**Small Population Trial Design Considerations**

Freda Cooner, Ph.D. and Ram Tiwari, Ph.D., Office of Biostatistics, Office of Translational Sciences, CDER/FDA

This chapter presents a brief overview of various aspects of a clinical trial for rare diseases, including regulatory requirements, a summary of trial designs, trial conduct considerations, and some relevant statistical issues arising from small population trials.

Rare Diseases and FDA

A rare disease is defined by the Orphan Drug Act[[1]](#endnote-1) of 1983 as a disorder or condition that affects less than 200,000 persons in the United States. However, most rare diseases affect far fewer persons. A natural challenge for a clinical development program in rare disease population is the small number of patients available for enrollment. The Orphan Drug Act provides incentives to financially help the development of such treatments; however, it does not create a statutory standard for the approval of orphan drugs that is different from the standard for drugs developed for common conditions.

Since the Orphan Drug Act, the FDA received over 4800 requests and over 3400 orphan product designations were granted. A majority of these designations were for drugs and the rest biologics. The top two therapeutic categories in Years 2013 and 2014 are oncology and neurology. Both the number of submissions and approvals for orphan products are increasing. To date, over 500 approvals were issued and 48 in 2014 alone.

The statutory requirement for all drug approvals is demonstration of substantial evidence of effectiveness in treating or preventing the condition and evidence of safety for that use. Substantial evidence of effectiveness should be obtained from one or more adequate and well-controlled studies, each convincing on its own. FDA acknowledges the difficulties for such requirements implementation in rare diseases, and provides flexibility because of the many types and intended uses of drugs. In particular, FDA "exercise[s] its scientific judgment" in determining the kind and quantity of data a sponsor is required to provide for individual drug development programs.[[2]](#endnote-2) This flexibility extends from early phase of development to design of adequate and well-controlled clinical studies required to demonstrate safety and effectiveness to support marketing approval.

Before considering detailed aspects of trial designs suitable for rare diseases, careful attention should be paid to endpoints selection. Choosing endpoints can be especially difficult for rare diseases due to lack of knowledge of disease progression. The endpoints for the clinical trial should be both feasible and clinically meaningful. They can be decided based on knowledge of either the nature of the rare disease derived from a natural history study or the pharmacological effects of the investigational drugs or interventions. Surrogate endpoints that are able to be accurately measured earlier in follow-up than clinical endpoints and are thought to be reasonably likely to predict clinical benefit are often used in rare disease trials. Trial designs should be suitable for the endpoints of interest.

Trial Designs

The limited number of patients in a rare disease population makes the size of the clinical trials conducted in that population small, and that in turn restricts the usage of inferential statistics. With the usual requirements for significance level and study power, small population trials with a conventional design may only be able to detect very large intervention effects. It is therefore necessary to practice flexibility in establishing evidence of effectiveness and safety in rare diseases, and there are many ways this can be accomplished. Historically, FDA has allowed the following program options to provide development of a rare disease therapy: 1) one adequate and well-controlled trial with supporting evidence from pharmacological data or other sources, 2) a single-arm, non-randomized, open-label trial, and 3) a single adequate and well-controlled trial with the significance level relaxed somewhat. Although these may be the only feasible approaches, careful consideration should be paid to the consequences, and other options should be explored, when possible.

A control group is critical to the ability to estimate the true treatment effect of an investigational therapy. There are three types of control groups typically considered for use in rare disease trials: historical controls, patient self-controls, concurrent controls.

The source of historical control data may be scarce on its own for rare diseases. Patient-level data are often not available or not of the same quality as data from prospectively designed clinical trials FDA usually requires to support approval. Moreover, some historical data are from a different group of patients whose disease status or diagnoses cannot be verified. The standard of care may vary significantly at different times and different locations. It should also be noted that patients under any standard of care may not represent the natural course of the disease. All these elements present difficulties in the use of patient registry data or natural history studies as a comparator.

Using patient-self as its control means either comparing the outcome to the baseline value for the same patient or comparing the outcomes from two periods of a cross-over design. Change from baseline values do not account for changes unrelated to treatment. One should also take into account the impact of regression to the mean, especially when only a more severe stage of the disease is represented in the study. Without a proper control or knowledge or the natural history of the disease, accurately measuring the treatment effect could be challenging and/or spontaneous improvement may be misinterpreted as evidence of a treatment effect.

The benefit of a concurrent control group is that its use in comparative analyses eliminates the potential for bias due to differences in patient populations, time periods, locations, and regression to the mean associated with the use of historical or patient self-controls. Furthermore, it allows consistent reporting of prognostic factors in data collection. Although the benefit is obvious, it is universally recognized that including a concurrent placebo control group may not always be ethical for rare disease trials. Other options, such as the use of a very low dose of the investigational treatment or standard of care as a concurrent control may be more acceptable.

Once the control group is selected, some clinical trial features, such as randomization and blinding should also be weighed against ethical and logistical considerations in rare disease clinical trials to ensure the interpretability of the trial results. Additional attention should be paid to the data collection process to ensure data quality because of the limited sources of data on rare diseases.

*Trial Design Options*

* Parallel group design

This is the traditional and most popular design for clinical trials. Either using placebo or an active control, along with randomization and blinding, this design has obvious benefits in both trial conduct and data analyses. However, it may not be feasible or ethical for a small trial as discussed above. Some features, such as interim futility analysis and group sequential design, can be incorporated to increase efficiency.

* Repeated measures design

This design collects repeated measurements of an endpoint from the same subjects in a longitudinal study. Although this design may increase the information collected from each patient in a trial, it may introduce more measurement error and may not be suitable for some endpoints, such as clinical response.

* Factorial design

This design is conventionally used for a combination therapy to understand the contribution of each component without conducting separate trials. It can reduce the number of subjects needed for the whole clinical development program. This design could potentially test the interaction of the components. However, it will lose the advantage in efficiency for that purpose.

* Cross-over design

Each patient receives all treatments or interventions, but randomly follows a specific sequence of treatments in this design. The advantages include a reduction in the number of study subjects required to achieve statistical significance. However, this design is only suitable for a treatment of short duration or intervention with an even shorter wash-out period. Withdrawal effects of certain biological products may also preclude the usage of this design.

* Enrichment strategy

Enrichment strategy includes several trial designs, such as randomized withdrawal design. The purpose is to pre-select subjects, who more likely will respond to a treatment or intervention with more measurable responses. This may be difficult for some rare diseases, of which we have very little knowledge. It was elaborated in the draft guidance on enrichment strategies[[3]](#endnote-3).

* N-of-1 design

With this design, the clinician administers more than one treatment to a single patient during multiple treatment periods to achieve a desired effect. This design may be more suitable for determining an optimal treatment algorithm for individual patients and is used only infrequently in drug development.

* Ranking/Selection design

Ranking and selection procedures are statistical techniques for comparison of the parameters for multiple study populations under the assumption that these parameters are not all the same (Gibbons, Olkin, and Sobel, 1979[[4]](#endnote-4)). The design may be beneficial in selecting a subgroup of a population that is more responsive to the treatment or intervention. It may be more efficient for a small trial. However, it requires more planning at the design stage and some background knowledge of the disease, neither of which rare disease trials may afford.

* Multiple treatment arm trial with shared control (master protocol)

This design brings multiple companies together to test multiple experimental treatments for the same disease using one master protocol. One challenge of this design is in the logistics of having the companies with potentially conflicting interests share resources and data.

* Adaptive designs

This design is referring to any trial design that incorporates adaptation of one or more design features during the conduct of the trial, e.g., inclusion/exclusion criteria, endpoint, sample size, etc. Some adaptations are commonly used, such as sample size re-estimation, while others, such as Bayesian adaptive design, will need more investigations. This is elaborated in the FDA draft guidance on adaptive designs[[5]](#endnote-5).

None of the above trial designs are universally accepted as the design for a small population trial. The design for a treatment or intervention should be carefully selected to be suitable for the biological background and to efficiently investigate the treatment or intervention under study.

Trial Conduct

With limited subjects' availability and sometimes a more complicated trial design, holding trial conduct to a rigorous standard is critical for small sample size trials, where achieving a high level of accuracy is essential. First and most important in trial conduct is to maximize data quality. This goal requires obtaining accurate and precise measurements of key variables through implementation of standardized data collection processes. The overarching goal is to minimize bias in data ascertainment and reduce the measurement error of key variables in order to increase the precision of estimates. Other aspects that may aid trial conduct are training and monitoring. Use of a well-organized data monitoring committee (DMC) to implement interim analyses can enhance trial conduct and efficiency. Moreover, central training of clinic personnel on data acquisition and data entry and clinical monitoring and auditing of data collection can reduce measurement error and variability.

The National Academies of Sciences (NAS) report on missing data[[6]](#endnote-6) strongly recommends the use of prevention strategies to reduce the amount of missing data in any clinical trial, and this recommendation is all the more important for small population trials. Depending on the choice of estimand, prevention strategies may include retrieving data from patients discontinuing treatment or dropping out, provided their informed consent allows such retrieval (LaVange and Permutt, 2015[[7]](#endnote-7)). With limited numbers of patients, accumulating all data needed for analysis from study subjects and minimizing missing data where possible is more important than with a conventional design. A properly designed trial and data collection plan, along with sufficient clinical staff training, may assist in patient retention.

Data Analysis

Once data are collected according to an appropriate trial design and standardized procedures, the small sample sizes make it essential to optimize data analysis strategies. Efficient statistical models and analysis methods should be considered for effect estimation and hypothesis testing. For example, use of repeated measurements models may have some advantages over analyses that are based only on the last observation in a trial. Similarly, time-to-event analyses may be considered in lieu of analyses of proportions responding. If the primary outcome is based on a continuous measurement and not the occurrence of a discrete event, then analyzing the outcome with a linear model (e.g., ANCOVA) will provide efficiencies compared to determining a threshold for response and then conducting a responder analysis. Covariate adjustment in a statistical model may also minimize variability of study outcomes and increase the ability to detect a clear signal of efficacy. The covariates to be included in the model should be both statistically and clinically meaningful and pre-specified in the protocol.

Sensitivity analyses should be planned to assess the impact of assumptions required for the analysis to be valid on the trial’s results. In particular, assumptions that are difficult to verify, e.g., assumptions about the missing data mechanism, should be evaluated with appropriate sensitivity analyses (Permutt, 2015[[8]](#endnote-8)).

For trials with small sample sizes, standard statistical methods based on large-sample theory do not necessarily apply, thus making the results and conclusions of the analysis less reliable. Also, modern trials often involve several subgroups, such as gene-mutation subgroups, with subgroup sample sizes even smaller. In such situations, Bayesian analysis methods may be appropriate as they allow borrowing information from homogeneous subgroups while discounting information from heterogeneous subgroups, regardless of subgroup sample sizes. In addition, when historical information is available; for example, through historical controls, early phase trials, or some observational studies, Bayesian methods naturally lead to the incorporation of the historical information through the application of Bayes Theorem in a way that allows for some of the historical data that are obsolete or less useful to be discounted.

Example

Consider patients randomized to control arm (C) and to the investigational drug arm (E), with the primary endpoint as mortality. Let



with



where or 1 depending on whether the patients were assigned to the control arm or to the experimental arm. Let  and assume that  follows a non-informative prior (which may be considered as a bivariate normal with diffuse covariance-matrix), . Using the likelihood function



and the Bayes Theorem, the posterior distribution of , is given by



where  denotes the data (from both the arms). If the historical data, say ,

are available with  being most recent and informative than and so on, with being the least informative, a power prior (Ibrahim and Chen, 2000[[9]](#endnote-9)) approach can be used to incorporate the historical data in the likelihood, and the posterior can be obtained using:



where  is the prior for  with support on  so that most recent historical data get higher weight than the older studies (See Gamalo et al. (2014) for such a prior[[10]](#endnote-10)).

Using a Markov chain Monte Carlo (MCMC) technique the (conditional) posteriors of  and , given the data, , can be obtained.

We propose the following decision rule:

* Reject the hypothesis

 versus 

if



for a large (pre-specified) value of such as 0.975, where denotes the MCMC samples on .

We use the Monte Carlo simulations to access the frequentists operating characteristics namely the type-I error and power:



where  and are the posterior probabilities based the Monte Carlo data , generated under the null and alternative hypotheses  and , respectively. Here, both  and  are taken to be large.

When the trial includes subgroups, so that



the above logistic model can be modified to



where one can assume a hierarchical model for ; for example, specified by 

with and the variance components and having independent diffuse inverse-gamma, half-Cauchy or uniform priors. The Bayesian methods for other endpoints such as the absolute risk difference or relative risk ratio can be derived in a similar manner.

If the number of gene-mutated groups is large, with small sample sizes, the odds-ratio estimates across groups may form different clusters. In that case, instead of using a bi-variate normal prior for , one may use a Dirichlet process prior (Fergusion, 1973[[11]](#endnote-11)).

The Bayesian approach described above can easily be extended to (Bayesian) adaptive designs and/or for non-inferirority trials.

Summary

Small population trial in rare diseases poses unique challenges in both trial design and data analyses, while it also provides opportunities for innovative designs and methods. The FDA not only provides regulatory flexibilities, but also has led great efforts in research and practice for rare disease clinical programs. With an increasing number of orphan products submissions, the FDA has cultivated early and frequent communication with the stakeholders to facilitate the regulatory review process. There remain unexplored area to discover and develop in small population trial design and analyses and this section has briefly laid out the FDA's current thinking.

References

1. Public Law 97-414, 96 Stat. 2049 (1983). Amended by Public Law 98-551 (1984) to add a numeric prevalence threshold to the definition of rare diseases. [↑](#endnote-ref-1)
2. 21 CFR 314.105. [↑](#endnote-ref-2)
3. FDA draft guidance for industry, 2012, Enrichment Strategies for Clinical Trials to Support Approval of Human Drugs and Biological Products. [↑](#endnote-ref-3)
4. Gibbons, J. D., I. Olkin, and M. Sobel, 1979, An introduction to ranking and selection, *The American Statistician*, 33: 185–192. [↑](#endnote-ref-4)
5. FDA draft guidance for industry, 2010, Adaptive Design Clinical Trials for Drugs and Biologics. [↑](#endnote-ref-5)
6. The Prevention and Treatment of Missing Data in Clinical Trials, 2010. [↑](#endnote-ref-6)
7. LaVange, L.M., and Permutt, T, 2015, A regulatory perspective on missing data in the aftermath of the NRC report, *Statistics in Medicine*, DOI: 10.1002/sim.6840 [↑](#endnote-ref-7)
8. Permutt, T, 2015 Sensitivity analysis for missing data in regulatory submissions, *Statistics in Medicine,* DOI: 10.1002/sim.6753 [↑](#endnote-ref-8)
9. Ibrahim, J.G., and Chen, M-H, 2000, Power prior distribution for regression models, *Statistical Science*, 25: 46-60. [↑](#endnote-ref-9)
10. Gamalo, M.A., Tiwari, R.C., and LaVange, L.M., 2014, Bayesian approach to the design and analysis of non-inferiority trials for anti-infective products, *Pharmaceutical Statistics*, 13: 25-40. [↑](#endnote-ref-10)
11. Ferguson, T. S., 1973, A Bayesian Analysis of Some Nonparametric Problems, *Ann. Statist*., 1: 209-230. [↑](#endnote-ref-11)