


Novel Approach to Movement Disorder Society–Unified Parkinson’s Disease Rating Scale Monitoring in Clinical Trials: Longitudinal Item Response Theory Models

Sheng Luo, PhD,¹  Haotian Zou, MS,² Christopher G. Goetz, MD,³ Dongrak Choi, MS,¹ David Oakes, PhD,⁴ Tanya Simuni, MD,⁵ and Glenn T. Stebbins, PhD^{3,*}

ABSTRACT: Background: Although nontremor and tremor Part 3 Movement Disorder Society–Unified Parkinson’s Disease Rating Scale items measure different impairment domains, their distinct progression and drug responsivity remain unstudied longitudinally. The total score may obscure important time-based and treatment-based changes occurring in the individual domains.

Objective: Using the unique advantages of item response theory (IRT), we developed novel longitudinal unidimensional and multidimensional models to investigate nontremor and tremor changes occurring in an interventional Parkinson’s disease (PD) study.

Method: With unidimensional longitudinal IRT, we assessed the 33 Part 3 item data (22 nontremor and 10 tremor items) of 336 patients with early PD from the STEADY-PD III (Safety, Tolerability, and Efficacy Assessment of Isradipine for PD, placebo vs. isradipine) study. With multidimensional longitudinal IRT, we assessed the progression rates over time and treatment (in overall motor severity, nontremor, and tremor domains) using Markov Chain Monte Carlo implemented in Stan.

Results: Regardless of treatment, patients showed significant but different time-based deterioration rates for total motor, nontremor, and tremor scores. Isradipine was associated with additional significant deterioration over placebo in total score and nontremor scores, but not in tremor score. Further highlighting the 2 separate latent domains, nontremor and tremor severity changes were positively but weakly correlated (correlation coefficient, 0.108).

Conclusions: Longitudinal IRT analysis is a novel statistical method highly applicable to PD clinical trials. It addresses limitations of traditional linear regression approaches and previous IRT investigations that either applied cross-sectional IRT models to longitudinal data or failed to estimate all parameters simultaneously. It is particularly useful because it can separate nontremor and tremor changes both over time and in response to treatment interventions.

Parkinson’s disease (PD) is a progressive neurodegenerative disorder characterized by distinctive motor and nonmotor manifestations. Symptom severity and progression in PD can be

evaluated using disease-specific clinical rating scales. Among them, the International Parkinson and Movement Disorder Society revision of the Unified Parkinson’s Disease Rating Scale (MDS-

¹Department of Biostatistics and Bioinformatics, Duke University, Durham, North Carolina, USA; ²Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA; ³Department of Neurological Sciences, Section of Movement Disorders, Rush University Medical Center, Chicago, Illinois, USA; ⁴University of Rochester Medical Center, Department of Biostatistics and Computational Biology, Rochester, New York, USA; ⁵Parkinson’s disease and Movement Disorders Center, Northwestern University Medical Center, Chicago, Illinois, USA

*Correspondence to: Dr. Glenn T. Stebbins, Department of Neurological Sciences, Section of Movement Disorders, Rush University Medical Center, 1725 West Harrison Street, Suite 755, Chicago, IL, USA; E-mail: gstebbin@rush.edu

Keywords: Parkinson’s disease, clinimetrics, disease progression, longitudinal data, Bayesian modeling. Relevant disclosures and conflicts of interest are listed at the end of this article.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 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.

Received 4 June 2021; revised 5 July 2021; accepted 9 July 2021.

Published online 00 Month 2021 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mdc3.13311

UPDRS) has become the most widely used scale for measuring parkinsonian symptoms in clinical and research practice. Specifically, MDS-UPDRS Part 3 or Motor Examination consists of a total of 33 items with 23 nontremor items (items 3.1–3.14) and 10 tremor items (3.15a–3.18). Each item is measured by a 5-point Likert scale (0–4), with higher values denoting increased motor severity.

To measure PD motor severity, the MDS-UPDRS Part 3 item values are typically added to obtain a sum score that is treated as a continuous variable and analyzed using regression models with the assumption of a normal distribution. However, this approach is problematic because (1) the sum score is discrete, asymmetric with ceiling or floor effects, and items have varying sensitivities to change (termed curvilinearity), and (2) as a sum, the value risks a loss of information because improvements and declines could balance out to a stable score despite significant impairment changes.¹ There is a low rate of change of MDS-UPDRS scores in early PD that, coupled with high variance, leads to large sample size requirements to test efficacy specifically in disease modification trials.²

Item response theory (IRT) models can be used to analyze data from rating scales such as the MDS-UPDRS^{3–6} in a way that addresses these highly pertinent clinical and statistical issues. By examining the relationship of items and scales to the underlying latent trait of PD severity, captured conceptually by the term *theta*, IRT approaches can identify curvilinearity among potential subdomains. For example, using a cross-sectional IRT analysis, our prior work demonstrated that a multidimensional model of the MDS-UPDRS Part 3 was superior to a unidimensional model, and 2 specific domains were identified from the statistical modeling.⁴ Specifically, nontremor and tremor items emerged with very different relationships to the latent variable or *theta* of overall PD impairment severity, suggesting that a fundamentally different pathophysiological basis underlies each domain.⁴ Longitudinal IRT models can extend this finding to investigate patterns of domain-specific disease progression, considering nontremor as 1 *theta* component and tremor as a separate *theta* component thereby providing unbiased parameter estimates in longitudinal data analyses.⁴ It is of particular clinical interest to quantify and compare rates of progression in nontremor and tremor domains in a clinical trial so that time and treatment effects can be evaluated separately but simultaneously. Our longitudinal IRT models are novel because they fully use the repeated measures from the same subject, simultaneously estimate all unknown parameters, and sufficiently account for the correlation between different *theta* components. In contrast, in prior studies involving PD and other neurological disorders, previous IRT investigations have had the following limitations: (1) employed cross-sectional IRT models on longitudinal data,^{6,7} (2) adopted a 2-step sequential parameter estimation process that failed to estimate all parameters simultaneously,^{8–10} and (3) modeled the disease progression separately on each of the latent variables but without regard to their internal relationship.^{7,11}

The purpose of this study was to apply these unique longitudinal IRT models to investigate change over time and in response to treatment for overall motor severity and for the multidimensional traits of nontremor and tremor domains captured

by the Part 3 MDS-UPDRS Motor Examination score in a clinical trial. We selected STEADY-PD III study that recorded the MDS-UPDRS and prospectively followed patients on placebo or the investigational drug during a period of up to 36 months.¹² This analytic method allows an evaluation of the rates of overall motor severity progression over time and in response to an intervention captured by changes within both unidimensional and multidimensional (nontremor and tremor) models. In addition, this approach allows us to assess the strength of the association between nontremor and tremor impairments.

Patients and Methods

Study Population

Longitudinal data on the 33 items of the MDS-UPDRS Part 3 Motor Examination performed in the STEADY-PD III study (ClinicalTrials.gov: NCT02168842) were used to illustrate the proposed techniques. The STEADY-PD III was a 36-month, multicenter, randomized, parallel-group, placebo-controlled phase 3 study of the efficacy of isradipine in early-stage PD. A total of 336 patients with de novo PD within 3 year of diagnosis and without the use of dopaminergic medications at enrollment were randomized to either isradipine ($n = 170$) or placebo ($n = 166$).¹² Although most patients enrolled on no medication, they were allowed to receive amantadine ($n = 26$) or anticholinergic agents ($n = 5$) if on stable doses at the time of study entry. The primary outcome was the change in the original Unified Parkinson's Disease Rating Scale (UPDRS) Parts 1 to 3 score from baseline to 36 months.¹³ In addition to the UPDRS, the MDS-UPDRS was measured at baseline and months 12, 24, and 36. MDS-UPDRS was also administered at the premature withdrawal visit or the visit when symptomatic treatment started. However, because these visits could occur at any time during the study, we did not include them unless they occurred within 10 days of the scheduled visits and thereby replaced the scheduled visit. There were no missing item scores in the MDS-UPDRS. The reason for dropouts at visits was the initiation of symptomatic treatment in addition to the placebo or isradipine study medication. Because the MDS-UPDRS Part 3 scores potentially were markedly changed by symptomatic treatment when patients were on medications, we did not include the MDS-UPDRS Part 3 score measured after symptomatic treatment started.

Statistical Analysis

Unidimensional Longitudinal IRT Model

The unidimensional analysis assumed that there was a single latent variable *theta* in IRT models, representing the underlying overall parkinsonian motor severity (higher value for more severe status) manifested by all 33 MDS-UPDRS Part 3 items. The unidimensional longitudinal IRT model consists of 2 levels. The first level, a graded-response measurement model (model 1 in the Supplemental Materials), quantifies the relationship between

the response of each item and theta. The probability of every score in each item was determined by 5 parameters: the discrimination parameter and the 4 location parameters. Higher value in the discrimination parameter suggests that this item is powerful for determining the individual's overall motor 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) regresses the overall motor severity, indicated as theta, on time in years and the time by treatment interaction, with a subject-specific random intercept. The random intercept was assumed to have standard normal distribution with mean 0 and variance 1. Figure S1 displays an illustrative path diagram of the unidimensional longitudinal IRT model applying to the 33 MDS-UPDRS Part 3 items.

We adopted the unidimensional longitudinal IRT model on the MDS-UPDRS Part 3 data at all visits to investigate the progression in overall motor severity and the effects of isradipine in comparison with placebo using 95% credible intervals (95% CI), the Bayesian equivalence of 95% confidence intervals, to determine significant effects. To compare the estimates of the discrimination and location parameters, we implemented a cross-sectional graded-response IRT model on the MDS-UPDRS Part 3 data at baseline. To ensure the parameters could be correctly estimated, if the number of patients in the higher categories (more severe status, specifically 3 or 4) was less than 10, these patients were consolidated as having a score of 3 or lower.

Multidimensional Longitudinal IRT Model

The multidimensional longitudinal IRT approach allows for more than 1 latent variable. Based on our previous findings,⁴ we considered the latent variable of disease severity, termed theta, as multivariate with tremor and nontremor function as separate theta components, the former captured by 23 nontremor items (items 3.1–3.14 measuring bradykinesia, rigidity, gait, and posture, with a total score range of 0–92) and the latter covering 10 tremor items (items 3.15a–3.18 measuring tremor, with a total score range of 0–40). The first level graded-response measurement models (models 3 and 4 in the Supplemental Materials) are similar to the unidimensional longitudinal IRT model. In the second-level structural models (models 5 and 6 in the Supplemental Materials), both the nontremor theta component and the tremor theta component were regressed on time in years and the time by treatment interaction with subject-specific random intercepts. The random intercepts were assumed to follow a standard normal distribution with a correlation coefficient. We adopted the multidimensional longitudinal IRT model on the MDS-UPDRS Part 3 data to investigate the progression in nontremor and tremor motor severities and the effects of isradipine in comparison to placebo. Figure S2 displays an illustrative path diagram of the multidimensional longitudinal IRT

model applying to the 23 nontremor items and 10 tremor items. Because the latent variables, or theta components, are dimensionless, the rates of progression in nontremor and tremor domains can be directly compared and tested, thereby providing additional clinically relevant information not available in the unidimensional longitudinal IRT model. Moreover, we implemented 2 separate unidimensional longitudinal IRT models on the longitudinal Part 3 nontremor and tremor items (referred to as the confirmational model). To compare the unidimensional and multidimensional longitudinal IRT models, and the confirmational model, we used the deviance information criterion (DIC), where lower values indicate better fit to the model. As sensitivity analysis, we also implemented separate linear mixed models on the MDS-UPDRS Part 3 sum score and on sum scores of Part 3 nontremor and tremor items.

Model Fitting Using Bayesian Inference

The analysis was conducted using Bayesian inference based on Markov Chain Monte Carlo posterior simulations implemented in Stan (version 2.26)¹⁴ via interface in the R statistical program (version 4.0.3).¹⁵ 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.¹⁶

Typically, IRT results are expressed as unitless theta values to allow for an equivalency between subdomains represented by different numbers of rating items. This allows for direct comparisons between various models regardless of differences in the number of items within a given domain. To facilitate clinical interpretation of the regression coefficients, which are expressed in theta values of the overall motor severity, and in theta component values of the nontremor and tremor domains, we can interpret them in terms of MDS-UPDRS point scores via simulation using the posterior samples from Bayesian inference. Although allowing a metric (MDS-UPDRS Part 3 point changes) that may provide more clinically meaningful impact for clinicians, these simulations are associated with increased variance that may lead to statistically insignificant results. Therefore, we have expressed the primary results in unitless theta or theta component values, but provided the simulation point-based results in the Supplemental Materials.

Results

From the original STEADY-PD III study of 336 patients with PD (170 on isradipine and 166 on placebo), we considered only visits before the initiation of any dopaminergic therapy, thereby using 670 observations (336 at baseline, 149 at year 1, 88 at year 2, and 69 at year 3 in addition to 28 at visits when symptomatic treatment started) with a mean length of 10.1 months of follow-up (standard deviation [SD], 12.1 months). Baseline characteristics of the cohort (Table 1) are derived from the previously

TABLE 1 Baseline demographic and disease characteristics

Characteristic	Isradipine	Placebo
Sample size, N	170	166
Age, years, mean (SD)	62.1 (8.7)	61.6 (9.3)
Sex, n (%)		
Male, n (%)	122 (71.8)	108 (65.1)
Female, n (%)	48 (28.2)	58 (34.9)
Non-Hispanic White race, n (%)	156 (91.8)	148 (88.0)
UPDRS score Part 3, mean (SD)	18.0 (7.3)	16.3 (6.53)
Hoehn and Yahr stage, mean (SD)	1.69 (0.46)	1.57 (0.51)
Modified Rankin score, mean (SD)	1.09 (0.31)	1.09 (0.33)
Systolic blood pressure seated, mm Hg, mean (SD)	128.1 (17.2)	127.7 (14.6)
Disease duration from diagnosis, months, mean (SD)	9.89 (8.1)	10.6 (9.4)

Abbreviations: SD, standard deviation; UPDRS, Unified Parkinson's Disease Rating Scale.

published primary article⁷: mean age 62.1 (SD, 8.7) and 61.6 (SD, 9.3) years and mean disease duration from diagnosis 9.9 (SD, 8.1) and 10.6 (SD, 9.4) months in the isradipine and placebo groups, respectively.¹² Figure 1 displays the bar plots (with SD bars) of sum of scores (upper left panel, Part 3 sum score; lower left panel, nontremor sum score; lower right panel, tremor sum score) over visit times.

With the unidimensional longitudinal IRT model using all 33 Part 3 items, there was a statistically significant time effect for both placebo and isradipine groups, demonstrating a deterioration in motor severity at the rate of 0.123 theta values per year (95% CI, 0.081–0.167; Table 2). Moreover, there was a differential treatment effect, isradipine being associated with additional significant deterioration over the placebo group at the added rate of 0.065 theta values per year (95% CI, 0.009–0.126). With these 2 rates (time and treatment) summed, the isradipine patients showed a total rate of deterioration of 0.188 theta values per year (95% CI, 0.147–0.231). Figure 2 (left panel) displays in theta values the different rates of progression of the 2 groups. Please note that at baseline when time was 0, the overall motor severity theta equaled to 0. Hence, the change from baseline of theta was equivalent to theta. Supplementary Section S2 presents the estimates of the discrimination and location parameters from the

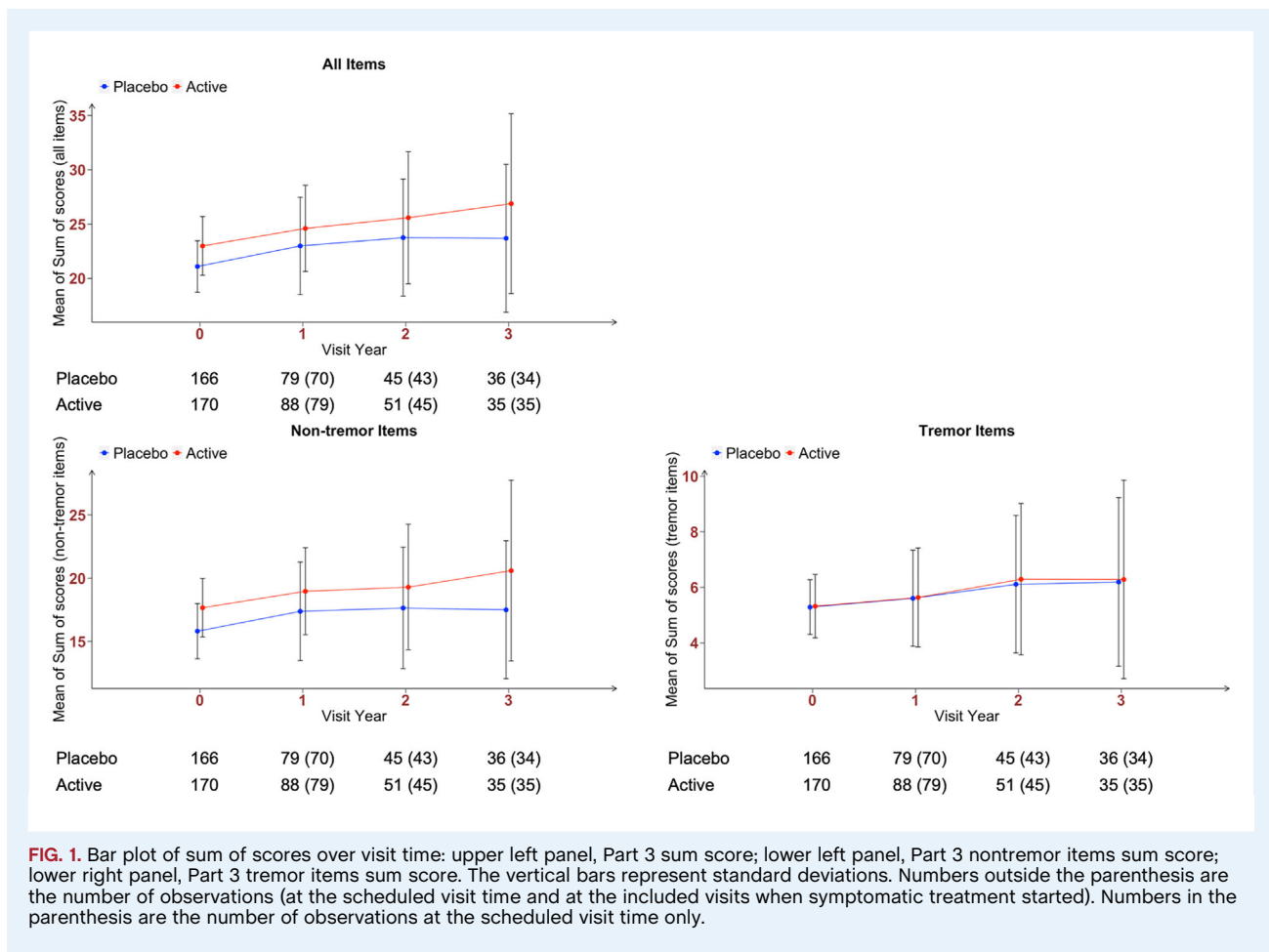


FIG. 1. Bar plot of sum of scores over visit time: upper left panel, Part 3 sum score; lower left panel, Part 3 nontremor items sum score; lower right panel, Part 3 tremor items sum score. The vertical bars represent standard deviations. Numbers outside the parenthesis are the number of observations (at the scheduled visit time and at the included visits when symptomatic treatment started). Numbers in the parenthesis are the number of observations at the scheduled visit time only.

TABLE 2 Estimates of the regression parameters from model 2 (see the Supplemental Materials) in the unidimensional longitudinal item response theory model

Parameters	Mean	SD	95% CI
Time	0.123	0.022	0.081–0.167
Isradipine × time	0.065	0.030	0.009–0.126

Abbreviations: SD, standard deviation; 95% CI, 95% credible intervals.

unidimensional longitudinal IRT model and their comparison with those from a cross-sectional graded-response IRT model on the MDS-UPDRS Part 3 data at baseline (Supplementary Section S3).

With the multidimensional longitudinal IRT model fitted to assess separately the 23 nontremor items and 10 tremor items, there was a significant time effect for both placebo and isradipine groups on deterioration for both nontremor motor severity and tremor motor severity. For nontremor motor severity, the rate of decline was 0.120 theta component values per year (95% CI, 0.076–0.165; Table 3), whereas for tremor motor severity, the rate of decline was 0.123 theta component values per year (95% CI, 0.053–0.197). The difference of -0.003 theta component values per year (95% CI, -0.091 to 0.082) between the nontremor and tremor deterioration was not statistically significant. There was, however, a significant time by treatment interaction with isradipine being associated with additional significant deterioration in nontremor motor severity over the placebo

group, giving the patients in isradipine group the rate of 0.195 theta component values per year (95% CI, 0.146–0.243). In comparison, isradipine was not associated with additional significant deterioration in tremor motor severity (0.046 theta component values per year; 95% CI, -0.056 to 0.149). Figure 2 (right panel) displays in theta component values the different rates of progression of the 2 groups in both the nontremor and tremor domains. It suggests that in the placebo group, the rate of progression in the nontremor domain was slightly slower than the tremor domain, whereas the opposite trend was observed in the isradipine group. Further highlighting the 2 separate domains, nontremor and tremor severity changes were positively but weakly correlated (correlation coefficient, 0.108; 95% CI, -0.138 to 0.399). Supplementary Section S4 presents the estimates of the discrimination and location parameters from the multidimensional longitudinal IRT model (Table S3). Supplementary Section S5 presents a comprehensive assessment of the multidimensional longitudinal IRT model performance using diagnostics based on residuals and simulation. Figures S3 to S6 suggest that the model provides excellent goodness of fit and model performance.

As confirmation, unidimensional longitudinal IRT models applied separately to the nontremor and tremor items (referred to as the confirmational model; Supplementary Section S6 and Tables S4–S7) were very close to the single multidimensional longitudinal IRT analysis. We also compared the goodness of fit in DICs and obtained the following results: unidimensional longitudinal IRT model $DIC = 38,285$, multidimensional

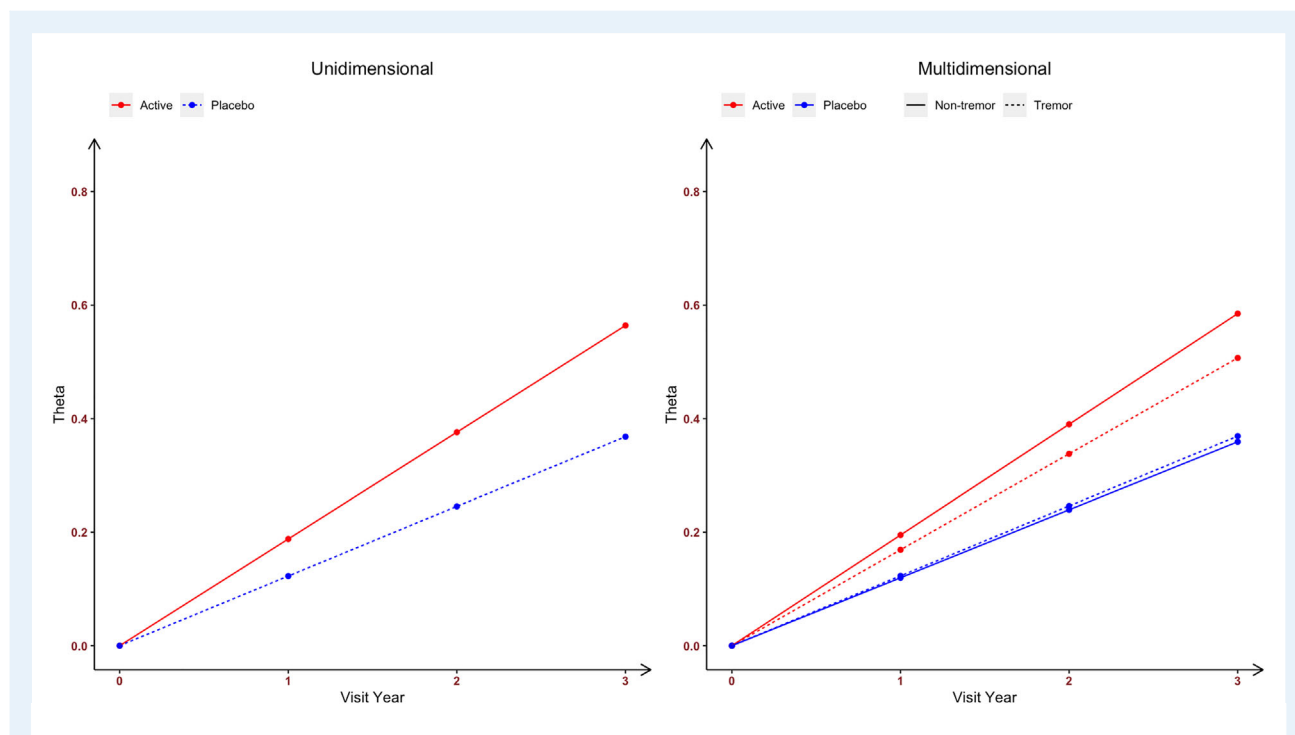


FIG. 2. The estimated rates of progression among patients in the isradipine and placebo groups in overall parkinsonian motor severity (left panel, from the unidimensional longitudinal item response theory model) and in nontremor and tremor motor severities (right panel, from the multidimensional longitudinal item response theory model).

TABLE 3 Estimates of the parameters from models 5 and 6 (see the Supplemental Materials) in the multidimensional longitudinal item response theory model

Parameters	Mean	SD	95% CI
Nontremor domain			
Time	0.120	0.023	0.076–0.165
Isradipine × time	0.075	0.032	0.011–0.138
Tremor domain			
Time	0.123	0.037	0.053–0.197
Isradipine × time	0.046	0.053	−0.056 to 0.149
Correlation coefficient	0.108	0.153	−0.138 to 0.399

Abbreviations: SD, standard deviation; 95% CI, 95% credible intervals.

longitudinal IRT model DIC = 37,366, and confirmational model DIC = 37,373, indicating a superior fit for the multidimensional longitudinal IRT model. Supplementary Section S7 presents the sensitivity analysis results of the linear mixed models on the MDS-UPDRS Part 3 sum score and on the sum scores of Part 3 nontremor and tremor items. Tables S8 and S9 suggest significant isradipine effects in increasing Part 3 sum score ($P = 0.047$) and in increasing nontremor sum score ($P = 0.029$), whereas Table S10 suggests small and insignificant isradipine effects in increasing tremor sum score ($P = 0.807$). These results are consistent with our findings using unidimensional and multidimensional longitudinal IRT models.

Discussion

It has been increasingly common to apply longitudinal IRT model in the study of complex diseases, for example, PD,^{7,8,11} Alzheimer's disease and dementia,^{10,17} multiple sclerosis,⁹ and schizophrenia.¹⁸ All of them have been contributory to the field, but none incorporating the full array of modeling techniques that we applied in a simultaneous manner. Buatois et al,⁸ Ueckert et al,¹⁰ and Novakovic et al⁹ fixed item-specific parameters to those estimated from a cross-sectional, not longitudinal, IRT model. Sheng et al⁷ and Gottipati et al¹¹ fit each of their identified latent variables separately and did not account for their correlation. Krekels et al¹⁸ analyzed the 3 subscales of the Positive and Negative Syndrome Scale as different domains and analyzed them separately without attention to potential correlations among each. Finally, Vandemeulebroecke et al¹⁷ applied a unidimensional longitudinal IRT model to the 14 main test items of the 2 neuropsychological test batteries and did not investigate the multidimensionality of the latent ability.

In contrast to all of these analyses, we consider our approach both novel and innovative because our strategy has been to integrate three simultaneous analyses not specifically conducted before in unison. Specifically, our results were derived from a unique application of longitudinal IRT modeling where item-specific discrimination and location parameters in the

measurement model and regression parameters in the structural models were estimated simultaneously. Next, our longitudinal IRT models fully accounted for the correlation among visits from the same subject. Lastly, our multivariate longitudinal IRT model simultaneously estimated and compared rates of progression in nontremor and tremor domains and modeled the correlation between domains. Moreover, we did not consider bifactor IRT models because the existing bifactor IRT models are only applicable for cross-sectional data and as these data unveil, it is no longer reasonable to assume that a single latent variable representing the underlying overall parkinsonism motor severity was best suited for PD.

In our view, this type of multidimensional longitudinal IRT analysis provides a new and highly focused statistical approach to detect changes that could have clinical implications obscured by traditional analytic methods. As shown in our analysis, both the time-based changes that occurred during the study and the treatment response to isradipine in comparison to placebo followed different patterns for tremor and nontremor impairments. The ability to detect relative improvements or exacerbations in 1 domain, even when the other domain does not change or even when the overall Part 3 sum score does not change, offers important advantages for patient safety and efficacy monitoring. Specifically, 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 a unidimensional analysis. Our approach can be viewed as a complement to usual analytic approaches (eg, linear mixed models) based on sum scores. It resolves several of their limitations such as ignoring varying item sensitivities to change and the sum score balancing out improvements and declines manifested in different items.

Furthermore, our approach has unique attributes that offer insights that the traditional analytic approaches cannot address. Our results support the multidimensional constructs implicit to PD with a superior fit over a unidimensional construct, as shown by the significant DIC comparisons. The multidimensional IRT analysis confirms that, in PD, nontremor and tremor changes occur together as part of natural disease progression, but they do not progress at the same pace.

As reported previously,⁴ statistical modeling has shown that the MDS-UPDRS Part 3 nontremor and tremor items have very different relations to the construct of PD severity, and this longitudinal study shows that they also have different patterns of treatment responses. Although the items included in each domain make clinical sense, they were derived purely from the statistical models.⁴ We view this scientific article as a methodological study. It is not our intent to ascribe or interpret clinical significance to these specific changes, but we emphasize the power of the statistical strategy to separate time and treatment effects by domains with this method in a clinical context. Whereas the domains of nontremor and tremor were statistically defined, not selected by us with bias or choice, the idea of exploring other domains relative to PD, especially those involving nonmotor elements, could be pursued with similar analyses of other inventories.

In our analysis, the results are presented in unitless theta or theta component values that can obscure a direct clinical interpretation in terms of MDS-UPDRS point changes. Our method, however, allows for simulation of MDS-UPDRS points from the change in theta or theta component values, although that simulation is associated with increased variance. Such simulations can permit the theta or theta component values to be interpreted in a clinically meaningful context (see the Supplemental Materials). For example, in our model, the simulated annual rate of MDS-UPDRS point changes for total Part 3 would be 0.971 points, but divergent for nontremor (0.784 points) and tremor (0.551 points) domains, suggesting 2 related but distinct pathophysiological mechanisms of disease progression.

In the original STEADY-PD III primary analysis,¹² the prespecified outcome measure focused on the original UPDRS, not the MDS-UPDRS, and we plan to collaborate with the STEADY-PD III Parkinson Study Group Investigators to apply this IRT approach to the original UPDRS. Although we did demonstrate a faster rate of impairment progression in the isradipine group, that observation should be interpreted cautiously as the analysis included data only to the time of initiation of dopaminergic therapy and did not explore progression during the full 3 years of the trial. Indeed, a recent post hoc analysis of the STEADY-PD III demonstrated a positive effect of isradipine on the time to initiation of dopaminergic therapy in the highest dose exposure group.¹⁹ A review of former “failed” studies could also examine the nontremor and tremor patterns again as separate domains using this approach to dissect whether related improvements or exacerbations were obscured by traditional statistical approaches.

As a confirmatory exercise, we showed that the multidimensional longitudinal IRT model had a superior fit as compared with the unidimensional longitudinal IRT model and the separate analyses of nontremor and tremor items. At the clinical level, these results suggest (1) it is reasonable to consider nontremor and tremor severities as 2 separate domains, in addition to an overall parkinsonian motor severity; and (2) the nontremor and tremor theta components are correlated, suggesting that the domains, although providing different information on the progression of disease and treatment effects, do reflect an overall measure of PD severity.

We fully acknowledge that 1 limitation is the small number of observations because the MDS-UPDRS was not part of the standard assessment at all visits in the STEADY-PD III study. Furthermore, we were examining predopaminergic intervention, and the

large number of patients, both in the isradipine and placebo assignment groups, required dopaminergic intervention before the full program completed.

In both unidimensional and multidimensional longitudinal IRT models, we have assumed linear progression in disease severity and only included random intercepts, not random slopes, due to the relatively small number of observations. In studies where MDS-UPDRS measurements are more frequent, for example, the Parkinson's Progression Markers Initiative study (<https://www.ppmi-info.org/>), the linearity assumption can potentially be relaxed by adding nonlinear terms or splines, and models with random slopes can be thoroughly investigated. As next steps, we plan to extend and further validate our findings accessing other clinical trial data sets with more MDS-UPDRS data points and possibly longer durations. We welcome collaborations with colleagues interested in developing new approaches to clinical trial design and data analysis.

Acknowledgments

The full listing of Parkinson Study Group authors of the original STEADY-PD III primary report can be found in reference 12.

Author Roles

(1) Research Project: A. Conception, B. Organization, C. Execution; (2) Statistical Analysis: A. Design, B. Execution, C. Review and Critique; (3) Manuscript Preparation: A. Writing of the First Draft, B. Review and Critique.

S.L.: 1A, 1B, 1C, 2A, 2C, 3A, 3B

H.Z.: 2A, 2B, 2C, 3B

C.G.G.: 1B, 2C, 3B

D.C.: 2A, 2B, 2C, 3B

D.O.: 2C, 3B

T.S.: 2C, 3B

G.T.S.: 1A, 1B, 1C, 3B

Disclosures

Ethical Compliance Statement: The authors confirm that the approval of an institutional review board was not required for this work. Informed consent was obtained from all patients in the original STEADY-PD III study. We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this work is consistent with those guidelines.

Funding Sources and Conflicts of Interest: The research of Sheng Luo was supported by National Institute on Aging (Grant R01AG064803). The Rush Parkinson's Disease and Movement Disorders Program is a designated Clinical Center of Excellence supported by the Parkinson Foundation. The authors have no conflicts of interest to report.

Financial Disclosures for the Previous 12 Months: S.L. reports the following: consulting and advisory board membership

with honoraria with the National Institutes of Health and CHDI Management, Inc.; grants and research from the National Institutes of Health, CHDI, International Parkinson and Movement Disorder Society, and Parkinson's Foundation; honoraria from St. Jude Children's Hospital; and a salary from Duke University. H.Z. reports a salary from University of North Carolina at Chapel Hill. C.G.G. reports the following: funding to Rush University Medical Center from the National Institutes of Health, Department of Defense, and The Michael J. Fox Foundation for research conducted by Dr. Goetz; presidential stipend from the International Parkinson and Movement Disorder Society paid to Rush University Medical Center as part of Dr. Goetz's salary; faculty stipends from the American Academy of Neurology; guest professorship honorarium provided by the University of Chicago and NorthShore University Health System; royalties from Elsevier Publishers and Wolters Kluwer Publishers; and a salary from Rush University Medical Center. D.C. reports a salary from Duke University. D.O. reports the following: data and safety monitoring board of University of Pennsylvania; grants/research from the National Institutes of Health, The Michael J. Fox Foundation, and Grey Matter Technology; royalties from Springer and Taylor and Francis; and a salary from University of Rochester Medical Center. T.S. reports the following: consulting or advisory board membership with honoraria from Acadia, Abbvie, Accordia, Adamas, Allergan, Amneal, Aptinyx, Denali, General Electric (GE), Kyowa, Neuroderm, Neurocrine, Sanofi, Sinopia, Sunovion, Roche, Takeda, Voyager, and US World Meds; grants/research from Biogen, Roche, Neuroderm, Sanofi, Sun Pharma, Abbvie, IMPAX, Prevaile, National Institute of Neurological Disorders and Stroke (NINDS), Michael J. Fox Foundation for Parkinson's Research (MJFF), and the Parkinson's Foundation; honoraria from Sanofi; and a salary from Northwestern University. G.T.S. reports the following: consulting and advisory board membership with honoraria from Acadia, Pharmaceuticals, Adamas Pharmaceuticals, Inc., Biogen, Inc., Ceregene, Inc., CHDI Management, Inc., Cleveland Clinic Foundation, Ingenix Pharmaceutical Services (i3 Research), MedGenesis Therapeutix, Inc., Neurocrine Biosciences, Inc., Pfizer, Inc., Tools-4-Patients, Ultragenyx, Inc., and the Sunshine Care Foundation; grants and research from the National Institutes of Health, Department of Defense, The Michael J. Fox Foundation for Parkinson's Research, Dystonia Coalition, CHDI, Cleveland Clinic Foundation, International Parkinson and Movement Disorder Society, and CBD Solutions; honoraria from the International Parkinson and Movement Disorder Society, American Academy of Neurology, The Michael J. Fox Foundation for Parkinson's Research, Food and Drug Administration, National Institutes of Health, and the Alzheimer's Association; and a salary from Rush University Medical Center. ■

References

- 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.
- Holden SK, Finseth T, Sillau SH, Berman BD. Progression of MDS-UPDRS scores over five years in de novo Parkinson disease from the Parkinson's progression markers initiative cohort. *Mov Disord Clin Pract* 2018;5:47–53.
- Arrington L, Ueckert S, Ahmadi M, Macha S, Karlsson MO. Performance of longitudinal item response theory models in shortened or partial assessments. *J Pharmacokinet Pharmacodyn* 2020;47:461–471.
- de Siqueira Tosin MH, 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.
- 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;36:135.
- 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 for greater trial efficiency [published online ahead of print 2021]. *Br J Clin Pharmacol*. <https://doi.org/10.1111/bcp.14777>.
- 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.
- Novakovic AM, Krekels EH, Munafo 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.
- Ueckert S, Plan EL, Ito K, Karlsson MO, Corrigan B, Hooker AC. Improved utilization of ADAS-cog assessment data through item response theory based pharmacometric modeling. *Pharm Res* 2014;31:2152–2165.
- Gottipati G, Karlsson MO, Plan EL. Modeling a composite score in Parkinson's disease using item response theory. *AAPS J* 2017;19:837–845.
- Parkinson Study Group STEADY-PD III Investigators. Isradipine versus placebo in early parkinson disease: a randomized trial. *Ann Intern Med* 2020;172(9):591–598.
- FAHN S, Elton R. *Recent Developments in Parkinson's Disease*. Vol 2. Florham Park, NJ: Macmillan Health Care Information; 1987:293–304.
- Stan Development Team Stan modeling language user's guide and reference manual 2017. <http://mc-stan.org/users/documentation/>
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2020.
- 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;11:1787–1809.
- 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.
- Krekels E, Novakovic AM, Vermeulen A, 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.
- Venuto CS, Yang L, Surmeier DJ, et al. Pharmacokinetics of isradipine in participants with Parkinson's disease from the Phase 3 STEADY-PD clinical trial. *Mov Disord Clin Pract* 2020;7:S17.

Supporting Information

Supporting information may be found in the online version of this article.

Table S1. Estimates of discrimination (Discrim) and item location parameters of Movement Disorder Society–Unified Parkinson's Disease Rating Scale Part 3 items from model 1 in the unidimensional longitudinal item response theory model.

Table S2. Estimates of discrimination (Discrim) and item location parameters for all Movement Disorder Society–Unified

Parkinson's Disease Rating Scale Part 3 items at baseline in a cross-sectional graded-response item response theory model.

Table S3. Estimates of discrimination (Discrim) and item location parameters of all Movement Disorder Society–Unified Parkinson's Disease Rating Scale Part 3 items from models 3 and 4 in the multidimensional longitudinal item response theory model.

Table S4. Estimates of discrimination (Discrim) and item location parameters of nontremor Movement Disorder Society–Unified Parkinson's Disease Rating Scale Part 3 items in the unidimensional longitudinal item response theory model.

Table S5. Estimates of the regression parameters for nontremor motor severity in the unidimensional longitudinal item response theory model.

Table S6. Estimates of discrimination (Discrim) and item location parameters of tremor Movement Disorder Society–Unified Parkinson's Disease Rating Scale Part 3 items in the unidimensional longitudinal item response theory model.

Table S7. Estimates of the regression parameters for tremor motor severity in the unidimensional longitudinal item response theory model.

Table S8. Regression parameters from the linear mixed model for Movement Disorder Society–Unified Parkinson's Disease Rating Scale Part 3 sum score and likelihood ratio test (LRT) of isradipine effect.

Table S9. Regression parameters from the linear mixed model for Movement Disorder Society–Unified Parkinson's Disease Rating Scale Part 3 nontremor items sum score and likelihood ratio test (LRT) of isradipine effect.

Table S10. Regression parameters from the linear mixed model for Movement Disorder Society–Unified Parkinson's Disease Rating Scale Part 3 tremor items sum score and likelihood ratio test (LRT) of isradipine effect.

Figure S1. A path diagram for the unidimensional longitudinal item response theory (IRT) model, where $\theta_i(t)$ is the overall motor severity of subject i at time t and it regresses on $X_i(t)$, a vector of covariates that impacts the progression of the motor severity. Items 3.1 to 3.18 are the 33 Movement Disorder Society–Unified Parkinson's Disease Rating Scale Part 3 items.

Figure S2. A path diagram for the multidimensional longitudinal item response theory (IRT) model, where $\theta_i^{non}(t)$ and $\theta_i^{trem}(t)$ are the nontremor and tremor theta components, respectively, of subject i at time t . Both severities regress on $X_i(t)$, a vector of covariates that impacts the progression of the nontremor and tremor theta components. Items 3.1 to 3.18 are the 33 Movement Disorder Society–Unified Parkinson's Disease Rating Scale Part 3 items.

Figure S3a. Mirror plot of observed (blue) versus simulated (red) portion of each category of items 3.1 to 3.4b (9 items) from all visits. The lines on the top of the red bar represent the standard error.

Figure S3b. Mirror plot of observed (blue) versus simulated (red) portion of each category of items 3.5a to 3.9 (9 items) from all visits. The lines on the top of the red bar represent the standard error.

Figure S3c. Mirror plot of observed (blue) versus simulated (red) portion of each category of items 3.10 to 3.16b (9 items) from all visits. The lines on the top of the red bar represent the standard error.

Figure S3d. Mirror plot of observed (blue) versus simulated (red) portion of each category of items 3.17a to 3.18 (6 items) from all visits. The lines on the top of the red bar represent the standard error.

Figure S4. Proportion of observed versus simulated data in each item category of all 33 Movement Disorder Society–Unified Parkinson's Disease Rating Scale Part 3 items from all visits. The line corresponds to the identify line.

Figure S5. Correlation between the residuals obtained using the multidimensional longitudinal IRT model, across 33 Movement Disorder Society–Unified Parkinson's Disease Rating Scale Part 3 items from all visits.

Figure S6. Visual predictive checks for the sum scores from all 33 Part 3 items (top panel), 23 nontremor items (middle panel), and 10 tremor items (bottom panel) comparing the median (purple lines) and 2.5th and 97.5th quantiles (orange lines) of the observed data (blue points) with the respective confidence intervals (shaded areas) based on the multidimensional longitudinal item response theory model.