JOURNAL OF CLINICAL RESEARCH BEST PRACTICES

Vol. 3, No. 6, June 2007 "Can You Handle the Truth?"

Bayesian Designs for Clinical Trials in Early Drug Development

By St. Clare Chung and Miklos Schulz

Introduction

A clinical trial is designed to answer key questions on safety and efficacy of a drug compound. Early-phase trials typically focus on the safety and efficacy aspects separately. This article discusses some innovative approaches to clinical trials that are well-suited to early clinical development. The methodology is primarily Bayesian, which allows for the use of prior and historical information.

Bayesian Statistics

In statistics, there are two schools of thought – Frequentist and Bayesian. Frequentist statistics are backwards-looking – given an existing set of data, what can we conclude? Bayesian statistics are forward-looking – given the data we have collected and plan to collect, what is the likelihood of a future conclusion?

Most traditional clinical trial designs have their roots in Frequentist statistics. When we report results with confidence intervals and p-values, we are describing Frequentist characteristics of the data.

Frequentist statistics utilize prior information formally only in the design of a clinical trial (e.g., sample size calculation) but not in the analysis of the data. In contrast, Bayesian statistics provide a formal mathematical method for combining prior information with current information at the design stage, during the conduct of the trial, and at the analysis stage.

Bayesian statistics has its roots in Bayes' Theorem (named after Thomas Bayes). Bayesian statistics provides a mathematical method for calculating the likelihood of a future event given knowledge from prior events. In doing so, it provides a means for quantifying uncertainty about an unknown parameter of interest. In a clinical trial, parameters of interest may be clinical safety and efficacy endpoints. When the level of uncertainty for a study hypothesis declines below an acceptable threshold, such as 5%, the hypothesis can be accepted. Spiegelhalter, et al., give a good comparison of Frequentist and Bayesian approaches in Table 1.

Table 1. Comparison of Frequentist vs. Bayesian Approaches

Issue	Frequentist	Bayesian
Prior information other than that in the study being analyzed	Informally used in design	Used formally by specifying a prior probability distribution
Interpretation of parameter of interest	A fixed state of nature	An unknown quantity that can have a probability distribution
Basic question	How likely is the data given a particular value of the parameter?	How likely is a particular value of the parameter given the data?
Presentation of	Likelihood functions, p-values	Plots of posterior distribution of

results	and confidence intervals	parameters, calculation of specific posterior probabilities of interest, and use of the posterior distribution in formal decision analysis
Interim analysis	p-values and estimates adjusted for number of analyses	Inference not affected by number of timing of analyses
Interim predictions	Conditional power analysis	Predictive probability of getting a firm conclusion
Dealing with subsets in trials	Adjusted p-values (e.g., Bonferroni)	Subset effects shrunk towards zero by a "skeptical" prior

Clinical Applications of Bayesian Designs

Adaptive designs refer to designs in which design factors can be adjusted during the trial. Design factors that can be adjusted include sample size, dropping ineffective treatment arms, randomization, stopping early rules, etc. The decision is usually based on accumulating data in the current trial but can also accommodate data from other ongoing trials. The Continual Reassessment Method, discussed below, is an example of an adaptive design. Other Bayesian designs can be found in articles by Thall, Sung & Estey (2002) and Thall & Wathen (2005). In the latter paper, the study was designed to address patient heterogeneity (covariates) and multi-stage therapies.

Benefits of adaptive designs include reduction in sample size, the ability to assess predictive probabilities of statistical significance and terminate early if there is no evidence of drug effect, reduction of total drug development time, and earlier availability of long term safety data.

Interest in Bayesian designs have increased for a number of reasons:

- Publication of FDA's draft "Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials"
- Availability of Bayesian designs that are flexible without reducing robustness
- Availability of computational resources to perform the required calculations
- Ability to explore multiple combination treatments in one study
- Effective decision-making with fewer subjects and shorter trials

In early-phase development, when effective decision-making is critical, Bayesian designs may offer a good alternative to traditional designs.

Phase I applications

A primary objective of many Phase I studies is to determine the maximum tolerated dose (MTD) of a new drug that becomes toxic at some level. The traditional design for a Phase I study that targets a 33% dose-limiting toxicity (DLT) (toxic for one-third of the subjects) has many name-incarnations such as the "1-in-3" or "3+3" design. The standard methodology in this type of design is as follows:

1. Treat 3 subjects at the starting dose level D_i
2. If 0 subjects experience dose-limiting toxicity (DLT), escalate to dose D_{i+1}
3. If 1 or more subjects experience DLT, treat 3 more subjects at dose level D_i
4. If 1 of 6 subjects experiences DLT, escalate to dose D_{i+1}

5. If 2 or more subjects experience DLT, MTD = D_{i-1}

Dose escalation stops when one-third of the subjects experience DLT at a given dose level; MTD is next lower dose level.

Although inherently simple to use, this approach lacks flexibility because it is "designed" to estimate MTD as a 33%-tile, which may not be the MTD target of studies in certain indications or subject populations. Other drawbacks of the design include:

- Dose levels must be selected in advance.
- Requirement that 3 subjects be exposed at each dose level, resulting in more subjects than probably necessary being treated at low-toxicity doses, i.e., inefficient use of subjects.
- The planned number of subjects may be exhausted before MTD is found.

Because the design is conservative, it usually underestimates MTD.

Continual Reassessment Method

The continual reassessment method (CRM) introduced by O'Quigley, Pepe and Fisher (1990) was one of the first clinical applications of Bayesian methodology for determining the MTD of a drug molecule. The doses given to subjects were determined by prior and historical data and data obtained from previously-dosed subjects:

1. Determine the range of doses to be explored.
2. Assign a probability of toxicity to each dose based on historical data or investigator input; these probabilities represent prior information and are the starting point for the search for the MTD.
3. Define a model that represents the dose-response relationship; e.g. a tangent hyperbolic or other logistic regression model.
4. Treat subjects at the starting dose, usually the lowest being considered.
5. Observe dose-limiting toxicities.
6. Calculate the next-best estimate of the MTD based on the prior information and the new results from the study.

Based on this approach, subjects are treated up to the dose that currently available evidence indicates to be the best estimate of the MTD. The CRM is flexible. It allows different numbers of subjects to be treated per dose and accommodates specific dose-limiting toxicity rates that are expected in different therapeutic areas.

Maximum Tolerated Schedule (MTS)

Phase I designs usually involve a single administration of the study drug. However, in some situations, investigators may administer a drug repeatedly and monitor effects of toxicity over time. In these cases, the objective of the trial may not be to determine the MTD but rather the MTS of a drug. Braun et al (2005) proposed a new method for accomplishing this objective.

Braun's method uses subject time to toxicity as the outcome, with the hazard of toxicity modeled as the sum of a sequence of hazards, each associated with one administration. The MTS that the subject may receive is based on the risk of toxicity occurring within a specified follow-up period that includes the maximum schedule being considered. Subject accrual, data monitoring, and adaptive decision-making are done continuously throughout the trial under a Bayesian formulation. Each time a new subject is accrued, the most recent data is used to evaluate criteria that define the optimal schedule to which the new subject is assigned.

Phase I/II Applications – Efficacy/Toxicity Tradeoffs

Thall et al. (2004) extended the Phase I paradigm to include not just safety, but also efficacy. Their argument was that Phase I design is too limited and ignores several important factors:

- There may be a lost opportunity to evaluate efficacy as well as safety.
- Assumes that the probability of response is monotone (steadily increasing) with dose.
- May be inefficient when the probability of toxicity is low for all doses but the probability of efficacy response increases with dose. In this case, the superior higher dose may not be found.

The approach of Thall et al. combines the objectives of Phase I and Phase II trials. Subject outcome in the trial is bivariate, consisting of the possible sets of outcomes for toxicity and response. Like the CRM method, the efficacy/toxicity tradeoff method is Bayesian in nature.

The investigator defines the following:

- Lowest acceptable level of response
- Highest acceptable level of toxicity
- 3 equally acceptable combinations of {response, toxicity}: π_1^*, π_2^* and π_3^*

The 3 desirable targets, π_1^*, π_2^* and π_3^*, set up the initial efficacy-toxicity trade-off contours (dimension reduction). The first target, π_1^*, represents the minimum probability of efficacy when toxicity is not expected (i.e. probability of toxicity = 0), while π_2^* is the maximum probability of toxicity when efficacy is certain (i.e. probability = 1). The final target π_3^* is an intermediate point between the smallest-efficacy and largest-toxicity probabilities. These three points define a target contour that is then used to specify desirable probabilities for any pair of probabilities, which in turn then determines the desirability of the doses under investigation. Once this structure has been defined, the dose-finding algorithm is as follows:

- Investigator chooses the starting dose x_i
- Dose x is acceptable if the dose has acceptable probabilities of efficacy and toxicity or if the dose is the lowest untried dose and has an acceptable probability of toxicity.
- Treat each group of subjects at the current most-desirable dose.
- Explore each dose at least once.
- If no dose is acceptable, stop the trial.
- At the end of the trial, select the most desirable dose.

Two key advantages to this approach are that the method (a) reliably finds safe doses with high efficacy and (b) stops if no dose is acceptable, either due to lack of efficacy or safety concerns.

Phase II Proof-of-Concept Studies

The proof-of-concept study is another good opportunity for innovative designs. Phase II proof-of-concept studies employ a small, targeted number of subjects to determine if there is enough evidence of clinical efficacy to warrant full-scale development. Traditional designs may not be appropriate for a number of reasons, such as:

- Unnecessarily exposes excessive subjects to an ineffective arm before concluding futility of the drug.

- Designs are based on a fixed sample size.
- Conclusions can only be drawn once the study is completed.

Designs that address these limitations include two-stage Simon designs, three-stage designs, optimal flexible two-stage designs, and adaptive two-stage designs. As their names suggest, these studies are implemented in stages. At each stage, the data from subjects in the study are examined and a decision is made to stop the study early or to enroll additional subjects into the next stage.

Summary

This article presents Bayesian approaches for designing studies in early drug development. The key difference between Bayesian designs and traditional designs based on the Frequentist approach is the use of prior or historical information. Bayesian designs are also more naturally geared toward decision-making because they use a benefit/risk approach for evaluating each stage of the study. As a result, Bayesian designs provide an efficient and effective approach to evaluating drug compounds during the early phases of drug development.

References

David J Spiegelhalter etc., An Introduction to Bayesian Methods in Health Technology Assessment, *BMJ*, 319, 508-511 (1999)

O'Quigley, J, Pepe, M, and Fisher, L. Continual reassessment method: A practical design for Phase 1 clinical trials in cancer. Biometrics 1990;46, 33–48.

Braun T, Yuan Z, Thall P. Determining the maximum tolerated schedule of a cytotoxic agent. Biometrics, 61, 335

Thall, PF, Cook, J. Dose-finding based on toxicity-efficacy trade-offs. Biometrics 2004; 60, 684–693.

Maurer W. Creative and innovative statistics in clinical research and development. Methods Inf. Meth. 2005;44(4):551-60

Authors

St. Clare Chung is Director of Biostatistics and Data Management at SciAn Services, a provider of large-scale data management, biostatistics and clinical research report-writing services. Contact her at 1.416.231.8008 or schung@scian.com.

Miklos Schulz, Ph.D. is President of SciAn Services, a provider of large-scale data management, biostatistics and clinical research report-writing services. Contact him at 1. 416.231.8008 or miklos.schulz@scian.com