

DSTs: Part 1

Delayed-Start Trials (Part 1): Design and Interpretation

The next major goal in neurodegenerative pharmacotherapeutics is to discover and clearly identify therapies that have disease-modifying effects on Parkinson's disease (PD) and Alzheimer's disease (AD). In the absence of a biological marker, the US Food and Drug Administration (FDA) has suggested that a delayed-start trial (DST) is the type of clinical study most likely to satisfactorily demonstrate disease modification. Not only are DSTs inherently complex, but their designs often incorporate extra caveats (eg, hierarchical endpoints) in attempts to prove additional hypotheses. In view of the central role that DSTs will be playing in the evaluation of therapeutics for neurodegenerative diseases, an understanding of their design is of particular importance. The methods used in DSTs, along with approaches to interpreting their results, are reviewed in this article. In the next issue of *Neura*, the use of DSTs in the assessment of therapies for neurodegenerative diseases will be further explored by examining previous trials (such as PRISMS and ADAGIO), with a glimpse into the ways in which the designs of such studies may be improved upon in the future.

Numerous agents have been found to exhibit neuroprotective effects in laboratory models of neurodegenerative diseases, but demonstrating efficacy in the clinical setting has proved to be a major challenge.¹ Inconclusive, albeit encouraging, results have emerged from phase III placebo-controlled, head-to-head, and combination treatment trials of putative disease-modifying therapies.² Such traditional trial designs are limited by several factors, such as the need to specify a clinical endpoint that is truly capable of measuring disease progression without being confounded by the symptomatic effects of the study treatment.¹ Another drawback is the need for a prolonged study duration, as the neuroprotective effects of a treatment cannot realistically be expected to manifest rapidly in slowly progressive diseases such as PD and AD.² The duration may be shortened somewhat if the statistical power to detect small changes in clinical presentation is increased through measures such as enrolling a greater number of patients, but recruitment problems may undermine the feasibility of this objective.² The delayed-start trial's design may offer a way of overcoming these obstacles.

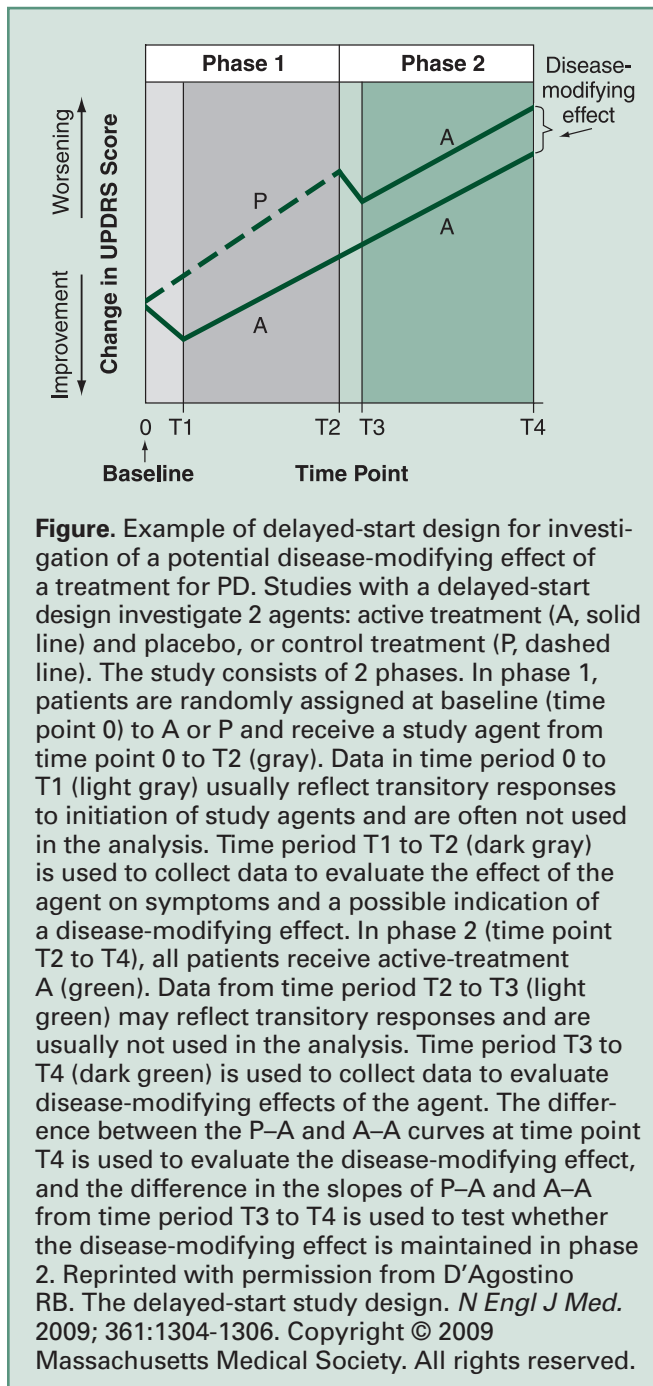
DST Design

First described in the mid-1990s, the DST consists of 2 phases.³ An example of a DST applied to the assessment of therapy for PD is shown in the Figure.⁴ In phase 1 (time point 1 to time point 2), patients are randomized to either the study drug (termed an “early start”) or placebo for a fixed length of time.⁴ The duration of phase 1 is chosen such that sufficient time is allowed for the effects of treatment on disease symptoms to become apparent.⁴ Significant differences between treatment groups at the conclusion of this phase may be due to symptomatic effects, disease-modifying effects, or both.^{1,3} In phase 2, patients in the

active treatment group continue to receive the study drug, and patients in the placebo group switch to this treatment (termed a “delayed start”).⁴ If the difference observed at the end of phase 1 disappears at the end of phase 2, this indicates that the drug has a solely symptomatic effect.^{2,5,6,7} Conversely, a sustained difference at the end of phase 2 argues in favor of a neuroprotective effect, provided both phase 1 and phase 2 are of sufficient duration.^{2,3,4,5,7} In this way, any symptomatic effects of a true disease-modifying therapy can be separated from its neuroprotective effects.⁵

In a recent review of DST methods, D'Agostino noted that the duration of the phases is just one aspect that must be considered in the design.⁴ Another issue concerns how much data should be entered into the analysis. As shown in the Figure, the typical DST design incorporates a lapse of time between baseline (time point 0, when treatment is initiated) and the official beginning of phase 1 (time point 1).⁴ Data from this period reflect temporary changes at the initiation of treatment and are not usually entered into the data analysis. Similarly, a transitory period is built into the design between the end of phase 1 (time point 2) and the beginning of phase 2 (time point 3). Changes during this period may reflect transitory responses and are not usually entered into the analysis. Decisions must be made regarding the appropriate durations of these so-called “data-not-used zones.”

An additional decision relates to the number of repeated outcome measurements taken during phase 1. Trial design must ensure that the statistical power (see Glossary) be sufficient to demonstrate differences in the slope of change in the early-start group vs the placebo group, should such a distinction actually exist.⁴ Likewise, the amount of data collected during phase 2 must provide sufficient power to demonstrate that the slopes of change in the delayed-start group and the early-start group do not



converge. The maintenance of a difference between the slopes would confirm that the disease-modifying effects of the active treatment persisted and did not decrease over the course of a given study.

DST Data Interpretation

D’Agostino points out that the statistical analyses to be used must be identified *a priori*.⁴ A minimum of 3 statistical tests should be performed. During phase 1, a test must be selected that will demonstrate a difference between the

early-start group and placebo group with respect to effects on symptoms (this is not necessarily a disease-modifying effect). In addition, the analytic method must incorporate an appropriate alpha error (for “false positives”) and adjustment for multiple testing. At least 2 statistical tests must be used for the analysis of data collected during phase 2. To confirm that any benefit of active treatment observed at the end of phase 1 persisted at the end of phase 2, a test must be performed to examine whether less worsening was apparent in the early-treatment group than in the delayed-treatment group during phase 2. A second test should be conducted to determine whether the effects in the early-treatment group and delayed-start group did or did not converge during phase 2, as mentioned above; this analysis is needed to determine whether the clinical effects of the treatment persisted and did not diminish over the time covered by the study. The differences in the slopes of the outcome data between groups can be tested via repeated-measures or random-effects statistical analyses. A noninferiority test of the slopes, with a preselected noninferiority margin, is the suggested approach for determining whether the difference between the curves remains constant. A constant difference between the curves indicates a fixed effect of the treatment on the underlying mechanism of disease (eg, disease modification).

If data are missing in phase 1, a test such as the mixed-effects model for repeated measures will likely be adequate for analyzing the data.⁴ Use of the last-observation-carried-forward approach, on the other hand, may be problematic. More data are apt to be missing in phase 2 because many patients randomized to placebo will probably drop out during phase 1 due to lack of treatment effect. One suggested solution is to design the study so that patients in the placebo group who want to drop out of phase 1 are permitted to proceed directly to phase 2.

The DST design described here assumes that the data obtained during the study will fit into linear models.⁴ Observers have pointed out that this assumption is probably valid when studying treatments for diseases that develop slowly during the early stages (such as PD), but it should nonetheless be specifically addressed when designing a DST.

Drawbacks of DSTs

The DST design is subject to several potential flaws and drawbacks that may affect the quality of the trial results. For instance, DSTs must enroll patients who can forego symptomatic therapy for the first half of the study, should they be randomized to placebo (ie, the delayed-start group). In the case of DSTs for PD, the length of phase 1 is usually 6 to 9 months.⁵ In light of the fact that a large percentage of individuals with PD already require symptomatic treatment

by the time they are diagnosed, the pool of potential patients available for recruitment into a DST may be limited.⁵ Moreover, suitable candidates for the study may very likely have more slowly progressive disease, which would reduce the generalizability of the results.⁵

Of particular interest is a hypothesis put forward by Schapira and Obeso that could complicate interpretation of DST trials in early PD.^{5,8} Specifically, these authors have proposed that early correction of basal ganglia functional abnormalities caused by dopaminergic cell loss and dopamine deficiency can support beneficial intrinsic compensatory physiologic responses and delay disease progression.⁸ Early symptomatic treatment may confer long-term benefits by influencing these mechanisms.^{1,5,8} If this hypothesis proves true, then any treatment capable of reducing symptoms in a DST involving patients with early untreated PD will have an apparent disease-modifying effect as well.⁵ Until the hypothesis can be confirmed or refuted, it will not be possible to accurately interpret the results of delayed-start treatment trials in patients at this stage of PD.⁵

The likelihood of missing data also constitutes a major concern in DSTs, particularly with regard to phase 2.⁴ The high probability of discontinuations due to lack of treatment effect among placebo recipients during phase 1 might cause an imbalance in the baseline characteristics of the remaining patients who enter phase 2—in effect, negating the randomization and undermining the integrity of the results.⁴ These potential problems must be anticipated in advance, so the study can be designed to accommodate differences in drop-out rates between treatment groups and to provide methods to offset these disparities.⁴ As mentioned, some observers have suggested handling this situation by allowing phase 1 dropouts to proceed immediately to phase 2.⁴

A further consideration is the fact that DSTs are powered to detect small differences in outcome measures (for instance, the Unified Parkinson's Disease Rating Scale [UPDRS]) that may be statistically significant but not necessarily clinically meaningful.^{2,5} It is quite likely that, following a DST showing a significant impact of a neuroprotective agent on an endpoint such as the UPDRS, regulatory authorities will require larger, longer-term, double-blind, parallel-group trials to confirm the clinical significance of such findings.^{2,5}

Conclusion

The ability to identify neuroprotective treatments for disorders that progress slowly but lead to marked disability (such as PD and AD) is imperative. Until recently, such efforts have been hampered by the shortcomings of tradi-

tional study designs, which fail to separate the symptomatic effects of a treatment from its disease-modifying effects. Although several aspects remain to be refined, the delayed-start design offers a promising method of evaluating potential disease-modifying therapies in the future.

Glossary of Statistical Terms

Alpha error: A statistical error made in testing a hypothesis when it is concluded that a result is positive when it really is not; often referred to as a “false positive.”

Last observation carried forward: A statistical method by which a dataset can be analyzed even though values are missing due to drop-outs; the last observation prior to drop-out is treated as though it were the observation at the last scheduled visit.

Mixed-effects model: A statistical model describing a situation in which both fixed and random effects are present (fixed-effects models assume that the data came from normal populations and may differ only with respect to means, whereas random-effects models assume that the data describe a hierarchy of different populations).

Noninferiority test: A method of statistical analysis used in clinical trials to determine whether a new treatment is equivalent to standard treatment.

Statistical power: The probability that a statistical test will reject the null hypothesis when the alternative hypothesis is true.

Slope: Rate of change as reflected in a curve derived by plotting data.

References

1. Olanow CW, Rascol O, Hauser R, et al. A double-blind, delayed-start trial of rasagiline in Parkinson's disease. *N Engl J Med*. 2009;361:1268-1278.
2. Rascol O. “Disease-modification” trials in Parkinson disease: target populations, endpoints and study design. *Neurology*. 2009;72(suppl 2):S51-S58.
3. Leber P. Observations and suggestions on antideementia drug development. *Alzheimer Dis Assoc Disord*. 1996;10(suppl 1):31-35.
4. D'Agostino RB. The delayed-start study design. *N Engl J Med*. 2009;361:1304-1306.
5. Clarke CE. Are delayed-start design trials to show neuroprotection in Parkinson's disease fundamentally flawed? *Mov Disord*. 2008;23:784-789.
6. Disease Progression Sub-Group. Protocols to demonstrate slowing of Alzheimer disease progression. Position paper from the International Working Group on Harmonization of Dementia Drug Guidelines. *Alzheimer Dis Assoc Disord*. 1997;11(suppl 3):50-53.
7. Kieburtz K. Issues in neuroprotection clinical trials in Parkinson's disease. *Neurology*. 2006;66(suppl):S50-S57.
8. Schapira AHV, Obeso J. Timing of treatment initiation in Parkinson's disease: a need for reappraisal? *Ann Neurol*. 2006;59:559-562.