The Journal of Rare DISORDERS

STATISTICAL CONSIDERATIONS FOR CLINICAL TRIALS IN RARE DISEASES

Demissie Alemayehu, PhD; Marcia Levenstein, DSc; and Charles Knirsch, MD, MPH
Pfizer Inc, New York, NY 10017

ABSTRACT

Proper design and reporting of clinical trials in rare diseases typically require nonstandard statistical approaches that may generally be optional in conventional large clinical trials. The small size of the target population, coupled with the challenges of recruiting eligible study subjects from mostly vulnerable populations, makes traditional design framework and large sample–based statistical methods nonoptimal to generate minimally acceptable evidentiary data for licensing and approval. In this article, we outline some key statistical points for consideration in the design, analysis, and reporting of such trials, with particular reference to adaptive designs, exact procedures, and the perils of small-sample inference.

INTRODUCTION

Drug development in rare diseases poses both practical and methodologic challenges and opportunities.[1] From a statistical perspective, most of the issues emanate from the small size of the target study population and the logistic problems associated with the conduct of the study. In some cases, the natural history of the disease under study may not be well established, adding to the complexity of sample-size determination, dose selection, definition of outcome measures, and specification of other study protocol components.

Although there have been numerous measures taken by regulatory agencies to encourage research in rare diseases,[2-5] the evidentiary standard to bring safe and effective medicines to these patients is still relatively high. Accordingly, there is a need to assess the adequacy of conventional statistical approaches to tackle the issues that arise in rare disease clinical trials, and to apply or develop novel design and analytic techniques that are appropriate in small trial settings.

In this article, we evaluate alternative design approaches and analytic techniques for rare diseases, and highlight steps that need to be taken to ensure proper dissemination of trial results. Alternative design schemes for small trials are provided, and nonstandard analytic approaches and the reporting of results are discussed.

DESIGN CONSIDERATIONS FOR SMALL TRIALS

In rare disease drug development, the traditional paradigm of randomized controlled trials involving fixed samples and rigid criteria for study conduct may not be optimal. Thus, alternative approaches that consider the accumulated data without compromising trial integrity should be considered. One such approach is the implementation of adaptive designs that permit flexibility to update various aspects of the study (including randomization scheme, number of treatment groups, and number and frequency of interim analyses) using prespecified and statistically sound criteria.[6,7]

From a drug development perspective, adaptive seamless designs can help to reduce the timelines for drug approval, combining data from different stages. Seamless phase 2/3 designs, for example, leverage data from the 2 phases to learn and confirm by treating both phases as a single study conducted in 2 stages.[8] In a typical application, such designs may be used in dose-finding trials, in which optimal doses selected based on the available information at interim are studied further in the second stage. For the final analysis, in which patients from both stages are included, appropriate statistical methods should be used to account for the combination of data from the 2 stages, because a naive combination of data from the 2 stages may result in inflated type I error rates. In addition, suitable measures should be taken to minimize the potential for bias from the information obtained from the analysis at the end of stage 2.

The utility of a seamless phase 2/3 trial is, of course, a function of the strength of the information obtained at the interim and the infrastructure that is put in place to ensure trial integrity. In particular, there should be a clear understanding of who performs the analysis (eg,Data and Safety Monitoring Board or sponsor), and how blinding is preserved (including blinding of patients, investigators, and trial sponsor staff) to minimize operational bias.

In early-phase drug development, dose selection and risk –benefit assessment may require modified versions of

the approaches used in nonrare disease trials. For example, adaptive dose-finding methods may be used to obtain information on dose response earlier in development and maximize the probability of technical success. One common approach is the so-called continual reassessment method (CRM), which unlike traditional cohort designs, permits selection of a dose level sequentially, updating the dose–toxicity relationship based on the patients' response data using Bayesian methods. However, careful attention should be paid to the implementation of the CRM in rare disease research, including the appropriateness of the dose–toxicity model in small samples, the target toxicity/response rate, and the stopping rules.[9,10]

In late-phase development, group sequential designs have increasingly been used to enhance decision making, as they allow for early termination of a trial either for efficacy or futility, based on interim analyses. However, most procedures are based on large-sample theory, and therefore may require appropriate modifications for a small sample.

Providing the rationale for sample-size calculations is always a challenge in rare disease research. When there is uncertainty about effect size or variability for sample-size determination, adjustments to the initial sample size may be made based on a review of accumulating data. In adaptive sample-size reestimation, a major issue is preservation of the overall type I error at the time of the final analysis. Some approaches have been proposed by Bauer and Kohne [11] and Lehmacher and Wassmer,[12] involving combining the *P* values obtained before and after the adaptation, with prespecified weights; by Cui, Hung, and Wang,[13] in which the Wald statistics are combined with prespecified weights; and by Chen, DeMets, and Lan,[14] permitting use of conventional test statistics in connection with increased sample size based on positive interim results. In small trials, when internal pilots are used to reassess nuisance parameters, appropriate adjustments should be made for possible bias.[15,16]

Because adequate knowledge about the rare disease under study may not be available at the design stage, it may also be essential to change other aspects of the protocol, including the test statistic, primary end point, and inclusion/exclusion criteria, based on interim data. In the context of regulatory review and approval, any such adaptation would require extreme caution, owing to the potential for bias. Accordingly, all expected changes to any aspect of the protocol should be prespecified, be adequately justified, and, to the extent possible, garner mutual agreement and understanding between the sponsor and regulatory agencies.

With proper implementation and execution, novel randomization schemes can also prove advantageous in small trials. For example, adaptive randomization permits flexible probability of treatment assignment based on patient characteristics and observed response to treatment.[17,18] In response adaptive randomization, in which allocation probability is based on responses observed in previous patients, one would expect that exposure to more efficacious treatment would be maximized. In covariate adaptive randomization, in which the allocation probability is a function of the covariate balance between groups, the effects of confounders are minimized, thereby permitting valid inference about treatment-effect differences.

In general, assessment of the operating characteristics of an adaptive trial may involve nontraditional procedures, including simulation under varying assumptions about effect size and other design features. In the context of rare diseases, some of the available methods may require appropriate modifications, taking into account the small number of events or subjects involved.[19]

Crossover designs, in which participants serve as their own control, are attractive options for rare diseases provided that the study involves a chronic condition and no carryover effect, and that a relatively rapid response to intervention is a realistic assumption. Such designs generally require fewer subjects and involve less variability than their parallel group counterparts. However, they may be unreliable in the face of potential carryover of treatment effects or if the disease under study is not stable over time. Moreover, they cannot be implemented in certain situations in which the intervention may not be reversible or the outcome of interest may be terminal, eg, all-cause mortality.

In recent years, use of enriched designs has been advocated as a viable option to enhance drug development by targeting a subgroup of the patient population. Enriched designs have the benefit of reducing variability or increasing effect size by carefully selecting homogenous and responsive subgroups, respectively. In rare disease research this is particularly relevant, because enrichment generally may increase the

chance of demonstrating treatment effect with limited sample size. One major limitation of such a strategy is the ability to establish generalizability of findings to the wider patient population. In addition, in rare diseases the enrichment may also lead to a reduction of the available patient pool, thereby posing recruitment and other logistic problems.

ANALYTIC STRATEGY

As in large clinical trials, implementation of a given statistical procedure is heavily dependent on the validity of the assumptions underlying the procedure. One key distinction that handicaps the analysts in a small trial is the inability to implement alternative approaches that rely on large-sample theory. For example, when distributional assumptions cannot be justified to implement parametric methods, one often resorts to nonparametric approaches in the analysis of large trials. In small trials, methods that rely on large-sample theory may not be a substitute when the validity of standard approaches is in question.

Given the small size of studies in rare diseases, exact procedures should be used, to the extent possible, in analyzing the data. Most statistical software packages, such as StatXact, provide exact hypothesis tests, exact confidence intervals, and exact power and sample-size calculations.

Hierarchical models are useful in small trials involving 2-stage sampling, including prospective longitudinal studies. With longitudinal data, analytic strategies based on the last-observation-carried-forward approach are generally inefficient. When model assumptions are satisfied, mixed-effect models for repeated measures (MMRM) for continuous responses, and marginal (eg, generalized estimating equations [GEE]) and random-effects (eg, generalized linear mixed models [GLMM]) approaches for categorical responses and count data, may be used to increase efficiency, especially in small trials. In the presence of missing values, the validity of these methods depends on the pattern of "missingness" (eg, the likelihood-based methods [MMRM and GLMM] and some weighted GEE models are applicable under "missing completely at random" (MCAR) and "missing at random" assumptions, while the usual GEE models require MCAR assumptions).[20]

In small clinical trials, Bayesian approaches may be advantageous for a number of reasons. With small samples, inferences are easy to formulate and solve by Bayesian methods. Sequential analyses are readily implemented in a Bayesian paradigm, without the need to adjust for multiplicity, because the posterior distribution can be updated using the current posterior distribution as the prior for the next update. Bayesian hierarchical models allow information to be combined from different sources, thereby gaining strength. However, in small samples, different prior opinions may lead to different conclusions, because the influence of the choice of priors on the posterior probability distribution is a function of sample size.[21] Therefore, the choice of priors should be considered carefully.

Lastly, in view of the uncertainties inherent in small clinical trials, the primary analysis results should be corroborated with alternative statistical analyses, to ensure the robustness of the results to departures from model assumptions.

DISSEMINATION OF RESULTS

Compared with large trials, investigations in rare diseases provide limited scientific evidence about the safety risk–benefit profile of a treatment option. Accordingly, the researchers involved in such endeavors should exercise extreme caution in interpreting and reporting the results of a small trial. In particular, the results of the study should be interpreted with fair balance, highlighting the limitations of the study and analytic strategy, the biological plausibility of the findings, and the consistency of the findings with established knowledge of the drug class.

The requirement to present the results with fair balance should not be viewed as an unnecessary measure intended to stifle innovation and research in rare diseases. In fact, it is an essential means of disseminating results, so that patients, regulatory agencies, health care providers, and other stakeholders can evaluate the strength of evidence when making informed decisions about the risks and benefits of a drug.

CONCLUDING REMARKS

In this article, we highlighted points to consider when designing, analyzing, and reporting a clinical trial in rare diseases. While some of the issues raised are generally

pertinent to most clinical trials, studies in rare diseases pose additional problems as a result of the features that are inherent in the limited size of the target population. To achieve regulatory and scientific objectives, it is therefore important to enhance and employ best statistical practices established for traditional trials while taking into account the special challenges associated with rare diseases, including the small patient numbers, paucity of knowledge about the disease, and recruitment and retention of study subjects.

ACKNOWLEDGMENTS

The authors would like to thank Jack Cook of Pfizer for helpful suggestions and feedback at various stages of manuscript development.

REFERENCES

1. Kesselheim AS, Myers JA, Avorn J. Characteristics of clinical trials to support approval of orphan vs nonorphan drugs for cancer. *JAMA.* 2011;305:2320-2326.

2. US Food and Drug Administration. Developing orphan products: FDA honors Rare Disease Day. http://www.fda.gov/downloads/ForIndustry/DevelopingProductsforRareDiseasesConditions/UCM202027.doc. Accessed January 20, 2014.

3. National Institutes of Health, NIH News, "Therapeutics for Rare & Neglected Diseases," http://www.ncats.nih.gov/research/rare-diseases/trnd/trnd.html. National Center for Advancing Translational Sciences. Therapeutics for Rare and Neglected Diseases. Accessed March 28, 2014.

4. Rubinstein YR, Groft SC, Bartek R, Brown K, Christensen RA, Collier E, et al. Creating a global rare disease patient registry linked to a rare diseases biorepository database: Rare Disease-HUB (RD-HUB) Contemp Clin Trials. 2010;31(5):394–404.

5. European Medicines Agency. Guideline on clinical trials in small populations. Doc. ref. CHMP/EWP/83561/2005. Published July 27, 2006. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003615.pdf. Accessed January 20, 2014.

6. Meurer WJ, Lewis RL, Berry DA. Adaptive clinical trials: a partial remedy for the therapeutic misconception? *JAMA*. 2012;307:2377-2378.

7. Chow SC, Chang M, Pong A. Statistical consideration of adaptive methods in clinical development. *J Biopharm Stat.* 2005;15:575-591.

8. Inoue LY, Thall PF, Berry DA. Seamlessly expanding a randomized phase II trial to phase III. *Biometrics*. 2002;58:823-831.

9. Garrett-Moyer E. The continual reassessment method for dose-finding studies: a tutorial. *Clin Trials*. 2006;3:57-71.

10. Ishizuka N, Ohashi Y. The continual reassessment method and its application: a Bayesian methodology for phase I cancer clinical trials. *Stat Med*. 2001;20:2661-2681.

11.Bauer P, Kohne K. Evaluation of experiments with adaptive interim analyses. *Biometrics*. 1994;50:1029-1041.

12. Lehmacher W, Wassmer G. Adaptive sample size calculations in group sequential trials. *Biometrics*.1999;55:1286-1290.

13. Cui L, Hung HM, Wang SJ. Modification of sample size in group sequential clinical trials. *Biometrics.* 1999;55:853-857.

14. Chen YH, DeMets DL, Lan KK. Increasing the sample size when the unblinded interim result is promising. *Stat Med.* 2004;23:1023-1038.

15. Proschan MA, Wittes J. An improved double sampling procedure based on the variance. *Biometrics*. 2000;56:1183-1187.

16. Kairalla JA, Muller KE, Coffey CS. Combining an internal pilot with an interim analysis for single degree of freedom tests. *Commun Stat Theory Methods*. 2010;39:3717-3738.

17. Karrison TG, Huo D, Chappell R. A group sequential, response-adaptive design for randomized clinical trials. *Control Clin Trials*. 2003;24:506-522.

18. Rosenberger WR, Lachin JM. The use of response-adaptive designs in clinical trials. *Control Clin Trials*. 1993;14:471-484.

19. Chow SC, Chang M. Adaptive design methods in clinical trials—a review. *Orphanet J Rare Dis*. 2008;3:11.

20. Mallinckrodt CH, Lane PW, Schnell D, et al. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. *Drug Information J.* 2008;42:303-319.

21. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian Data Analysis. London, UK: Chapman & Hall; 1995.

Address Correspondence To:
Demissie Alemayehu, PhD (corresponding author)
235 East 42nd Street/219-7-47
Pfizer Inc.
New York, NY 10017
alemad@pfizer.com
(212) 573-2084