Contemporary Clinical Trials 47 (2016) 40–48

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

Methods for flexible sample-size design in clinical trials: Likelihood, weighted, dual test, and promising zone approaches

Weichung Joe Shih [a,*], Gang Li [b], Yining Wang [b]

[a] Department of Biostatistics, Rutgers School of Public Health, Rutgers University, Piscataway, NJ 08854, United States
[b] Janssen Pharmaceutical Research and Development, Raritan, NJ 08869, United States

ARTICLE INFO

Article history:
Received 24 August 2015
Received in revised form 29 November 2015
Accepted 3 December 2015
Available online 7 December 2015

Keywords:
Sample size re-estimation
Two-stage designs
Flexible clinical trials
Conditional power
Adaptive design
Group-sequential boundary

ABSTRACT

Sample size plays a crucial role in clinical trials. Flexible sample-size designs, as part of the more general category of adaptive designs that utilize interim data, have been a popular topic in recent years. In this paper, we give a comparative review of four related methods for such a design. The likelihood method uses the likelihood ratio test with an adjusted critical value. The weighted method adjusts the test statistic with given weights rather than the critical value. The dual test method requires both the likelihood ratio statistic and the weighted statistic to be greater than the unadjusted critical value. The promising zone approach uses the likelihood ratio statistic with the unadjusted value and other constraints. All four methods preserve the type-I error rate. In this paper we explore their properties and compare their relationships and merits. We show that the sample size rules for the dual test are in conflict with the rules of the promising zone approach. We delineate what is necessary to specify in the study protocol to ensure the validity of the statistical procedure and what can be kept implicit in the protocol so that more flexibility can be attained for confirmatory phase III trials in meeting regulatory requirements. We also prove that under mild conditions, the likelihood ratio test still preserves the type-I error rate when the actual sample size is larger than the re-calculated one.

1. Introduction

Consider a randomized parallel-arm clinical trial. Assume that the observations are normally distributed with means μ_x and μ_y respectively for the two arms and a common variance σ^2. Assume σ^2 is known, and for simplicity, let $\sigma^2 = 1$. Denote $\delta = \mu_x - \mu_y$. We are interested in testing the null hypothesis H_0: $\delta = 0$ versus the one-sided alternative hypothesis H_a: $\delta > 0$. For a traditional fixed sample size design, one would calculate the estimated sample size n_0 for the protocol. When the study is completed, the two-sample test statistic uses the data with the actual sample size, say n^*. In practice, we should have $n^* \approx n_0$, and conditioning on the actual n^* is valid as long as the reason for the minor difference between them has nothing to do with the data itself. For a conventional group-sequential trial, where early stop for efficacy is a feature of the design, the maximal information size n_0 is given in the protocol. The information time/fraction for the interim analyses and the critical values for the interim tests are calculated based on n_0, although this designed n_0 may not, and often does not, coincide with the actually realized information n^*, on which the final test statistic is based. For the situation when early stop occurs due to convincing interim data, conditioning on the observed $n^* < n_0$ for the final test is taken into account in the adjusted critical value.

Now, we consider clinical trials with a flexible sample-size design. For simplicity and practicality, we consider just two stages. In such trials, the outcome information from the first stage influences the sample size of the second stage, thus also influences the sample size of the whole study with both stages combined. Put aside for a moment the issue that the second stage sample size is dependent on the first-stage data and just from the viewpoint of magnitude, a flexible sample-size design is obviously different from the fixed sample-size design; in fact, the designed n_0 can be very different from the actual n^*. It also differs from the conventional group-sequential design in that the initially planned n_0 is not maximal, and that the final sample size can (and often does) exceed n_0. The problem of how to construct a proper sample-size flexible design is the subject of this paper. Many authors have discussed this topic; see, e.g., [1–6] for commentaries and references cited therein. We review and comment on the likelihood ratio test [3,7–9], contrasting it with the weighted test [3,10], the dual test [3,11–13], and the promising zone approach [12,14]. We explore the relationship among these methods by examining in detail the promising zone construction and discuss their merits as useful methods for flexible sample-size trials. All four methods control the type-I error rate and have included the estimation of treatment effect in the literature that we will not review here. We agree that, while the rule of the design is

* Corresponding author at: Department of Biostatistics, Rutgers School of Public Health, Rutgers University, 683 Hoes Lane West, Piscataway, NJ 08854, United States.
E-mail address: w.joe.shih@rutgers.edu (W.J. Shih).

http://dx.doi.org/10.1016/j.cct.2015.12.007
1551-7144/

flexible, the structure of the decision process must be pre-specified for the protocol to follow so that there is no ambiguity for the regulatory agencies to examine the validity of the statistical procedure [15–17]. We delineate what is necessary to specify in the study protocol and what can be kept implicit to attain more flexibility for confirmatory phase III trials in meeting regulatory requirements. We give numerical scenarios to illustrate the application of these methods. Discussion of more regulatory and operational issues can be found in [17–19].

2. General framework for flexible sample-size design

Following the notation in Section 1, a clinical trial is designed with an initial sample size of n_0 patients per group given in the protocol. At the first/interim stage when data from n_1 patients in each group are available, we calculate the sample means $\bar{x}_1$ and $\bar{y}_1$ of the two groups, and let $\hat{\delta}_1 = \bar{x}_1 - \bar{y}_1$ and $Z_1 = \frac{\hat{\delta}_1}{se(\hat{\delta}_1)}$. It is common for a sequential design to consider possible early termination of a study for either futility or efficacy at the interim stage. For given constants h and k, we plan to (i) reject H_0 and terminate the trial if $z_1 > k$, (ii) accept H_0 and terminate the trial if $z_1 < h$, or (iii) continue the trial to the second (final) stage if $h \le z_1 \le k$. For the path (iii), the task is to determine an additional n_2 number of patients per group and a critical value c for the final test so that the overall type I error rate is preserved at the prescribed level α. In the literature, different forms of the final test and associated formula for n_2 and c have been discussed. Throughout the following, we write $n_2(z_1)$ and n_2 interchangeably; the former emphasizes the fact that n_2 depends on z_1 for flexible sample-size designs.

3. The likelihood ratio test

The likelihood ratio test (LRT) for flexible sample-size design was discussed initially in Li et al. [7,8]. It was an improvement over Proschan and Hunsberger [20] by relaxing the need for their special error functions or a function of combining p-values as in Bauer and Köhne [21]. All that is needed is the conditional power function itself. Müller and Schäfer [22] noted that their conditional rejection probability function corresponds to the conditional probability function of [20] for the 2-stage design case. Hence [22] is consonant with [7,8]. Denote $\hat{\delta}_2 = \bar{x}_2 - \bar{y}_2$ and $Z_2 = \frac{\hat{\delta}_2}{se(\hat{\delta}_2)}$ based on the second stage samples. Z_2 is defined only if the study continues. At the end of the trial, the Wald test statistic is $Z = \frac{n_1(\bar{X}_1 - \bar{Y}_1) + n_2(\bar{X}_2 - \bar{Y}_2)}{\sqrt{2(n_1+n_2)}} = \frac{\sqrt{n_1}Z_1 + \sqrt{n_2(Z_1)}Z_2}{\sqrt{n_1+n_2(Z_1)}}$ based on $n = n_1 + n_2$ patients per group. References [7,8] showed that the one-sided LRT is $Z > c$ for some constant c and derived solutions for n_2 and c via the conditional power approach.

The conditional probability for the final likelihood ratio test to be significant is

$$CP_\delta(n_2, c|z_1) \equiv \mathrm{P}(Z > c|z_1, \delta) = 1 - \Phi\left[\frac{c\sqrt{2(n_1+n_2)} - z_1\sqrt{2n_1} - n_2\delta}{\sqrt{2n_2}}\right]. \quad (1)$$

The conditional power (1) for given n_2 and c is conditioning on two quantities: the assumed treatment effect size δ for Stage 2 data and the observed $Z_1 = z_1$ from Stage 1 data. The treatment effect size can be based on several considerations and is up to the choice of the researcher. For example, it can be the originally hypothesized value, observed value at Stage 1 or the lower bound of a confidence interval, or some combination of them, perhaps even with other external information or opinion of a clinical meaningful effect that needs to be detected. When a design aims to provide a conditional power (CP) of $1 - \beta_1$ for detecting the current trend $\delta = \hat{\delta}_1$ at the final stage given the interim result $h \le z_1 \le k$, Li et al. [7] derived

$$n_2(z_1) \ge \max\left\{\min\left(n_{2\max}, \left[\left(\frac{c + z_{\beta_1}}{z_1}\right)^2 - 1\right]n_1\right), n_{2\min}\right\}, \quad (2)$$

where $n_{2\max}$ is the maximum resource-allowable and $n_{2\min}$ is the minimum sample sizes for the second stage (usually $n_{2\min} = n_0 - n_1$), and c is solved in

$$1 - \Phi(h) - \alpha = \int_h^k \Phi\left[\frac{c\sqrt{n_1 + n_2(z_1)} - z_1\sqrt{n_1}}{\sqrt{n_2(z_1)}}\right]\phi(z_1)dz_1 \quad (3a)$$

by numerical integration, for the given set of design parameters $\alpha, \beta_1, h, k, n_{2\max}$ and $n_{2\min}$. $\Phi(.)$ is the cumulative distribution function and $\phi(.)$ is the density function of the standard normal, and $z_{\beta_1} = \Phi^{-1}(1 - \beta_1)$.

3.1. Comments

3.1.1. Point 1

Futility is usually regarded as an internal business decision for the manufacturers, thus health agencies often view the boundary h non-binding to the manufacturer. In this case, we replace h by $-\infty$ in Eq. (3a) as an option; that is, set

$$1 - \alpha = \int_{-\infty}^k \Phi\left[\frac{c\sqrt{n_1 + n_2(z_1)} - z_1\sqrt{n_1}}{\sqrt{n_2(z_1)}}\right]\phi(z_1)dz_1. \quad (3b)$$