Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Time-to-Event Analyses of UPDRS and MDS-UPDRS Motor Examination and Functional Outcomes in Early Parkinson's Disease Clinical Trials

Andrew McGarry
mcgarry-andrew@cooperhealth.edu
Cooper University Hospital
Christopher R Meyer
Abli Pharmaceuticals
Karl Kieburtz
University of Rochester
Jordan Dubow
Clintrex Research Corporation
Milton H Werner
Abli Pharmaceuticals
Charles Venuto
University of Rochester

Article

Keywords:

Posted Date: June 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10021801/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at npj Parkinson's Disease on September 23rd, 2026. See the published version at https://doi.org/10.1038/s41531-026-01578-3.

Abstract

Clinical trials in untreated Parkinson's Disease test therapeutics aimed at improving underlying disease pathophysiology, but inevitably risk the confounding of study results by treatment initiation. Milder clinical features, relative functional preservation and slow progression also raise questions as to how a meaningful benefit can be demonstrated in the period before treatment begins. Time-to-event (TTE) outcomes may be a valuable approach to detect benefit before treatment. To explore the feasibility of detecting large, potentially more meaningful exam-based changes before treatment and how they correlate to functional change, we analyzed four early untreated PD interventional studies and one observational study over a 12-month period. Exam and functional worsening were assessed at varying thresholds using the Unified Parkinson's Disease Rating Scale (UPDRS) or its Movement Disorders Society revision (MDS-UPDRS) Parts 3 and 2, respectively. At the largest thresholds, the percentage of persons achieving untreated worsening ranged from 32%-42% on part 3 (≥ 7 points) and 27%-34% on part 2 (≥ 4 points). As thresholds were relaxed on part 3 and part 2, the proportion of participants achieving worsening increased across all cohorts. Among those crossing Part 3 thresholds, a substantial proportion also experienced concordant Part 2 worsening, with higher Part 3 thresholds generally associated with greater functional worsening. These TTE analyses in early PD trials support the feasibility of 12-month studies powered for large worsening thresholds in exam scores.

Introduction

Clinical trials enrolling participants with early Parkinson's Disease (PD), before the initiation of motor symptom treatment, offer certain advantages for testing therapeutics. These persons are generally earlier in their biological course and plausibly have less neuronal dysfunction or loss, such that strategies aimed at improving the underlying disease pathology might have a greater chance to be effective. Debilitating non-motor features in early PD are typically less prevalent, and milder motor symptoms help with practical considerations like trial adherence and sustained participation. However, the milder phenotype, relative functional preservation and slow progression in early PD raise concerns as to how a clinically meaningful benefit can be demonstrated. Sensitive exam-based metrics may be able to detect change in motor performance, but lessened motor impairment may be of uncertain clinical importance depending on the magnitude of effect and the features that improve. There is no consensus in the literature, nor formally documented among regulators, regarding how much change on examination would represent a clinically meaningful treatment effect; indeed, changes in examination findings do not necessarily or predictably translate into how a person feels or functions. It may be argued that favorably altering exam findings alone cannot be taken as persuasive evidence of meaningful benefit, although experience with levodopa clearly shows that large magnitude changes on exam do correlate to improvement in function. Alternatively, functional outcomes in untreated patients could be interpreted as inherently meaningful, even for benefits of smaller magnitude than those seen with dopaminergic therapy. However, improvement in function might be subject to the same critique as examination-based findings: depending on what improves or by how much, positive change may not be judged of equal value to all patients. Ideally, an exam-based benefit of specific magnitude would correlate with favorable change in a functional outcome, and could be accepted as comprehensively representing an improvement that matter to patients.

Critical decisions on key outcomes and study design for this population, then, must balance the "how much" of exam change and the "so what" of functional change. These are ideally measured before the advent of treatment. While careful screening and selection may help postpone treatment in a study measuring change from baseline to a fixed timepoint, treatment initiation will inevitably occur and require consideration analytically as an intercurrent event. The temporal profile of change in impairment that may motivate treatment considerations can be heterogeneous even in carefully considered participants with early disease, such that rapid deterioration with or without an obvious precipitating event may affect some and not others. These occurrences may be difficult or impossible to predict and complicate the usual approach of analyzing mean group score changes at a given timepoint. The study population, while homogenous clinically, may be a mixture of persons with relatively stable compensatory physiology and others who may imminently exhaust some component of it during the study. As such, there is an inherent challenge when testing innovations meant to hinder or pause pathophysiology in a slowly progressive disease with available symptomatic treatments: longer studies may simultaneously promote success (prolonged exposure to therapeutics in a more amenable biological state) and risk failure (measurable benefit may be confounded by treatment initiation before the end of the study).

In that context, time-to-event (TTE) outcomes represent a potential way to minimize these risks in early untreated PD trials. TTE designs use a pre-specified event or status change as the primary endpoint, as seen in examples from oncology, like time to death or disease progression, or from cardiovascular studies following major adverse cardiac events [1, 2]. A study period that is sufficiently long to have enough events for effectively comparing treatment groups can be specified. Appearance of the defined event at any time lessens the risk of prolonged participation incurring important intercurrent events that could impact data interpretation at a fixed timepoint. Given the detectable and often robust effects that dopamine replacement has on both exam features and function, TTE designs may be particularly useful. Selection of the event of interest would still need to manage the tension between sensitive phenotypic observations and their relationship to meaningful functional improvements as seen by patients, families, clinicians, and regulators.

The Unified Parkinson's Disease Rating Scale (UPDRS) and Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) have been the preferred instruments in PD clinical trials for decades [3, 4]. Each has an exam-based element completed by a certified rater (Motor Examination) and a functional assessment (Activities of Daily Living (ADL) and Motor Aspects of Experiences of Daily Living (M-EDL), respectively). The ADL scale is also completed by the rater, whereas the M-EDL is a patient-reported outcome. Several FDA-approved PD medications have evidence of efficacy demonstrated with these outcomes at fixed timepoints [5–9], while other experimental agents using them failed to demonstrate efficacy [10]. More recently, TTE approaches have been implemented in PD trial designs for MDS-UPDRS Part II, Part III, and Part II + III in early and stably

treated populations [11–13]. Here, we sought to explore 1) the frequency of exam-based changes of varying magnitude in early PD studies using the UPDRS and MDS-UPDRS, 2) how commonly these were observable before initiation of treatment, and 3) what changes were present in functional outcomes when motor worsening thresholds were met.

Methods

Clinical Studies

We analyzed data from four multicenter, randomized, double-blind clinical trials (SURE-PD3 [14], STEADY-PD III [15], FS-1 [16], and FS-TOO [17]) and one multicenter observational study (PPMI [18]) enrolling individuals with early PD. All participants provided informed consent to participate in these studies, and Institutional Review Boards (IRBs) at participating sites provided ethical approval for study activity. Conduct of all studies was in accordance with the Declaration of Helsinki and International Conference on Harmonization Good Clinical Practice Guidelines. As this work is a secondary analysis of those studies, no IRB evaluation was undertaken and there is no ethical approval number associated with this work. Protocol ID numbers for the individual studies are listed below.

All studies recruited individuals with a recent diagnosis of PD who were not yet requiring symptomatic dopaminergic therapy at baseline, although each protocol defined allowable symptomatic medications differently. All trials collected longitudinal clinician-rated measures of parkinsonian disability using either the UPDRS or the MDS-UPDRS, with assessment schedules determined by the parent protocols (**Supplemental Table 1**).

SURE-PD3 (NCT02642393): This was a phase 3, randomized (1:1), placebo-controlled trial testing inosine as a urate-elevating therapy for up to two years in early PD [14]. The trial enrolled 298 participants who had (1) dopaminergic deficit on DaT imaging and (2) serum urate < 5.8 mg/dL who had not yet started dopaminergic antiparkinsonian medication, although stable MAO-B inhibitor use was permitted at enrollment. The primary outcome was the rate of change in MDS-UPDRS total score (I-III) prior to initiation of dopaminergic therapy. Dopaminergic therapy was defined as starting levodopa, dopamine agonists, amantadine, or anticholinergics; dose increases or initiation of MAO-B inhibitors were also counted as symptomatic therapy events. Participants were followed regularly through 24 months; however, the study was stopped early for futility at a prespecified interim analysis.

STEADY-PD III (NCT02168842): This was a phase 3, randomized (1:1), double-blind, placebo-controlled trial evaluating isradipine over 36 months in early PD [15]. The study enrolled 336 participants diagnosed within 3 years who were not expected to require dopaminergic therapy for at least 3 months after baseline. Stable amantadine or anticholinergic use was allowed if initiated 8–12 weeks before enrollment. The primary endpoint was change in original UPDRS Parts I-III from baseline to 36 months. Participants who initiated symptomatic dopaminergic therapy during follow-up were assessed in the medication ON state for the primary analysis; OFF state scores were collected and analyzed separately. Primary and secondary outcomes showed no significant effect of isradipine treatment compared to placebo.

FS-1 (NCT00063193): This was a phase 2, randomized (1:1:1), double-blind, multicenter futility trial testing creatine and minocycline versus placebo in early PD (n = 200) [16]. Participants were diagnosed within 5 years, did not require symptomatic therapy, and were excluded if they had atypical parkinsonism, prior PD surgery, or had used study-related compounds prior to enrollment. The primary outcome was the change in total UPDRS (Parts I-III) from baseline until either (1) the time site investigators judged disability sufficient to require symptomatic therapy, or (2) 12 months. Investigators were trained on the symptomatic initiation endpoint using case vignettes, and retrospectively each decision to initiate therapy was reviewed by members of the Steering Committee. Neither creatine nor minocycline could be rejected as futile. Subsequently, creatine was studied in a phase 3, randomized, placebo-controlled double-blind study but was terminated early due to futility during a planned interim analysis.

FS-TOO (NCT00076492): This was a phase 2, randomized (1:1:1), double-blind futility trial of GPI-1485 and coenzyme Q10 versus placebo in early PD (n = 213) [17]. The inclusion criteria, futility design, and primary outcome (change in UPDRS I-III to time of symptomatic therapy initiation or 12 months) mirrored FS-1. Neither GPI-1485 or coenzyme Q10 could be rejected as futile.

PPMI (NCT01141023)

The Parkinson's Progression Marker Initiative (PPMI) is an ongoing, international, multicenter observational study launched in 2010 to characterize PD progression and identify clinical, imaging, and biospecimen-based biomarkers of disease onset and progression [18]. PPMI enrolls individuals with recent-onset PD, at-risk individuals based on clinical features or genetic variants, and neurologically healthy controls. For this analysis, we focused exclusively on participants enrolled under Protocol Version 1, restricting the sample to the idiopathic (sporadic) PD cohort. Key inclusion criteria for PD participants included the presence of at least two cardinal motor signs (resting tremor, bradykinesia, or rigidity), baseline Hoehn and Yahr stage I-II, evidence of a dopamine transporter deficit on SPECT imaging, and no anticipated need for symptomatic therapy within six months of baseline.

Time-to-Event Analyses

Time-to-event analyses evaluated the time from the baseline study visit to the earliest occurrence of motor worsening of varying magnitudes or initiation of symptomatic antiparkinsonian therapy, with follow-up restricted to approximately 12 months. This duration was felt to reflect a reasonable balance between observing meaningful, measurable clinical progression while plausibly delaying the initiation of treatment as long as possible for a

putative study design. Motor worsening was assessed using clinician-rated UPDRS or MDS-UPDRS Part III scores (Motor Examination). Functional worsening was assessed using UPDRS clinician-rated functional scores (ADL) and patient-reported MDS-UPDRS Part II scores (M-EDL). Separate analyses were conducted for predefined worsening thresholds. For Part III, thresholds of ≥ 7-, ≥ 6-, ≥ 5-, and ≥ 4-point worsening from baseline were evaluated; for Part II, thresholds included ≥ 4-, ≥ 3-, and ≥ 2-point worsening. For each threshold, the event time was defined as the earliest study visit at which the participant's score met or exceeded the specified increase relative to baseline. Initiation of symptomatic therapy was defined as the first use of levodopa, dopamine agonists, amantadine, anticholinergics, or MAO-B inhibitors. Because study protocols differed, increases in MAO-B inhibitor dose (SURE-PD3) and increases in amantadine or anticholinergic dose (STEADY-PD III) were also counted as treatment initiation events.

For each participant, we recorded the earliest date of motor worsening or symptomatic therapy initiation; if neither event occurred, the participant was censored at the earlier of study withdrawal (if occurring ≤ 12-month window) or the last observed clinic visit within the 12-month window. A small number of participants met both worsening and treatment initiation criteria on the same day; these were treated as events of motor worsening given worsening would have been identified first. Event timing was recorded at day-level resolution for the clinical trial datasets, whereas the PPMI de-identified data were provided with visit timing only at month-level resolution.

Kaplan-Meier methods were used to estimate event-free survival for each worsening threshold, overall and stratified by randomized treatment group, and, depending on the study, either baseline MAO-B inhibitor use (SURE-PD 3) or amantadine/anticholinergic use (STEADY-PD III). Results from placebo groups are shown in the main analysis below, with active arms from each study available in supplementary material (results were similar). All analyses were implemented in R.3.1 (R Foundation for Statistical Computing, Vienna, Austria) using the survival package, with graphical output generated using survminer and base R tools.

Part II Threshold Attainment at the Time of Part III Threshold Crossing

For each participant, we identified the earliest visit at which they crossed pre-specified Part III worsening thresholds (≥ 4, ≥ 5, ≥6, or ≥ 7-point change from baseline). For each person crossing their respective Part III threshold, we then determined whether they also crossed a pre-specified Part II threshold (≥ 2, ≥ 3, or ≥ 4 points from baseline) at any visit up to and including the time of the Part III crossing. The result of this analysis is the observed proportion of persons meeting a given Part III threshold who had also crossed each Part II threshold by that same time point (i.e., co-attainment). Thus, participants who did not cross a given Part III threshold were not included in that threshold's co-attainment calculation.

For this analysis, participants were censored at symptomatic treatment initiation and only (MDS)-UPDRS observations collected prior to initiation were eligible for determining threshold crossing. Any visits occurring after initiation were excluded from the analysis.

Sample Size Calculations

Sample size calculations were performed for each study cohort across all composite threshold definitions as trial outcomes. The required number of events was first derived using the Schoenfeld formula for log-rank tests, assuming 1:1 active-to-placebo randomization, two-sided $\alpha = 0.05$, and 80% and 90% power to detect hazard ratio of 0.70 (30% reduction in hazard). Total sample size was then calculated from the required event count based on the average event probability across arms. Control arm event probabilities were derived from the observed cumulative proportions of participants reaching each composite endpoint in the source cohorts (visit schedules differed slightly across cohorts but are implicitly captured in these observed proportions). Treatment arm event probabilities were derived under the proportional hazards assumption. Total sample sizes were inflated to account for dropout rates ranging from 5% to 20%, assuming non-informative censoring. Dropout was defined as withdrawal for reasons unrelated to disease progression.

Results

Demographics

Baseline data for placebo participants from each study are shown in Table 1. Male participants were more common across the studies (range 55.0%-70.4%), participants were predominantly white (89.2–97.3%), and mean time since diagnosis across studies was less than 1 year (0.56–0.82). Motor examination scores at baseline reflected mild features for UPDRS part III (15.6–16.3) and MDS-UPDRS part III (20.7–22.5) with mild functional disability (ADL range of 5.4-6.0 or M-EDL range of 5.8–6.1). 41.6% of participants in SURE-PD3 were on stable MAO-B inhibition, and 8.4% of STEADY-PD III participants at baseline were on stably administered doses of amantadine/anticholinergics.

Part III Threshold Composite Worsening

Across all randomized clinical trials and the observational PPMI cohort, the proportion of placebo participants experiencing the endpoint, defined as either Part III motor worsening or symptomatic therapy initiation, increased consistently as lower worsening thresholds were applied (Table 2). In general, motor worsening accounted for the majority of events, while treatment initiation contributed a smaller share that declined progressively as the motor-worsening threshold became less stringent. In PPMI, event rates were comparable to those observed in the SURE-PD 3 and STEADY-PD III

studies (80.8% at ≥ 7 points; 89.2% at ≥ 4 points) (Table 2). Because PPMI is non-interventional, no comparisons were performed between the active and placebo arm, but the pattern of relatively high worsening rates and modest treatment-initiation frequencies mirrored those in the placebo arms of the clinical trials.

The FS-1 and FS-TOO cohorts compared to the other studies had lower proportions of participants reaching the endpoint across thresholds (e.g., 56.3–65.7% at ≥ 7 points; ~78% at ≥ 4 points) (Table 2). This lower event rate was driven primarily by fewer participants meeting the symptomatic treatment initiation component rather than by Part III motor worsening.

Part II Threshold Worsening

For functional measures, event rates followed the expected pattern of increasing frequency as lower worsening thresholds were applied (Table 3). When focusing on worsening-related scoring threshold events, the proportions with worsening exceeded 50% consistently across studies only when the threshold was relaxed to 2 points. Across the UPDRS-based studies, STEADY-PD III consistently had higher event rates than FS-1 and FS-TOO at each threshold, largely because treatment initiation contributed a greater share of events in STEADY-PD III (Table 3). Regarding studies using MDS-UPDRS, event rates in PPMI were consistently higher than in SURE-PD3.

Active/Placebo Comparison

For the Part III composite endpoints, no statistically significant differences between active and placebo arms were observed in any of the clinical trials. In STEADY-PD III, composite event rates were numerically lower in the isradipine arm than in the placebo arm across thresholds; however, these differences did not reach statistical significance in the Kaplan-Meier analyses (**Supplementary Table S2, Supplementary Figures S1-S4**).

Similarly, for Part II-based composite endpoints, no significant differences were observed between “active” and placebo arms for any of the clinical trials (**Supplementary Table S3, Supplementary Figures S5-S8**).

Use of Allowable Symptomatic PD Medications at Baseline Comparison

In the STEADY-PD III and SURE-PD3 cohorts, baseline symptomatic PD medication use was not associated with a significant difference in time to reaching the Part III worsening thresholds or initiating PD treatment composite outcome (**Table S4, Figures S9-S10**). This was consistent in STEADY-PD III for participants taking amantadine or anticholinergics at baseline (log rank p-values ≥ 0.61 across thresholds), and in SURE-PD3 for those taking MAO-B inhibitors (log rank p-values ≥ 0.33).

Baseline symptomatic PD medication use was similarly not associated with a significant difference in time to reaching composite outcome of Part II threshold worsening or PD treatment initiation (Table S5, Figures S11-S12).

Functional Threshold Attainment at Motor Threshold Crossing

We examined time-aligned MDS-UPDRS Part II threshold attainment across five cohorts to determine what proportion of participants experiencing Part III threshold worsening had also crossed Part II worsening threshold by the same follow-up time point. Across all cohorts and both analyses, higher Part III worsening thresholds were associated with greater proportions also meeting Part II thresholds, suggesting that participants with more pronounced motor worsening were more likely to have also reported functional decline by that time (Fig. 1). For the most severe threshold combination examined (Part III ≥ 7, Part II ≥ 4), the proportion of Part III attainees who also met the Part II threshold ranged from 20% (SURE-PD3) to 50% (FS1). For the least severe combination (Part III ≥ 4, Part II ≥ 2), this proportion ranged from 40% (SURE-PD3) to 53% (FS-TOO). We also summarized the continuous Part II change from baseline at the time each Part III threshold was first crossed among untreated visits (Table S6). Mean changes were variable across cohorts, with larger changes observed in UPDRS trials than MDS-UPDRS cohorts.

Sample Size Calculations

For each study analyzed, the total sample size required to detect 30% slowing for various time-to-event thresholds are shown, assuming 1:1 active/placebo randomization, two-sided alpha of 0.05 and power of 80% when considering Part III (Table 4A) or Part II (Table 4B) threshold changes as the outcomes of interest (90% power calculations in Tables S7A-B). Dropout rates from 5 to 20% are shown. This analysis assumes dropouts would be for reasons other than treatment initiation, as this was accounted for in the combined event of interest (threshold met/treatment).

Discussion

The present analysis sought to examine motor deterioration and the associated functional change over 12 months in early PD studies using a TTE methodology for potential application in trial design. The frequencies of varying examination-based and functional thresholds were attained independently, and functional scores were also examined at the time motor worsening thresholds were met. Worsening while untreated was more common than starting treatment across all five studies for each Part III threshold of 4, 5, 6, or 7 points, with the exceptions of the 7-point threshold in the active and placebo arms of STEADY PD 3 and the minocycline arm in FS-1.The percentage of persons achieving untreated worsening remained substantial for the largest threshold (≥ 7 points), and 6-point worsening in untreated participants was also relatively common.

These observations have several important implications. Even for larger magnitude worsening events (6 or 7 points in Part III), a substantial portion of participants in the examined trials met these thresholds without treatment and usually had associated functional decline, suggesting these thresholds are reasonable targets for therapeutic interventions tested with TTE analyses. Our sample size calculations suggest Phase 2 or 3 studies using TTE outcomes are plausible to conduct. We note that initiation of treatment was part of our outcome of interest and factored into the identified event (the earliest date of motor worsening or starting treatment), and so dropout rates in these studies were likely due to reasons other than starting dopaminergic medication. Since initiation of treatment can be expected to be the predominant intercurrent event in an early untreated PD study, smaller rates in our tables may be more realistic (5–10%). The observed frequencies for larger thresholds being met without treatment also suggest that common features and procedures for early studies already conducted seem applicable towards future studies testing the occurrence of larger magnitude events. Typical early PD inclusion/exclusion criteria, site considerations in the selection of suitable candidates for longer studies, and the ingredients comprising judgements around when to start treatment together appear to make TTE designs using higher magnitude thresholds plausible to conduct for Parkinson's Disease.

Part III declines of 6 or 7 points were generally associated with worsening on functional outcomes at the times of meeting these thresholds, but with variability in functional scales as might be expected. For time-aligned analyses of specific functional (Part II) thresholds being met (2-, 3-, and 4-point decline) coincident with worsening on examination, more impaired thresholds (i.e., 4-points) were consistently less common than lesser ones (i.e., 2- or 3- points) at each Part III threshold across studies. These patterns were generally similar irrespective of treatment status, suggesting larger decrements on examination carry some level of intrinsic functional impairment. More than half of each cohort examined declined by at least 2 functional points at the highest 7-point examination threshold, with comparable percentages at 6-point thresholds. Associated 4-point worsening, the most severe threshold for functional impairment, was not rare at the time of 7- or 6-point Part III worsening (20–50% in time-aligned analyses). Our data show sufficiently visible motor worsening tends to contemporaneously reveal itself in functional disability of participants, suggesting a useful intersection of "how much" and "so what" may exist for higher Part III thresholds. Given the coincident functional disability, it is reasonable to consider that forestalling or preventing Part III changes of larger magnitude may also ameliorate or delay functional disability, a provision expected to be meaningful for patients, families, and clinicians. These data support the feasibility of expecting an opportunity for concomitant functional improvement with higher Part III thresholds in TTE analyses.

While the time-aligned analyses support a correlation between exam worsening and functional impairment, we note that the full range of Part II scores associated with each threshold of Part III worsening are in a distribution that includes no worsening and even functional improvement. Part II does not ask participants to disentangle other possible influences on functional capability, will reflect the inherent variation present in daily activity across persons even in large samples, and depends on the attention and awareness of participants with a progressive neurodegenerative process commonly involving behavioral and cognitive change even in early disease. Nonetheless, the alignment of exam abnormalities and self-reported functional disability at these larger examination-based thresholds in early PD datasets suggests that 1) powering a study in early PD around delaying time to 6-or 7-point change has the potential for demonstrating benefit in a highly ranking functional secondary outcome like Part II to support clinical meaningfulness, or 2) some combination of 6- or 7-point change on Part III with Part II as a composite outcome would seem practical to consider and potentially offer further advantages for sample sizes and regulatory interpretability. Future analyses will examine the possibility of a composite outcome using Parts II and III in more detail.

We note that our analyses explored time to the initial events of interest and did not assess the persistence of these thresholds being met. As forestalling or preventing the achievement of worsening would serve as important evidence of efficacy, we think it reasonable to include timely provisions for confirmation of the worsening event of interest in any such study design to support a more robust interpretation. While one might expect less variation for larger examination-based thresholds that clearly reflect impairment, data suggest Part III in untreated persons can vary within a day depending on time of measurement, as well as over several weeks in practically-defined OFF states or years in both ON and OFF states (19–21). Provisions to ensure the same rater, same time of the day for repeat examination, queries for any subsequent circumstances that may influence examination results (intercurrent illness, starting medication, etc.), and relatively short time between exams may help minimize variation in confirmatory exams. We also note that delaying time to worsening examination-based or functional thresholds may carry a different interpretation than typical mean differences measured at fixed timepoints. Studies designed to demonstrate a slower tempo of defined worsening between groups may not necessarily reflect smaller thresholds of worsening that could be meaningful, or even improvement from baseline that may be present in some participants. While one might expect beneficial effects of an effective agent to be persistent, it does not follow that forestalling a worsening threshold would invariably prevent subsequent, potentially substantial decline before the end of a study, and so it may not necessarily be clear how durable treatment effect may be without additional outcomes to provide context. Time-to-event outcomes may be supplemented by other outcomes reflecting mean change between groups at fixed timepoints, or with sensitivity analyses incorporating various thresholds of worsening to broadly assess the study population.

In summary, TTE analyses examining ≥ 6-point changes and ≥ 7-point changes in Part III from several early PD studies suggest using these as primary outcomes in interventional studies is reasonable, and that larger changes tend to carry concomitant alterations in function that could support the clinical meaningfulness of examination-based improvement. As such, showing benefit on a potentially useful combination of phenotype and function seems attainable in a 12-month study design. Analysis of MDS-UPDRS Part II TTE thresholds also suggest realistic sample sizes for changes that may be seen as clinically meaningful. Our data suggest TTE analyses in early PD trials have promise to help conduct longer studies while accounting for the expected arrival of symptomatic treatment.

Declarations

Acknowledgements

The authors are grateful to the participants, their families, investigators, and coordinators from the clinical trials analyzed in this work.

Funding

This work was funded by Abli Pharmaceuticals without a formal grant mechanism, and so has no associated grant number.

Contributions

AM, KK, MW, CM, and CV conceived the study. AM and CV drafted the manuscript. All authors reviewed the manuscript and contributed to the interpretation of the results. All authors read and approved the final manuscript.

References

1. Le-Rademacher J, Wang X. Time-To-Event Data: An Overview and Analysis Considerations. J Thorac Oncol. 2021 Jul;16(7):1067-1074. doi: 10.1016/j.jtho.2021.04.004. Epub 2021 Apr 19. PMID: 33887465.
2. Wang Y, Fuerte-Hortigon A, Chetty S, Garah M, Sarmiento R, Barber PA, Saposnik G, Lee DS, Singnurkar A, Swartz RH, Varickanickal J, Field TS. Types and Rates of Major Adverse Cardiovascular Events in Antithrombotic Trials: A Systematic Review and Meta-Analysis. Can J Cardiol. 2025 Dec;41(12):2629-2643. doi: 10.1016/j.cjca.2025.08.358. Epub 2025 Oct 25. PMID: 41137817.
3. Fahn S, Elton RL, UPDRS Program Members. Unified Parkinson's disease rating scale. In: Fahn S, Marsden CD, Goldstein M, Calne DB, editors. Recent developments in Parkinson's disease, Vol. 2. Florham Park, NJ: Macmillan Healthcare Information; 1987. p 153–163, 293–304.
4. Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, Poewe W, Sampaio C, Stern MB, Dodel R, Dubois B, Holloway R, Jankovic J, Kulisevsky J, Lang AE, Lees A, Leurgans S, LeWitt PA, Nyenhuis D, Olanow CW, Rascol O, Schrag A, Teresi JA, van Hilten JJ, LaPelle N; Movement Disorder Society UPDRS Revision Task Force. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. Mov Disord. 2008 Nov 15;23(15):2129-70. doi: 10.1002/mds.22340. PMID: 19025984.
5. https://www.accessdata.fda.gov/drugsatfda_docs/label/2008/020667s014s017s018lbl.pdf. Accessed January 19, 2026.
6. https://www.accessdata.fda.gov/drugsatfda_docs/label/2008/020658s018s020s021lbl.pdf Accessed January 19, 2026.
7. https://www.accessdata.fda.gov/drugsatfda_docs/label/2009/021485s008,%20s012lbl.pdf Accessed January 19, 2026.
8. https://www.accessdata.fda.gov/drugsatfda_docs/label/2017/201971orig1s000lbl.pdf Accessed January 19, 2026.
9. https://www.accessdata.fda.gov/drugsatfda_docs/label/2018/209184s000lbl.pdf Accessed January 19, 2026.
10. Mari Z, Mestre TA. The Disease Modification Conundrum in Parkinson's Disease: Failures and Hopes. Front Aging Neurosci. 2022 Feb 28;14:810860. doi: 10.3389/fnagi.2022.810860. PMID: 35296034; PMCID: PMC8920063.
11. https://clinicaltrials.gov/study/NCT06976268?term=biohaven&cond=Parkinsons%20Disease&rank=1 Accessed January 19, 2026.
12. https://clinicaltrials.gov/study/NCT04777331?term=roche&cond=Parkinsons%20Disease&rank=1 Accessed January 19, 2026.
13. https://clinicaltrials.gov/study/NCT05348785?term=luma&cond=Parkinsons%20Disease&rank=1 Accessed January 19, 2026.
14. Parkinson Study Group SURE-PD3 Investigators; Schwarzschild MA, Ascherio A, Casaceli C, Curhan GC, Fitzgerald R, Kamp C, Lungu C, Macklin EA, Marek K, Mozaffarian D, Oakes D, Rudolph A, Shoulson I, Videnovic A, Scott B, Gauger L, Aldred J, Bixby M, Ciccarello J, Gunzler SA, Henchcliffe C, Brodsky M, Keith K, Hauser RA, Goetz C, LeDoux MS, Hinson V, Kumar R, Espay AJ, Jimenez-Shahed J, Hunter C, Christine C, Daley A, Leehey M, de Marcaida JA, Friedman JH, Hung A, Bwala G, Litvan I, Simon DK, Simuni T, Poon C, Schiess MC, Chou K, Park A, Bhatti D, Peterson C, Criswell SR, Rosenthal L, Durphy J, Shill HA, Mehta SH, Ahmed A, Deik AF, Fang JY, Stover N, Zhang L, Dewey RB Jr, Gerald A, Boyd JT, Houston E, Suski V, Mosovsky S, Cloud L, Shah BB, Saint-Hilaire M, James R, Zauber SE, Reich S, Shprecher D, Pahwa R, Langhammer A, LaFaver K, LeWitt PA, Kaminski P, Goudreau J, Russell D, Houghton DJ, Laroche A, Thomas K, McGraw M, Mari Z, Serrano C, Blindauer K, Rabin M, Kurlan R, Morgan JC, Soileau M, Ainslie M, Bodis-Wollner I, Schneider RB, Waters C, Ratel AS, Beck CA, Bolger P, Callahan KF, Crotty GF, Klements D, Kostrzebski M, McMahon GM, Pothier L, Waikar SS, Lang A, Mestre T. Effect of Urate-Elevating Inosine on Early Parkinson Disease Progression: The SURE-PD3 Randomized Clinical Trial. JAMA. 2021 Sep 14;326(10):926-939. doi: 10.1001/jama.2021.10207. PMID: 34519802; PMCID: PMC8441591.
15. Parkinson Study Group STEADY-PD III Investigators. Isradipine Versus Placebo in Early Parkinson Disease: A Randomized Trial. Ann Intern Med. 2020 May 5;172(9):591-598. doi: 10.7326/M19-2534. Epub 2020 Mar 31. PMID: 32227247; PMCID: PMC7465126.
16. NINDS NET-PD Investigators. A randomized, double-blind, futility clinical trial of creatine and minocycline in early Parkinson disease. Neurology. 2006 Mar 14;66(5):664-71. doi: 10.1212/01.wnl.0000201252.57661.e1. Epub 2006 Feb 15. PMID: 16481597.
17. NINDS NET-PD Investigators. A randomized clinical trial of coenzyme Q10 and GPI-1485 in early Parkinson disease. Neurology. 2007 Jan 2;68(1):20-8. doi: 10.1212/01.wnl.0000250355.28474.8e. PMID: 17200487.
18. Parkinson Progression Marker Initiative. The Parkinson Progression Marker Initiative (PPMI). Prog Neurobiol. 2011 Dec;95(4):629-35. doi: 10.1016/j.pneurobio.2011.09.005. Epub 2011 Sep 14. PMID: 21930184; PMCID: PMC9014725.

19. van Wamelen DJ, Urso D, Ray Chaudhuri K. How Time Rules: Diurnal Motor Patterns in de novo Parkinson's Disease. J Parkinsons Dis. 2021;11(2):695-702. doi: 10.3233/JPD-202352. PMID: 33523019.
20. Evers LJW, Krijthe JH, Meinders MJ, Bloem BR, Heskes TM. Measuring Parkinson's disease over time: The real-world within-subject reliability of the MDS-UPDRS. Mov Disord. 2019 Oct;34(10):1480-1487. doi: 10.1002/mds.27790. Epub 2019 Jul 10. PMID: 31291488; PMCID: PMC6851993.
21. Janssen Daalen J, van der Heiden M, Meinders M, Post B. Motor Symptom Variability in Parkinson's Disease: Implications for Personalized Trial Outcomes? Mov Disord. 2025 May;40(5):975-979. doi: 10.1002/mds.30133. Epub 2025 Feb 5. PMID: 39907497; PMCID: PMC12089913.

Tables

Table 1. *Baseline characteristics of participants from each study

	STEADY-PD III (n=166)	**FS-1** (n=67)	**FS-TOO** (n=71)	**SURE-PD3** (n=149)	**PPMI** (n=406)
Age, years	61.6 (9.3)	60.7 (9.8)	60.1 (10.6)	63.6 (9.4)	61.9 (9.5)
Sex, n (%)					
Female	58 (34.9%)	27 (40.3%)	21 (29.6%)	67 (45.0%)	140 (34.5%)
Male	108 (65.1%)	40 (59.7%)	50 (70.4%)	82 (55.0%)	266 (65.5%)
Race, n (%)					
American Indian/Alaska	1 (0.6%)	0	1 (1.4%)	0	1 (0.2%)
Asian	5 (3.0%)	1 (1.5%)	2 (2.8%)	1 (0.7%)	8 (2.0%)
Black/African American	6 (3.6%)	2 (3.0%)	4 (5.6%)	0	7 (1.7%)
Native Hawaiian/Pacific Island	1 (1.2%)	0	0	0	0
White	148 (89.2%)	63 (94.0%)	64 (90.1%)	145 (97.3%)	378 (93.1%)
More than one race specified	2 (1.2%)	1 (1.5%)	0	1 (0.7%)	10 (2.5%)
Other	2 (1.2%)	0	0	0	1 (0.2%)
Unknown/Not Reported	1 (1.2%)	0	0	2 (1.3%)	1 (0.2%)
Ethnicity, n (%)					
Hispanic/Latino	6 (3.6%)	2 (3.0%)	2 (2.8%)	4 (2.7%)	9 (2.2%)
Not Hispanic/Latino	160 (96.4%)	65 (97.0%)	69 (97.2%)	143 (96.0%)	397 (97.8%)
Unknown/Not Reported	0	0	0	2 (1.3%)	0
Years Since PD Diagnosis	0.82 (0.78)	0.64 (0.59)	0.70 (0.90)	0.71 (0.62)	0.56 (0.54)
(MDS-)UPDRS Part II[1]	5.4 (3.3)	6.0 (3.5)	5.7 (3.3)	6.1 (4.4)	5.8 (4.3)
(MDS-)UPDRS Part III[2]	16.3 (6.5)	15.6 (7.0)	16.1 (6.9)	22.5 (9.2)	20.7 (9.3)
Symptomatic Medication, n (%)	14 (8.4%)	NA	NA	62 (41.6%)	NA

*Presented as mean (standard deviation) unless otherwise specified

[1]UPDRS used for STEADY-PD III, FS-1, and FS-TOO

[2]MDS-UPDRS used for SURE-PD3 and PPMI

Table 2. Participants Reaching (MDS-)UPDRS Part III Threshold Changes or Treatment Initiation for each Study

Scale	Study	Part III Threshold			
		7-point Composite, n (%)	6-point Composite, n (%)	5-point Composite, n (%)	4-point Composite, n (%)
		(7-point / Tx)	(6-point / Tx)	(5-point / Tx)	(4-point / Tx)
UPDRS Part III	**STEADY-PD III**	129 (77.7%)	131 (78.9%)	138 (83.1%)	143 (86.1%)
	Placebo (n=166)	(36.1% / 41.6%)	(46.4% / 32.5%)	(54.2% / 28.9%)	(62.0% / 24.1%)
	FS-1	44 (65.7%)	45 (67.2%)	46 (68.6%)	52 (77.6%)
	Placebo (n=67)	(47.8% / 17.9%)	(49.3% / 17.9%)	(55.2% / 13.4%)	(67.2% / 10.4%)
	FS-TOO	40 (56.3%)	44 (62.0%)	49 (69.0%)	55 (77.5%)
	Placebo (n=71)	(35.2% / 21.1%)	(42.3% / 19.7%)	(52.1% / 16.9%)	(62.0% / 15.5%)
MDS-UPDRS Part III	**SURE-PD 3**	118 (79.2%)	124 (83.2%)	127 (85.2%)	133 (89.3%)
	Placebo (n=149)	(46.3% / 32.9%)	(57.0% / 26.2%)	(61.1% / 24.2%)	(66.4% / 22.8%)
	PPMI	328 (80.8%)	335 (82.5%)	349 (86.0%)	362 (89.2%)
	Sporadic PD (n=406)	(47.5% / 33.3%)	(52.5% / 30.0%)	(59.9% / 26.1%)	(67.2% / 21.9%)

Tx = initiation of symptomatic treatment

Table 3. Participants Reaching (MDS-)UPDRS Part II Threshold Changes or Treatment Initiation for each Study

Scale	Study	Part II Threshold		
		4-point Composite, n (%)	3-point Composite, n (%)	2-point Composite, n (%)
		(4-point / Tx)	(3-point / Tx)	(2-point / Tx)
UPDRS Part II	**STEADY-PD III**	128 (77.1%)	138 (83.1%)	149 (89.8%)
	Placebo (n=166)	(37.9% / 39.2%)	(48.8% / 34.3%)	(65.7% / 24.1%)
	FS-1	41 (61.2%)	45 (67.2%)	51 (76.1%)
	Placebo (n=67)	(32.8% / 28.4%)	(40.3% / 26.9%)	(56.7% / 19.4%)
	FS-TOO	42 (59.1%)	49 (69.0%)	54 (76.1%)
	Placebo (n=71)	(39.4% / 19.7%)	(53.5% / 15.5%)	(63.4% / 12.7%)
MDS-UPDRS Part II	**SURE-PD 3**	97 (65.1%)	110 (73.8%)	126 (84.6%)
	Placebo (n=149)	(26.8% / 38.3%)	(40.9% / 32.9%)	(59.7% / 24.8%)
	PPMI	320 (78.8%)	334 (82.3%)	355 (87.4%)
	Sporadic PD (n=406)	(45.3% / 33.5%)	(54.2% / 28.1%)	(66.0% / 21.4%)

Tx = initiation of symptomatic treatment

Table 4A. Total sample sizes needed for (MDS-)UPDRS part III thresholds with various dropout rates (80% Power)

Study	Part III Threshold	Total Sample Size (No Dropout)	Total Sample Size (5% Dropout)	Total Sample Size (10% Dropout)	Total Sample Size (15% Dropout)	Total Sample Size (20% Dropout)
STEADY-PD III	7-point	348	368	388	410	436
STEADY-PD III	6-point	342	360	380	404	428
STEADY-PD III	5-point	322	340	358	380	404
STEADY-PD III	4-point	308	326	344	364	386
FS-1	7-point	418	440	466	492	524
FS-1	6-point	408	430	454	480	510
FS-1	5-point	398	420	444	470	498
FS-1	4-point	348	368	388	410	436
FS-TOO	7-point	494	520	550	582	618
FS-TOO	6-point	446	470	496	526	558
FS-TOO	5-point	396	418	440	466	496
FS-TOO	4-point	348	368	388	410	436
SURE-PD 3	7-point	340	358	378	400	426
SURE-PD 3	6-point	320	338	356	378	400
SURE-PD 3	5-point	312	330	348	368	390
SURE-PD 3	4-point	294	310	328	346	368
PPMI Sporadic PD	7-point	332	350	370	392	416
PPMI Sporadic PD	6-point	324	342	360	382	406
PPMI Sporadic PD	5-point	308	326	344	364	386
PPMI Sporadic PD	4-point	294	310	328	346	368

Sample sizes shown are for 1:1 active: placebo randomization, two-sided alpha 0.05, 80% power, and 30% slowing for treatment effect

Table 4B. Total sample sizes needed for (MDS-)UPDRS part II thresholds with various dropout rates (80% Power)

Study	Part II Threshold	Total Sample Size (No Dropout)	Total Sample Size (5% Dropout)	Total Sample Size (10% Dropout)	Total Sample Size (15% Dropout)	Total Sample Size (20% Dropout)
STEADY-PD III	4-point	350	370	390	412	438
STEADY-PD III	3-point	322	340	358	380	404
STEADY-PD III	2-point	292	308	326	344	366
FS-1	4-point	450	474	500	530	564
FS-1	3-point	408	430	454	480	510
FS-1	2-point	356	376	396	420	446
FS-TOO	4-point	468	494	520	552	586
FS-TOO	3-point	396	418	440	466	496
FS-TOO	2-point	356	376	396	420	446
SURE-PD 3	4-point	422	446	470	498	528
SURE-PD 3	3-point	368	388	410	434	460
SURE-PD 3	2-point	314	332	350	370	394
PPMI Sporadic PD	4-point	342	360	380	404	428
PPMI Sporadic PD	3-point	324	342	360	382	406
PPMI Sporadic PD	2-point	302	318	336	356	378

Figures

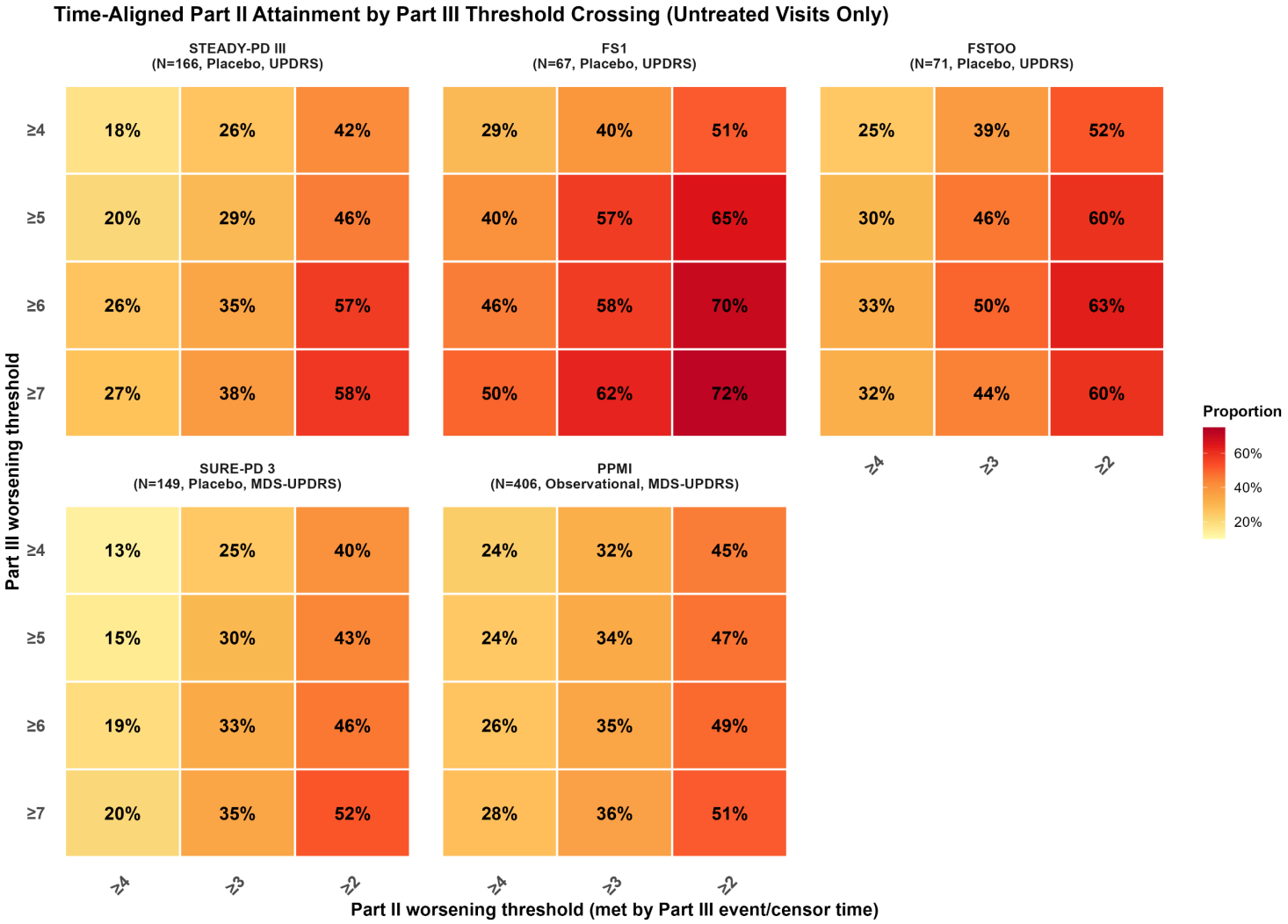


Figure 1

Heat maps showing time-aligned concordance between Part II and Part III worsening across five cohorts for untreated analysis. Colors indicate the proportion of participants meeting each Part II threshold by the time of Part III threshold crossing. Numbers show percentages.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- PDProgressionsupplementalfinal6.12.2026.docx