

Clinical Trials for Rare Lung Diseases

Jeffrey Krischer

Abstract Clinical trial designs for rare lung diseases must meet the same rigorous standards as do designs for trials for diseases that occur with much more frequency. However, there are many different types of study designs; some of which require only a fraction of the number of subjects required to the randomized controlled trial, which is often considered the gold standard.

Alternate designs can address those issues by the use of external or historical controls or with participants serving as their own control. In the case of external or historical controls, all patients to be recruited on a proposed study would receive the new or experimental therapy and their outcomes would be compared to a population that had already been treated by a standard therapy. If historical data are valid and available, this is a very efficient design because it requires fewer patients to be accrued. The downside of such a design is that the selection of historical controls must be made with extreme caution so as not to bias the study results.

A design that avoids this problem is the use of concurrent controls for which participants can serve as their own control. Such designs are desirable if there is less within patient variability in a treatment response than there is between-patient variability. In such cases, outcome estimates will have less variance and the study design will require less accrual. Examples of these designs include cross-over designs and “N-of-1” designs. A design that is well suited to rare events and rare diseases is the case-control design. In such a design, individuals in whom a certain outcome has been observed (disease severity or particular event) are matched to controls who did not have such an outcome and then the two groups are compared with respect to a particular intervention or exposure. Such designs can be developed from prospective as well as retrospective data collection perspectives.

Examples of prospectively randomized designs include cross-over designs as well as factorial designs. In the former, participants are randomized to a treatment arm for a period at the end of which the outcome is assessed and then “crossed over” to the other treatment. The cross-over design makes the same assumptions as do “N-of-1” trials where participants are randomized to pairs of therapies given in random sequence and a washout period is assumed to eliminate the affect of the treatment after the intervention

is withdrawn. Factorial designs essentially involve a double randomization in which two questions are asked in the same participant population.

Finally, designs for ranking and selection procedures are often helpful and generally require a smaller sample size than randomized controlled trials. Ranking statistics are often used when information about underlying parametric distributions is unknown. It could be argued that less is learned in such an experimental design and a subsequent experiment is required to measure the actual difference between treatment outcomes.

There are many approaches to the design of a trial and many of them can achieve certain economies in terms of the required number of participants that need to be enrolled. However, the options are not without their drawbacks and require investigators to make a number of assumptions.

Keywords: clinical trials, bias, sample size, randomized control trials, historical controls, cross-over designs, N-of-1 designs, case-control designs, factorial designs

Introduction

The challenges of designing clinical trials for rare diseases have been recognized by many investigators (1–4) and the issues apply to much more common disorders as well, in that it is preferable to be able to answer a study question with the fewest number of subjects enrolled, irrespective of the number of subjects available. If two alternative treatments are to be compared in a trial, then a scientific and ethical imperative is to discover which is superior so as to minimize the number of subjects given the inferior treatment. Even in the case of trials designed to establish equivalency between two or more treatments, the imperative is to find the one with the least side effects, least cost, or least inconvenience, while maintaining the same degree of efficacy with the fewest number of subjects exposed to the more toxic therapy.

There are a number of alternative study designs that can be considered in the context of rare diseases (Table 2.1). These same designs are available for more common diseases and, for the most part, clinical trial designs for rare diseases must meet the same rigorous standards as do designs for trials for diseases that occur with much more frequency. They must ask important scientific questions, minimize bias, and have appropriate likelihood of achieving a scientifically acceptable answer. Indeed, there are no designs for rare diseases that are not applicable to any other category of diseases. However, there are many different types of study designs; some of which require only a fraction of the number of subjects required to the randomized controlled trial, which is often considered the gold standard.

To begin, it is helpful to consider that a study is, in its most abstract form, an experiment designed to draw a conclusion about which the scientific community, the population of affected individuals, and the population at large can agree. To the extent possible, a study should be free of bias in that its conduct and results are not affected by factors other than the specific study question. The more evidence that a study is bias-free, the stronger one's conviction about the study results can be.

A randomized controlled trial is considered the gold standard because inherent in its design is the minimization of bias. Thus, the results are often considered as the strongest evidence in testing a hypothesis. However, randomized controlled trials are

Table 2.1 Alternative clinical trial designs.

-
- Prospective cohort design
 - Historical controls design
 - Parallel group design
 - Case-control design
 - Cross-over design
 - N-of-1 design
 - Factorial design
 - Ranking and selection design
 - Randomized controlled trials
 - Bayesian designs
 - Decision analysis-based design
 - Randomized withdrawal design
 - Early escape design
 - Group sequential design
 - Adaptive design
 - Risk-based allocation design
-

not easy to do in that many potential participants object to the concept of randomization and many investigators feel that randomization, in of it itself, is unethical (5). Randomization requires that the investigator and the subject consider themselves in the state of equipoise in that they truly feel that the treatment received from either arm of a randomized trial is equivalent unless proven otherwise. This is difficult for participants who want to believe that their treatment will be based upon what is best for them and not the “flip of a coin” and difficult for physicians who also think that they are ethically bound to provide the “best” treatment. Equipoise is made the more difficult since trials are often developed because an investigator feels that an experimental therapy is better and they wish to test that hypothesis in a rigorous fashion. Many subjects object to the trials if they have a likelihood of being assigned a potentially inferior arm (i.e., have a likelihood of not receiving the experimental therapy) or randomized to a placebo.

There are other sources of bias that should also be considered in addition to study design. Bias can result from the conduct of a study as well as its reporting in the literature. In the former, bias can result from the selection of subjects enrolled into a study, allocation to the arms of a study, differences in follow-up, or in ascertainment of study end points. The interpretation of study results from a trial conducted at a single institution might be affected by the types of cases that are referred to that institution for enrollment, if they are not representative of the general population of individuals affected by a certain disorder. For example, methods developed for the identification of rare mucociliary clearance disorders tested at a major referral center might give very different results if they were to be tested in the setting of a primary care practice since the population evaluated at the referral center can be very different. Differences in the study populations could affect the calculations of the sensitivity and specificity or a diagnostic test or its interpretation since the detection of rarer conditions generally requires a high level of specificity as compared to more common conditions to be scientifically and societally acceptable.

Bias that results from subject follow-up or ascertainment of study end points can arise insidiously and be very difficult to control. If a study is designed with historical controls or literature controls then follow-up practices may not be reported or differ in

some unknown way from the contemplated study. Even using concurrent controls may be biased if one treatment group is followed more closely than another leading to earlier recognition of study end points. A difference in the drop-out rates between study arms that is correlated with the study end point can introduce bias. For example, subjects who feel that their condition is not improving may withdraw from a trial and, as a result, the subjects available to evaluate at study end may be the remaining few who experienced a favorable outcome.

It is also recognized (6) that bias can come from study reporting in the scientific literature. Studies with positive outcomes are more likely to be published than are studies with negative outcomes. Thus, the historical or background information upon which a study is based might be biased in a particular direction. For this reason there are now national registries of clinical trials such that trials are registered when they are opened (to provide an accounting of the total universe of open trials in a particular field) rather than when the results are known and only a subset published. Not all of these types of bias are easily recognized, nor controlled, by investigators.

A Hierarchy of Study Designs

While the randomized controlled clinical trial is regarded as the standard for trial designs, such trials designs are not always applicable in a given setting and there are alternatives to be considered. Most have to do with the selection of control groups to which the experimental intervention is to be compared.

Historical Controls. One approach is the use of external or historical controls. In the case of external or historical controls, all patients to be recruited on a proposed study would receive the new or experimental therapy and their outcomes would be compared to a population that had already been treated by a standard therapy. This results in considerable savings in terms of the number of patients to be accrued, even though the total number of patients may be substantial fraction of the total needed in a randomized controlled trial. For example, such a study would require less than half the number of patients to be treated compared to a randomized trial if only a moderate number of historical controls patients were available. Testing a question of a 20% difference in response rates, assuming the availability of data on 50 historical controls and a historical response rate of 40% would require only 74 patients prospectively treated by an experimental agent (total 124 patients) as compared to a prospective randomized trial which would require 153 patients.

If historical data are valid and available, this is a very efficient design because it requires fewer patients to be accrued prospectively and the newly accrued subjects would all be offered the experimental intervention. The downside of such a design is that the selection of historical controls must be made with extreme caution so as not to bias the study results. Often it is very difficult to know whether bias has been introduced by factors that have not been reported in the historical series or through changes in clinical practice that may affect clinical assessments or outcomes.

Concurrent Controls. A design that avoids this problem is the use of concurrent controls in which participants can serve as their own control. Such designs are desirable if there is less within patient variability in a treatment response than there is between-patient variability. In such cases, outcome estimates will have less variance and the study design will require less accrual. Examples of these designs include cross-over

designs and “N-of-1” designs. These study designs are applicable, however, only in the situation where there is a relatively rapid response to the intervention, the response disappears relatively soon after the intervention is withdrawn and the participant’s overall condition does not change over the periods of time in which the intervention occurred or the intervention has been withdrawn. (That is, the condition or the severity of the disease does not change over time.) These designs work well for chronic diseases, but there are many settings in which this assumption cannot be justified or even tested.

Case–Control Designs. A design that is well suited to rare events and rare diseases is the case–control design. In such a design, individuals in whom a certain outcome has been observed (disease severity or particular event) are matched to controls that did not have such an outcome and then the two groups are compared with respect to a particular intervention or exposure. Such designs can be developed from prospective as well as retrospective data collection perspectives. Retrospective data collection is particularly efficient since one can identify just the cases where the events have occurred and matched them to a control where a particular event of interest has not occurred. But it suffers because of the reliance on the quality of historical data. Yet, such designs can be particularly useful in rare diseases in which there is a long lag time between genotype and phenotypic expression. Again the problem is the same as in the case of historical controls where investigators have to be extremely careful in selecting appropriate controls. Therefore, this design is not ranked as high as the randomized controlled trial in terms of the strength of evidence, because of this potential bias.

Cross-Over, “N-of-1,” and Factorial Designs. There are a number of different designs which can be employed even when treatment arms are prospectively randomized to reduce sample size requirements. Examples include cross-over designs as well as factorial designs. In the former, participants are randomized to a treatment arm for a period at the end of which the outcome is assessed and then the subjects are “crossed over” to the other treatment. The cross-over design makes the same assumptions as do “N-of-1” trials where participants are randomized to pairs of therapies given in random sequence and a washout period is assumed to eliminate the effect of the treatment after the intervention is withdrawn (6, 7). Cross-over designs use the same patients twice and effectively halve the number of patients that must be enrolled. “N-of-1” designs use the same patients a number of times (generally up to 5) and are even more efficient. The repeated evaluation of a therapy for the same subject also allows the treating physician to draw conclusions about the efficacy of the intervention for a single patient which is very appealing as well.

Factorial designs are similar to cross-over designs but differ importantly in that they essentially involve a double randomization in which two questions are asked in the same participant population. This essentially results conducting two studies at the same time in the same patient population with a sample size savings of an appropriate 50% for both. The sample size requirement for each study is unchanged, however. This type of design also assumes that there is no interaction between the two treatments. By interaction we mean that the effect of treatment A over its comparison group (placebo) is in the same direction regardless of whether the patient received treatment B or not. Again there is an assumption being made that is hard to verify.

Ranking and Selection Designs. Designs for ranking and selection procedures are often helpful and generally require a smaller sample size than randomized controlled trials (8). In ranking and selection designs, the objective is to maximize the likelihood

of selecting the better therapy from a number of therapies as opposed to designing a trial that actually compares therapy directly and measures how much better one is as compared to another. Ranking statistics are also used when information about underlying parametric distributions are unknown. It could be argued that less is learned in such an experimental design and a subsequent experiment is required to measure the actual difference between treatment outcomes. That's because a randomized clinical trial design is to detect a minimally clinical significance between treatments, whereas the ranking statistics only seek to determine which treatment has the better response rate. Yet, the sample size savings can be appreciable as compared to a randomized control trial with less than 25% of the needed accrual to answer almost the same question with the same statistical power.

Randomized Trials. It should also be noted that the choice of end points in a randomized trial can also affect the sample size requirement. For example, a study designed to detect a change in the percentage of cases that respond (a binary outcome, yes or no) to a given treatment versus and alternative (control) treatment will generally have a larger sample size requirement than a study that seeks to detect a 20% change in the value of a continuous outcome measure (e.g., %FEV1). This depends somewhat on the distribution of the outcome measure and its variability (standard deviation) among patients treated on the same (experimental or control) treatment.

There are also some options when designing studies that have time-until-event outcomes, in which the study seeks to determine which treatment delays or prevents the occurrence of an outcome of interest. This might be a study of time until disease progression or overall survival. In these types of study designs, it is the person-years of follow-up that can have a substantial effect of the sample size requirement. For example, a study may take several years to accrue and the study end point is to be assessed at a certain time after the last patient has been accrued. All those patients accrued before the last patient will have been followed for a variable, but longer, period of time. The sample size calculation takes this into account, utilizing all the follow-up data that is available on every patient. If the duration of follow-up is extended for all patients, then the total amount of person-years of follow-up is increased and the sample size is decreased to measure the same effect size. Maintaining the original sample size has the effect of increasing the study power to detect a planned difference in outcome or being able to detect a smaller difference than planned with the original study power.

Interim Analyses. Another consideration in study design is the provision for interim analyses. Interim analysis plans can accompany any type of study design. They generally focus on one or more of the following determinations: (1) are the outcomes observed on the control arm of a trial close to the original planning parameters? (2) do the early results indicate an difference so large as to warrant stopping the study? or (3) do the early results indicate that no difference will be detected if the study would continue as planned?

When studies are designed with control arms, they generally cite data from the literature to estimate the natural history of the disease under standard care assumptions. There is some risk, of course, that the population reported in the literature is unlike that to be prospectively accrued, there may be differences in non-study-related care or outcome ascertainment. Should any of these occur, then the study planning parameters may not hold and there maybe a reason to reconsider the sample size in light of the treatment effect to be measured. Another situation that can occur is when there a large differences that emerge between study arms such that it becomes unethical to continue

to expose the enrolled study participants to the inferior treatment or to offer the possibility of treatment assignment to an inferior treatment for new subjects to be accrued. To make this determination, the study monitoring group must have a high degree of certainty that the difference is real and not simply the randomness of the order in which better or worse outcomes are observed. This high degree of certainty means that the likelihood of falsely concluding there is a difference between the alternative treatments when, in fact, there really is not, is the Type 1 error associated with the study and corresponds to the p value. Thus studies recommended for early termination due to emerging differences generally require much more stringent p values (of the order of $p = 0.001$) than the level of significance for which the overall study is planned (say $p = 0.05$) (9, 10). Terminating a study for lack of a difference between the treatment arms is the mirror of the situation and such a recommendation is based upon the power of the study to detect a difference should there really be one (11). Interim analyses in which there is very little chance of falsely concluding that there is no difference have very low Type 2 error (which is $1 - \text{the study power}$). Many studies are designed to have 80% power (20% Type 2 error) at study conclusion and interim analyses that conclude the “futility” of continuing would generally do so if the Type 2 error was much greater. There are a number of software packages that are available for calculating stopping rules for interim monitoring designs (12, 13).

Summary

There are many approaches to the design of a trial and many of them can achieve certain economies in terms of the required number of participants that need to be enrolled. However, the options are not without their drawbacks and require investigators to make a number of assumptions, many of which cannot be verified or even tested. It is clear that careful consideration needs to be made regarding those assumptions to find the study design that fits the research question the best. However, in doing so it may be possible to select a clinical trial design that is well suited for a specific rare disease and the clinical question that is to be answered.

References

1. Evans CH Jr, Ilstad ST (eds.) Small Clinical Trials: Issues and Challenges. Washington, DC: National Academy Press; 2001.
2. Lagakos S. Clinical trials and rare diseases. *N Engl J Med* 2003;348(24):2455–6.
3. Wilcken B. Rare diseases and the assessment of intervention: What sorts of clinical trials can we use? *J Inherit* 2001;24:294–8.
4. Venance S, Herr BE, Griggs RC. Challenges in the design and conduct of therapeutic trials in channel disorders. *Neurotherapeutics* 2007;4(2):199–204.
5. Kodish E, Lantos JD, Siegler M. The ethics of randomization. *CA Cancer J Clin* 1991;41(3):180–6.
6. Guyatt G, Sackett D, Taylor W, Chong J, Roberts R, Pugsley S. Determining optimal therapy – randomization trials in individual patients. *N Engl J Med* 1986;314(14):889–92.
7. Reitberg DP, Weiss SL, Rio ED. Advances in single-patient trials for drug treatment optimization and risk management. *Drug Inf* 2005;39:119–24.
8. Simon R, Wittes RE, Ellenberg SS. Randomized PHASE II clinical trials. *Cancer Treatment Rep* 1985;69(12):1375–81.

9. Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika* 1983;70(3):659–63.
10. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. *Biometrika* 1977;64:191–9.
11. Goldman B, Crowley J. Interim futility analysis with intermediate endpoints. *Clinical Trials* 2008;5:14–22.
12. PASS [Computer software]. UT: NCSS: Kaysville, 2008.
13. East (Version 5) [Computer software]. Cambridge, MA: Cytel.



<http://www.springer.com/978-1-58829-963-5>

Molecular Basis of Pulmonary Disease

Insights from Rare Lung Disorders

McCormack, F.X.; Panos, R.J.; Trapnell, B.C. (Eds.)

2010, XI, 438 p., Hardcover

ISBN: 978-1-58829-963-5

A product of Humana Press