A two-stage design with two co-primary endpoints

James X. Song
100 Tice Blvd, Woodcliff Lake, NJ 07677, USA

ABSTRACT

Two-stage designs are commonly used in phase II oncology trial to mitigate the risk of exposing patients to an inefficacious drug. Typically, the decision of moving into stage 2 enrollment is made based on response rate in stage 1 patients; and trials are designed in the hypothesis testing framework. When the primary objective of a trial involves more than one efficacy endpoints it is desirable to extend the two-stage design to a setting that accommodates two hypotheses while controlling overall type I and II errors ($\alpha$ and $\beta$). In this manuscript, we propose a simple method of stopping boundaries of both hypotheses simultaneously that satisfy $\alpha$ and $\beta$ constrains using binomial distribution. Several design characteristics of these selected boundaries are further examined in order to choose the most desirable design based on an objective function. Simulation is used to confirm the results. A trial design in metastatic breast cancer where both response rate and health-related quality of life are of interest is used as an example of the application of the proposed method. In conclusion, the proposed design is an extension of Simon Two-Stage Design. It can be applied to phase II oncology trials with two independent co-primary efficacy endpoints.

© 2015 The Author. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The objective of a phase II oncology trial is to assess the antitumor activity of a new drug and thus to determine whether it warrants further study in the phase III setting. It is common not to include a control arm in these trials in order to gain evidence on the efficacy and safety as quickly as possible and to focus on a relatively rapidly observable endpoint such as tumor objective response rate (ORR). An interim futility analysis is often desirable to limit the number of patients exposed in case of low activity.

Several authors [1–4] proposed two-stage designs in which a fixed number of patients are accrued in stage 1 ($n_1$) and in total ($n$). The study is continued into stage 2 only when interim boundary ($r_1$) is crossed. Typically, one tests the null hypothesis $H_0$: $p = p_0$ against the alternative $H_A$: $p = p_A$, where $p$ is the probability of response, $p_0$ is the response rate below which one considers the drug insufficiently active and $p_A$ is the assumed response rate of the new drug. In a well-known two-stage design, Simon proposed method of choosing $n_1$, $n$, $r_1$ and boundary at final analysis ($r$) via binomial distribution when $p_0$, $p_A$, type I and II errors ($\alpha$ and $\beta$) are specified. A numerical search is conducted to find all designs that have actual $\alpha$ and $\beta$ no more than specified levels. Among them, Simon suggested the design that minimizes the expected sample size under $p_0$ (optimal) and the one that minimizes the maximum sample size (minimax).

A number of extensions to Simon Two-Stage Design have been proposed. Green and Dahlberg [5] proposed several approaches to adapting interim stopping rules when the actual sample size is not the planned size. Jung et al. [6] developed a family of two-stage designs that are admissible according to a Bayesian loss function in which Simon optimal and minimax designs are special cases. Sargent et al. [7] extended hypothesis testing in two-stage design to three possible outcomes: reject $H_0$, reject $H_A$ or reject neither. Jones and Holmgren [8] modified the design to allow an opportunity to assess whether a drug is active in a disease population as a whole or in a targeted subset of biomarker positive patients. Bayesian two-stage designs have been proposed by a number of authors. Sam-bucini [9] proposed method of deriving an adaptive two-stage design, where the second stage sample size is not selected in advance, but depends on the first stage results via Bayesian predictive design. Tan and Machin [10] proposed new Bayesian designs which allow for the incorporation of relevant prior information and the presentation of the trial results in a manner which is more intuitive and helpful.

In oncology trials, besides ORR, efficacy outcomes such as disease control rate (DCR), defined as percentage of patients with best overall response of stable disease or better; and
progression-free survival rate at a specific time point (e.g., 12-week PFS) are also useful measures of anticancer activities. Various measures of health-related quality of life (HRQL), including European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30), Functional Assessment of Cancer Therapy – General (FACT-G) and their disease specific modules are widely accepted as endpoints in clinical trials. Consequently, achieving improvement or maintaining stability in HRQL measures by a specific time point could also be considered as clinical benefit. In addition, dichotomous outcomes based on pharmacodynamic markers, dosing feasibility or a subgroup of special interest are often included in one of the primary objectives of the phase II trials. Therefore, it is desirable to extend Simon Two-Stage Design to a setting that accommodates two null hypotheses. In this manuscript, we propose a two-stage design that includes a 2nd hypothesis, H0: p' = p0 vs. H0A: p' = pA, such that both hypothesis tests are carried out within two-stage design framework while controlling overall type I and II errors.

2. Methods

In the current design, we describe a method of searching for boundaries of the 1st hypothesis (r1 and r) and the 2nd hypothesis (r1' and r') simultaneously using binomial distribution. The trial will move into stage 2 if the number of successes in the first endpoint (x1) is greater than r1 or the number of successes in the second endpoint (y1) is greater than r1' in the first n1 patients. In the final analysis of n patients, the first H0 will be rejected if more than r successes are observed in the first endpoint (x); and the second H0 will be rejected if more than r' successes are observed in the second endpoint (y). To better illustrate the proposed approach, we assume n1 and n are given since both are often pre-fixed due to operational constrain and logistic convenience.

This design differs from combining two independent Simon Two-Stage Designs since a different interim decision rule is used. Instead of evaluating each endpoint independently and accepting H0 when it fails to pass interim boundary, hypothesis tests in both endpoints are carried out at final analysis as long as stage 1 boundary is crossed in at least one endpoint. This testing strategy is more intuitive and flexible than making separate interim decisions within each endpoint. Otherwise, one may face the dilemma of accepting H0 when failing to pass stage 1 boundary whereas the final boundary is crossed.

According to the decision rules described above, the probability of accepting H0 is P(X1 ≤ r1 ∩ Y1 ≤ r1') + P(X ≤ r ∩ Y ≤ r'|X1 > r1 ∩ Y1 > r1'). Under the independence assumption of two-points, it is expressed as

\[
B(r_1, p, n_1)B(r_1', p', n_1) + \sum_{x=0}^{\min(n_1, r_1)} \sum_{y=0}^{\min(n_1, r_1')} b(x, p, n_1)B(r - x, p, n - n_1) \\
- n_1 b(y, p', n_1)B(r' - y, p', n) \\
- n_1 \text{ for } (x, y) \in \{x > r_1 \text{ or } y > r_1'\}
\]  

(1)

where b and B denote the binomial density and cumulative functions. Using Eq. (1), given n1, n, p, p', any sets of boundaries r1, r, r1' and r' that satisfy a and b can be obtained.

For a selected design, the probability of accepting H0 of a particular endpoint can be calculated. For example, P(accept H0: p = p0) = P(X ≤ r1 ∩ Y1 ≤ r1') + P(X ≤ r1 ∩ Y1 > r1' - P(X1 ≤ r1 ∩ Y1 ≤ r1') + P(X ≤ r1 ∩ Y1 > r1' - P(X1 ≤ r1 ∩ Y1 ≤ r1') + P(X ≤ r1 ∩ Y1 > r1' - P(X1 ≤ r1 ∩ Y1 ≤ r1')). Under the independence assumption, it is written as

\[
B(r_1, p, n_1)B(r_1', p', n_1) + \sum_{x=0}^{\min(n_1, r_1)} b(x, p, n_1)B(r - x, p, n - n_1) \\
+ \sum_{x=0}^{r_1} b(x, p, n_1)B(r - x, p, n - n_1)\{1 - B(r_1', p', n_1)\}
\]  

(2)

The type I error in the first hypothesis, z1, is calculated according to Eq. (2) when p = p0 and p' = p0'. The corresponding type II error, β1, is calculated when p = pA and p' = pA'. The type I and II errors in the second hypothesis (z2 and β2) can be obtained similarly.

Further statistical criteria need to be defined using type I and II errors of each endpoint calculated in Eq. (2) in order to select an appropriate design from those satisfying Eq. (1). Here, we regard an ideal design as one with comparable powers in the two hypothesis tests both at interim and final analyses, while controlling type I errors in each endpoint. Since overall type I error of each endpoint is controlled in Eq. (1) (i.e., z2), additional type I error control is achieved by excluding designs with unacceptable high type I errors at interim analysis (∝11 and ∝21). Next, a function of type II errors at the interim analysis (β11 and β21) and overall type II errors (β1 and β2) is evaluated. Such an objective function can be defined as

\[
S(β_1, β_2, β_{11}, β_{21}) = w(β_1^2 + β_2^2)^{1/2} + (1 - w)(β_{11}^2 + β_{21}^2)^{1/2}
\]  

(3)

where w is a constant weight between 0 and 1. Hence, S(β1, β2, β11, β21) is a weighted function of square root of sum of square of type II errors at final and square root of sum of square of type II errors at interim analysis. The goal is to identify designs that minimize type II errors of each endpoint at final and interim analyses. w can be set between ½ and 1 (e.g. 2/3) since typically minimizing the first part of function S is of most importance, therefore, carrying more weight. The design with the minimal S is recommended.

3. An example

The method was motivated by a planned phase II trial in metastatic breast cancer. ORR and percentage of patients without deterioration in Global Health Status of EORTC QLQ-C30 (GHS) in the first two cycles of treatment were two key efficacy variables of interest. We set up H0: ORR = 5% vs. HA: ORR = 15% and H0: GHS = 45% vs. HA: GHS = 60%. Sample sizes were determined beforehand as n1 = 15 and n = 55. Using Eq. (1), 636 sets of boundaries satisfying overall α ≤ 0.05 and β ≤ 0.2 were identified. Upon further inspection, some of the designs showed undesirable characteristics such as extreme high type I or II errors in one hypothesis test and extreme low probability of early termination after the first stage under null hypotheses. Therefore, designs with overly high type I errors at interim analysis (∝11 > 0.7 or ∝21 > 0.7) were excluded. To further select a sensible design, S(β1, β2, β11, β21) with w = 2/3 was calculated for the remaining 253 designs. The one with smallest value of S was chosen (r1 = 0, r = 6, r1' = 7 and r' = 31). With these boundaries, the probabilities of early stopping under the null and alternative hypotheses were estimated as 0.3028 and 0.0186, respectively. The expected sample sizes under the null and alternative hypotheses were 43 and 54, respectively. Within each test, ∝1 = 0.0183 and ∝11 = 0.2701 in testing the 1st hypothesis of ORR; and ∝2 = 0.0311 and β2 = 0.3440 in testing the 2nd hypothesis of GHS. The overall trial-wise α and β were 0.0488 and 0.1001, respectively.

Simulation was used to confirm the results. One hundred thousand trials were run using random samples generated based
on the parameters specified in the example. The same decision rules described in the Methods section were applied to simulate the trial results. The simulation results were very close to the ones described above (within 1% error).

4. Conclusions

A simple approach for modification of the Simon Two-Stage Design is proposed to accommodate phase II trials where two co-primary endpoints are of interest. Sample sizes in stage 1 and total are pre-specified to simplify the decision making. But the algorithm can be easily expanded to find designs for given \( \alpha \) and \( \beta \) in a fixed range of \( n_1 \) and \( n \). As illustrated in the example, it is recommended to first identify all the designs that satisfy trial-wise \( \alpha \) and \( \beta \) in Eq. (1). Subsequent selection of the stopping boundary should be made based on other operating characteristics that are calculated using exact binomial probabilities. Since two hypothesis tests are included, admissible designs including Simon optimal and minimax designs developed in one hypothesis testing [6] are not applicable. Instead of balancing between expected sample size and maximum sample size, it is desirable to have comparable type I or II errors between two tests, so each hypothesis can be properly tested. For trials having different treatment effects between the two endpoints, as shown in the example, it may not be an easy objective to achieve. Given the numerous parameters involved, we first screened out the designs with high type I errors at interim, then attempted to define an objective function \( S(\beta_1, \beta_2, \beta_{11}, \beta_{21}) \) to select the design that minimizes the type II errors. One may calculate \( S(\beta_1, \beta_2, \beta_{11}, \beta_{21}) \) with several different weights in order to better assess its impact; and therefore to choose the most appropriate design. If the primary objective of the design is to stop the trial early in case of ineffective treatment (type I error control), rather than maximizing powers, a different strategy of design selection should be employed. One might first screen out the designs with unacceptable high type II errors at interim analysis, then select a design with minimal value of \( S(\alpha_1, \alpha_2, \alpha_{11}, \alpha_{21}) \), where \( S(\alpha_1, \alpha_2, \alpha_{11}, \alpha_{21}) \), a function of type I errors, can be set up similar as \( S(\beta_1, \beta_2, \beta_{11}, \beta_{21}) \).

The approach assumes independence of two test statistics in the endpoints. In clinical trial, it is not uncommon that the two efficacy endpoints of interest are correlated. Future work will include extension of proposed method to two correlated endpoints.

Acknowledgment

The author wants to thank two anonymous reviewers for their helpful comments.

References

[1] E.A. Gehan, The determination of the number of patients required in a preliminary and follow-up trial of a new chemotherapeutic agent, J. Chronic Dis. 13 (1961) 346–353.
[2] Y.J. Lee, M. Staquet, R. Simon, R. Catane, F. Muggia, Two stage plans for patient accrual in phase II cancer clinical trials, Cancer Treat. Rep. 63 (1979) 1721–1726.
[3] T.R. Fleming, One sample multiple testing procedure for phase II clinical trials, Biometrics 38 (1982) 143–151.
[4] R. Simon, Optimal two-stage designs for phase II clinical trials, Control. Clin. Trials 10 (1989) 1–10.
[5] S.J. Green, S. Dahlberg, Planned versus attained design in phase II clinical trials, Stat. Med. 11 (1992) 853–862.
[6] S.H. Jung, T. Lee, K.M. Kim, S.L. George, Admissible two-stage design for phase II cancer clinical trials, Stat. Med. 23 (2004) 561–569.
[7] D.J. Sargent, V. Chan, G.M. Goldberg, A three-outcome design for phase II clinical trials, Control. Clin. Trials 22 (2001) 117–125.
[8] C.J. Jones, E. Holmgren, An adaptive simon two-stage design for phase 2 studies of target therapies, Contemp. Clin. Trials 28 (2007) 654–661.
[9] V.A. Sambucini, Bayesian predictive strategy for an adaptive two-stage design in phase II clinical trials, Stat. Med. 29 (2010) 1430–1442.
[10] S.B. Tan, D. Machin, Bayesian two-stage designs for phase II clinical trials, Stat. Med. 21 (2002) 1991–2012.