severities, but slower progression rates for both parts 2 and 3 severities, as compared with stage 1. Moreover, we detected differential changes over time and as functions of baseline H&Y stage 1 or 2 in both parts 2 and 3 domains of the MDS-UPDRS.

Furthermore, we have identified negative correlation between baseline severity and its progression in each part, suggesting that patients who were more severe in motor impact on function or in motor signs at baseline tend to progress slower. Our modeling results suggest that patients who are clinically worse in functions due to motor impact tend to have worse motor examination and the progression rates in both parts 2 and 3 were positively correlated.

The application of longitudinal IRT models to the study of neurological diseases has increased recently. To our knowledge, there are 13 prior papers of relevance to our work using longitudinal IRT models: eight dealing with PD^{10-15,28,29} and the others dealing with other neurological or psychiatric conditions.^{27,30-33} All of them have made significant contributions to the field. Among the PD studies, using cross-sectional IRT modeling, Regnault et al.¹¹ analyzed the pooled data of MDS-UPDRS parts 2 and 3 from baseline, year 1, and year 2, whereas Gottipati et al.¹⁰ applied the models developed earlier¹⁵ to the baseline UPDRS data from two clinical trials. Via longitudinal IRT models, Buatois et al.¹³ and Chen et al.²⁹ analyzed data of MDS-UPDRS and UPDRS, respectively, and concluded that IRT is a powerful tool to capture the heterogeneous nature of the rating scales and it markedly reduces sample size. Sheng et al.¹² and Chael et al.²⁸ analyzed the data of part 3 of MDS-UPDRS and UPDRS, respectively, using two separate IRT models with the latent variable representing tremor and non-tremor symptoms. Similarly, Gottipati et al.¹⁵ analyzed all three parts of MDS-UPDRS using separate IRT models with the latent variable representing patient-reported responses, sided responses, and non-sided responses. Last, Arrington et al.¹⁴ applied a unidimensional longitudinal IRT model to motor-related items in MDS-UPDRS parts 2 and 3 and assessed the impact of reducing the number of items. Outside of PD, Ueckert et al.³⁰ used IRT to study Alzheimer's disease, and Novakovic et al.³² used the technique in multiple sclerosis and concluded that IRT modeling is an effective tool to study disease progression with item-level data. In schizophrenia research, Krekels et al.³¹ fit separate longitudinal IRT models for three subscales of the Positive and Negative Symptom Scale as different domains. Finally, in dementia research, Vandemeulebroecke et al.²⁷ and Bascoul-Molleivi et al.³³ applied a unidimensional longitudinal IRT model to 14 items of two neuropsychological test batteries and 30 items of Quality of Life Questionnaire (QLQ-C30) and demonstrated its values as a powerful approach for progressive diseases.

Multidimensional longitudinal IRT modeling, as applied in this work, is a novel statistical method because

it simultaneously integrates three modeling components in unison. Specifically, it fully accounts for the correlation among visits from the same participant. It simultaneously estimates the item-specific parameters and regression parameters and it concurrently estimates and compares rates of progression in multiple domains and model the correlations among them. Moreover, the latent variables in the multidimensional IRT model are unitless and this allows us to directly compare the rates of progression and H&Y stage effects in parts 2 and 3 domains. The Bayesian modeling technique we adopted is flexible in fitting complex models and it can readily account for missing data via sampling from their posterior distributions, in spite of the additional modeling complexity and computational time. In our future work, we plan to extend the Bayesian modeling to account for missing item scores.

The course of PD progression is variable within individual patients and heterogeneous between individuals. In modeling both motor impact on function and severity of motor symptoms captured by the parts 2 and 3 MDS-UPDRS scores, respectively, our multidimensional longitudinal IRT analysis adequately addresses PD progression variability by incorporating both random intercepts and random slopes, which allow each patient to have individualized severity and progression rate in each domain. We may further extend the model and include the inter-occasion random effects to account for the variability in the course of PD progression. In this study, the multidimensional longitudinal IRT model suggests that among subjects with baseline H&Y stage 1, the rate of decline in part 2 is slower than part 3. In clinical trials involving patients with early PD with H&Y stage 1, it may be challenging to capture the progression rate for part 2 with the traditional analysis methods based on sum of scores.

We acknowledge that the number of latent variables in our IRT models are prespecified and it is based on the scale's original intent to provide separate assessments of patients' perception of motor impact on function and examined motor severity. The inclusion of multiple latent variables in each part may further improve the model.³⁴ It is worth noting that the linear progression rates in both parts 2 and 3 were assumed based on the linear progression patterns displayed in [Figure 1](#), whereas the longitudinal IRT models can readily account for the nonlinear progression, if it exists. The potential impact of symptomatic treatment on progression rates was not modeled. The main purpose of this work was to compare unidimensional IRT with multidimensional IRT analysis on a harmonized dataset. As noted by others,¹³ documentation of medication dose and last medication intake is sometimes absent from study data which prevents a fine-grained analysis of symptomatic medication effect. Expansion of the clinical database with additional studies that use MDS-UPDRS and with defined

ON/OFF assessment protocols will allow for further investigation of potentially differential progression patterns in both “ON” and “OFF” medication states using multidimensional IRT. Finally, the analysis dataset only consists of patients with PD with H&Y stages 1 and 2 at baseline. The results may not apply to patients with more severe PD.

AUTHOR CONTRIBUTIONS

H.Z., V.A., G.T.S., M.L.T.M., J.M.C., D.S., T.S., and S.L. wrote the manuscript. V.A., G.T.S., M.L.T.M., J.M.C., A.P., D.S., T.S., and S.L. designed the research. H.Z. and S.L. performed the research. H.Z. analyzed the data. S.L. contributed new reagents/analytical tools.

ACKNOWLEDGMENTS

The Critical Path Institute’s CPP Consortium is funded by Parkinson’s United Kingdom and the following industry members: AbbVie, Biogen, Cerevel, Denali, GSK, MSD, Takeda, Sanofi, Roche, IXICO, Cereval, Clario, and UCB. The authors also acknowledge additional CPP member organizations, including the Parkinson’s Disease Foundation, The Michael J. Fox Foundation, the Davis Phinney Foundation, The Cure Parkinson’s Trust, PMD Alliance, the University of Oxford, University of Cambridge, Newcastle University, University of Glasgow, as well as the NINDS, US Food and Drug Administration, and the European Medicines Agency. We also acknowledge The Michael J. Fox Foundation for sponsoring of PPMI. Data were obtained from the Parkinson’s Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data). The PPMI is sponsored and partially funded by The Michael J. Fox Foundation for Parkinson’s Research and funding partners, including AbbVie, Avid, Biogen, Bristol-Myers Squibb, Convance, GE Healthcare, Genentech, GSK, Lilly, Lundbeck, MSD, Meso Scale Discovery, Pfizer, Piramal, Roche, Sanofi Genzyme, Servier, TEVA, UCB, and Golub Capital. For up-to-date information on the study, visit www.ppmi-info.org. We would also like to recognize the scientific leadership of CPP advisors Karl Kieburtz, Tanya Simuni, Michael Schwarzschild, and Jesse Cedarbaum. Data used in the preparation of this article were obtained from the CPP Unified Clinical Database. C.P.P. acknowledges the contributions of UK investigators Michele Hu, Donald Grosset, Caroline Williams Gray, Rachael Lawson, and David Burn for their role in contributing data from PD cohort studies to the CPP Unified PD database.

FUNDING INFORMATION

The research of Sheng Luo was supported by National Institute on Aging (grant number: R01AG064803, P30AG072958, and P30AG028716) and the Critical Path for Parkinson’s Consortium (M.L.T.M., A.P., V.A., and D.S.).

The Rush Parkinson’s Disease and Movement Disorders Program is a designated Clinical Center of Excellence supported by the Parkinson Foundation.

CONFLICT OF INTEREST

The authors declared no competing interests for this work.

ORCID

Haotian Zou  <https://orcid.org/0000-0002-3595-8716>

Varun Aggarwal  <https://orcid.org/0000-0002-8406-0961>

Martijn L. T. M. Müller  <https://orcid.org/0000-0002-1133-7202>

REFERENCES

- McFarthing K, Rafaloff G, Baptista MAS, Wyse RK, Stott SRW. Parkinson’s disease drug therapies in the clinical trial pipeline: 2021 update. *J Parkinsons Dis*. 2021;11:891-903.
- Hauser RA. Help cure Parkinson’s disease: please don’t waste the Golden Year. *npj Park Dis*. 2018;4:1-2.
- Goetz CG, Tilley BC, Shaftman SR, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS): scale presentation and clinical testing results. *Mov Disord Off J Mov Disord Soc*. 2008;23:2129-2170.
- Marek K, Jennings D, Lasch S, et al. The Parkinson progression marker initiative (PPMI). *Prog Neurobiol*. 2011;95:629-635.
- Malek N, Swallow DM, Grosset KA, et al. Tracking Parkinson’s: study design and baseline patient data. *J Parkinsons Dis*. 2015;5:947-959.
- Tan AH, Lim SY, Mahadeva S, et al. *Helicobacter pylori* eradication in Parkinson’s disease: a randomized placebo-controlled trial. *Mov Disord*. 2020;35:2250-2260.
- Martinez-Fernández R, Mánuez-Miró JU, Rodríguez-Rojas R, et al. Randomized trial of focused ultrasound subthalamotomy for Parkinson’s disease. *N Engl J Med*. 2020;383:2501-2513.
- Athauda D, Maclagan K, Skene SS, et al. Exenatide once weekly versus placebo in Parkinson’s disease: a randomised, double-blind, placebo-controlled trial. *Lancet*. 2017;390:1664-1675.
- Olanow CW, Factor SA, Espay AJ, et al. Apomorphine sublingual film for off episodes in Parkinson’s disease: a randomised, double-blind, placebo-controlled phase 3 study. *Lancet Neurol*. 2020;19:135-144.
- Gottipati G, Berges AC, Yang S, Chen C, Karlsson MO, Plan EL. Item response model adaptation for analyzing data from different versions of Parkinson’s disease rating scales. *Pharm Res*. 2019;36:1-14.
- Regnault A, Boroojerdi B, Meunier J, Bani M, Morel T, Cano S. Does the MDS-UPDRS provide the precision to assess progression in early Parkinson’s disease? Learnings from the Parkinson’s progression marker initiative cohort. *J Neurol*. 2019;266:1927-1936.
- Sheng Y, Zhou X, Yang S, Ma P, Chen C. Modelling item scores of Unified Parkinson’s Disease Rating Scale Part III for greater trial efficiency. *Br J Clin Pharmacol*. 2021;87:3608-3618.
- Buatois S, Retout S, Frey N, Ueckert S. Item response theory as an efficient tool to describe a heterogeneous clinical rating scale

- in de novo idiopathic Parkinson's disease patients. *Pharm Res*. 2017;34:2109-2118.
14. Arrington L, Ueckert S, Ahamadi M, Macha S, Karlsson MO. Performance of longitudinal item response theory models in shortened or partial assessments. *J Pharmacokinet Pharmacodyn*. 2020;47:461-471.
 15. Gottipati G, Karlsson MO, Plan EL. Modeling a composite score in Parkinson's disease using item response theory. *AAPS J*. 2017;19:837-845.
 16. Wang J, Luo S. Multidimensional latent trait linear mixed model: an application in clinical studies with multivariate longitudinal outcomes. *Stat Med*. 2017;36:3244-3256.
 17. Luo S, Zou H, Goetz CG, et al. Novel approach to Movement Disorder Society–Unified Parkinson's Disease Rating Scale monitoring in clinical trials: longitudinal item response theory models. *Mov Disord Clin Pract*. 2021;8:1083-1091.
 18. Hoehn MM, Yahr MD. Parkinsonism: onset, progression, and mortality. *Neurology*. 1998;50:318.
 19. Yarnall AJ, Breen DP, Duncan GW, et al. Characterizing mild cognitive impairment in incident Parkinson disease: the ICICLE-PD study. *Neurology*. 2014;82:308-316.
 20. Szweczyk-Krolikowski K, Tomlinson P, Nithi K, et al. The influence of age and gender on motor and non-motor features of early Parkinson's disease: initial findings from the Oxford Parkinson disease center (OPDC) discovery cohort. *Parkinsonism Relat Disord*. 2014;20:99-105.
 21. Parkinson Study Group STEADY-PD III Investigators. Isradipine versus placebo in early Parkinson disease: a randomized trial. *Ann Intern Med*. 2020;172:591-598.
 22. Bluett B, Togasaki DM, Mihaila D, et al. Effect of urate-elevating inosine on early Parkinson disease progression: the SURE-PD3 randomized clinical trial. *JAMA*. 2021;326:926-939.
 23. Fox J-P. *Bayesian Item Response Modeling: Theory and Applications*. Springer; 2010.
 24. Stan Development Team. *Stan Modeling Language Users Guide and Reference Manual*, v.2.22.1. 2021.
 25. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2013.
 26. Wang J, Luo S, Li L. Dynamic prediction for multiple repeated measures and event time data: an application to Parkinson's disease. *Ann Appl Stat*. 2017;11:1787.
 27. Vandemeulebroecke M, Bornkamp B, Krahnke T, Mielke J, Monsch A, Quarg P. A longitudinal item response theory model to characterize cognition over time in elderly subjects. *CPT Pharmacometrics Syst Pharmacol*. 2017;6:635-641.
 28. Chae D, Chung SJ, Lee PH, Park K. Predicting the longitudinal changes of levodopa dose requirements in Parkinson's disease using item response theory assessment of real-world Unified Parkinson's Disease Rating Scale. *CPT Pharmacometrics Syst Pharmacol*. 2021;10:611-621.
 29. Chen C, Jönsson S, Yang S, Plan EL, Karlsson MO. Detecting placebo and drug effects on Parkinson's disease symptoms by longitudinal item-score models. *CPT Pharmacometrics Syst Pharmacol*. 2021;10:309-317.
 30. Ueckert S, Plan EL, Ito K, et al. Improved utilization of ADAS-cog assessment data through item response theory based pharmacometric modeling. *Pharm Res*. 2014;31:2152-2165.
 31. Krekels EHV, Novakovic AM, Vermeulen AM, Friberg LE, Karlsson MO. Item response theory to quantify longitudinal placebo and paliperidone effects on PANSS scores in schizophrenia. *CPT Pharmacometrics Syst Pharmacol*. 2017;6:543-551.
 32. Novakovic AM, Krekels EHV, Munafa A, Ueckert S, Karlsson MO. Application of item response theory to modeling of expanded disability status scale in multiple sclerosis. *AAPS J*. 2017;19:172-179.
 33. Bascoul-Mollevis C, Barbieri A, Bourcier C, et al. Longitudinal analysis of health-related quality of life in cancer clinical trials: methods and interpretation of results. *Qual Life Res*. 2021;30:91-103.
 34. Tosin MH de s, Goetz CG, Luo S, Choi D, Stebbins GT. Item response theory analysis of the MDS-UPDRS motor examination: tremor vs. nontremor items. *Mov Disord*. 2020;35:1587-1595.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Zou H, Aggarwal V, Stebbins GT, et al. Application of longitudinal item response theory models to modeling Parkinson's disease progression. *CPT Pharmacometrics Syst Pharmacol*. 2022;11:1382-1392. doi:[10.1002/psp4.12853](https://doi.org/10.1002/psp4.12853)