HIGH LIGHT
CME

Optimizing the ongoing search for new treatments for Parkinson disease

Using futility designs

B.C. Tilley, PhD; Y.Y. Palesch, PhD; K. Kieburtz, MD; B. Ravina, MD; P. Huang, PhD; J.J. Elm, MA; K. Shannon, MD; G.F. Wooten, MD; C.M. Tanner, MD, PhD; and G.C. Goetz, MD, on behalf of the NET-PD Investigators*

Abstract—Many agents are being considered for treatment of Parkinson disease (PD). Given the large number of agents and the limited resources to evaluate new agents, it is essential to reduce the likelihood of advancing ineffective agents into large, long-term Phase III trials. Futility design methodology addresses this goal. The authors describe how a single-arm Phase II futility study uses a short-term outcome to compare a treatment group response to a predetermined hypothesized or historically based control response. The authors present advantages and limitations of futility designs along with examples derived from the data archive of a large Phase III efficacy study of treatments to delay PD progression, the Deprenyl And Tocopherol Antioxidative Therapy Of Parkinsonism (DATATOP) trial. Using the same control progression rate and treatment effect assumptions used to power the original DATATOP trial, the authors calculated the number of subjects needed to conduct two 12-month futility studies. DATATOP was designed to enroll 800 patients. Using data on 124 consecutive subjects randomized into each of the DATATOP treatment groups, the authors identified tocopherol as futile and deprenyl as worthy of further study. Using Phase II information, DATATOP could have been simplified from a 2 × 2 factorial design to a comparison of deprenyl vs placebo. While not testing efficacy, futility designs provide a strategy for discarding treatments unlikely to be effective in Phase III. A limitation is the dependence on historical data or hypothesized outcomes for untreated controls. Futility studies may decrease the time to identify treatments unworthy of further pursuit and reduce subjects' exposure to futile treatments.

NEUROLOGY 2006;66:628–633

Over the last 15 years many drugs have been studied as potential therapeutic agents for Parkinson disease (PD), and thousands of patients have been enrolled in PD trials. While such agents as rasagiline[1,2] and coenzyme Q[3] have shown promise, no therapies have been proven to modify PD progression.[4] Given the resources required for traditional Phase III efficacy studies and the large number of agents available for study in PD, there is a need to select promising agents efficiently. Single-arm Phase II studies address this need.[5-13]

Phase II studies generally follow Phase I toxicity studies, and provide some additional safety information prior to proceeding to Phase III efficacy studies. An alternative combines Phase I and Phase II into a single design.[12] Phase II studies may focus on identifying agents with potential efficacy for testing in randomized Phase III studies[7-13] or on discarding drugs without promise.[5,6] These latter studies, termed "futility studies" because they identify agents that are "futile" for further development, are the focus of this article.

Futility study methodology, often used in evaluation of cancer treatments,[5-7,14] has been extended to

Editorial, see page 626
See also pages 660 and 664

*The NET-PD Investigators are listed in the Appendix.

From the Department of Biostatistics, Bioinformatics & Epidemiology (B.C.T., Y.Y.P., P.H., J.J.E.), Medical University of South Carolina, Charleston; Department of Neurology (K.K.), University of Rochester, NY; NINDS (B.R.), National Institutes of Health, Bethesda, MD; Department of Neurological Sciences (K.S., G.C.G.), Rush-Presbyterian-St. Luke's Medical Center, Chicago, IL; Department of Neurology (G.F.W.), University of Virginia Health Sciences Center, Charlottesville; and The Parkinson's Institute (C.M.T.), Sunnyvale, CA.

Supported by NIH/NINDS grants U01NS043127 and U01NS43128.

Disclosure: The authors report no conflicts of interest.

Received April 13, 2005. Accepted in final form November 28, 2005.

Address correspondence and reprint requests to Dr. Barbara C. Tilley, Department of Biostatistics, Bioinformatics & Epidemiology, Medical University of South Carolina, P.O. Box 250835, 135 Cannon Street, Suite 303, Charleston, SC 29425; e-mail: tilleybc@musc.edu

Table 1 *Effect of changes in assumptions on sample size for Phase II and Phase III trials*

Historical placebo outcome (p*)	Hypothesized minimal treatment benefit (Δ)	Hypothesized treatment outcome ($p_{tx} = p^* - \Delta$)	Alpha (α)	Power ($1 - \beta$)	Total required sample size for Phase II†	Minimum no. of failures required in Phase II to reject the null hypothesis	Total required sample size for phase III‡
0.425	0.1	0.325	0.1	0.85	124	47	336
0.425	0.1	0.325	0.05	0.85	165	64	420
0.425	0.1	0.325	0.1	0.90	153	58	400
0.425	0.1	0.325	0.05	0.90	198	76	491
0.425	0.2	0.225	0.1	0.85	28	10	78
0.425	0.2	0.225	0.05	0.85	36	13	97
0.425	0.2	0.225	0.1	0.90	35	12	93
0.425	0.2	0.225	0.05	0.90	44	15	114
0.425	0.05	0.375	0.1	0.85	514	207	1,379
0.425	0.05	0.375	0.05	0.85	686	279	1,723
0.425	0.05	0.375	0.1	0.90	629	252	1,643
0.425	0.05	0.375	0.05	0.90	818	330	2,016

† Sample size for test that a proportion equals null value (normal approximation) in a single arm Phase II trial.

$$n = \frac{\left(Z_{1-\alpha}\sqrt{p_{tx}(1-p_{tx})} + Z_{1-\beta}\sqrt{p^*(1-p^*)}\right)^2}{(p_{tx}-p^*)^2}$$

‡ Sample size for two-sided chi-square test of proportions in a two arm Phase III trial (placebo vs treatment).

$$n = \frac{\left[Z_{1-\alpha/2}\sqrt{2\bar{p}(1-\bar{p})} + Z_{1-\beta}\sqrt{p^*(1-p^*) + p_{tx}(1-p_{tx})}\right]^2}{(p^*-p_{tx})^2}, \text{ where } \bar{p} = (p^* + p_{tx})/2$$

stroke trials[15] and applied in a trial of IV/IA rt-PA.[16] More recently Phase II futility studies have been used in evaluating treatments for PD. In hopes of identifying new therapeutic agents to modify PD progression, the National Institute of Neurologic Disorders and Stroke (NINDS) is sponsoring a series of Phase II and Phase III trials, i.e., Neuroprotection Exploratory Trials in Parkinson's Disease (NET-PD). This article explains the application of futility designs to PD trials. To illustrate the differences between futility and efficacy studies, we provide examples using data from the Deprenyl And Tocopherol Antioxidative Therapy Of Parkinsonism (DATATOP) trial.[17,18]

Futility design: Conceptual framework. A futility study compares the outcome in a single, treated arm against a predetermined threshold value reflective of a clinically meaningful change observable over a relatively short period of time. For example, the observed outcome in a futility trial may be defined as the proportion of treated patients who fail treatment (p_{tx}); for PD trials, failure is often defined as those patients who require levodopa. The proportion of patients expected to fail in the untreated group is defined as p*. This p* is usually obtained from historical data, but can be based on an investigator's best judgment or the consensus of clinical experts. The reduction in failures considered clinically meaningful, defined as (Δ), is set before Phase II begins, and may be derived from the same sources as p*. If the observed proportion of failures in the treatment group is greater than the predefined threshold (p* - Δ), based on a statistical test, then the treatment would be considered futile to evaluate further for efficacy. Both p* and Δ are used to estimate sample size for Phase II (table 1). The approach used is similar to that used in developing sample size estimates for Phase III studies. For a continuous outcome such as change in Unified Parkinson's Disease Rating Scale (UPDRS), the threshold can be the hypothesized mean change in the historical control (p*) group reduced (lower score is better) by the expected treatment effect (Δ). Determination of the threshold value is important as too small a Δ for a given control value, p*, might allow an ineffective treatment to be carried on to Phase III, whereas too large a Δ could exclude drugs that might demonstrate efficacy in Phase III.

The hypotheses being tested in a single-arm Phase II futility study with a binary outcome are as follows:

$$H_0: p_{tx} \leq p^* - \Delta \text{ vs } H_A: p_{tx} > p^* - \Delta$$

If we reject the null hypothesis, we conclude that the proportion of patients on treatment (p_{tx}) who fail is greater than the predetermined threshold ($p^* - \Delta$), and it is futile to proceed to a Phase III efficacy trial. Conversely, if we fail to reject the null hypothesis, we did not observe enough failures to conclude futility, and we would consider further testing of the treatment in a Phase III efficacy trial. Regardless of

Table 2 *Interpretation of alpha and beta under Phase II futility and Phase III efficacy studies**

	Type I error (alpha)	Type II error (beta)
Phase II futility study (tentative result)	Recommending an effective treatment should not move forward	Recommending an ineffective treatment should move forward
Traditional Phase III efficacy trial (definitive result)	Declaring an ineffective treatment to be effective	Declaring an effective treatment to be ineffective

* Adapted from reference 15.

the results of the Phase II study, we cannot conclude we have demonstrated efficacy.

The futility hypotheses differ from hypotheses specified in traditional Phase III trials of efficacy, and Type I (alpha) and Type II (beta) error probabilities are interpreted differently. Table 2 compares the interpretation of these error probabilities in Phase II futility study vs Phase III efficacy designs.[15] In Phase II futility trials, alpha and beta are set relative to the questions under investigation. We do not want to miss an effective agent; consequently we are less concerned about failing to reject the null hypothesis (falsely concluding an ineffective treatment is possibly effective). Additionally the sample size must be kept small. Thus, we set beta (false positive) greater than alpha (false negative) and set alpha at a value greater than 0.05. If we fail to reject the null hypothesis, we may proceed to a Phase III trial of efficacy, generally with smaller error probabilities (values of beta and alpha) and larger sample sizes. In the case of a binary outcome such as starting levodopa (yes, no), the statistical test of the futility hypothesis can be the one-sample test for comparing proportions (normal approximation).[19] For a continuous outcome the test can be a one-sample *t* test.

Ideal criteria for choosing historical controls have been published, and should be considered when choosing Phase II control values.[20,21] If all the criteria are not met or if the historical control response is in doubt, inclusion of a small calibration group[22] of concurrent placebo patients could strengthen the futility design. The calibration group is not utilized for direct comparison between a treatment group and placebo, but provides limited data on the validity of the hypothesized threshold value ($p^* - \Delta$).[22]

Data source. We used data from DATATOP, a Phase III trial conducted by the Parkinson Study Group,[17] to construct a Phase II futility study as an example of futility methodology. DATATOP enrolled patients with recent onset of PD, not requiring levodopa therapy. Patients were randomized to placebo (N = 199), tocopherol + placebo (N = 202), deprenyl + placebo (N = 202), or tocopherol + deprenyl (N = 197), in a 2 × 2 factorial design. Although 24 months follow-up was planned, treatment codes were revealed after an average of 14 months follow-up. In the original study, tocopherol failed to show efficacy whereas deprenyl was considered efficacious. However, interpretation of these results remains controversial.[23]

The following example of a constructed Phase II study uses data from the groups receiving either tocopherol alone or deprenyl alone. The groups receiving only placebo or both tocopherol and deprenyl are excluded as the objective was to test the effect of each treatment separately in a constructed single-arm Phase II study. Comparisons were made to the hypothesized control rate used in the design of the original Phase III trial rather than to the concurrent DATATOP placebo group.

Example 1. For this example we used the DATATOP primary outcome, need to initiate levodopa therapy, in order to determine what would have been concluded had a Phase II study been conducted. As a short term outcome, in order to make such a Phase II futility study practical, we used DATATOP data only through 12 months of follow-up. Treated DATATOP patients who needed levodopa therapy by the end of 12 months follow-up were considered failures; treated patients who continued through 12 months without dopaminergic therapy were considered successes.

To determine our threshold for failure, we utilized the original estimates (based on a university practice) on which DATATOP study investigators developed their Phase III sample size.[24] The DATATOP study investigators assumed the placebo proportion reaching endpoint (onset of the need for levodopa therapy) would be 85% in 2 years.[17] DATATOP investigators assumed an absolute decrease of 10% as the minimally clinically meaningful difference (i.e., 75% of patients in the treatment arm would fail). Assuming the proportion of failures is linear over time, we assumed 42.5% of untreated patients would need levodopa therapy at 1 year. We chose the futility threshold (maximum allowable proportion of failures for an efficacious treatment) to be 32.5%, based on the same minimally clinically meaningful difference and assuming a strong early benefit of treatment. We then created two single-arm Phase II futility studies: in one study all patients received tocopherol and in one study all patients received deprenyl.

We set alpha to 0.10, tolerating a 10% chance of rejecting an effective treatment, and we set beta at 15%, accepting a greater chance of carrying an ineffective treatment forward to Phase III testing. We computed the sample size needed (N) for a Phase II futility study to be 124 (assuming one-sided alpha = 0.1, power = 0.85, historical placebo, $p^* = 0.425$, difference between treatment and placebo, $\Delta = 0.1$).

The formula used to calculate this sample size is given in table 1.

Using the first 124 patients enrolled in the tocopherol (+placebo for deprenyl) arm, the proportion of patients needing levodopa therapy (failures) in the first year was 0.42. Comparing this proportion to the futility threshold ($p^* - \Delta$) of 0.325, the p value is <0.01, implying that tocopherol is futile to take forward into a Phase III trial. If the Δ had been chosen to be 0.15 rather than 0.1, the required sample size would be 53 and the p value would be 0.09. If the Δ had been chosen to be 0.2, the required sample size would be 28 and p value would be 0.005. Both results would imply futility.

Similarly, selecting the first 124 patients enrolled in the deprenyl arm, the proportion of failures is 0.14; compared to the futility threshold of 0.325, the p value is > 0.99. Thus for the deprenyl arm, we fail to reject the null hypothesis and cannot declare deprenyl futile. Based on this analysis, a Phase III trial could be conducted to assess efficacy.

Example 2. Table 1 shows the sample size requirements for a variety of scenarios, varying alpha, power, and the hypothesized treatment benefit. We also compare the sample size required for a single-arm Phase II study to a Phase III study with a treatment and control arm. For a Phase II study, using alpha = 0.05 and the same power, placebo proportion of failures, and threshold as used in Example 1, the sample size increases from 124 to 165 (33%). A Phase III trial with power = 85% and the traditional alpha = 0.05 requires a sample size of 420 per group or 840 total.

Discussion. Because DATATOP used a factorial design, the investigators were able to test both deprenyl and tocopherol using just 800 patients, approximately the same number that would be needed to compare just one treatment to placebo. In the factorial design DATATOP investigators assumed no interaction between deprenyl and tocopherol (i.e., no synergy and no detrimental effect of receiving the two treatments together). If this interaction had been present, the study would have been underpowered as only the arms with a drug alone vs placebo alone could have been used to test treatment effects. By using the Phase II design, the elimination of tocopherol could have simplified the final study design to a two-group comparison. While the Phase II sample size was large (N = 124) this is a reduction from the 400 needed to compare either tocopherol or deprenyl directly with concurrent placebo. In the Phase II design the tocopherol group would have been followed only 12 months at a decreased cost to the study.

If we had failed to reject the null hypothesis for both agents, there would have been an increased cost of the additional two groups of 124 patients followed for 1 year in addition to the 800 patients required for the factorial design. However, other gains, such as more precision in sample size estimation for Phase III, might have justified this extra expense. The primary advantage of the futility design is the rejection of agents without promise, and the reduction in the number of subjects exposed to treatment with an agent that has no reasonable hope of modifying disease progression. Recent reports on high doses of tocopherol emphasize the importance of limiting seemingly innocuous exposure.[25,26]

The proposed futility design would not have detected synergy, particularly a situation where we reject both drugs but would not have rejected the combination. If synergy is hypothesized, then it would be better to conduct the Phase II trial with the combination. Where there is an interest in testing multiple doses, the same Phase II approach could be used for each dose with the sample size based on the hypothesized threshold for each dose. If the effect size of interest is the same for all doses, then the sample sizes for each Phase II study would be identical. Others have proposed selection designs where the goal is to choose the best treatment to take forward.[27,28] In our setting where multiple drugs and/or multiple doses may be tested, we are not trying to find the best treatment nor are we requiring that any drug be chosen.

In general, although Phase II studies gather efficacy data, they are not suitably powered to test an efficacy hypothesis, and failing to reject the null hypothesis does not guarantee a positive Phase III trial. Additionally, failure to reject the null hypothesis is only one criterion that investigators must review in selecting agents for large-scale studies. Patient compliance, cost, availability, other ongoing studies of the same treatment, or other procedural issues may make an agent that successfully passed a futility study impractical for a Phase III trial. If multiple candidates for Phase III are identified in Phase II studies, the previous criteria may be used to choose the subset of drugs to be carried forward. Phase II studies also provide some information on safety and toxicity. Safety information would be considered in the decision to proceed to Phase III; however, the true test of safety would come from a Phase III study and postmarketing surveillance. Calibration controls may provide descriptive safety comparisons, but are limited by small sample size. In contrast, futility studies are designed to test the null hypothesis of futility with strong statistical power.

Futility designs utilize historically derived data to establish a threshold for comparison with the observed results. Generally authors of texts on clinical trials discourage the use of historical controls to study efficacy due to concerns that changes in disease management, new therapies, and new methods for assessing therapeutic responses occur over time.[29-31] However, these same authors recognize the usefulness of single armed studies for Phase II. For example, the use of historical controls has been described "As a rapid, relatively inexpensive method of

obtaining initial impressions regarding a new therapy. . .” (p. 50).[29]

After the Phase II study is completed, secondary sensitivity analyses allow investigators to assess the impact of the assumptions regarding the control rate (p*), particularly if a calibration group is included. If the calibration group response is worse than hypothesized, it may be necessary to repeat the Phase II study with a new hypothesized control rate in order to avoid the severe error of missing an effective treatment. If the calibration group response is better than hypothesized, repetition of the study is not required if the investigator is willing to accept the greater possibility that an ineffective drug is being carried forward. Generally, decisions about proceeding to Phase III primarily would be based on the preset threshold, the safety issues, and the other considerations described above.

The sample size for the Phase II futility study we created from DATATOP data was relatively large. If patients are being recruited over an extended period of time, it is possible to do interim analyses with stopping guidelines within futility studies and stop the trial early if a futility threshold is reached before full enrollment. This approach, often applied to Phase III trials,[32] has been adapted to Phase II trials.[8] However, some sequential designs require rapid ascertainment of outcomes, often not practical in PD Phase II trials.[7,33] If it becomes apparent that we will fail to reject the null hypothesis, it is advisable to continue to the end of Phase II and obtain more information on the magnitude of treatment effects, side effects, compliance, and procedural issues involved in trial conduct. The additional information on the treatment to be studied in Phase III should result in a more efficient, rigorous Phase III design.

Another issue for Phase II trials in PD is the choice of an acceptable short-term outcome to identify agents that may modify PD progression. Previously these authors investigated the properties of several short-term outcomes.[34] In PD, short-term outcomes are usually only surrogates for long-term outcomes of interest and have the limitations described by other authors for such surrogates.[35] NET-PD used a futility threshold based on change in the Unified PD Rating Scale (UPDRS) from baseline to 12 months or start of symptomatic therapy, whichever came first, estimated from DATATOP placebo data. Using the same NET-PD Phase II studies threshold and applying it to the DATATOP treatment groups, we again would have found tocopherol to be futile ($p < 0.001$), and we would have failed to find evidence of futility for deprenyl ($p = 0.74$).

The futility design, shown to be valuable in PD and stroke, could be applied to any neurologic disease (including Alzheimer disease, amyotrophic lateral sclerosis, and multiple sclerosis), where a short-term outcome can be identified. There is substantial flexibility in choosing the short-term outcome, but the sample size will depend on the magnitude and variability of the short-term response to treatment. As with PD, the choice of an historical standard for comparison in Phase II will depend upon what is known about the changes in disease definition or standards of care over time. Known changes may require some adjustment to the historical control rates before finalizing the design.

Acknowledgment

The authors thank NINDS, DATATOP Steering Committee, and the Parkinson Study Group for permission to reanalyze their data.

Appendix: NET-PD Investigators and Coordinators

Brad Racette, MD, Patricia Deppen: *Washington University School of Medicine, St. Louis, MO;* Richard B. Dewey, Jr., MD, Brigid Hayward: *University of Texas Southwestern Medical School, Dallas;* Burton Scott, MD, Joanne Field, BSN, RN: *Duke University, Durham, NC;* Julie Carter, RN, MN, ANP, Matthew Brodsky, MD, Pamela Andrews: *Oregon Health & Science University, Portland;* Bala Manyam, MD, Jacqueline Whetteckey: *Scott & White Clinic/Texas A&M University, Temple, TX;* Jayaraman Rao, MD, Maureen Cook, RN, BSN: *Louisiana State University Health Sciences Center, New Orleans;* Michael J. Aminoff, MD, FRCP, Chadwick Christine, MD, Jessie Roth, RN: *University of California San Francisco;* Martha Nance, MD, Sotirios Parashos, MD, Susan Peterson, RN: *Struthers Parkinson's Center, Golden Valley, MN;* Kathleen Shannon, MD, Jeana Jaglin, RN, CCRC: *Rush University Medical Center, Chicago, IL*; Carlos Singer, MD, Marian A. Perez, AA, Anita Blenke, PA-C, MS: *University of Miami, FL;* Robert Hauser, MD, Theresa McClain, ARNP, Summer Wolfrath, MSPH: *University of South Florida, Tampa;* Ted Dawson, MD, PhD, Becky Dunlop, RN: *Johns Hopkins University, Baltimore, MD;* Rajesh Pahwa, MD, Kelly Lyons, PhD, Amy Parsons, RN, BSN: *University of Kansas Medical Center, Kansas City;* Maureen Leehey, MD, Jacci Bainbridge, PharmD: *University of Colorado Health Sciences Center, Aurora;* Lisa Shulman, MD, William Weiner, MD, Katharine Pabst, CRNP, MS, MPH: *University of Maryland, Baltimore;* Rodger Elble, MD, PhD, Charlene Young, RN, MSN, CFNP: *Southern Illinois University, Springfield;* Kapil Sethi, MD, Buff Dill, BS, ED: *Medical College of Georgia, Augusta;* Wayne Martin, MD, Germaine McInnes, RN: *University of Alberta Glenrose Rehab Hospital, Edmonton, AB, Canada;* Vincent P. Calabrese, MD, Peggy Roberge, RN: *Hunter Holmes McGuire Veterans Medical Center, Richmond, VA;* Robert Hamill, MD, Clare Homan, MA: *University of Vermont, Burlington;* Ivan G. Bodis-Wollner, MD, Elizabeth Hayes, RN: *State University of New York Downstate Medical Center, Brooklyn;* G. Webster Ross, MD, Stephanie Terashita, RN: *Pacific Health Research Institute, Honolulu, HI;* G. Frederick Wooten, MD, Joel Trugman, MD, Margaret F. Keller, RN, MS, CCRC: *University of Virginia, Charlottesville;* Jacob I Sage, MD, Debbie Caputo, RN, MSN: *UMDNJ Robert Wood Johnson Medical School, New Brunswick, NJ;* John Fang, MD, Dorothy Shearon, RN: *Vanderbilt University Medical Center, Nashville, TN;* Danna Jennings, MD, Tammie Kelsey, LPN: *Institute for Neurodegenerative Disorders, New Haven, CT;* Joanne Wojcieszek, MD, Joann Belden, RN: *Indiana University School of Medicine, Indianapolis;* Andrew Feigin, MD, Margaret Marie Cox, RN, BSN: *North Shore University Hospital, Manhasset, NY;* Ray L. Watts, MD, Natividad Stover, MD, Rebecca McMurray, RN, MSN: *University of Alabama, Birmingham;* Roger Albin, MD, Kristine Wernette, RN, MS: *University of Michigan, Ann Arbor;* Andrew Siderowf, MD, Sue Reichwein, CCRC: *University of Pennsylvania, Philadelphia;* David Simon, MD, Daniel Tarsy, MD, Lisa Scollins, NP: *Beth Israel Deaconess Medical Center, Boston, MA;* Ron Tintner, MD, Germaine Hunter, RN, CCRC: *Baylor College of Medicine, Houston, TX;* Peter LeWitt, MD, Maryan Deangelis, RN: *William Beaumont Hospital, Southfield, MI;* Richard S. Burns, MD, Lynn Marlor, RN: *Barrow Neurologic Institute, Phoenix, AZ;* Charles Adler, MD, PhD, Marlene Lind, RN: *Mayo Clinic Scottsdale, AZ;* Jay Gorell, MD, Shana Krstevska, MD, Marilyn Flewellen, RN: *Henry Ford Health System, Detroit, MI;* Jay Schneider, MD, Stephanie Sendek: *Thomas Jefferson University, Philadelphia, PA;* Stephen Gollomp, MD, Gwyn Vernon, MSN, CRNP: *Thomas Jefferson University/Lankenau Hospital, Wynnewood, PA;* David Coffey, MD, Pauline LeBlanc, BS: *Dartmouth Hitchcock Medical Center, Lebanon, NH;* Mark F. Lew, MD, Allan Wu, MD, Connie Kawai, RN, BSN, CCRC: *University of Southern California, Los Angeles;* Ryan Uitti, MD, Margaret Turk, RN: *Mayo Clinic Jacksonville, FL;* James Bower, MD, Susan Torgrimson, RN: *Mayo Clinic Rochester, MN;* Joseph Handler, MD, Kathleen Greenbaum, RN: *Neurology Associates PA, Wilmington, DE:* Marwan Sabbagh, MD, Zoran Obradov, CRC: *Sun Health Research Institute, Sun City, AZ.*

References

1. Parkinson Study Group. A controlled, randomized, delayed-start study of rasagiline in early Parkinson disease. Arch Neurol 2004;61:561–566.
2. Parkinson Study Group. A randomized placebo-controlled trial of rasagiline in levodopa-treated patients with Parkinson disease and motor fluctuations: the PRESTO study. Arch Neurol 2005;62:241–248.

3. Shults CW, Oakes D, Kieburtz K, et al. Effects of coenzyme Q10 in early Parkinson disease: evidence of slowing of the functional decline. Arch Neurol 2002;59:1541–1550.
4. Ravina B, Fagan S, Hart R, et al. Neuroprotective agents for clinical trials in Parkinson's disease: a systematic assessment. Neurology 2003; 60:1234–1240.
5. Herson J. Statistical aspects in the design and analysis of Phase II clinical trials. In: Buyse M, Staquet M, Sylvester R, eds. Cancer clinical trials: methods and practice. Oxford: Oxford University Press, 1984; 239–257.
6. Herson J. Predictive probability early termination plans for Phase II clinical trials. Biometrics 1979;35:775–783.
7. Simon R. Optimal two-stage designs for phase II clinical trials. Control Clin Trials 1989;10:1–10.
8. Fleming TR. One sample multiple testing procedure for Phase II clinical trials. Biometrics 1982;38:143–151.
9. Shuster J. Optimal two-stage designs for single arm Phase II cancer trials. J Biopharm Stat 2002;12:39–51.
10. Chang MN, Therneau TN, Wieand HS, Cha SS. Designs for group sequential Phase II clinical trials. Biometrics 1987;43:865–874.
11. Lin Y, Shih WJ. Adaptive two-stage designs for single-arm Phase IIA cancer clinical trials. Biometrics 2004;60:482–490.
12. Thall PF, Cook JD. Dose-finding based on efficacy-toxicity trade-offs. Biometrics 2004;60:684–693.
13. Schoenfeld D. Statistical considerations for pilot studies. Int J Radiat Oncol Biol Phys 1980;6:371–374.
14. Green SJ, Dahlberg S. Planned versus attained design in phase II clinical trials. Stat Med 1992;11:853–862.
15. Palesch YY, Tilley BC, Sackett D, Johnston K, Woolson R. Applying a phase II futility study design to therapeutic stroke trials. Stroke 2005; 36:2410–2414.
16. IMS Investigators. Combined intravenous and intra-arterial recanalization for acute ischemic stroke: The Interventional Management of Stroke (IMS) Study. Stroke 2004;35:904–911.
17. Parkinson Study Group. DATATOP: a multicenter controlled clinical trial in early Parkinson's disease. Arch Neurol 1989;46:1052–1060.
18. Parkinson Study Group. Impact of deprenyl and tocopherol treatment on Parkinson's disease in DATATOP subjects not requiring levodopa. Ann Neurol 1996;39:29–36.
19. D'Agostino RB, Chase W, Belanger A. The appropriateness of some common procedures for testing the equality of two independent binomial populations. American Statistical Association 1988;42:198–202.
20. Pocock SJ. The combination of randomized and historical controls in clinical trials. J Chronic Dis 1976;29:175–188.
21. Gehan EA. The evaluation of therapies: historical control studies. Stat Med 1984;3:315–324.
22. Herson J, Carter SK. Calibrated phase II clinical trials in oncology. Stat Med 1986;5:441–447.
23. Landau WM. Clinical neuromythology IX. Pyramid sale in the bucket shop: DATATOP bottoms out. Neurology 1990;40:1337–1339.
24. Goetz CG, Tanner CM, Shannon K. Progression of Parkinson's disease without levodopa. Neurology 1987;37:695–698.
25. Miller ER, III, Pastor-Barriuso R, Dalal D, Riemersma RA, Appel LJ, Guallar E. Meta-analysis: high-dosage vitamin E supplementation may increase all-cause mortality. Ann Intern Med 2005;142:37–46.
26. Lonn E, Bosch J, Yusuf S, et al. Effects of long-term vitamin E supplementation on cardiovascular events and cancer: a randomized controlled trial. JAMA 2005;293:1338–1347.
27. Liu PY, Dahlberg S, Crowley J. Selection designs for pilot studies based on survival. Biometrics 1993;49:391–398.
28. Simon R, Wittes RE, Ellenberg SS. Randomized phase II clinical trials. Cancer Treat Rep 1985;69:1375–1381.
29. Friedman L, Furberg C, DeMets D. Fundamentals of clinical trials. 2nd ed. Littleton, MA: PSG Publishing Company, Inc., 1985;47–51.
30. Green S, Benedetti J, Crowley J. Clinical trials in oncology. 1st ed. London: Chapman & Hall, 1997;143:43–49.
31. Piantadosi S. Clinical trials: a methodologic perspective. 1st ed. New York: John Wiley & Sons, Inc., 1997;21:155–158.
32. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics 1979;35:549–556.
33. Ensign LG, Gehan EA, Kamen DS, Thall PF. An optimal three-stage design for phase II clinical trials. Stat Med 1994;13:1727–1736.
34. Elm JJ, Goetz CG, Ravina B, et al. A responsive outcome for Parkinson's disease neuroprotection futility studies. Ann Neurol 2005;57:197–203.
35. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med 1989;8:431–440.