Appendix III – Simulation based Diagnostics
Single Latent Variable Longitudinal IRT Model – Simultaneous Approach

This appendix contains simulation based diagnostic plots evaluating the performance of the single latent variable IRT model using simultaneous approach. It consists of two types of plots: i) mirror plots and ii) visual predictive checks at individual item- and total score-level.

**Mirror plots:** ‘Disability’ (the latent variable) values were simulated from assumed normal distribution (which is centered around a mean of 0 and variance of 1 \(N(0, \omega^2=1)\)). For these disability values, the probabilities for each response category were estimated based on the ICC model parameters (as shown in the Figure – 5 from the manuscript) and the responses were simulated. These simulated responses (green) were then compared with the observed data (orange) and the error bars around the simulated responses represent the variability from 200 repetitions of the simulations. The figures are presented in the chronological order of the items. It should be noted that items 60 and 61 are binary and item 62 is 0 – 5 ordered categorical item while the rest were rated between 0 – 4.

Overall, there seems good to be good agreement between the observations and simulations, even after accounting for the variability from the 200 repetitions of simulations.

**Visual predictive check (VPC) at total score level:** The total score was calculated as the sum of the 68 items. Monte Carlo simulations of 200 datasets were generated using the single latent variable longitudinal model using simultaneous approach. The 95% confidence intervals for the median (red shaded area), 2.5\(^{th}\) and 97.5\(^{th}\) percentiles (grey shaded area) of the total scores from the simulations were compared with the median (solid blue line), 2.5\(^{th}\) and 97.5\(^{th}\) percentiles (dashed blue line) of the total scores in the observed data (red points) respectively.

While the median and 2.5\(^{th}\) percentiles in the observed total scores were in good agreement with the respective 95% confidence intervals simulations for all the time points; except for the 12 month time point, the 97.5\(^{th}\) percentile in the observed data were also within the respective confidence intervals.
Visual predictive check (VPC) at individual item level: Monte Carlo simulations of 200 datasets were generated using the final three latent variable longitudinal model with mixture. The VPCs for these simulations are shown in following figures, presented in the chronological order of the items and depict the longitudinal aspects at individual item level. The observed proportion of subjects (blue circles) was compared to 95% confidence intervals (blue shaded area) obtained from the simulations and in general, they were in good agreement for most of the items.
Figure III-1: Observed responses (orange) and simulated responses (green) and error bars (black whiskers) on the simulated responses represent the variability from 200 repetitions of the simulations.
Figure III - 2: Observed responses (orange) and simulated responses (green) and error bars (black whiskers) on the simulated responses represent the variability from 200 repetitions of the simulations.
Figure III - 3 Observed responses (orange) and simulated responses (green) and error bars (black whiskers) on the simulated responses represent the variability from 200 repetitions of the simulations.
Figure III - 4: Observed responses (orange) and simulated responses (green) and error bars (black whiskers) on the simulated responses represent the variability from 200 repetitions of the simulations.
Figure III - 5: Visual predictive check of the total MDS-UPDRS scores for single latent variable IRT model:

Observed data (red points) – median (solid blue line), 2.5th and 97.5th percentile (dashed blue line) was compared to 95% confidence interval around median (red shaded area) and respective percentiles (grey shaded area).
PsN generated vpc item wise pdf