

# Item Response Theory as an Efficient Tool to Describe a Heterogeneous Clinical Rating Scale in De Novo Idiopathic Parkinson's Disease Patients

Simon Buatois, Sylvie Retout, Nicolas Frey, Sebastian Ueckert

▶ **To cite this version:**

Simon Buatois, Sylvie Retout, Nicolas Frey, Sebastian Ueckert. Item Response Theory as an Efficient Tool to Describe a Heterogeneous Clinical Rating Scale in De Novo Idiopathic Parkinson's Disease Patients. Pharmaceutical Research, 2017, 34 (10), pp.2109-2118. 10.1007/s11095-017-2216-1 . inserm-01563224

**HAL Id: inserm-01563224**

**https://inserm.hal.science/inserm-01563224**

Submitted on 17 Jul 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Item response theory as an efficient tool to describe a heterogeneous clinical rating scale in de novo idiopathic Parkinson's disease patients

Simon Buatois[1,2,3*], Sylvie Retout[1], Nicolas Frey[1], Sebastian Ueckert[4]

(1) Roche Pharma Research and Early Development, Clinical Pharmacology, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd, Grenzacherstrasse 124, 4070 Basel, Switzerland

(2) IAME, UMR 1137, INSERM, F-75018 Paris, France

(3) INSTITUT ROCHE, Roche S.A.S., 30, cours de l'île Seguin, 92650 Boulogne-Billancourt, France

(4) Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden

*Corresponding author: 30, cours de l'île Seguin, 92650 Boulogne-Billancourt, France

Email: simon.buatois@roche.com

Telephone number : +33678489218

## ABSTRACT

**Purpose:** This manuscript aims to precisely describe the natural disease progression of Parkinson's disease (PD) patients and evaluate approaches to increase the drug effect detection power.

**Methods:** An item response theory (IRT) longitudinal model was built to describe the natural disease progression of 423 *de novo* PD patients followed during 48 months while taking into account the heterogeneous nature of the MDS-UPDRS scale. Clinical trial simulations were then used to compare drug effect detection power from IRT and sum of item scores based analysis under different analysis endpoints and drug effects.

**Results:** The IRT longitudinal model accurately describes the evolution of patients with and without PD medications while estimating different progression rates for the subscales. When comparing analysis methods, the IRT-based one consistently provided the highest power.

**Conclusion:** IRT is a powerful tool which enables to capture the heterogeneous nature of the MDS-UPDRS.

# KEY WORDS

MDS-UPDRS, Item response theory, Parkinson's disease, Pharmacometrics, Drug effect


# ABBREVIATIONS

| | |
|---|---|
| CV | Coefficient of Variation |
| EOT | End Of Trial |
| IIV | Inter-Individual Variability |
| IRT | Item Response Theory |
| LRT | Likelihood Ratio Test |
| MDS-UPDRS | Movement Disorder Society UPDRS |
| MJFF | Michael J. Fox Foundation |
| NONMEM | NON-linear Mixed Effect Model |
| PD | Parkinson's Disease |
| PPMI | Parkinson's Progression Markers Initiative |
| RSE | Relative Standard Error |
| SD | Standard Deviation |
| SIS | Sum of Item Scores |
| SWEDD | Scans Without Evidence of Dopaminergic Deficit |
| UPDRS | Unified Parkinson's Disease Rating Scale |
| VPC | Visual Predictive Check |

# INTRODUCTION

Parkinson's disease (PD) is one of the most common neurodegenerative brain disorders (1). It affects 1% of individuals older than 60 years (2) and 7 to 10 million people worldwide. PD is a complex and heterogeneous disease (3) characterized by a progressive degeneration of dopaminergic neurons of the nigrostriatal pathway leading to motor symptoms. Degenerative lesions are also found with varying degrees of severity, depending on the patient, in other subcortical nuclei resulting in both motor and non-motor symptoms (4).

As no specific biomarker exists (5), the diagnosis of PD is mainly based on the clinical information provided by the patient, that is, the motor signs and symptoms (e.g. bradykinesia, rigidity, and tremor). However, according to recent publications, non-motor symptoms appear nearly 20 years (6) before its diagnosis (e.g. constipation, depression and hyposomnia). The Unified Parkinson's Disease Rating Scale (UPDRS) is a clinical rating scale used as a surrogate measure of the neuronal impairment in PD (7). It is composed of 55 items categorized as motor and non-motor symptoms. Based on nonlinear mixed effect models, longitudinal models have been developed to estimate the average population PD status, its interindividual variability and the sources of this variability (8). Such features are particularly meaningful in PD, where the clinical expression varies largely among patients. Using longitudinal models, Holford *et al.* were able to determine the evolution of the total score over time as well as the impact of various PD medications (8).

There is currently no cure for PD. The existing therapies (carbidopa-levodopa, dopamine agonists, monoamine oxidase B inhibitors, catechol-O-methyltransferase inhibitors, anticholinergics and amantadine) have symptomatic effects and, for some of them, disease modifying effects (8,9). With a high number of failures in PD clinical trials and among the several challenges in the field of the neurological diseases, questions have been raised concerning the sensitivity and the accuracy of the clinical endpoint (9,10).

The UPDRS scale was recognized as having a lack of sensitivity for some clinically relevant PD symptoms (11). Dichotomous yes/no questions were, for example, considered as insufficient to measure the degree of severity of the symptoms (e.g. sleep problems and gastrointestinal plaints). Moreover, some signs and symptoms were not captured in the original scale (e.g. urinary problems and dopamine dysregulation syndrome). A revised version, the Movement Disorder Society-UPDRS (MDS-UPDRS), has been proposed in 2007 (12). The MDS-UPDRS scale consists of 65 items measuring the disturbance of non-motor experiences of daily living (items 1.1 to 1.13, e.g. constipation, sleep disorders, depression) as well as motor experiences (2.1 to 2.13 and 3.1 to 3.18 items, e.g. tremor, rigidity, bradykinesia.) and motor complications due to PD medications (6 items,

e.g. dyskinesia, On/Off states). All items are now anchored with five categories, from 0 (lack of impairment) to 4 (severe impairment) and leading to a global MDS-UPDRS score range from 0 to 260.

In addition, the disease modeling performed so far focused mostly on the sum of item scores (SIS). It assumes, therefore, an equal rate of disease progression for the all items of the scale, and thus, a similar degenerative mechanism for the motor, non-motor and tremor symptoms. However, recent analyses of the symptoms have led some to question this hypothesis (7,10,13,15–18) .

Several factor analyses demonstrated that non-motor, motor and tremor symptoms are relatively independent (12,15). When analyzed independently, tremor items have been shown to progress more slowly than all other motor features (9,13,14). Non-motor symptoms occur a decade before the other symptoms (16). Finally, unlike motor symptoms, PD medications have a limited impact on non-motor symptoms. These arguments underline the notion of a composite scale which cannot be summarized by a single index (i.e. the total MDS-UPDRS score) (12). As a result, pharmaceutical companies are increasingly moving toward an analysis of a subset of the scale (17). This subset is often selected per subscale (e.g. motor experiences), and not per sensitivity of each item to a potential disease modifying drug effect, leading to a potential loss of power.

Item response theory (IRT), also known as latent trait theory, assumes that each item, as a function of its difficulty and power of discrimination, explains an underlying latent variable (i.e. the neuronal disability) and therefore, captures more precisely the heterogeneity of the scale (18). Recent publications in the field of central nervous system diseases (19,20) have demonstrated the use of IRT to leverage the information provided by clinical rating scales.

Combined to a longitudinal model, IRT can be used to characterize the natural disease progression of PD patients while respecting the composite nature of the MDS-UPDRS scale. It was therefore the first focus of this work to establish an IRT-based longitudinal model to capture the natural history of the MDS-UPDRS score. To do so the potential effect of medication had to be taken into account to avoid an underestimation of the progression, as highlighted by Gottipato et al.

In the second part of the manuscript, the resulting longitudinal model and clinical trial simulations are used to answer the question whether a SIS-based analysis for a most sensitive subset of the assessment or an IRT-based analysis can increase the power to detect a drug effect?

# MATERIALS & METHODS

## PPMI database

Data used in this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data). For up-to-date information on the study visit www.ppmi-info.org. PPMI is an observational study with the primary objective to identify biomarkers of disease progression. It is sponsored and partially funded by The Michael J. Fox Foundation for Parkinson's Research (MJFF). Subjects were enrolled from over 30 sites across the USA, Europe, and Australia. The data used in this article were downloaded on febuary the 13$^{th}$, 2015, while the PPMI study was ongoing. It consisted of 423 *de novo* PD patients, 196 control subjects, and 64 Subjects with scans without Evidence of Dopaminergic Deficit (SWEDD). The average follow-up time was 27.3 months. The minimum follow-up time was 0 month and observations after 48 months were discarded from the analysis. Table I represents a summary of some demographic characteristics of the PPMI database.

At baseline, all patients were treatment-naïve and were not expected to require PD medications within at least 6 months. Nonetheless, according to the protocol, PD medications, without any restriction on the number or the type, might be initiated at any time based on the discretion of the patient or treating physician. The disease progression was assessed using the MDS-UPDRS scale at visits, every 3 to 6 months (12 months for the control group). For patients on PD medication, the assessments were performed before the administration of the next dose.

## Disease progression model

### *Modeling hypothesis and MDS-UPDRS dataset simplification*

First, due to the composite nature of the scale, we subdivided it into three different subscales. Their corresponding scores will be referred to as subscores. The MDS-UPDRS total score is based on the sum of item scores of parts I, II and III. Consequently, subscores were, measured as the sum of

items score related to each of the motor, non-motor and tremor subscales. According to the MDS, the first 13 items (part I) of the scale were considered as surrogate measures of the non-motor disability. Based on PPMI recommendations, Items 2.10 ("Tremor") and from 3.15a to 3.18 ("Postural tremor", "Kinetic Tremor", "Rest Tremor", "Constancy of rest Tremor") were considered as surrogate measures of the tremor disability. All remaining items of part II and III were considered as surrogate measures of the motor disability (12). Finally motor complication items due to PD medications (part IV) were no included in this analysis as also recommended by the PPMI recommendation. Items of the different subscales are reported in online supplementary material A.

Second, in the assessment of MDS-UPDRS scale, some tasks are repeated for all or part of the limbs (e.g. resting tremor of the right limb, left limb). Knowing that the patient's dominant side (21) was reported in the study, right and left repetitions were reassigned to the dominant or non-dominant side of the subject (e.g. resting tremor of the dominant lower limb, non-dominant lower limb).

### *IRT model*

An ordered categorical model was used to describe the data (19,22). The probability for a patient *i* to obtain a score greater or equal to *k* is expressed as a function of the latent variable ($D_{i,v}$), where $v$ represents the motor, non-motor or tremor subscales. The link between the probability for different response k, with k=0 to 4, and the latent variable is modulated using item-specific parameters, as follows:

$$P(Y_{ij} \geq k) = \frac{e^{a_j(D_{i,v} - b_{j,k})}}{1+e^{a_j(D_{i,v} - b_{j,k})}} \qquad [1]$$

$$P(Y_{ij} = k) = P(Y_{ij} \geq k) - P(Y_{ij} \geq k+1) \qquad [2]$$

Where $a_j$ and $b_{j,k}$ denote, respectively, the slope and the difficulty parameter of the item *j*. Difficulty parameters of the same item were constrained to be non-decreasing for increasing category severity ($b_{j,k+1} \geq b_{j,k}$). While $a_j$ and $b_{j,k}$ were treated as fixed effect parameter, $D^0_{i,v}$, the latent variable at baseline, was modeled as a random effect correlated between the different latent variables and assumed to follow a standard normal distribution in the population (19).

Due to the targeted population of the study, there was a lack of observations in the highest categories of most items. Therefore, difficulty parameter associated to $P(Y_{ij} \geq k)$ (as expressed in Eq 1), were estimated as long as more than 2.5% of the subjects answer the response $k$. Consequently, if less than 2.5% of the subjects were classified in the highest observed category, this category was merged with the previous one. Moreover, if less than 2.5% of the subjects received a rating greater than 0 for one given item, that item was ignored during the analysis (in this study: "hallucinations and psychosis", "Features of dopamine dysregulation syndrome" and "Freezing of gait").

Data used to estimate the item specific parameters came from untreated subjects from 0 to 24 months. To take into account the disease progression of the patients over time, an unstructured latent variable model was used to estimate a separate mean for each visit (0, 3, 6, 9, 12, 18 and 24 months).

### *Longitudinal model*

In a second step, to allow for inter and extrapolation, the unstructured latent variable model was replaced by a longitudinal model. Item-specific parameters were fixed to the values as obtained in the analysis described in the section "IRT model" and the time course of the motor, non-motor and tremor disability was modeled through a baseline value ($D_v^0$ ) and a linear disease progression with a constant rate of change over time ($\alpha_v$). Both parameters were modeled as subscale specific:

$$D_{i,v}(t) = D_{i,v}^0 + \alpha_{i,v} \cdot t \qquad [3]$$

As the medication type and dose received by the patient was not recorded, a combined treatment effect was estimated using only the patient initiation treatment time. Three symptomatic effects ($E$) with different onset dynamics were evaluated on each of the tremor, non-motor and motor disability. The models (equation [5], [6] and [7]) are all variants of an offset (or symptomatic) drug effect on disease progression. Equations [6] and [7] assume an increase of $E^0$ over time to mimic the clinical impact of additional PD medications, with, respectively, a linear increase β or an exponential increase reaching an asymptote at a rate $k_{eq}$ in function of time since the administration of the first PD medication $t_{DM}$.

$$D_{i,v}(t) = D_{i,v}^0 + \alpha_{i,v} \cdot t + E_{i,v}(t) \qquad [4]$$

with,

$$E_{i,v}(t) = E_{i,v}^0 \qquad [5]$$

$$E_{i,v}(t) = E_{i,v}^0 + \beta_{i,v} \cdot t_{DM} \qquad [6]$$

$$E_{i,v}(t) = E_{i,v}^0 + \beta_{i,v} \cdot (1 - e^{-k_{eq} \cdot t_{DM}}) \qquad [7]$$

An additive inter-individual variability (IIV) was estimated for $D_{i,v}^0$, $\alpha_{i,v}$, $E_{i,v}^0$ and $\beta_{i,v}$

It should be noted that the parameters characterize the time course of the latent variables ($D_{i,v}(t)$) and not the scale itself. Therefore, simulations were used to convert the estimates back to the MDS-UPDRS scale. To do so, for each item observation of the PPMI database, a random number drawn from a uniform distribution between 0 and 1 was compared to the estimated probabilities $P(Y_{ij} \geq k)$ to assign a corresponding category of severity; a simulated total score was then derived as the sum of those item categories. Then, this simulation was repeated 100 times to predict score proportions per patient and per time.

### *Estimation & validation method*

The parameters were estimated using the conditional Laplacian estimation method in NONMEM version 7.3 (23).

The capacity of the IRT model to describe the data was assessed using graphical diagnostics based on 100 simulated PPMI trial replications. The observed proportions of response in each of the 4 categories were compared to the simulated ones at the item level (scatter plot) and the mean observed proportions of response in each categories, over the items, were compared to the simulated ones at the subscale level (mirror plots). For the mirror plots, patients were subdivided into two disability subgroups. Patients with a disability lower or equal to the median value were assigned to the mild subgroup and the remaining ones to the severe subgroup.

The appropriateness of the IRT-longitudinal model to describe the data was assessed using visual predictive checks (VPCs) of the motor, non-motor, and tremor subscores. 100 PPMI trial replications were simulated from the model, and the simulated data were compared to the observations.

## Impact of analysis method and analysis endpoint

In an MDS-UPDRS-based POC trial, both the analysis method as well as the analysis endpoint constitute two separate aspects that can be optimized to increase the power of the trial. Regarding the endpoint, the analysis could be based on the full MDS-UPDRS assessment or on a selection of items that is expected to be most sensitive given the assumed drug effect and the planned trial duration. For the analysis method, one could either use a SIS or and IRT-based analysis. The following sections describe how these four different approaches were compared using clinical trial simulations.

### *Simulation scenarios*

The study design assumed a short 6 month trial with 100,200,400 and 600 patients that were equally randomized to either the placebo or treatment group. The disease model developed in the disease progression section was used as the simulation model. Regarding the drug effect, instead of a combined treatment effect resulting from the association of multiple PD medications, a hypothetical disease modifying effect ($\gamma_v$) was introduced as follows:

$$D_{i,v}(t) = D_{i,v}^0 + \alpha_{i,v}(1 - \gamma_v \cdot x_{DE}) \cdot t$$

Where, $x_{DE}$ is a treatment indicator covariate (0 in the placebo group and 1 in the treatment group). Two scenarios were investigated: 1) a homogeneous drug effect profile with a 50% reduction of the subject-specific disease progression ($\gamma_1 = \gamma_2 = \gamma_3 = 0.5$) and 2) a heterogeneous drug effect profile with a 50%, 20% and 30% reduction of the subject-specific dis**ea**se progression for the motor, non-motor and tremor subscales ($\gamma_1 = 0.5,\ \gamma_2 = 0.2,\ \gamma_3 = 0.3$).

### *Analysis methods*

For a fair comparison, both SIS and IRT-based analysis methods utilized only data from the baseline and the end of trial (EOT) visit.

#### SIS-based analysis:

For the SIS-based analysis each simulated trial was analysed with a linear mixed effect model using:

$$\tilde{y}_i = y_i^0 + \alpha \cdot (1 - \gamma \cdot x_{DE}) \cdot x_{visit}$$

Where $\tilde{y}_i$ is the assessment summary score of patient $i$, $y_i^0$, $\alpha$ and $\gamma$ are the parameters for baseline, progression and treatment, and $x_{visit}$, $x_{trt}$ are dummy coded covariates ($x_{visit}$= 0 at baseline and $x_{visit}$= 1 at EOT, $x_{DE}$= 0 in the placebo and $x_{DE}$= 1 in the treatment group).

The significance of the drug effect was tested using a log-likelihood ratio test (LRT) with 1 degree of freedom for $\gamma$ ($H0 : \gamma = 0$).

IRT-based analysis:

The IRT-based analysis utilized the item-level data and the IRT model described in equations 1 and 2. The latent variable $D_{i,v}$ of patient *i* for the three subscales were described with a linear mixed effect model using:

$$\widetilde{D}_{i,v} = D_{i,v}^0 + \alpha_v \cdot (1 - \gamma \cdot x_{DE}) \cdot x_{visit}$$

Where $D_{i,v}^0$ and $\alpha_v$ are the subscale specific parameters for baseline and progression.

The significance of the drug effect was tested using a LRT with 1 degree of freedom for $\gamma$ ($H0 : \gamma = 0$).

## *Analysis endpoints*

### Total number of items

All items included in the disease progression section were used for this endpoint.

### Optimized selection of items

A selection of items specific to each scenario was used for this endpoint. Selection is based on an optimization algorithm which assumes *a priori* knowledge of the drug effect profile. Under scenario 2, an additional optimized selection of items was performed assuming an incorrect prior knowledge of the drug effect profile. To that end, the expected drug effect profile was arbitrarily set to a 20%, 30% and 60% reduction of the individual disease progression for the motor, non-motor and tremor subscales, respectively.

The optimization algorithm worked in an iterative manner, starting from an empty set of items and choosing a new item at each iteration by testing all the remaining ones. The algorithm aimed at selecting, at each iteration, the item that maximizes the probability of correctly identifying the treatment effect to be larger than the placebo effect. To calculate that probability, score change from

baseline at EOT in both groups was approximated as a normally distributed random variable with mean ($\mu_c$) and variance ($\omega_c$).

The per group, $\mu_c$ and $\omega_c$ were calculated, through 100000 Monte Carlo simulations, as the sum of the item level means and variances (while taking into account the correlation between items), such as:

$$\mu_c = \sum_{1}^{N} z\bar{y}_n$$

$$\omega_c = z^T \Omega z$$

Where $z$ denotes an indicator column vector of $1, \dots, n$ items, $\bar{y}_n$ $the$ item level means and $\Omega$ the variance covariance matrix. For each combination $c$:

$$z_n = 1 \; if \; n \in c$$
$$z_n = 0 \; if \; n \notin c$$

Finally the quantile function was used to calculate the probability of the score difference, between the placebo and treatment arms, to be larger than 0. The algorithm stopped when none of the remaining items led to an increase in the optimized probability. The R code of the optimization algorithm can be found in the supplementary material B.

### *Technical implementation*

Under each scenario and each sample size, power was calculated as the number of times where a statistically significant difference was found according to the analytical method and the analytical endpoint out of 1000 trial replicates.

# RESULTS

## Disease progression model

In total, the IRT model consists of 184 item-specific parameters which were estimated from the observations between baseline and 24 months of subjects without PD medication. According to the modeling hypothesis, these parameters describe 56 MDS-UPDRS items which consist of 2 to 5 categories.

The longitudinal model was estimated on *de novo PD* patients with and without PD medications from baseline to 48 months. Latent variables time course were characterized by a baseline and a linear disease progression rate. Parameters of the longitudinal model and their relative standard errors (RSE) are reported in Table II. Parameters are expressed in the latent variable scale. To ease the interpretation of those values; simulations were used to predict the median baseline value and the median rate of disease progression per subscale. The estimated median baseline values equal 0.62, 0.45 and 0.45 unit/item for the motor, non-motor and termor subscales, respectively. The median natural disease progression rate correspond to an increase of 11.25 unit/year for the total score and 0.26, 0.11 and 0.11 unit/item/year for the motor, non-motor and tremor subscales, respectively. The estimated rate of disease progression in the motor subscale was higher compared to the others. Finally, positive correlations were estimated between the different latent variables for the individual baselines and disease progression rates.

Adding drug effect significantly improved the fit according to the LRT. Drug effects were best described using a constant offset [5] with a linear increase [6] and an exponential increase [7] for the non-motor, tremor and motor disability, respectively.

## Model Performance

Figure 1 and 2 underline the model's ability to reproduce the score of a specific population, in this case *de novo* PD patients at baseline, at both the item and subscale levels. The observed mean proportion of responses in each category of the 3 subscales were adequately reproduced by the model (Figure 1). Futhermore, at the item level, the symbols were homogeneously distributed around the identity line, highlighting the adequacy between the observed and simulated proportion of patients per category of each item (Figure 2).

The predictive performances of the longitudinal model for each subscore are reported in Figure 3. The 2.5, median and 97.5 percentiles of the observations were mostly included in the predicted confidence interval of the 2.5, median and 97.5 percentiles, representing a satisfactory ability of the model to reproduce the data. After 12 months, the model slightly over predicts the 97.5 percentile and under predicts the 2.5 percentile of the motor subscale. Additional plots highlighting the adequacy of

the model to describe the data at both the item and total score levels can be found in the online supplementary materials C and D.

## Impact of analysis method and analysis endpoint

The result of the subset optimization algorithm are visualized in Figure 4. Assuming *a priori* knowledge of the drug effect profiles, Figure 4 represents the probability of correctly identifying the treatment effect to be larger than the placebo effect after each iteration of the optimization algorithm. Under the hypothesis of a homogeneous drug effect (optimisation scenario 1) , the optimized set consisted of 39 items including 2 non-motor items, 6 tremor items and 31 motor items. Based on the heterogeneous drug effect (optimization scenario 2), there were 33 items in the optimized set. Two of them were tremor items and the remaining 31 were motor items. For the scenario 1 and 2, once the optimized set of items was reached, the probability was gradually decreasing after each iteration of the optimization algorithm. Finally, under optimization scenario 3, the optimized set consist of 1 non-motor items, 9 tremor items and 25 motor items. A list of the items selected under each of the optimization scenarios is reported in online supplementary material B.

Figure 5 compares the power to detect a drug effect for different analysis methods and analysis endpoints. It illustrates that using a SIS-based analysis (the reference approach), power can be significantly increased by focusing on a subset of most informative rather than all items. However, from Figure 6, where incorrect prior knowledge of the drug effect profile was used for the item selection, it is clear that this increase in power can only be achieved if the prior assumptions are correct.

According to figure 5, regardless of the simulation scenario and analysis endpoint, an IRT-based analysis consistently had the highest power.. An analysis of an optimized set of items doesn’t increase the power of the IRT based analysis. IRT allowed leveraging of the information provided by the scale without any *a priori* knowledge of the drug effect and led to a decrease of 25% and 23% in the required number of patient to achieve 80% power (for scenarios 1 and 2, respectively), compared to SIS.

# DISCUSSION

The first focus of this work was to establish an IRT-based longitudinal model to capture the natural history of the MDS-UPDRS score. An IRT-based modeling on three latent variables for the motor, non-motor and tremor disability was used to capture the heterogeneity of a clinical rating scale. The choice of three latent variables corroborates the literature (9,13,14). In Parkinson disease clinical trials, the initiation time of PD medications is closely related to the rate of disease progression. Ignoring the observations of patients under treatment may lead to an underestimation of the disease progression. Moreover, knowing that most of the patients will require PD medications within 6 to 9 months, it would dramatically reduce the information related to the disease progression. By combining IRT and longitudinal models we were able to describe the evolution of patients with and without PD medications over time while estimating different progression rates for the subscales. The estimated progression of the total MDS-UPDRS was in line with the value reported in the literature on the UPDRS scale (8). Furthermore, in agreement with the results of Vu *et al* (9) , a statistically significant disease progression was found for each of the subscales with a higher rate for the motor subscale. Finally, the estimated positive correlations between subscales for the individual baselines and disease progression rates underline the concept of a single pathology as the origin of these processes.

The second part of this work used the final longitudinal IRT model to investigate the influence of the analysis method on power to detect a drug effect under different conditions. Our findings highlight that defining an *a priori* optimized set of items is one way to increase the power of a SIS-based analysis. This paper presents an optimization algorithm that allows to select the most sensitive items in an assessment if a SIS-base analysis is planned. However, it is important to point out that these approaches are dependent on *a priori* knowledge of the drug effect profile and it was shown that no increase in power, or even a decrease, is to be expected if the assumed profile is wrong. An IRT-based analysis, in constrast to that, consistently delivered the highest power without any assumption about the drug effect profile. It should be highlighted that both SIS and IRT-based analysis were performed in a pre-specifiable manner.

Results should be interpreted with caution as they are associated with some assumptions and limitations. The work was based on an ongoing study which may introduce a bias in the analysis of the data (24). In fact, due to the lack of information we were not able to differentiate the subjects who withdrew from the study, from the censored patients which could introduce a bias in the analysis of the data if the dropout doesn't occur completely at random.

Regarding the disease progression, different functions can be used to approximate the natural disease progression. Simple models assume a linear disease progression rate over time and other, more complicated models, describe an asymptotic state (8). In this manuscript, the short time of the study justifies the use of a linear function to describe the natural disease progression. In fact, according to Holford *et al.* (8), the linear and Gompertz models provides similar predictions of the natural UPRDS progression from 0 to 4 years. One should note that the disease progression was assumed to be linear under the neuronal disability scale. When converting the predictions to the MDS-UPDRS scale, the score is capped to the maximal value of the scale.

As the medication type and dose received by the patient was not recorded, a combined treatment effect was estimated using only the patient initiation treatment time. To mimic the clinical impact of additional PD medications equations [6] and [7] assume an increase of the symptomatic effect over time. However, due to the lack of recorded information regarding the PD medications, these variants of a symptomatic effect would not be distinguishable from a disease modifying effect. The structure of the drug effect model as well as the estimated parameters describe the observations correctly, however, they cannot be used for extrapolation beyond 48 months.

Another limitation concerns the heuristic nature of the greedy algorithm. It approximates a global optimum by a series of local optimal choices leading to an optimized set of items and not the optimum; however, an extensive search is prohibitive.

In order to carry out a fair comparison, SIS and IRT-based were analyzed under the same conditions. Therefore, both methods estimated a disease modifying effect common to all items even if the IRT model could do differently.

In PD, clinicians are increasingly using combination therapies to treat the different components of the disease with the ultimate goal of slowing the rate of disease progression. However, the recommendations or tools to combine drugs in an efficient manner are limited. This work could be also used as a tool to determine the drug effect profile of different PD medications and therefore help clinicians to combine drugs efficiently.

# CONCLUSION

Based on MDS-UPDRS data from the PPMI database, a longitudinal IRT model was implemented. The model acknowledges the composite nature of the scale through separate latent variables, disease progression rates and drug effects for the motor, non-motor and tremor items. Based on clinical trial simulations, the model was then used to investigate the impact of the analysis method on the power to detect a drug effect. Regardless of the drug effect profile and analysis endpoint, IRT-based analysis provides an equivalent or higher power compared to a SIS-based analysis without any need of "*a priori*" selection of the most informative items. It is a powerful tool which allows to yield information from the MDS-UPDRS scale, and therefore maximizes the likelihood of bringing new medicines to Parkinson's disease patients.

# ACKNOWLEDGEMENTS

This work was financed by F-Hoffmann- La Roche Ltd. We are grateful to the PPMI – a public-private partnership – funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including Abbvie, Inc.; Eli Lilly and Company and its subsidiary company Avid Radiopharmaceuticals; Biogen Idec Inc.; Bristol-MyersSquibb Company; Covance, Inc.; F-Hoffmann- La Roche Ltd and its subsidiary company Genetech, Inc.; GE Healthcare; Servier; GlaxoSmithKline plc.; H. Lundbeck A/S; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; Pfizer Inc.; Piramal Imaging; and UCB. for providing access to the PPMI cohort.

# REFERENCES


1. Pringsheim T, Jette N, Frolkis A, Steeves TDL. The prevalence of Parkinson's disease: a systematic review and meta-analysis. Mov Disord Off J Mov Disord Soc. 2014 Nov;29(13):1583–90.

2. Samii A, Nutt JG, Ransom BR. Parkinson's disease. Lancet Lond Engl. 2004 May 29;363(9423):1783–93.

3. Hughes AJ, Daniel SE, Blankson S, Lees AJ. A clinicopathologic study of 100 cases of Parkinson's disease. Arch Neurol. 1993 Feb;50(2):140–8.

4. Braak H, Ghebremedhin E, Rüb U, Bratzke H, Del Tredici K. Stages in the development of Parkinson's disease-related pathology. Cell Tissue Res. 2004 Oct;318(1):121–34.

5. Frasier M, Kang UJ. Parkinson's Disease Biomarkers: Resources for Discovery and Validation. Neuropsychopharmacology. 2014 Jan;39(1):241–2.

6. Kalia LV, Lang AE. Parkinson's disease. Lancet Lond Engl. 2015 Aug 29;386(9996):896–912.

7. Ramaker C, Marinus J, Stiggelbout AM, Van Hilten BJ. Systematic evaluation of rating scales for impairment and disability in Parkinson's disease. Mov Disord Off J Mov Disord Soc. 2002 Sep;17(5):867–76.

8. Holford NHG, Chan PLS, Nutt JG, Kieburtz K, Shoulson I, Parkinson Study Group. Disease progression and pharmacodynamics in Parkinson disease - evidence for functional protection with levodopa and other treatments. J Pharmacokinet Pharmacodyn. 2006 Jun;33(3):281–311.

9. Vu TC, Nutt JG, Holford NHG. Progression of motor and nonmotor features of Parkinson's disease and their response to treatment. Br J Clin Pharmacol. 2012 Aug;74(2):267–83.

10. Perlmutter JS. Assessment of Parkinson Disease Manifestations. Curr Protoc Neurosci Editor Board Jacqueline N Crawley Al. 2009 Oct;CHAPTER:Unit10.1.

11. Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease. The Unified Parkinson's Disease Rating Scale (UPDRS): status and recommendations. Mov Disord Off J Mov Disord Soc. 2003 Jul;18(7):738–50.

12. Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, 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 Off J Mov Disord Soc. 2008 Nov 15;23(15):2129–70.

13. Jankovic J, Kapadia AS. Functional decline in Parkinson disease. Arch Neurol. 2001 Oct;58(10):1611–5.

14. Louis ED, Tang MX, Cote L, Alfaro B, Mejia H, Marder K. Progression of parkinsonian signs in Parkinson disease. Arch Neurol. 1999 Mar;56(3):334–7.

15. Zetusky WJ, Jankovic J, Pirozzolo FJ. The heterogeneity of Parkinson's disease: clinical and prognostic implications. Neurology. 1985 Apr;35(4):522–6.

16. Chaudhuri KR, Schapira AHV. Non-motor symptoms of Parkinson's disease: dopaminergic pathophysiology and treatment. Lancet Neurol. 2009 May;8(5):464–74.

17. Poewe W, Hauser RA, Lang A, ADAGIO Investigators. Effects of rasagiline on the progression of nonmotor scores of the MDS-UPDRS. Mov Disord Off J Mov Disord Soc. 2015 Apr;30(4):589–92.

18. Baker FB. The Basics of Item Response Theory. 2nd edition. College Park, Md.: ERIC Clearinghouse on Assessment and Evaluation; 2001.

19. 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(8):2152–65.

20. Gottipati G, Karlsson MO, Plan EL. Modeling a Composite Score in Parkinson's Disease Using Item Response Theory. AAPS J. 2017 Feb 28;

21. Hoehn MM, Yahr MD. Parkinsonism: onset, progression and mortality. Neurology. 1967 May;17(5):427–42.

22. Kjellsson MC, Zingmark P-H, Jonsson EN, Karlsson MO. Comparison of proportional and differential odds models for mixed-effects analysis of categorical data. J Pharmacokinet Pharmacodyn. 2008 Oct;35(5):483–501.

23. Beal S, Sheiner LB, Boeckmann A, Bauer RJ. NONMEM User's Guides. (1989–2009). Icon Development Solutions, Ellicott City, MD USA; 2009.

24. Vu TC, Nutt JG, Holford NHG. Disease progress and response to treatment as predictors of survival, disability, cognitive impairment and depression in Parkinson's disease. Br J Clin Pharmacol. 2012 Aug;74(2):284–95.

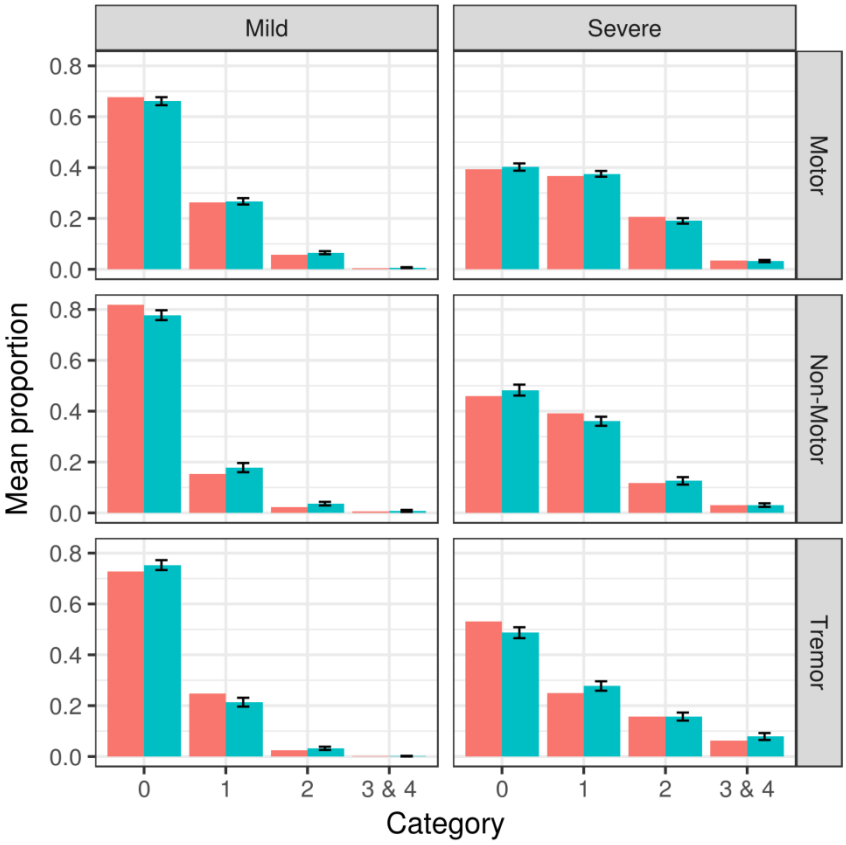



Figure 1: Mirror plot of observed (red) versus simulated (blue) count of subjects for each subscore Motor, Non Motor and Tremor and for two disability subgroups. The lines represent the standard deviation with one hundred repetitions of the simulations

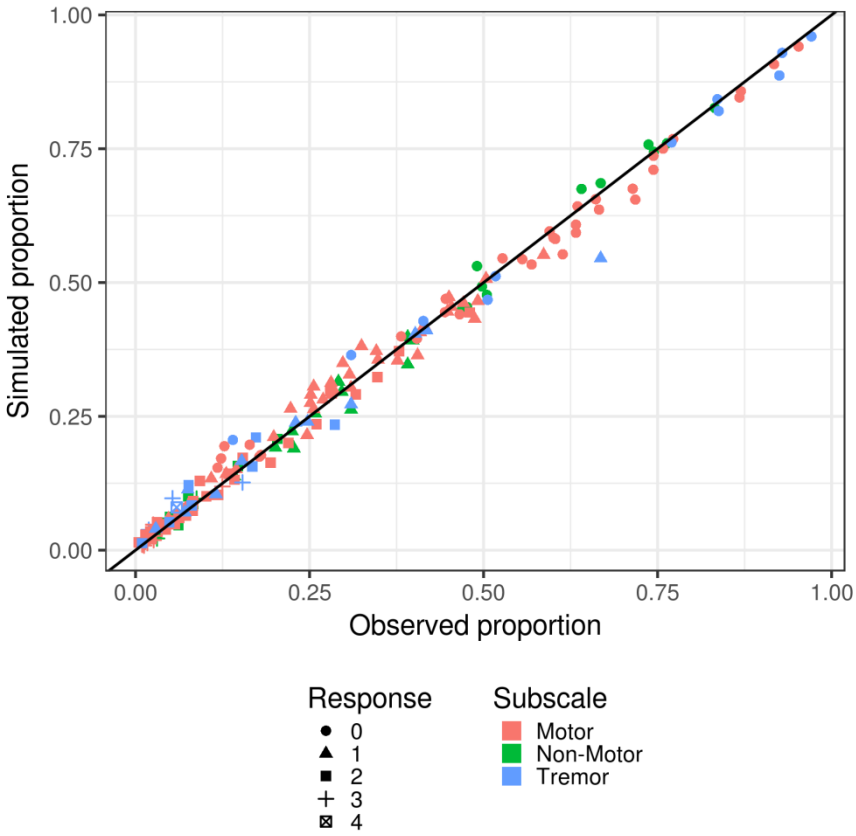



Figure 2: Proportion of observed versus simulated data in each item category. The line corresponds to the identity line.

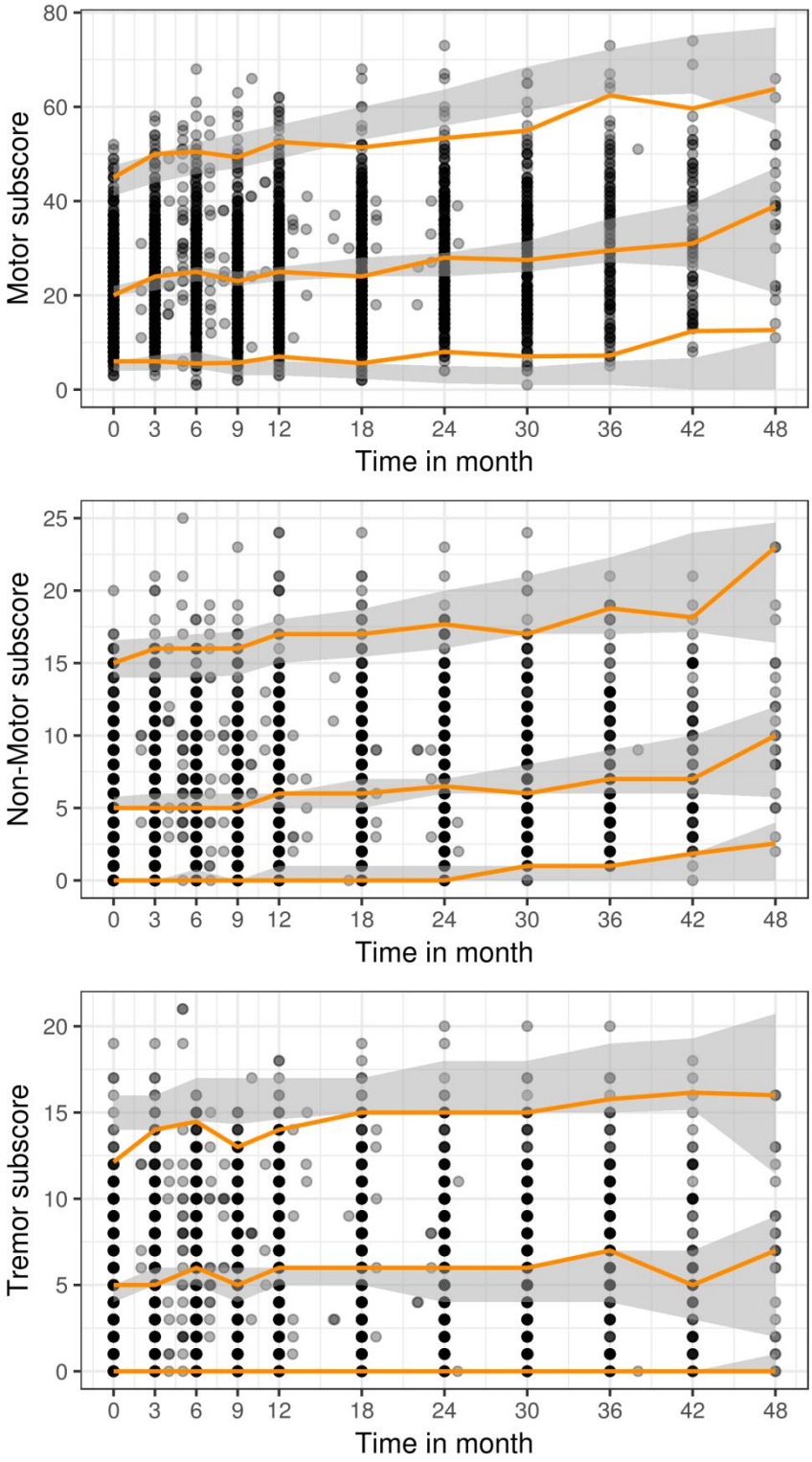



Figure 3: Visual predictive Check of the final total MDS-UPDRS time course model for motor (top), non-motor (middle) and tremor (bottom) subscores. The orange curves represent, the 2.5, Median, and 97.5 percentiles of the observations and the grey area the 95% confidence interval simulated from the model.

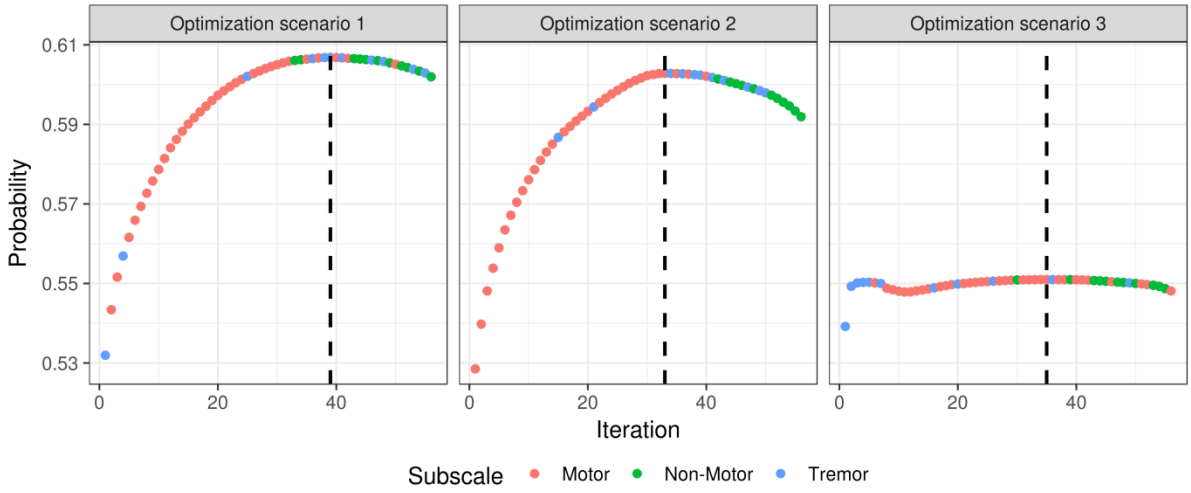



Figure 4: Probability of correctly identifying the treatment effect to be larger than the placebo effect as a function of the optimization algorithm iterations and assuming either a homogeneous (left) or two heterogeneous (center and right) drug effect profiles. The dashed line corresponds to the optimized number of items.

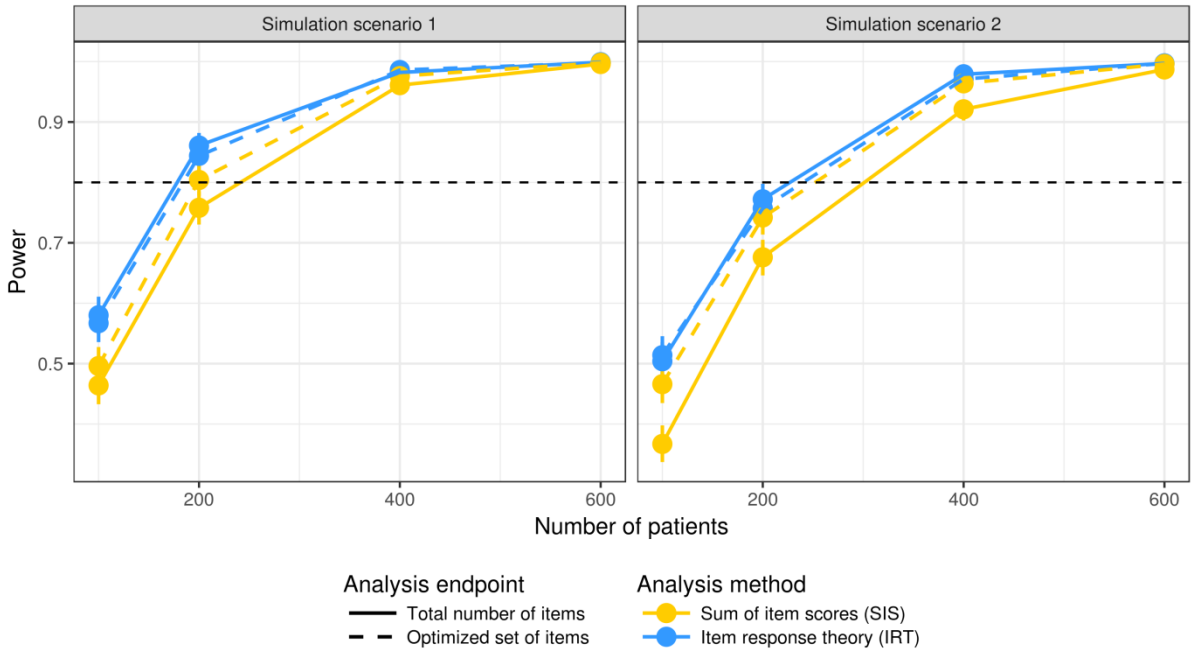



Figure 5: Drug effect detection power as a function of the number of subjects based on a homogeneous (left) or a heterogeneous (right) drug effect profile. Each curve corresponds to the combination of one of the two analysis methods with one of the two analysis endpoints. Vertical lines indicate the 95 confidence interval.

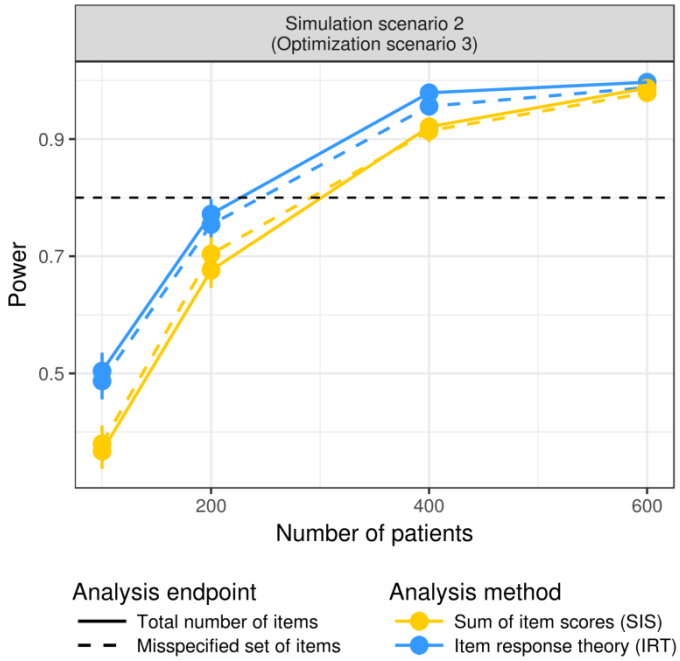



Figure 6: Drug effect detection power as a function of the number of subjects based on a heterogeneous drug effect profile and assuming an incorrect prior knowledge of the drug effect profile. Each curve corresponds to the combination of one of the two analysis methods (SIS or IRT) with one of the two analysis endpoints (all items or a misspecified set of items).Vertical lines indicate the 95 confidence interval.

Table I: Characteristics of PPMI database at baseline.

| Group | Age (Year) | | Duration of disease (Month) | | MDS-UPDS (Point) | |
|---|---|---|---|---|---|---|
| | Median | SD[a] | Median | SD[a] | Median | SD[a] |
| PD | 63 | 9.7 | 4 | 6.5 | 31 | 13.1 |
| Control | 62.5 | 11.2 | . | . | 3 | 4.4 |
| SWEDD | 63 | 10 | 3.5 | 8 | 26 | 17.4 |

[a] Standard Deviation.

Table II: Estimated population parameter values of the IRT-longitudinal model and their uncertainty for respectively the motor, non-motor and tremor subscales.

| Parameters | Latent variable ($v$) | Value (RSE[a]) | |
|---|---|---|---|
| | | Fixed effect | ω[b] |
| $D^0\ (v)$ | Motor | 0.66 (4%) | 0.374 (7%) |
| | Non-Motor | 0.26 (17%) | 0.83 (8%) |
| | Tremor | 0.54 (8%) | 1.0 (9%) |
| $\alpha\ (v/yr)$ | Motor | 0.45 (6%) | 0.049 (18%) |
| | Non-Motor | 0.23 (9%) | 0.042 (18%) |
| | Tremor | 0.26 (10%) | 0.0354 (23%) |
| $E^0\ (v)$ | Motor | -0.34 (9%) | 0.151 (13%) |
| | Non-Motor | -0.18 (20%) | 0.075 (30%) |
| | Tremor | -0.26 (12%) | 0.098 (20%) |
| $\beta\ (v)$ | Motor | -1.4 (10%) | 1.22 (22%) |
| $\beta\ (v/yr)$ | Tremor | -0.31 (12%) | 0.082 (27%) |
| $k_{eq}\ (yr^{-1})$ | Motor | 0.34 (3%) | . |
| Correlations for ${D^0}_i$ | (Motor, Non-Motor) | 0.49[c] | . |
| | (Motor, Tremor) | 0.51[c] | |
| | (Non-Motor, Tremor) | 0.28[c] | |
| Correlations for $\alpha_i$ | (Motor, Non-Motor) | 0.6 (14%) | . |
| | (Motor, Tremor) | 0.43 (19%) | |

[a] Relative Standard Error is the standard error of the parameter divided by the estimate.
[b] Variance of the random effect.
[c] Estimate from the IRT model.
[i] individual