ARTICLE

Detecting placebo and drug effects on Parkinson's disease symptoms by longitudinal item-score models

Chao Chen[1] | Siv Jönsson[2] | Shuying Yang[1] | Elodie L. Plan[2] | Mats O. Karlsson[2]

[1]Clinical Pharmacology Modelling and Simulation, GlaxoSmithKline, London, UK

[2]Department of Pharmacy, Uppsala University, Uppsala, Sweden

Correspondence
Elodie L. Plan, Department of Pharmacy, Uppsala University, Box 580, SE-75123 Uppsala, Sweden.
Email: elodie.plan@farmaci.uu.se

Abstract

This study tested the hypothesis that analyzing longitudinal item scores of the Unified Parkinson's Disease Rating Scale could allow a smaller trial size and describe a drug's effect on symptom progression. Two historical studies of the dopaminergic drug ropinirole were analyzed: a cross-over formulation comparison trial in 161 patients with early-stage Parkinson's disease, and a 24-week, parallel-group, placebo-controlled efficacy trial in 393 patients with advanced-stage Parkinson's disease. We applied item response theory to estimate the patients' symptom severity and developed a longitudinal model using the symptom severity to describe the time course of the placebo response and the drug effect on the time course. Similarly, we developed a longitudinal model using the total score. We then compared sample size needs for drug effect detection using these two different models. Total score modeling estimated median changes from baseline at 24 weeks (90% confidence interval) of −3.7 (−5.4 to −2.0) and −9.3 (−11 to −7.3) points by placebo and ropinirole. Comparable changes were estimated (with slightly higher precision) by item-score modeling as −2.0 (−4.0 to −1.0) and −9.0 (−11 to −8.0) points. The treatment duration was insufficient to estimate the symptom progression rate; hence the drug effect on the progression could not be assessed. The trial sizes to detect a drug effect with 80% power on total score and on symptom severity were estimated (at the type I error level of 0.05) as 88 and 58, respectively. Longitudinal item response analysis could markedly reduce sample size; it also has the potential for assessing drug effects on disease progression in longer trials.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
Large trials are required for detecting drug effects on the symptoms of Parkinson's disease, as measured by the total score of the Unified Parkinson's Disease Rating Scale instrument. Differentiating disease-modifying drugs from symptom-relief drugs is a major challenge.

Chao Chen and Siv Jönsson contributed equally to this work.

WHAT QUESTION DID THIS STUDY ADDRESS?
To what extent can analyzing item scores, instead of the total score, reduce the sample size of placebo-controlled trials?
WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?
Longitudinal modeling showed an approximately 30% sample-size reduction for detecting the treatment effect observed in a placebo-controlled trial of ropinirole.
HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?
Item-score analysis, where applicable, can reduce the time and cost of developing drugs for Parkinson's disease. Used in the longitudinal modeling of trials of adequate duration, it also has the potential to more efficiently estimate a drug's effect on the symptom progression rate.

INTRODUCTION

Parkinson's disease (PD) affects approximately 1% of the population aged older than 65 years.[1] It is a slowly progressive neurodegenerative disease manifested by a broad range of symptoms.[2] Diagnosis occurs with the onset of motor symptoms, for example, bradykinesia, muscular rigidity, rest tremor, and postural and gait impairment, but the disease begins years before diagnosis, with nonmotor symptoms such as constipation and rapid eye movement sleep behavior disorder. Following diagnosis, additional nonmotor features such as pain, fatigue, and mild cognitive impairment develop. In advanced PD, further motor and nonmotor symptoms develop, for example, postural instability with frequent falls, freezing of gait, urinary symptoms, orthostatic hypotension, and dementia. Accordingly, patients experience increasing clinically significant disability over time as a result of symptoms and long-term complications of dopaminergic therapy, including fluctuations, dyskinesia, and psychosis.[2] Available treatments relieve symptoms without clear evidence of reducing the rate of disease progression.[3]

The most widely used clinical measurements to quantify the disability and impairment of patients with PD over time are the Unified Parkinson's Disease Rating Scale (UPDRS) and, lately, the newer version, the Movement Disorder Society–sponsored revision of the UPDRS (MDS-UPDRS).[4,5] The scales assess the consequences of the disease in areas such as mental status, daily living, motor functionality, treatment complications, and adverse effects by scoring 44 questions (items) mostly from 0 to 4, with higher scores reflecting more severe disease. The items are composed of the following components: (I) mentation, behavior, and mood; (II) activities in daily living; and (III) motor examination (including left/right assessment).

The common approach for evaluating treatment effects is to analyze the sum of all item scores, that is, the total score. A key assumption for this approach is that all items have equal importance in reflecting the overall disability.[6] In contrast, item response theory (IRT) offers an alternative. It estimates a patient's disease severity (or disability level) as a latent variable directly from the item level scores and with the acknowledgment that each item bears varying importance for informing the disease severity. Several pharmacometric models applying IRT for disease areas using composite scales have been developed,[7–12] and a comprehensive tutorial presenting the framework of the IRT in the pharmacometric context is available.[13] Simulations using the UPDRS and Alzheimer's Disease Assessment Scale–Cognitive Subscale (a standard efficacy end point for clinical trials in Alzheimer's disease) have shown promising evidence that longitudinal modeling of a composite end point analyzed using IRT could enhance the study power.[7,12]

The efficacy of an investigative drug is typically assessed using the end-of-treatment observations in a placebo-controlled study. However, multiple assessments made during the treatment present an opportunity for longitudinal analysis to describe the time course of treatment response, which can be important for understanding whether a drug slows down the disease progression. Longitudinal modeling is also advocated by the pharmacometric community and supported by regulatory bodies for its potential to enhance trial efficiency.[14–16]

In a previous report of simulation-based longitudinal modeling of PD, where IRT analysis demonstrated the potential for higher study power than total score analysis, a hypothetical drug effect was assumed to reduce the linear disease progression with an immediate onset, and the placebo effect was not accounted for.[7] Because a drug effect typically takes time to be observed and a placebo effect is inevitable in clinical trials, it is useful to capture these aspects in a trial simulation. Furthermore, in the previous report, the total score was derived from item scores that were simulated from the item response model (IRM); as such, the comparison could potentially favor the IRT analysis.

Given this background, we explored a framework of longitudinal modeling for understanding the time course of the drug effect (symptom and slowdown of disease progression) and placebo effect on PD and compared the study power of two longitudinal modeling approaches—analyzing total score versus disease severity estimated by IRT—using historical clinical trial data.

METHODS

Studies analyzed

We used the UPDRS data from two published studies of ropinirole, a nonergot dopamine agonist. One was a crossover comparison of immediate-release and extended-release formulations in patients with early-stage PD,[17] and the other was a parallel comparison of ropinirole to placebo as an adjunct therapy to levodopa (L-dopa) in patients with advanced-stage PD at individually titrated doses between 6 and 24 mg/day, where symptoms were measured at 11 visits over 24 weeks starting from baseline.[18] The analysis included the baseline data from the first trial and all data from the second trial. The postbaseline data from the first trial were not included; the lack of placebo arm made those data unsuitable for comparing the time course between the drug and placebo treatments. Both trials were approved by the relevant institutional review boards. Details on study designs, inclusion/exclusion criteria, and assessment schedules were reported in the original publications.[17,18] For summaries of patient characteristics and the data included in the current analysis, see Table 1.

Pharmacometric analyses

Initially, we developed two longitudinal models to describe the placebo/drug treatment effects in the advanced-PD study on UPDRS total scores (total score model [TSM]) and on the disease severity estimated using the IRT (IRM). Based on these models, we performed simulations investigating the power for detecting the drug–placebo treatment difference. Model estimation and simulations were performed by means of nonlinear mixed effect modeling using NONMEM version 7.3.[19,20]

TABLE 1 Summary of patient characteristics and data used

Patient baseline characteristics							
	Age, years, mean (min-max)	Weight, kg, mean (min-max)	Sex, count	Prior concomitant drugs, count			
				Selegiline	Amantadine	Anticholinergics	Levodopa
Early PD, $N = 161$	61 (37–84)	77 (50–137)	Female 70 Male 91	No 114 Yes 47	No 132 Yes 29	No 143 Yes 18	No 161 Yes 0
Advanced PD, $N = 391$	66 (34–87)	74 (42–144)	Female 145 Male 246	No 302 Yes 89	No 300 Yes 91	No 381 Yes 10	No 0 Yes 391

UPDRS item score data[a]									
Early-PD study, $N = 161$		Advanced-PD study, $N = 391$							
Baseline		Baseline placebo		Baseline ropinirole		Longitudinal placebo		Longitudinal ropinirole	
N	n	N	n	N	n	N	n	N	n
161	7004	190	8330	201	8765	189	22,882	200	25,186

UPDRS total score data[a]									
Early-PD study, $N = 144$		Advanced-PD study, $N = 389$							
Baseline		Baseline placebo		Baseline ropinirole		Longitudinal placebo		Longitudinal ropinirole	
N	n	N	n	N	n	N	n	N	n
144	144	179	179	185	185	187	484	199	542

The analysis included baseline data from the early-PD study, and both baseline and longitudinal data from the advanced-PD study. The total score is the sum of the scores of 44 items.

Abbreviations: max, maximum; min, minimum; PD, Parkinson's disease; UPDRS, Unified Parkinson's Disease Rating Scale.

[a]N = number of patients; n = number of item scores.

Item response model

The methodology for applying the IRT to pharmacometric modeling has been described previously in detail.[13] The UPDRS is designed to measure the PD disease severity through the individual item scores, and the IRT is well suited for the analysis of such composite scale data. Essentially, the more severe a patient's condition is, the more likely the patient would have a higher score for an item in the instrument (e.g., 0, 1, 2, etc. for normal, mild, moderate, etc., respectively). In other words, the probability for a patient to have a certain score on a given item is a function of that patient's disease severity. The IRT thus uses the probability functions for all scores of all items collectively to estimate the disease severity for each patient at a given time. We applied a previously reported IRT model for UPDRS, which was developed using the baseline and placebo data from the two studies included here. The model included three latent severity variables reflecting certain aspects of the symptom severity and estimated through items that were (i) patient reported (items 1–17), (ii) clinician assessed and nonsided (items 18, 19, 20, 22, 27–31), and (iii) clinician assessed and sided (items 20, 21, 22, 23–26, evaluated for the right and left sides) for which a mixture component estimated the proportion of two subpopulations depending on their more disabled (dominant) side at baseline. Correlations between the latent variables were estimated within each study. Further information can be found in the original publication.[6]

Longitudinal models

The conceptual framework guiding this analysis is illustrated in Figure 1, which shows the longitudinal models of disease progression, the effect of a placebo treatment, the effect of drugs that cause symptom relief without affecting the progression, and the effect of drugs that slow down the progression.

It is well established that the symptom progression of PD, as measured by the UPDRS, is approximately linear over time over a relatively short period such as the duration of a clinical trial.[21–23] Many reports have shown that during a double-blind trial, the placebo treatment typically improves the symptoms to a small extent without altering the progression rate—effectively causing a slight downward shift of the underlying disease progression—and such effect takes a short time to stabilize.[18,24–26] The onset of this effect was modeled as an exponential function of time as supported by the data pattern; a greater onset rate meant an earlier achievement or stabilization of the placebo effect. Equation (1) describes *Symptom* in terms of either the UPDRS total score or each of the latent symptom severity variables estimated by the IRT model, for placebo-treated patient i at time t. In this equation, *Baseline*, *Progression rate*, *Placebo*, and *Onset rate* are the baseline, the progression rate over time, the placebo effect, and the onset rate of the placebo effect, respectively, for patient i.

$$\begin{aligned} \text{Symptom}_i(t) = {} & \text{Baseline} + (\text{Progression rate}) \cdot t \\ & + (\text{Placebo}) \cdot \left(1 - e^{-\text{Onset rate} \cdot t}\right) \end{aligned} \quad (1)$$

Conceivably, a drug may relieve the symptoms further, also without altering the progression rate. In a placebo-controlled trial, this drug effect would be reflected by a further shift of the time course, compared to the placebo effect. Equation (2) describes *Symptom*, also in terms of
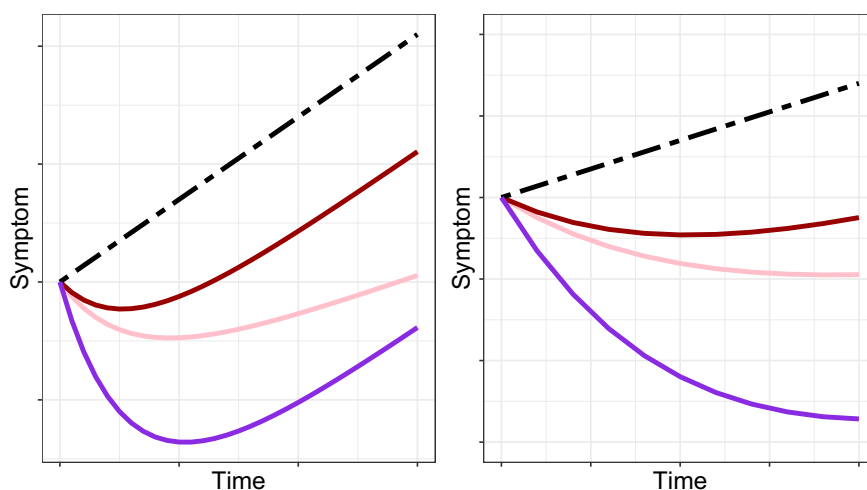

FIGURE 1 Longitudinal model with arbitrary scales to illustrate the importance of trial duration for correctly describing drug effects on the worsening of a symptom over time. The black dotted line indicates an underlying natural progression, which is not observable in a trial. The red line indicates observations for placebo treatment. The purple line indicates observations for symptom-relief agents. The pink line indicates observations for the progression slow-down agents. The left panel shows a trial of adequate duration. The right panel shows a shorter trial is inadequate for correctly describing these effects

either the UPDRS total score or each of the latent symptom severity variables estimated by the IRT model, for drug-treated patient i at time t. In this equation, *Baseline*, *Progression rate*, *Placebo*, *Drug,* and *Onset rate* are the baseline, the progression rate over time, the placebo effect, the drug effect, and the onset rate of the drug effect, respectively, for patient i.

$$\text{Symptom}_i(t) = \text{Baseline} + (\text{Progression rate}) \cdot t + (\text{Placebo} + \text{Drug}) \cdot \left(1 - e^{-\text{Onset rate} \cdot t}\right) \quad (2)$$

Alternatively, the drug may slow down the progression, so patients treated by the drug would show a different underlying progression rate while maintaining the placebo effect on the symptoms.[27] This is described in Equation (3), where *Drug* is the drug effect on the progression rate, and all other parameters are the same as in Equation 2.

$$\text{Symptom}_i(t) = \text{Baseline} + (\text{Progression rate} + \text{Drug}) \cdot t + (\text{Placebo}) \cdot \left(1 - e^{-\text{Onset rate} \cdot t}\right) \quad (3)$$

Therefore, the effect of the symptom-relief drugs is reflected as a downward parallel shift from the time course of placebo, whereas the effect of the progression-altering drugs is reflected as an eventual divergence from the placebo time course (Figure 1).

We fitted the longitudinal models described in Equations (1) to (3) using both the IRM (where Symptom_i represents each of the symptom severity variables estimated by IRT) and the TSM (where Symptom_i represents UPDRS total score) to explore whether ropinirole altered the symptom progression rate or simply showed greater symptom relief than placebo. Description on the individual level was ensured through a random effect associated with each of the symptom severity variables in the IRM and a random effect associated with some of the fixed effect parameters in the TSM. The IRM included multiple latent variables estimated from different UPDRS items, and we estimated the placebo and drug effects on each latent variable separately. The fit was performed on all data, the *Drug* parameter was assumed null for the placebo data and estimated for the drug data.

Evaluation of study power

For a comparison of power between the IRM and the TSM, we generated the power curves using both longitudinal models by employing a modified Monte Carlo mapped power (MCMP) approach.[20,28] Compared with alternative methods,[29,30] the MCMP approach offers greater computational efficiency by drawing samples from the results of model fitting to a single (large) dataset, instead of requiring model fitting to a large number of sample data sets.

In the MCMP, a stochastic simulation is preceded by the estimation of full and reduced models, followed by hypothesis testing using the likelihood ratio test (LRT), where the overall likelihood is substituted by the sum of individual contributions to the likelihood. In NONMEM, this corresponds to the sum of each individual's contribution to the objective function value (iOFV), the OFV being approximately proportional to minus twice the natural logarithm of the likelihood of the data.[19] In our case, the MCMP was modified such that the "observed" iOFVs were used (no simulation) based on the real data.

Thus, we fitted two models to the data from both ropinirole-treated and placebo-treated patients in the advanced-PD trial—a full model that included a drug effect for the ropinirole arm and a reduced model not including any drug effect but otherwise identical—to generate the iOFV for each patient. The difference in iOFV between the full and reduced models (ΔiOFVs) was calculated for each patient. We then drew 10,000 treatment-stratified virtual trial data sets of ΔiOFV values for a wide range of samples sizes. The LRT was then applied to the sum of the ΔiOFVs from all patients in that trial ($\Sigma\Delta$iOFVs), which is approximately χ^2 distributed. For each sample size, the proportion of the virtual trials that showed a difference in the $\Sigma\Delta$iOFVs corresponding to $p = 0.05$ or 0.1 for the χ^2 distribution would represent the study power at that p value. The actual critical $\Sigma\Delta$iOFVs values used in the MCMP were derived through the randomization test in Perl speaks NONMEM applied to the observed data of the placebo patients letting them be assigned randomly to active and placebo treatments.[20,31] The critical values for the IRM were 11.62 ($p = 0.05$) and 8.10 ($p = 0.1$), and the corresponding values for the TSM were 4.00 and 2.76.

RESULTS

Understanding the time course of placebo and drug effects

When the longitudinal IRM was fitted to the estimated disease severity, the disease *Progression rate* could not be estimated with adequate precision. This was somewhat expected because of the short treatment duration of 24 weeks and the small data set, considering the slow progression and high variability in the data (Figure 2). Effectively, without the *Progression rate*, the model describing the placebo effect and drug effect time courses took the form of Equation (4). In this model, the placebo produced a gradual effect of symptom relief, and the drug produced a greater relief.

$$\text{Symptom}_i(t) = \text{Baseline} + (\text{Placebo} + \text{Drug}) \cdot \left(1 - e^{-\text{Onset rate} \cdot t}\right) \quad (4)$$
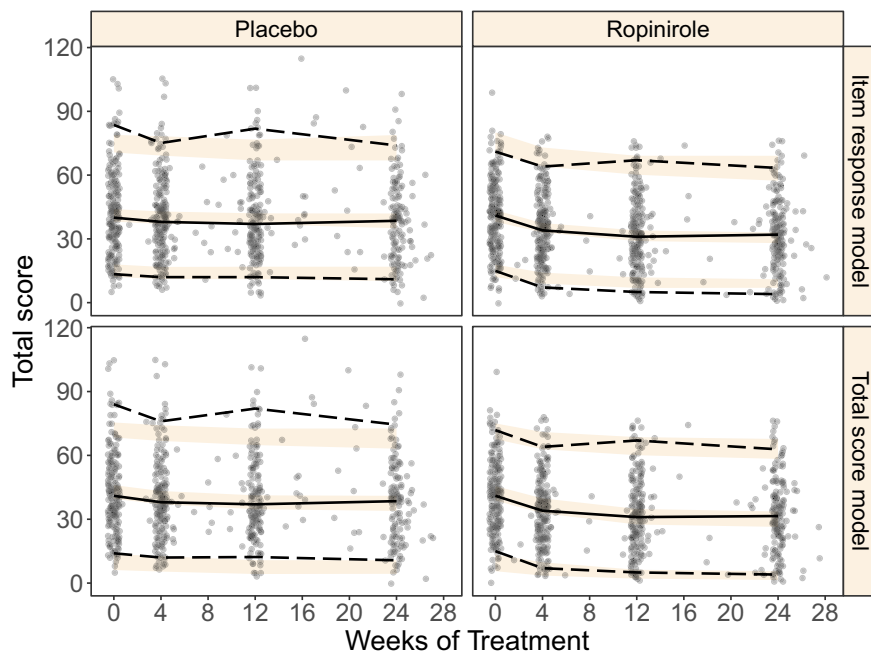
FIGURE 2 Visual predictive checks for total scores simulated from the item response model (upper) and total score model (lower). The lines describe the 5th, 50th, and 95th percentiles of the observations; the bands are the model-simulated 90% confidence intervals of the corresponding percentiles
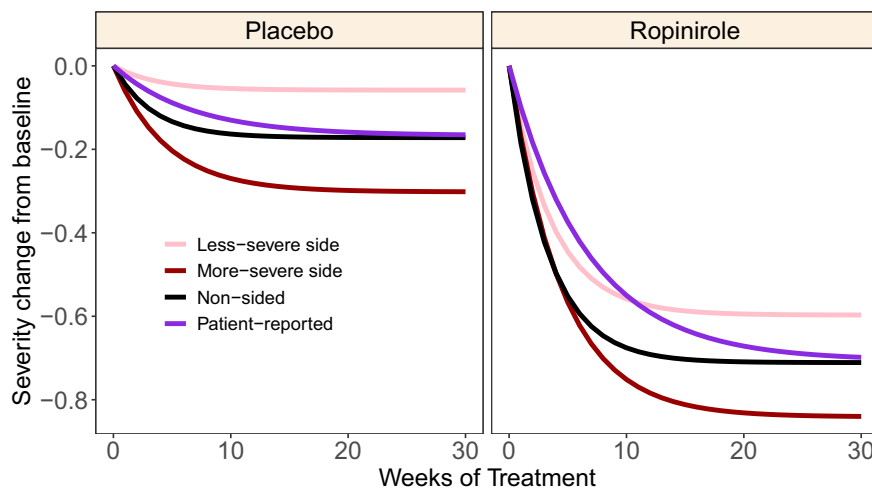

FIGURE 3 The model-estimated speed and magnitude of change from baseline for the latent variables. See the "Item Response Model" section in "Methods" for descriptions of the multiple latent variables

The data supported the estimation of separate parameters for the *Placebo* effect and the *Onset rate* on the four latent variables (Figure 3). The *Placebo* effects were −0.167, −0.172, −0.300, and −0.059 for the latent variables reflecting patient-reported items, clinician-assessed nonsided items, clinician-assessed items on the initially more severe side, and clinician-assessed items on the initially less severe side, respectively. The *Onset rate* of the *Placebo* and *Drug* effects for these latent variables varied slightly at $0.15-0.3$ week^{-1} (0.15, 0.30, 0.27, and 0.22 week^{-1} for patient-reported, nonsided, initially less-severe side, and initially more-severe side items,

respectively). The data were insufficient to estimate separate *Drug* effects on the different latent variables; a common symptomatic effect of −0.539 was estimated for all latent variables

Similarly, when the longitudinal TSM was fitted to the observed total score data, the *Placebo* effect was successfully estimated, but the disease progression rate could not be described ($p > 0.05$). The model parameters of *Placebo* and symptom-relief *Drug* effects were estimated to be −3.7 and −6.5 respectively, with an *Onset rate* of 0.13 week^{-1}.

Both longitudinal TSM and IRM exhibited adequate predictive performance when compared with the total UPDRS
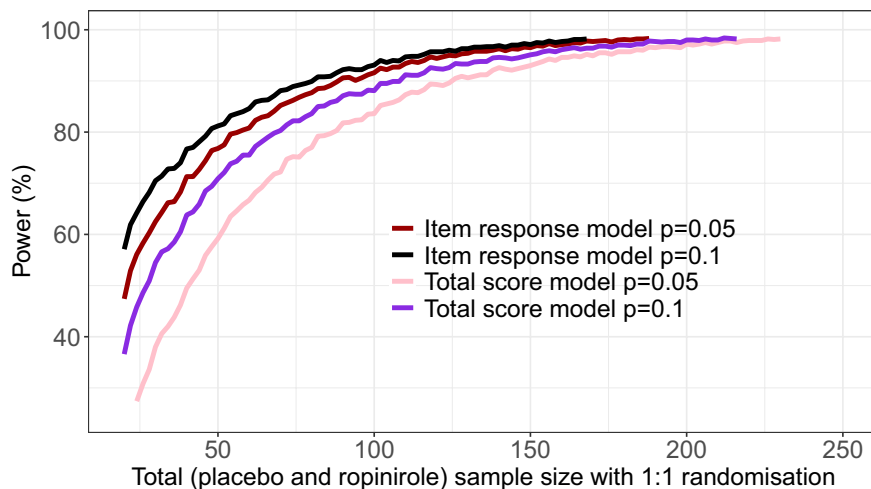
FIGURE 4 Power for detecting the drug effect (significance levels 0.1 and 0.05, one degree of freedom) based on analyses using a longitudinal item response model or total score model

observations (Figure 2). The TSM-estimated median (90% confidence interval) change in total UPDRS at week 24 for placebo and ropinirole groups were −3.7 (−5.4 to −2.0) and −9.3 (−11 to −7.3), respectively. Similarly, the IRM-simulated corresponding values were −2.0 (−4.0 to −1.0) and −9.0 (−11 to −8.0).

The estimated model parameters with their precision and corresponding NONMEM control streams are presented in the Supplementary Material.

Evaluation of study power

The sample size comparison for detecting a drug effect by the IRM versus the TSM analysis generated by the modified MCMP, in a parallel study of equal randomization, is shown in Figure 4. For the same study power, applying the IRM needed fewer patients than applying the TSM. At the significance level of 0.05, the sample sizes required for 80% power to detect a drug effect by the TSM and IRM were found to be 88 and 58 patients (34% smaller for the IRM), and the corresponding sample sizes for 90% power would be 126 and 90 patients (29% smaller for the IRM). At the less stringent significance level of 0.1, the sample sizes required for 80% power to detect a drug effect by the TSM and IRM would be 70 and 48 patients (31% smaller for the IRM).

DISCUSSION

There is a pressing need for drugs that can slow down functional deterioration in PD. Differentiating drugs that change the disease progression trajectory from drugs that relieve the symptoms without altering the progression trajectory

is a challenge for drug developers and approvers alike. For example, there is inconsistent evidence whether L-dopa, the standard-of-care pharmacotherapy for PD, slows down the disease progression.[24] Based on the paradigm shown in Figure 1, we conducted longitudinal modeling to test whether ropinirole, an approved dopaminergic agent, showed evidence of altering the disease progression rate. Both the longitudinal TSM and IRM adequately estimated the placebo effect and its onset rate. The models also successfully quantified the magnitude of the drug effect; however, it is important to point out that the parameters estimated from the short trial in patients with advanced PD may not be applicable to other patient populations or trial settings.

Neuroimaging data from patients with PD have suggested possible neuroprotection by ropinirole.[32] However, neither model could estimate the underlying progression rate and hence were unable to test whether the drug altered this progression rate. This was not unexpected: the treatment duration was 24 weeks—most likely too short to observe any notable progression, as noted by the original trial publication.[18]

These findings highlighted the importance of the treatment duration: a short trial does not provide enough information to describe the underlying progression rate; it could even provide misleading information on a placebo or drug effect. For example, observations in the early phase of the trial for a symptom-relief drug described in Figure 1 might erroneously show a divergence between the placebo and drug responses, suggesting a change in the progression rate. Similarly, observation times are also important for differentiating the nature of the drug effect: infrequent observations in the early phase of a trial could also lead to unreliable data for this differentiation.[33] Crucially, the claim of whether a drug changes the disease progression should not rely solely on the slope of the symptom progression; it must also be supported by biological evidence.[33]

The differential progression among different aspects of the symptoms in PD is well recognized.[21,22,34] It is conceivable that a drug or placebo treatment can relieve the symptoms or slow down the progression of these different aspects to varying extent. The IRM used in this investigation included multiple latent variables, representing patient-reported, nonsided, and sided items.[6,8] Data from the advanced-PD study suggested that the effect of placebo treatment varied for the different latent variables (Figure 3). The data could not differentiate the drug effect on the respective latent variables. These observations speak to the utility of item response analysis in understanding the differentiated symptom profile in an integrated fashion and assessing the placebo or drug effect on this profile; they also prompt the need for further investigations in larger and longer trials in more diverse patients.

Sample size requirement for detecting a drug effect depends on the size of the effect as well as the assessment schedule and duration of the trial. Given the nature of the data set that we used—the observed magnitude of symptom relief and the observation frequency and schedule—our analysis suggested reduced sample size with the IRM compared with the TSM. For example, the sample sizes saving for detecting the drug effect with 80% power at $p = 0.05$ would be 34%.

The potential sample size savings by IRM analysis of PD trials has been previously reported.[7] Buatois et al.[7] built a linear longitudinal IRM using 48-month MDS-UPDRS observational (instead of clinical trial) data. They then used the model to simulate longitudinal item–level data for a hypothetical drug that would reduce the slope of the progression rate, derived the total scores from the item scores (in absence of real data), and demonstrated notable potential savings in sample size by item response analysis compared with total score analysis for detecting the effect of a hypothetical drug on the progression slope. The exercise did not consider the placebo effect in a trial.

In comparison, the placebo time course captured by our model would be essential for simulation-based design of future trials: it informs the assessment time requirement for understanding the nature of the drug effect and for optimizing trial efficiency. Furthermore, deriving total scores from an IRM and subsequently comparing the efficiency between the TSM and the IRM could conceivably bias in favor of the IRM approach. Our work using real total score data during both placebo and drug treatments in a historical trial conceivably allowed a more reliable and fairer sample size comparison for the two approaches, complementing and confirming their findings.

In summary, this work generated strong evidence that item response analysis of the UPDRS could significantly reduce the sample size for drug effect detection compared with the standard total score approach. We also propose that longitudinal modeling could be used to explore whether a drug slows down symptom progression. Trials for drugs aimed at slowing down neurodegeneration are large and lengthy because of slow disease progression, a large placebo effect, and high end-point variability. Better understanding of these emerging methodologies through broad application will be highly valuable in the pursuit of disease-modifying treatments, especially benefiting trials for small patient populations.

CONFLICT OF INTEREST

AUTHOR CONTRIBUTIONS

C.C. and S.J. wrote the manuscript. C.C., M.O.K., and E.L.P. designed the research. C.C., S.J., S.Y., E.L.P., and M.O.K. performed the research. S.J. analyzed the data.

ORCID

Chao Chen https://orcid.org/0000-0002-9791-4727
Siv Jönsson https://orcid.org/0000-0001-8240-0865
Shuying Yang https://orcid.org/0000-0001-9841-2342
Elodie L. Plan https://orcid.org/0000-0002-2255-3904
Mats O. Karlsson https://orcid.org/0000-0003-1258-8297

REFERENCES

1. Nussbaum RL, Ellis CE. Alzheimer's disease and Parkinson's disease. *N Engl J Med*. 2003;348:1356-1364.
2. Kalia LV, Lang AE. Parkinson's disease. *Lancet*. 2015;386:896-912.
3. Elkouzi A, Vedam-Mai V, Eisinger RS, Okun MS. Emerging therapies in Parkinson disease - repurposed drugs and new approaches. *Nat Rev Neurol*. 2019;15:204-223.
4. 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 clinimetric testing results. *Mov Disord*. 2008;23:2129-2170.
5. Ramaker C, Marinus J, Stiggelbout AM, Van Hilten BJ. Systematic evaluation of rating scales for impairment and disability in Parkinson's disease. *Mov Disord*. 2002;17:867-876.
6. 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.
7. 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.
8. Gottipati G, Karlsson MO, Plan EL. Modeling a composite score in Parkinson's disease using item response theory. *AAPS J*. 2017;19:837-845.
9. Krekels E, 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.
10. 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.

11. Schindler E, Friberg LE, Lum BL et al. A pharmacometric analysis of patient-reported outcomes in breast cancer patients through item response theory. *Pharm Res*. 2018;35:122.

12. 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.

13. Ueckert S. Modeling composite assessment data using item response theory. *CPT Pharmacometrics Syst Pharmacol*. 2018;7:205-218.

14. Committee for Medicinal Products for Human Use (CHMP) EMA. Guideline on clinical investigation of medicinal products in the treatment of Parkinson's disease (EMA/CHMP/330418/2012 rev. 2); 2012.

15. Barrett JS, Fossler MJ, Cadieu KD, Gastonguay MR. Pharmacometrics: a multidisciplinary field to facilitate critical thinking in drug development and translational research settings. *J Clin Pharmacol*. 2008;48:632-649.

16. Karlsson KE, Vong C, Bergstrand M, Jonsson EN, Karlsson MO. Comparisons of analysis methods for proof-of-concept trials. *CPT Pharmacometrics Syst Pharmacol*. 2013;2:e23.

17. Stocchi F, Hersh BP, Scott BL, Nausieda PA, Giorgi L, Ease PDMSI. Ropinirole 24-hour prolonged release and ropinirole immediate release in early Parkinson's disease: a randomized, double-blind, non-inferiority crossover study. *Curr Med Res Opin*. 2008;24:2883-2895.

18. Pahwa R, Stacy MA, Factor SA et al. Ropinirole 24-hour prolonged release: randomized, controlled study in advanced Parkinson disease. *Neurology*. 2007;68:1108-1115.

19. NONMEM 7.3.0 Users Guides (1989-2013). [computer program]. Hanover, MD, USA: ICON Development Solutions; 2009.

20. Keizer RJ, Karlsson MO, Hooker A. Modeling and simulation Workbench for NONMEM: tutorial on Pirana, PsN, and Xpose. *CPT Pharmacometrics Syst Pharmacol*. 2013;2:e50.

21. 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.

22. Saunders-Pullman R, Mirelman A, Alcalay RN et al. Progression in the LRRK2-asssociated Parkinson disease population. *JAMA Neurol*. 2018;75:312-319.

23. Simuni T, Siderowf A, Lasch S et al. Longitudinal change of clinical and biological measures in early Parkinson's disease: Parkinson's progression markers initiative cohort. *Mov Disord*. 2018;33:771-782.

24. Fahn S, Oakes D, Shoulson I et al. Levodopa and the progression of Parkinson's disease. *N Engl J Med*. 2004;351:2498-2508.

25. Olanow CW, Rascol O, Hauser R et al. A double-blind, delayed-start trial of rasagiline in Parkinson's disease. *N Engl J Med*. 2009;361:1268-1278.

26. Verschuur CVM, Suwijn SR, Boel JA et al. Randomized delayed-start trial of levodopa in Parkinson's disease. *N Engl J Med*. 2019;380:315-324.

27. Ploeger BA, Holford NH. Washout and delayed start designs for identifying disease modifying effects in slowly progressive diseases using disease progression analysis. *Pharm Stat*. 2009;8:225-238.

28. Vong C, Bergstrand M, Nyberg J, Karlsson MO. Rapid sample size calculations for a defined likelihood ratio test-based power in mixed-effects models. *AAPS J*. 2012;14:176-186.

29. Kowalski KG, Hutmacher MM. Design evaluation for a population pharmacokinetic study using clinical trial simulations: a case study. *Stat Med*. 2001;20:75-91.

30. Lee PI. Design and power of a population pharmacokinetic study. *Pharm Res*. 2001;18:75-82.

31. Wählby U, Jonsson EN, Karlsson MO. Assessment of actual significance levels for covariate effects in NONMEM. *J Pharmacokinet Pharmacodyn*. 2001;28(3):231-252.

32. Whone AL, Watts RL, Stoessl AJ et al. Slower progression of Parkinson's disease with ropinirole versus levodopa: The REAL-PET study. *Ann Neurol*. 2003;54(1):93-101.

33. Chen C. Opportunities and pitfalls in clinical proof-of-concept: principles and examples. *Drug Discov Today*. 2018;23:776-787.

34. Prasad S, Saini J, Yadav R, Pal PK. Motor asymmetry and neuromelanin imaging: Concordance in Parkinson's disease. *Parkinsonism Relat Disord*. 2018;53:28-32.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.