Published in final edited form as:

Commun Stat Simul Comput. 2018 ; 47(4): 1157–1167. doi:10.1080/03610918.2017.1307398.

Optimal inference for Simon's two-stage design with over or under enrollment at the second stage

Guogen Shan^{1,*} and John J. Chen²

¹Epidemiology and Biostatistics Program School of Community Health Sciences, University of Nevada Las Vegas, Las Vegas, NV 89154

²Biostatistics Core, John A. Burns School of Medicine, University of Hawaii, Honolulu, Hawaii 96813

Abstract

Simon's two-stage designs are widely used in clinical trials to assess the activity of a new treatment. In practice, it is often the case that the second stage sample size is different from the planned one. For this reason, the critical value for the second stage is no longer valid for statistical inference. Existing approaches for making statistical inference are either based on asymptotic methods or not optimal. We propose an approach to maximize the power of the study while maintaining the type I error rate where the type I error rate and power are calculated exactly from binomial distributions. The critical values of the proposed approach are numerically searched by an intelligent algorithm over the complete parameter space. It is guaranteed that the proposed approach is at least as powerful as the conditional power approach which is a valid but non-optimal approach. The power gain of the proposed approach can be substantial as compared to the conditional power approach. We apply the proposed approach to a real Phase II clinical trial.

Keywords

Adaptive design; Clinical trials; Conditional power; Sample size; Simon's two-stage design

1. Introduction

Simon's two-stage designs [1] have been widely used in Phase II clinical trials, especially in Oncology studies, to assess the activity of a new treatment or a therapy in a one-arm study. The primary endpoint of such studies is the clinical response that often includes complete response and partial response according to the guidelines, such as the standard Response Evaluation Criteria in Solid Tumors (RECIST) guidelines [2]. Simon's designs provide the sample size and the critical values for both stages, and the study has to be followed precisely with the planned first stage and final sample sizes in order to make proper statistical inference regarding the activity of the treatment.

*Corresponding author. Epidemiology and Biostatistics Program, Department of Environmental and Occupational Health, School of Community Health Sciences, University of Nevada Las Vegas, Las Vegas, NV 89154, USA. address: guogen.shan@unlv.edu.

In practice, the number of enrolled patients may be deviated from the planned study due to patient dropout, limited study time, budget cut, or other reasons. It is common that the actual sample size of a study for the second stage is different from the planned one while the first stage is followed as planned [3, 4, 5, 6, 7, 8, 9]. For example, a Phase II clinical trial of Imatinib in AIDS-associated Kaposi's Sarcoma (AIDS-KS) [10], the first stage is conducted as originally planned. When the trial moves to the second stage, investigators chose to enroll more patients in order to achieve enough evaluable patients after accounting for the possible dropout during the study period. At the end of study, no dropout actually occurred, thus the study is over enrolled in the second stage. When the study does not followed the design precisely, the original critical values are no longer valid for proper statistical inference.

In order to address the over or under enrollment at the second stage in practice, Green and Dahlberg [11] was among the first to compare several approaches to modify the critical values when the study is not followed exactly. They also presented a simple approach motivated by a clinical study from Southwest Oncology Group, but the type I error rate is not guaranteed. Wu and Shih [12] investigated multiple scenarios where the actual study deviates from the pre-specified Simon's two-stage design by using conditional error functions. Additional interim analysis is involved in this approach, thus, the type I error rate needs to be redefined in order to control for the overall type I error rate. Later, Koyama and Chen [3] proposed a proper statistical inference for Simon's two-stage design when the second stage sample size is different from that planned. They used the concept of conditional power to determine the new critical values with the new second stage sample size while the type I error rate is respected. Recently, Zeng et al. [13] proposed a normal approximation to maximize the power of a study with the type I error rate respected. The type I error rate and power from the study of Zeng et al. [13] were calculated using the asymptotic approach. Li et al. [14] proposed a Bayesian approach for a two-stage design when the attained sample sizes in both stages deviate from the original design. This approach was shown to approximately maintain the frequentist properties of the design.

In this article, we propose an approach to maximize the power of a study whose second stage sample size is difference from the planned one from Simon's design. The study will be terminated in the first stage for futility only as Simon's. The type I error rate and power are computed from the exact binomial distributions of data. The remaining of this article is organized as follows. In Section 2, we introduce the conditional power approach by Koyama and Chen [3], and propose the optimal approach to maximize power of a study. A Phase II clinical trial from a AIDS-associated Kaposi's Sarcoma study is used to illustrate the application of the proposed approach in Section 3. In Section 4, we compare the performance of the proposed approach and the conditional power approach under various conditions. In Section 5 we discuss the findings in this article and some future work.

2. Optimal statistical inference

Simon's two-stage design is widely used in clinical trials to assess the activity of a new treatment by comparing the estimated response rate to a historical response rate of the standard therapy, π_0 . The hypotheses are often formulated as $\pi \leq \pi_0$ against the alternative

$\pi \geq \pi_1$, where π_1 is the anticipated response rate of the new treatment. At a given type I and II error rates, (α, β) , Simon's two-stage design is indexed with four numbers,

$$(R_1/n_1, R_t/n_t)$$

to test the aforementioned hypothesis, where n_1 and n_t are the planned sample sizes for the first stage and both stages combined, and R_1 and R_t are the associated critical values to make the statistical inference at the first stage and the end of the study. When R_1 or more responders among n_1 patients are observed in the first stage, the study goes to the second stage with additional $n_2 = n_t - n_1$ patients. Out of total n_t patients, if R_t or more number of responses are obtained, the null hypothesis will be rejected and the trial moves forward to the next phase. Therefore, the rejection region of Simon's two-stage design is defined as

$$\Omega = \{(x_1, x_t): x_1 \geq R_1, \text{ and } x_t \geq R_t\},$$

where x_1 and x_t are the number of responses from the first stage and both stages combined. It follows that the number of responses in the second stage is $x_2 = x_t - x_1$.

In practice, it is common to find a study with the second stage sample size being different from that planned using Simon's two-stage approach [3]. One typical example is a multi-center clinical trial where the enrollment information shared with each center may not be updated timely. This could result in the over enrollment in the study. Another example is a study with a high drop-out rate during the second stage, leading to the under enrollment situation. When over or under enrollment in the study occurs in the second stage, the planned critical value, R_t , is no longer valid due to the sample size change. Several approaches have been developed for the proper statistical inference. For example, Koyama and Chen [3] used the concept of conditional power to determine the new critical values under the new second stage sample size, n'_2 , while the type I error rate is controlled for. We first review this conditional approach, then propose the optimal approach for the hypothesis testing in such situations.

2.1 Conditional power approach

Conditional error function for observing x_2 responses out of n_2 patients given x_1 responses from the first stage can be expressed as

$$\varphi(x_2|n_2, X_1 = x_1, \pi) = P_\pi(X_2 \geq x_2|n_2, X_1 = x_1) = \sum_{i=x_2}^{n_2} \binom{n_2}{i} \pi^i (1-\pi)^{n_2-i},$$

where X_1 and X_2 are the number of responses in the first stage and the second stage, respectively. The study will be terminated at the first stage if $x_1 < R_1$, thus $\varphi(x_2|n_2, X_1 = x_1, \pi) = 0$ when $x_1 < R_1$. For Simon's design, the final critical value R_t is the same regardless the number of responses observed from the first stage as long as it is greater

than or equal to R_1 . It follows that the critical value for the second stage is, $R_2(x_1) = R_t - x_1$ when x_1 patients respond to the treatment in the first stage. It is obvious that $R_2(x_1)$ is a decreasing function of x_1 . When $x_1 < R_1$, the study is stopped for futility in the first stage, and the second stage critical value $R_2(x_1)$ is set as 0.

In the case with over or under enrollment in the second stage with a new sample size, n'_2 , instead of the originally planned sample size n_2 , new critical values for the second stage need to be determined. Simon's design controls for the type I error rate as

$$\sum_{x_1=R_1}^{n_1} b(x_1|n_1, \pi_0) \times \varphi(R_2(x_1)|n_2, X_1 = x_1, \pi_0) \leq \alpha,$$

where $R_2(x_1) = R_t - x_1$, and $b(x_1|n_1, \pi_0)$ is the probability density function for a binomial distribution. The general idea of Koyama and Chen [3] is to bound each new conditional error function, $\varphi(R'_2(x_1)|n'_2, X_1 = x_1, \pi_0)$, by its associated value from the original design, specifically,

$$\varphi(R'_2(x_1)|n'_2, X_1 = x_1, \pi_0) \leq \varphi(R_2(x_1)|n_2, X_1 = x_1, \pi_0), \quad 1$$

for each x_1 from R_1 to n_1 . To increase the power of the study, the smallest $R'_2(x_1)$ value is identified for each x_1 under the new second stage sample size n'_2 . It is easy to show that the study with the new critical values guarantees the type I error rate. It has been pointed out that $R'_2(x_1)$ is generally a non-increasing function of x_1 as in Simon's design [13]. The new final critical value is computed as $R'_t(x_1) = R'_2(x_1) + x_1$, which is not always a constant as observed in the traditional Simon's design [13]. With this new second stage sample size n'_2 , the new critical values $R'_t(x_1), x_1 = R_1, R_1 + 1, \dots, n_1$, are then used to make the inference.

2.2. Optimal approach

In the conditional error function, the number of responses from the second stage, X'_2 , follows a binomial distribution with parameters, n'_2 and π . This is a discrete distribution, and the conditional error function $\varphi(R'_2(x_1)|n'_2, X_1 = x_1, \pi_0)$ could be much less than $\varphi(R_2(x_1)|n_2, X_1 = x_1, \pi_0)$, that would then lead to the test being conservative with the actual type I error rate being much less than the nominal level. We propose an approach to numerically search for the critical values, $R'_t(x_1), (x_1 = R_1, R_1 + 1, \dots, n_1)$, over all possible values. The critical value set, $(R'_t(R_1), R'_t(R_1 + 1), \dots, R'_t(n_1))$, with the maximum power is the one referred to as the optimal critical value set. The range for these critical values are from R_1 to $n_1 + n'_2$. Then, the naive grid search has to compute type I error rate and power for a

total of $(n_1 + n'_2 - R_1 + 1)^{n_1 - R_1 + 1}$ times. The computation cost becomes expensive quickly as the sample size increases in the study.

In order to overcome the computational intensity, we utilize the branch-and-bound intelligent algorithm [15] to search for the optimal critical set. This algorithm numerically search for the optimal critical value set over all possible sets. The optimal set is the one with the maximum power given the constraints. Power is the target function, which needs to be maximized,

$$f = \sum_{x_1=R_1}^{n_1} b(x_1|n_1, \pi_0) \times \varphi(R'_2(x_1)|n'_2, X_1 = x_1, \pi_1)$$

Two constraints considered here are: the type I error rate

$$g = \sum_{x_1=R_1}^{n_1} b(x_1|n_1, \pi_0) \times \varphi(R'_2(x_1)|n'_2, X_1 = x_1, \pi_0) \leq \alpha$$

and the monotonic property of the second stage critical values

$$R'_2(R_1) \geq R'_2(R_1 + 1) \geq \dots \geq R'_2(n_1)$$

The first constraint is used to control for the type I error rate of the study. It should be noted that the type I error rate and power in this article are computed exactly using binomial distributions, not based on asymptotic distributions. The second constraint is generally satisfied as in other existing approaches [3, 13].

The branch-and-bound algorithm is an algorithm for solving a one-dimensional optimization problem. The parameters in the target function are $n_1 - R_1 + 1$ by $(n_1 + n'_2 - R_1 + 1)$ dimensions. Therefore, we need to reduce the parameter space to one dimension in order to use the algorithm. We calculate all possible attainable conditional error functions for x_1 from R_1 to n_1 at $\pi = \pi_0$ and π_1 . These conditional error functions at $\pi = \pi_0$ and π_1 are the new parameters, which have a one-to-one relationship with the original parameters. The new onedimensional parameter is used in the search algorithm. As the name of the algorithm implies that two procedures are included in the branch-and-bound algorithm, the branching procedure and the bounding procedure. They are performed iteratively in order to find the maximum of the objective function. The branching procedure is used to split the problem into two subproblems at the current node. In this problem, each x_1 value from R_1 to n_1 is considered as a node. The other procedure, the bounding procedure, is used to discard the subproblems that do not lead to the optimal solution based on the aforementioned two constraints. Discarding the unpromising subproblems in each branching procedure is the key of the branch-and-bound algorithm in finding the optimal design efficiently.

This algorithm is able to identify the optimal critical values, $R'_l(x_1)$, $(x_1 = R_1, R_1 + 1, \dots, n_1)$, such that the power of the study is maximized while the type I error rate is met and the monotonic property is satisfied. We use a real Phase II clinical trial to illustrate the application of this new approach in following section.

3. Example

We consider a Phase II clinical trial of Imatinib in AIDS-associated Kaposi's Sarcoma (AIDS-KS) [10] to illustrate the application of the optimal approach for the situation with over enrollment in the second stage. AIDS-KS study was a multicenter clinical trial with the primary objective to estimate the response rate to Imatinib. Simon's optimal two-stage design was used for designing the clinical trial with $\pi_0 = 20\%$ and $\pi_1 = 50\%$ at the significance level of $\alpha = 0.05$ to attain 90% power. The 20% response rate under the null hypothesis was estimated from the control group in a randomized, double-blinded, placebo-controlled Phase III clinical trial [16]. The desired response rate, $\pi_1 = 50\%$, is the one expected from Imatinib in AIDS-KS.

Simon's two-stage optimal design is computed as

$$(R_1/n_1, R'_l/n_l) = (3/10, 8/22)$$

The overall response is defined as a complete response or a partial response according the RECIST guidelines [2]. As designed, ten patients will be enrolled in the first stage of the study. With three or more responses are observed at the end of the first stage, the trial moves on to the second stage with additional enrollment of $n_2 = n - n_1 = 12$ patients. The activity of Imatinib in AIDS-KS at the end of the trial will be confirmed if 8 or more responses are observed from the total 22 patients, otherwise, no further investigation of Imatinib in AIDS-KS is warranted. In practice, the number of patients enrolled in the first stage and the second stage were $n_1 = 10$ and $n'_2 = 20$, respectively. In the second stage, 20 patients were enrolled in order to guarantee at least additional 12 evaluable patients after excluding patients who might be lost to follow up or discontinue intervention. Three responses were observed from the first stage, and at the end of the study, ten responses were observed from the total 30 patients. The study was over enrolled due to additional patients in the second stage. The number of responses observed were more than the originally critical value $R_l = 8$. Since the more patients were treated and followed than originally planned, the critical value $R_l = 8$ is no longer valid.

The final response rate is estimated as $10/30 = 33\%$, with the confidence interval provided based on an approach for a binomial proportion. Only the final results, $X = 10$ and $n = 30$, were used in the hypothesis testing in the article of Koon et al. [10]. This estimated response rate is biased [17]. One obvious reason is that the results from the first stage are not utilized in the statistical inference. If the 95% one-side confidence interval is used for statistical inference by using the final results, $X = 10$ and $n = 30$, it is computed as 0.194 to 1.000 based on Blaker's exact approach [18]. Then, $\pi_0 = 20\%$ is covered in the confidence

interval, and one fails to reject the null hypothesis at the significance level of $\alpha = 0.05$. This approach does not fully use the information from the study, e.g., the first stage results, the number of responses from the second stage.

The proposed optimal approach is considered for the hypothesis testing, and this approach uses the results from both stages. Table 1 presents the critical values $R'_2(x_1)$ using the proposed optimal approach, for an over enrolled study with $n'_2 = 20$ in the second stage instead of $n_2 = 12$ as originally planned. As 3 responses observed in the first stage, the associated second stage critical value is $R'_2(X_1 = 3) = 7$ and the final critical value $R'_t(X_1 = 3) = 10$. That is, with 10 or more responses observed at the end of the study, the activity of Imatinib in AIDS-KS is proved. In the current study, out of the total 30 patients, 10 responses are sufficient enough to demonstrate the activity of the drug.

4. Numerical study

We compare the performance of the proposed optimal approach and the conditional approach by Koyama and Chen [3] (referred to be as the KC approach) under various conditions. The first stage critical value is adopted from Simon's design for both approaches. The second stage critical value for the KC approach is computed from Equation (1), which is computational easy but is not optimal. The proposed optimal approach determines the second stage critical value by enumerating all possible critical values based on the intelligent search algorithm as described in Section 2.2. Table 2 presents the critical values for Simon's optimal design with design parameters $(\alpha, \beta, \pi_0, \pi_1) = (0.05, 0.2, 0.2, 0.4)$, $(R_1/n_1, R_t/n_t) = (4/13, 13/43)$ and the second stage sample size $n'_2 = 1.5n_2 = 45$. The trial stops at the first stage when X_1 is less than or equal to 3. When this occurs, no additional patients will be enrolled in the second stage for both approaches, therefore, the associated critical values for the second stage and final stage are zero. Both approaches provides the critical values for the second stage when 4 or more responses are observed from the first stage. For example, in the case with 4 responses observed in the first stage, 13 or more responses from $n'_2 = 45$ patients are needed in order to claim the activity of the treatment by the KC approach, while 12 or more responses for the proposed approach.

We investigate the scenarios with the targeted response rate, π_1 15%, 20%, or 30% higher than the historical response rate, π_0 . Type I and II error rates are set as $\alpha = 0.05$ and $\beta = 0.2$, respectively. For each design parameter setting, $(\alpha, \beta, \pi_0, \pi_1)$, Simon's optimal and minimax two-stage designs are provided in the form as $(R_1/n_1, R_t/n_t)$. When the second stage sample size is changed from the original n_2 to n'_2 , we first determine the critical values for both approaches, then calculate the actual type I error rate and power for them based on the computed critical values. Table 3 presents the actual type I error rate and power of these two competitive approaches when $n'_2 = 1.5n_2$. Since the proposed approach finds the optimal critical value set such that power of the study is maximized, and the critical value set from the KC approach is only a candidate of the optimal critical value set, it is guaranteed that the power of the proposed approach is higher than or equal to that of the KC approach. Among these 18 cases studied, the KC approach and the proposed approach have the same power in

only two cases. With additional $0.5n_2$ subjects in the study, it would be reasonable to expect that power of the study should be more than the original level of 80%, however, we find that the KC approach has power less than 80% in two cases and another three cases with power very close to the desired power. It is noticeable that the KC approach is generally conservative as the actual type I error rates being much less than the nominal level, which could lead to the unsatisfied power.

With doubling the number of patients in the second stage, $n'_2 = 2n_2$, the power of the study using the KC approach is improved to all over 80% (Table 4). The actual type I error rate is still much lower than α for the KC approach, and power of the KC approach is still much less than that of the proposed approach. After investigating the over enrollment cases, we also study the cases of under enrollment in the second stage with $n'_2 = 2/3n_2$. The results of the KC approach and the proposed approach are presented in Table 5. Power gain of the proposed approach as compared to the KC approach, ranges from 0% to 17.1%. In the minimax design for $(\alpha, \beta, \pi_0, \pi_1) = (0.05, 0.2, 0.2, 0.5)$, the proposed approach has power of 77.4% while the KC approach only has power of 60.3%. Lastly, we present the results of both approaches when the study is conducted as designed with $n'_2 = n_2$ in Table 6. It would be easy to show that the design from the KC approach would be the same as that from Simon's. The proposed approach generally has the same design as the KC approach except for one case. In the case for the minimax design with design parameters $\pi_0 = 0.1$ and $\pi_0 = 0.3$ in Table 6, power of the new approach is higher than the KC approach, and they have the sample power for all the other cases. The critical values are $R_1 = 2$ and $R_1 = 6$ for the first stage and the two stages combined, respectively. Therefore, $R'_2(X_1) = 6$ in the KC approach when $X_1 = 2$ with the actual type I error rate of 0.033. In the new proposed approach, a design with $R'_2(X_1 = 2) = 5$ and $R'_2(X_1) = 6$ when $X_1 = 3$, has the actual type I error rate of 0.048, and its power is 82.6%. Due to the discreteness of binomial distributions, the actual type I error rate could be already over the nominal level with a small decrease in the critical value.

5. Discussion

The proposed approach provides the optimal critical values for the second stage outcome given the first stage result. R statistical software program [19] is used to compute these optimal critical values, and it is available upon request from the first author. The power of the study will be maximized over all possible critical values for each first stage response by using the branch-and-bound intelligent algorithm. The monotonic ordering the second stage critical value is intuitive and consistent with existing approaches [3, 13, 5, 20, 21, 22, 5, 23, 24]. The power gain of the proposed approach as compared to the KC approach is generally substantial. As pointed out by the reviewer, since the proposed approach is generally more powerful than the KC approach, another constraint for the relationship between the critical values from two approaches can be potentially added in the design search: $R'_2(X_1)$ of the proposed approach is less than or equal to that of the KC approach. This would be an interesting topic for future research.

In addition to the critical value for make statistical inference, confidence interval is another important inference procedure. Often, one sided confidence interval is computed due to the one sided hypothesis for the problem [3, 10]. Obviously, it is not appropriate to compute the confidence interval only based on the final number of responses and the total sample size. A traditional approach by inverting the hypothesis test based on the central limit theorem could be used to construct the confidence interval. This is an asymptotic approach. We would consider the exact approach [18], or the exact approach based on an inductive function [25]. This would be our future work to develop exact confidence intervals for the response rate in a two-stage design setting.

Acknowledgment

We would like to thank the Associate Editor and the referee for their valuable comments and suggestions that helped to improve this manuscript. Shan's research is supported by grants from the National Institute of General Medical Sciences 5U54GM104944, P20GM109025, and P20GM103440 from the National Institutes of Health. Chen's research is partially supported by U54MD007584, G12MD007601, P20GM103466, and U54GM104944 from the National Institutes of Health.

References

- [1]. Simon R Optimal two-stage designs for phase II clinical trials. *Controlled clinical trials* 3 1989; 10(1):1–10. URL <http://view.ncbi.nlm.nih.gov/pubmed/2702835>. [PubMed: 2702835]
- [2]. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, Verweij J, Van Glabbeke M, van Oosterom AT, Christian MC, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *Journal of the National Cancer Institute* Feb 2000; 92(3):205–216, doi:10.1093/jnci/92.3.205. URL <http://dx.doi.org/10.1093/jnci/92.3.205>.
- [3]. Koyama T, Chen H. Proper inference from Simon's two-stage designs. *Statistics in medicine* 7 2008;27(16):3145–3154. URL <http://view.ncbi.nlm.nih.gov/pubmed/17960777>. [PubMed: 17960777]
- [4]. Shan G Exact Statistical Inference for Categorical Data. 1 edn., Academic Press, 2015 URL <http://www.worldcat.org/isbn/0081006810>.
- [5]. Shan G, Ma C. Unconditional tests for comparing two ordered multinomials. *Statistical methods in medical research* 2 2016; 25(1):241–254, doi:10.1177/0962280212450957. URL [PubMed: 22700600]
- [6]. Shan G, Chen JJ, Ma C. Boundary problem in Simon's two-stage clinical trial designs. *Journal of Bio-pharmaceutical Statistics* 2 2016; :nulldoi:10.1080/10543406.2016.1148716. URL <http://dx.doi.org/10.1080/10543406.2016.1148716>.
- [7].] Shan G, Zhang H, Jiang T. Minimax and admissible adaptive two-stage designs in phase II clinical trials. *BMC Medical Research Methodology* 8 2016; 16(1):90+, doi:10.1186/s12874-016-0194-3. URL .10.1186/s12874-016-0194-3. [PubMed: 27485595]
- [8]. Shan G, Ma C, Hutson AD, Wilding GE. Randomized Two-Stage Phase II Clinical Trial Designs Based on Barnard's Exact Test. *Journal of Biopharmaceutical Statistics* 8 2013; 23(5):1081–1090, doi:10.1080/10543406.2013.813525. URL 10.1080/10543406.2013.813525. [PubMed: 23957517]
- [9]. Shan G, Wilding GE, Hutson AD, Gerstenberger S. Optimal adaptive two-stage designs for early phase II clinical trials. *Statist. Med.* 4 2016; 35(8):1257–1266, doi:10.1002/sim.6794. URL <http://dx.doi.org/10.1002/sim.6794>.
- [10]. Koon HB, Krown SE, Lee JY, Honda K, Rapisuwon S, Wang Z, Aboulafia D, Reid EG, Rudek MA, Dezube BJ, et al. Phase II trial of imatinib in AIDS-associated Kaposi's sarcoma: AIDS Malignancy Consortium Protocol 042. *Journal of clinical oncology* 2 2014; 32(5):402–408. URL <http://view.ncbi.nlm.nih.gov/pubmed/24378417>. [PubMed: 24378417]

- [11]. Green SJ, Dahlberg S. Planned versus attained design in phase II clinical trials. *Statistics in medicine* 5 1992; 11(7):853–862. URL <http://view.ncbi.nlm.nih.gov/pubmed/1604065>. [PubMed: 1604065]
- [12]. Wu Y, Shih WJ. Approaches to handling data when a phase II trial deviates from the pre-specified Simon's two-stage design. *Statistics in medicine* 12 2008; 27(29):6190–6208, doi:10.1002/sim.3426. URL <http://dx.doi.org/10.1002/sim.3426>. [PubMed: 18800338]
- [13]. Zeng D, Gao F, Hu K, Jia C, Ibrahim JG. Hypothesis testing for two-stage designs with over or under enrollment. *Statistics in medicine* 3 2015; URL <http://view.ncbi.nlm.nih.gov/pubmed/25809924>.
- [14]. Li Y, Mick R, Heitjan DF. A Bayesian approach for unplanned sample sizes in phase II cancer clinical trials. *Clinical trials (London, England)* 6 2012; 9(3):293–302, doi: 10.1177/1740774512443429. URL <http://dx.doi.org/10.1177/1740774512443429>.
- [15]. Wolsey LA. *Integer Programming*. 1 edn., Wiley-Interscience, 1998 URL <http://www.worldcat.org/isbn/0471283665>.
- [16]. Noy A, Scadden DT, Lee J, Dezube BJ, Aboulafia D, Tulpule A, Walmsley S, Gill P. Angiogenesis Inhibitor IM862 Is Ineffective Against AIDS-Kaposi's Sarcoma in a Phase III Trial, but Demonstrates Sustained, Potent Effect of Highly Active Antiretroviral Therapy: From the AIDS Malignancy Consortium and IM862 Study Team. *Journal of Clinical Oncology* 2 2005; 23(5):990–998, doi:10.1200/jco.2005.11.043. URL <http://dx.doi.org/10.1200/jco.2005.11.043>. [PubMed: 15598977]
- [17]. Whitehead J On the bias of maximum likelihood estimation following a sequential test. *Biometrika* Dec 1986; 73(3):573–581, doi:10.1093/biomet/73.3.573. URL 10.1093/biomet/73.3.573.
- [18]. Blaker H Confidence curves and improved exact confidence intervals for discrete distributions. *Can J Statistics* Dec 2000; 28(4):783–798, doi:10.2307/3315916. URL 10.2307/3315916.
- [19]. Shan G, Wang W. ExactCIdiff: An R Package for Computing Exact Confidence Intervals for the Difference of Two Proportions. *The R Journal* 2013; 5(2):62–71.
- [20]. Shan G Exact sample size determination for the ratio of two incidence rates under the Poisson distribution. *Computational Statistics* 2016; 31(4):1633–1644, doi:10.1007/s00180-016-0654-6. URL <http://dx.doi.org/10.1007/s00180-016-0654-6>.
- [21]. Shan G, Zhang H. Exact unconditional sample size determination for paired binary data (letter commenting: J Clin Epidemiol. 2015;68:733–739.). *Journal of Clinical Epidemiology* 2016; In press.
- [22]. Shan G, Zhang H, Jiang T, Peterson H, Young D, Ma C. Exact p-Values for Simon's Two-Stage Designs in Clinical Trials 2016; 8(2):351–357, doi:10.1007/s12561-016-9152-1. URL
- [23]. Shan G Exact confidence limits for the response rate in two-stage designs with over or under enrollment in the second stage. *Statistical Methods in Medical Research* 2016; In press.
- [24]. Shan G, Moonie S, Shen J. Sample size calculation based on efficient unconditional tests for clinical trials with historical controls. *Journal of biopharmaceutical statistics* 2016; 26(2):240–249. URL <http://view.ncbi.nlm.nih.gov/pubmed/25551261>. [PubMed: 25551261]
- [25]. Wang W On construction of the smallest one-sided confidence interval for the difference of two proportions. *The Annals of Statistics* Apr 2010; 38(2):1227–1243, doi:10.1214/09-aos744. URL 10.1214/09-aos744.

Table 1:

The new critical values $R'_2(x_1)$ for the clinical trial with over enrollment in the second stage, $n'_2 = 20$ instead of $n_2 = 12$ as originally planned.

x_1	2	3	4	5	6	7	8	9	10
$R'_2(x_1)$	0	7	6	5	4	3	2	1	1
$R'_t(x_1)$	0	10	10	10	10	10	10	10	11

Critical values of the KC approach and the proposed approach when the second stage sample size is $n'_2 = 1.5n_2 = 45$ instead of $n_2 = 30$ from Simon's design with parameters $(\alpha, \beta, \pi_0, \pi_1) = (0.05, 0.2, 0.2, 0.4)$.

Table 2:

First stage response: x_1	3	4	5	6	7	8	9	10	11	12	13
Critical value: KC approach											
$R'_2(x_1)$	0	13	12	11	9	8	7	6	4	3	1
$R'_l(x_1) = R'_2(x_1) + x_1$	0	17	17	17	16	16	16	16	16	15	14
Critical value: Optimal approach											
$R'_2(x_1)$	0	12	12	11	10	9	8	7	6	5	4
$R'_l(x+1) = R'_2(x_1) + x_1$	0	16	17	17	17	17	17	17	17	17	17

Table 3:

The actual type I error rate and power of a study with over enrollment in the second stage with $n'_2 = 1.5n_2$.

π_0	π_1	Design (R_1/n_1 , R_2/n_2)	n_2	n'_2	KC approach		Optimal approach	
					TIE	Power	TIE	Power
0.05	0.2	Minimax (1/13,4/27)	14	21	0.025	0.817	0.047	0.850
		Optimal (1/10,4/29)	19	28	0.033	0.828	0.033	0.828
0.1	0.25	Minimax (3/22,8/40)	18	27	0.024	0.808	0.047	0.866
		Optimal (3/18,8/43)	25	37	0.034	0.818	0.048	0.830
0.2	0.35	Minimax (7/31,16/53)	22	33	0.038	0.826	0.049	0.845
		Optimal (6/22,20/72)	50	75	0.042	0.822	0.049	0.824
0.1	0.3	Minimax (2/15,6/25)	10	15	0.025	0.832	0.049	0.857
		Optimal (2/10,6/29)	19	28	0.026	0.811	0.049	0.832
0.2	0.4	Minimax (5/18,11/33)	15	22	0.037	0.827	0.049	0.842
		Optimal (4/13,13/43)	30	45	0.038	0.816	0.048	0.820
0.3	0.5	Minimax (7/19,17/39)	20	30	0.035	0.834	0.048	0.854
		Optimal (6/15,19/46)	31	46	0.042	0.826	0.050	0.832
0.05	0.35	Minimax (1/6,3/12)	6	9	0.006	0.803	0.030	0.888
		Optimal (1/4,3/16)	12	18	0.014	0.811	0.047	0.820
0.1	0.4	Minimax (2/8,4/13)	5	7	0.014	0.762	0.045	0.853
		Optimal (1/4,4/15)	11	16	0.032	0.841	0.032	0.841
0.2	0.5	Minimax (3/9,7/17)	8	12	0.018	0.794	0.036	0.855
		Optimal (3/8,7/18)	10	15	0.023	0.805	0.047	0.837

Table 4:

The actual type I error rate and power of a study with over enrollment in the second stage with $n'_2 = 2n_2$.

π_0	π_1	Design (R_1/n_1 , R_2/n_2)	n_2	n'_2	KC approach		Optimal approach	
					TIE	Power	TIE	Power
0.05	0.2	Minimax (1/13,4/27)	14	28	0.021	0.842	0.047	0.896
		Optimal (1/10,4/29)	19	38	0.029	0.854	0.049	0.862
0.1	0.25	Minimax (3/22,8/40)	18	36	0.028	0.864	0.048	0.894
		Optimal (3/18,8/43)	25	50	0.038	0.843	0.049	0.851
0.2	0.35	Minimax (7/31,16/53)	22	44	0.039	0.865	0.050	0.885
		Optimal (6/22,20/72)	50	100	0.040	0.831	0.049	0.833
0.1	0.3	Minimax (2/15,6/25)	10	20	0.023	0.861	0.047	0.909
		Optimal (2/10,6/29)	19	38	0.033	0.838	0.048	0.843
0.2	0.4	Minimax (5/18,11/33)	15	30	0.033	0.847	0.046	0.867
		Optimal (4/13,13/43)	30	60	0.037	0.826	0.045	0.828
0.3	0.5	Minimax (7/19,17/39)	20	40	0.034	0.865	0.050	0.885
		Optimal (6/15,19/46)	31	62	0.042	0.841	0.050	0.843
0.05	0.35	Minimax (1/6,3/12)	6	12	0.010	0.874	0.044	0.912
		Optimal (1/4,3/16)	12	24	0.025	0.82	0.025	0.820
0.1	0.4	Minimax (2/8,4/13)	5	10	0.024	0.846	0.049	0.860
		Optimal (1/4,4/15)	11	22	0.029	0.859	0.028	0.858
0.2	0.5	Minimax (3/9,7/17)	8	16	0.027	0.86	0.036	0.880
		Optimal (3/8,7/18)	10	20	0.028	0.837	0.050	0.848

Table 5:

The actual type I error rate and power of a study with under enrollment in the second stage with $n'_2 = 2/3n_2$.

π_0	π_1	Design (R_1/n_1 , R_2/n_2)	n_2	n'_2	KC approach		Optimal approach	
					TIE	Power	TIE	Power
0.05	0.2	Minimax (1/13,4/27)	14	9	0.022	0.663	0.044	0.717
		Optimal (1/10,4/29)	19	12	0.021	0.646	0.048	0.674
0.1	0.25	Minimax (3/22,8/40)	18	12	0.021	0.660	0.047	0.774
		Optimal (3/18,8/43)	25	16	0.034	0.705	0.042	0.740
0.2	0.35	Minimax (7/31,16/53)	22	14	0.037	0.693	0.049	0.743
		Optimal (6/22,20/72)	50	33	0.038	0.730	0.050	0.758
0.1	0.3	Minimax (2/15,6/25)	10	6	0.014	0.637	0.046	0.775
		Optimal (2/10,6/29)	19	12	0.033	0.706	0.046	0.751
0.2	0.4	Minimax (5/18,11/33)	15	10	0.033	0.703	0.050	0.753
		Optimal (4/13,13/43)	30	20	0.038	0.740	0.040	0.745
0.3	0.5	Minimax (7/19,17/39)	20	13	0.031	0.690	0.049	0.744
		Optimal (6/15,19/46)	31	20	0.031	0.699	0.050	0.748
0.05	0.35	Minimax (1/6,3/12)	6	4	0.011	0.729	0.049	0.780
		Optimal (1/4,3/16)	12	8	0.015	0.747	0.015	0.747
0.1	0.4	Minimax (2/8,4/13)	5	3	0.018	0.698	0.018	0.698
		Optimal (1/4,4/15)	11	7	0.017	0.666	0.017	0.666
0.2	0.5	Minimax (3/9,7/17)	8	5	0.012	0.603	0.042	0.774
		Optimal (3/8,7/18)	10	6	0.023	0.661	0.038	0.747

Table 6:

The actual type I error rate and power of a study followed by Simon's design with $n'_2 = n_2$.

π_0	π_1	Design ($R_1/n_1, R_2/n_2$)	n_2	n'_2	KC approach		Optimal approach	
					TIE	Power	TIE	Power
0.05	0.2	Minimax (1/13,4/27)	14	14	0.042	0.801	0.042	0.801
		Optimal (1/10,4/29)	19	19	0.047	0.801	0.047	0.801
0.1	0.25	Minimax (3/22,8/40)	18	18	0.040	0.803	0.040	0.803
		Optimal (3/18,8/43)	25	25	0.048	0.800	0.048	0.800
0.2	0.35	Minimax (7/31,16/53)	22	22	0.050	0.802	0.050	0.802
		Optimal (6/22,20/72)	50	50	0.049	0.800	0.049	0.800
0.1	0.3	Minimax (2/15,6/25)	10	10	0.033	0.802	0.048	0.826
		Optimal (2/10,6/29)	19	19	0.047	0.805	0.047	0.805
0.2	0.4	Minimax (5/18,11/33)	15	15	0.046	0.801	0.046	0.801
		Optimal (4/13,13/43)	30	30	0.050	0.800	0.050	0.800
0.3	0.5	Minimax (7/19,17/39)	20	20	0.045	0.804	0.045	0.804
		Optimal (6/15,19/46)	31	31	0.050	0.803	0.050	0.803
0.05	0.35	Minimax (1/6,3/12)	6	6	0.018	0.822	0.018	0.822
		Optimal (1/4,3/16)	12	12	0.027	0.803	0.027	0.803
0.1	0.4	Minimax (2/8,4/13)	5	5	0.031	0.802	0.031	0.802
		Optimal (1/4,4/15)	11	11	0.043	0.818	0.043	0.818
0.2	0.5	Minimax (3/9,7/17)	8	8	0.034	0.806	0.034	0.806
		Optimal (3/8,7/18)	10	10	0.039	0.800	0.039	0.800