Study design considerations for digital assessments in Parkinson's disease clinical trials

Jennie S. Lavine*, Anthony D. Scotina, Elena S. Izmailova, Larsson Omberg
Koneksa Health Inc, New York, NY, USA

*jennie.lavine@koneksahealth.com

MT33

Background

- A key challenge for detecting efficacy in Parkinson's disease (PD)-modifying therapy trials is a lack of easy-to-implement, sensitive outcome measures to detect changes in disease progression.
- Both in-clinic and at-home assessments have excellent test-retest reliability when administered well (Martinez-Martin 2013, Burq 2022) .
- Nonetheless, progression rate remains challenging to estimate due to PD's slow and variable progression dynamics, natural day-to-day fluctuations (Evers 2019) and the effects of dopaminergic therapy (Holden 2018).
- Digital at-home assessments can be implemented frequently and have a potential to assess progression in shorter times and/or with smaller sample sizes than traditional in-clinic outcome measures.
- Here, we develop recommendations for digital assessment implementation in PD clinical trials.

Methods

- We simulate clinical trials (see Figure 1) of DMTs from a previously developed and parameterized model of PD progression informed by PPMI data (Evers 2019, Figure 2 & Table 1).
- The parameters we use represent rates for Hoehn & Yahr stages 1 & 2, with the majority of the population on dopaminergic therapy.
- We consider three study designs to detect disease progression within one year (Figure 2, Box 1).

Results

- For study design 1, parallelling **in-clinic assessments, ~220 or 640** participants are required to detect a treatment effect with a **90% or 50% effective DMT,** respectively, at 80% power (Figure 3, magenta).
- Including a **burst of 7 daily assessments** (study design 2) reduces the necessary sample size to **~ 160** (90% effect) **or 480** (50% effect; Figure 3, cyan).
- For scenario (3), with **48 evenly spaced assessments**, only **~ 70** (90% effect) **or 190** (50% effect) participants are required (Figure 3, blue).
- Increasing the number of assessments per burst beyond ~5 does not increase power (Fig 4A).
- Increasing the number of evenly-spaced measurements per year shows a continuous increase in power up to at least 40 assessments, and, for less effective DMTs, beyond that (Fig 4B).

Conclusions

We find that frequent measurements in PD greatly increase the ability to detect a treatment effect with fewer participants. Including a burst of measurements at the beginning and end decreases the necessary sample size by ~ 25%. Weekly at-home measurements provides far greater benefit, reducing necessary sample sizes by ~70% from the in-clinic design and close to 60% from the burst design. Further work to understand the sensitivity of these results to non-linear disease progression, effects of a DMT that change with time, and varying characteristics of at-home assessments is warranted.

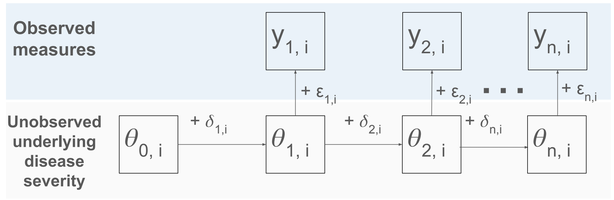
Figure 1: Schedule of assessments for study designs

Box 1: Endpoints

- For study design 1, a *difference in change from baseline* of the measure value (e.g., MDS-UPDRS part III total score) by study arm (i.e., placebo vs DMT) is assessed via analysis of variance (ANOVA) comparing the assessment at day 1 to the assessment at day 365.
- For study design 2, a *difference in change from baseline* of the measure value by study arm is assessed via ANOVA comparing the median of 7 assessments on days 1 - 7 to the median of 7 assessments on days 359 - 365.
- For study design 3, a *difference in the rate of change* by study arm is assessed via a linear mixed effects model with measure value as the response variable, fixed effects for weeks-from-baseline, study arm and their interaction, and a random intercept for participant.

Figure 2: Model formulation & simulation framework

A



$$\theta_{0,i} \sim N(\mu, \sigma_s); \quad \delta_{t,i} \sim N(\tau, \sigma_T); \quad \varepsilon_{t,i} \sim N(0, \sigma_m)$$

B

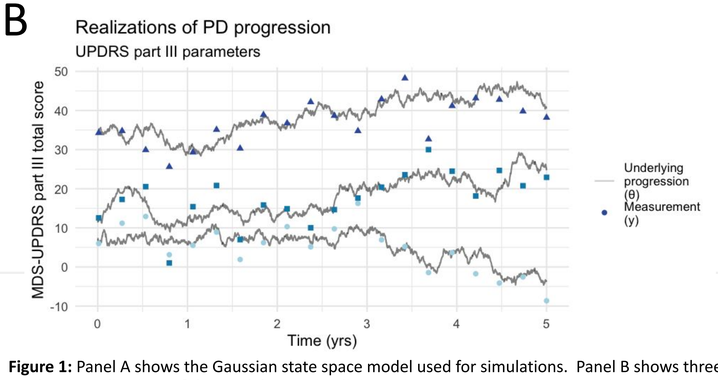


Figure 1: Panel A shows the Gaussian state space model used for simulations. Panel B shows three stochastic realizations of the model, using in-clinic MDS-UPDRS part III parameters (see Table 1). Underlying disease progression is represented by gray lines in panel B and $\theta_{t,i}$ in panel A. The variability in progression rates between individuals and across time within individuals arises from the variability in the trend, simulated by $\delta_{t,i}$, and result in the unobserved underlying disease states (gray lines). The observed measurements (i.e., MDS-UPDRS scores, digital assessment scores, etc) are represented by $y_{t,i}$ in panel A and point in panel B. The vertical distance between the gray line and its associated points represent the variability induced by the measurement process, $\varepsilon_{t,i}$.

Table 1: Parameters in power simulations

Parameter	μ	σ_S	τ	σ_T	σ_M
Value	20.3	8.9	3	5.6	3.9

All parameter values taken from Evers (2019) for MDS-UPDRS part III in the off state except for the trend, which is increased from 2.63 to 3 to reflect an early population with a few more treatment-naive individuals. For comparison, the annual progression rate for Part III off state in only treatment naive individuals has been estimated to be at least 4 points per year (Holden 2018).

Figure 3: Power to detect a difference between study arms

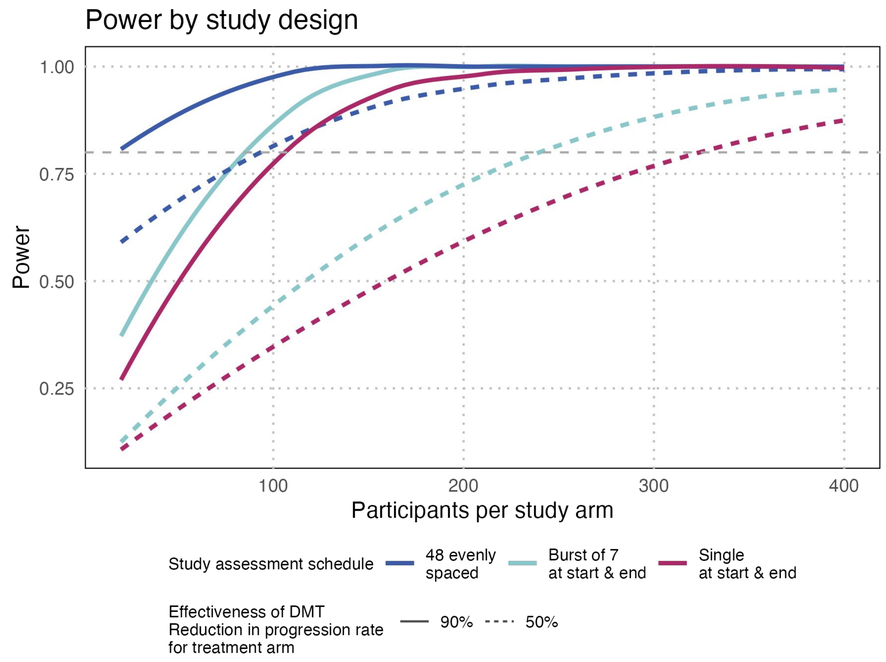


Figure 3: Simulations of clinical trials for DMTs across three study designs and two drug efficacies. Solid lines represent a drug that reduces disease progression by 90%, dashed lines by 50%. More frequent assessments increase power (blue and cyan above magenta), and evenly spaced frequent assessments provide a large additional benefit beyond bursts (blue curves above cyan).

Figure 4: Effect of increasing frequency on sample size needed for 80% power

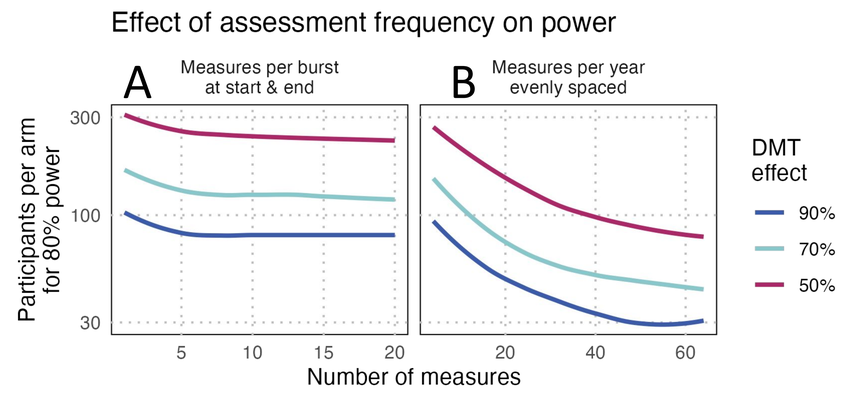


Figure 4: The effect of adding more assessments is shown here. In the burst design (panel A), including more than 5 or 6 assessments provides little additional benefit ('elbow' in curves at ~ 6). When using evenly spaced assessments, including more than 40 or 45 (i.e., approximately weekly) shows little additional benefit.

Citations

1. Burq, M., Rainaldi, E., Ho, K. C., Chen, C., Bloem, B. R., Evers, L. J. W., Helmich, R. C., Myers, L., Marks, W. J., Jr., & Kapur, R. (2022). Virtual exam for Parkinson's disease enables frequent and reliable remote measurements of motor function. NPJ Digit Med, 5(1), 65. https://doi.org/10.1038/s41746-022-00607-8
2. Evers, L. J. W., Krijthe, J. H., Meinders, M. J., Bloem, B. R., & Heskes, T. M. (2019). Measuring Parkinson's disease over time: The real-world within-subject reliability of the MDS-UPDRS. Mov Disord, 34(10), 1480-1487. https://doi.org/10.1002/mds.27790
3. Holden, S. K., Finseth, T., Sillau, S. H., & Berman, B. D. (2018). 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, 5(1), 47-53. https://doi.org/10.1002/mdc3.12553
4. Martinez-Martin, P., Rodriguez-Blazquez, C., Alvarez-Sanchez, M., Arakaki, T., Bergareche-Yarza, A., Chade, A., Garretto, N., Gershanik, O., Kurtis, M. M., Martinez-Castrillo, J. C., Mendoza-Rodriguez, A., Moore, H. P., Rodriguez-Violante, M., Singer, C., Tilley, B. C., Huang, J., Stebbins, G. T., & Goetz, C. G. (2013). Expanded and independent validation of the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS). J Neurol, 260(1), 228-236. https://doi.org/10.1007/s00415-012-6624-1