Published in final edited form as:

Stat Biosci. 2016 ; 8(2): 351–357. doi:10.1007/s12561-016-9152-1.

Exact p-values for Simon's two-stage designs in clinical trials

Guogen Shan¹, Hua Zhang², Tao Jiang^{2,*}, Hanna Peterson¹, Daniel Young¹, and Changxing Ma³

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

²Zhejiang Gongshang University, Hangzhou, Zhejiang, China, 310018

³Department of Biostatistics, University at Buffalo, Buffalo, NY 14214

Abstract

In a one-sided hypothesis testing problem in clinical trials, the monotonic condition of a tail probability function is fundamentally important to guarantee that the actual type I and II error rates occur at the boundary of their associated parameter spaces. Otherwise, one has to search for the actual rates over the complete parameter space, which could be very computationally intensive. This important property has been extensively studied in traditional one-stage study settings (e.g., non-inferiority or superiority between two binomial proportions), but there is very limited research for this property in a two-stage design setting, e.g., Simon's two-stage design. In this note, we theoretically prove that the tail probability is an increasing function of the parameter in Simon's two-stage design. This proof not only provides theoretical justification that p-value occurs at the boundary of the parameter space, but also helps to reduce the computational intensity for study design search.

Keywords

Boundary; Mathematical induction; Monotonic condition; Simon's two-stage design

1 Introduction

Two-stage designs are widely used in practice to protect participants when the investigated treatment is indeed ineffective. When the outcome is binary, Simon's optimal two-stage designs (Simon, 1989) are the most popular designs that have been frequently used in research, such as cancer studies (Siefker-Radtke et al., 2013), AIDS research studies (Zheng et al., 2012) and gastroesophageal research studies (Katz, Gerson, & Vela, 2013).

In these studies, the new treatment is tested by conducting a one-arm study, and its activity is confirmed when a large number of responses is observed. The associated hypotheses are presented as

*Corresponding author. Department of Statistics, Zhejiang Gongshang University, Hangzhou, Zhejiang, China, 310018 jtao@263.net.

$$H_0: p \leq p_0, \text{ against } H_a: p \geq p_1,$$

where p_0 and p_1 represent the unacceptable and acceptable response rates for the new treatment, respectively. The acceptable response rate, p_1 , is the target response rate that the new treatment could reach, and p_0 can be estimated from historical data. In such studies, a treatment with a high response rate is often preferable; thus $p_1 > p_0$ is assumed.

In a two-stage design, suppose X_i is the number of responses from the i -th stage, $i = 1$ and 2 . Then, the tail area is defined as the collection of samples with a larger X_1 and a larger $X_1 + X_2$ as compared to the observed sample. The tail probability is calculated as the sum of probabilities for all sample points in the tail area. Since X_1 and X_2 follow binomial distributions with the success rate of p , the tail probability is a function of the parameter p . By the definition of p-value, it should be computed as the worst case scenario of the probability over the null space: $\{p: p \leq p_0\}$. Due to the complexity of the tail probability in a two-stage setting, it is difficult to show that the tail probability is an increasing function of p . In practice, some researchers assume that the maximum of the tail probability occurs at the boundary without any theoretical proof, and others simply change the one-sided hypothesis to a simple hypothesis to test the response rate at two values: p_0 and p_1 . Obviously, these approaches to simplify the problem are not suitable to make proper statistical inference for a study in a two-stage setting.

In this work, we are going to prove the monotonic condition of the tail probability in a two-stage setting: the tail probability is an increasing function of the parameter p . The monotonic condition in a one-stage setting has been studied extensively for binomial distributions (Röhmel, 2005; Röhmel & Mansmann, 1999; Shan & Ma, 2016). Recently, Shan, Chen, and Ma (2016) partially prove this condition in a two-stage setting.

This research is motivated by the work by Tsai, Chi, and Chen (2008) who developed a new exact confidence interval for the response rate in a two-stage design by using the Clopper-Pearson exact approach (Clopper & Pearson, 1934). This exact approach has to be used in conjunction with a method to order the sample space. They ordered the sample space by the sufficient test statistic $X_t = X_1 + X_2$, which is the total number of responses (Jung & Kim, 2004). They showed that the cumulative distribution function of X_t is a decreasing function of p . This property must be satisfied in order to construct the exact interval based on the Clopper-Pearson approach (Tsai et al., 2008). They have only one random variable in the problem, but we have two random variables, X_1 and X_2 in the p-value calculation. The rest of this note is organized as follows. In Section 2, we prove the monotonic condition of the tail probability in a two-stage setting by using mathematical induction. Section 3 provides some discussion and remarks.

2 Main results

In Simon's two-stage design, four design parameters need to be determined (R_1, n_1, R, n_2) given α, β, p_0 and p_1 . It should be noted that the design parameters are derived from the exact probability calculation for the actual type I and II error rates by using binomial

distributions. R_1 and R are the boundary values for the number of responses using n_1 and $n = n_1 + n_2$ subjects from the first stage and both stages combined, respectively. If $X_1 < R_1$, then a study will be terminated due to non-sufficient activity of the treatment. Otherwise, an additional n_2 subjects will be considered in stage two. The final conclusion will be made by comparing the total number of responses X_t and R . A new treatment is considered promising for further investigation when $X_t \geq R$.

The null hypothesis will be rejected when $X_1 \geq R_1$ and $X_t \geq R$, and the rejection region is presented as

$$\Phi = \{(X_1, X_2) : X_1 \geq R_1, (X_1 + X_2) \geq R\}.$$

It follows that the actual type I error (TIE) is defined as the maximum of the tail probability, $P(R_1, n_1, R, n_2 | p) = P((X_1, X_2) \in \Phi | p)$ over the null space $\{p : p \leq p_0\}$,

$$\max_{p \in H_0} P(R_1, n_1, R, n_2 | p). \quad (1)$$

The tail probability $P(R_1, n_1, R, n_2 | p)$ can be calculated exactly by using binomial distributions as

$$P(R_1, n_1, R, n_2 | p) = \sum_{X_1=R_1}^{\min(n_1, R-1)} b(X_1, n_1, p) [1 - B(R - X_1 - 1, n_2, p)] + \sum_{X_1=\min(n_1, R-1)+1}^{n_1} b(X_1, n_1, p),$$

where $b(\cdot)$ and $B(\cdot)$ are the probability density function and cumulative distribution function of a binomial distribution, respectively.

Due to the intense computation of the TIE in Equation (1), it is usually assumed that the actual TIE rate is obtained at the boundary of the null space H_0, p_0 . This note fills the gap to prove the monotonic condition of the tail probability, $P(R_1, n_1, R, n_2 | p)$, for Simon's two-stage design.

Theorem 2.1

The monotonic condition of the tail probability in a two-stage setting, $P(R_1, n_1, R, n_2 | p)$, is satisfied for any given critical values R_1 and R where $0 \leq R_1 \leq n_1$ and $R_1 \leq R \leq n_1 + n_2$. In other words, $P(X_1 \geq R_1, (X_1 + X_2) \geq R | p)$ is an increasing function of p .

Proof—The number of responses from the second stage, X_2 , follows a binomial distribution with parameters (n_2, p) . Then, X_2 can be expressed as $X_2 = B_1 + B_2 + \cdots + B_{n_2}$, where B_i follows a Bernoulli distribution with the success rate of p , $i = 1, 2, \dots, n_2$, and these B_i 's are independent from each other. Therefore, the target function can be rewritten as

$$P(X_1 \geq R_1, (X_1 + X_2) \geq R | p) = P(X_1 \geq R_1, (X_1 + B_1 + B_2 + \cdots + B_{n_2}) \geq R | p)$$

We use mathematical induction (Tsai et al., 2008) for this proof. The first step is to prove that $P(X_1 \geq R_1, (X_1 + B_1) \geq R|p)$ is an increasing function of p when $n_2 = 1$. There are only two possible outcomes of B_1 , 1 with the probability of p , and 0 with the probability of $1 - p$. We have

$$\begin{aligned} P(X_1 \geq R_1, (X_1 + B_1) \geq R) &= P(X_1 \geq R_1) \times P(B_1 = 0) + P(X_1 \geq R_1, X_1 \geq R - 1) \times P(B_1 = 1) \\ &= P(X_1 \geq R_1) \times (1 - p) + P(X_1 \geq \max(R_1, R - 1)) \times p \end{aligned}$$

Based on the above equation and the constraint $R \leq R_1$, we consider two cases according to $R = R_1$ and $R = R_1 + 1$. In the first case with $R = R_1$, the tail probability can be written as follows,

$$P(X_1 \geq R_1, (X_1 + B_1) \geq R) = P(X_1 \geq R) \times (1 - p) + P(X_1 \geq R) \times p = P(X_1 \geq R).$$

Its derivative is calculated as

$$\frac{\partial P(X_1 \geq R)}{\partial p} = \binom{n_1}{R - 1} (n_1 - R + 1) p^{R-1} (1 - p)^{n_1 - R}.$$

Since $R = R_1 \leq n_1$, $(n_1 - R + 1)$ is always positive. The binomial coefficient $\binom{n_1}{R - 1}$ is positive when $R \leq 1$, and zero when $R = 0$. It follows that the derivative is positive when $R \leq 1$ and zero when $R = 0$. In practice, a study design with $R_1 = R = 0$ is not realistic because any study under these critical values is guaranteed to be successful regardless the number of responses observed. After excluding this unrealistic case, the derivative is always positive, then $P(X_1 \geq R_1, (X_1 + B_1) \geq R)$ is an increasing function of p .

In the second case with $R = R_1 + 1$, we have

$$P(X_1 \geq R_1, (X_1 + B_1) \geq R) = P(X_1 \geq R) \times (1 - p) + P(X_1 \geq R - 1) \times p.$$

In the Appendix C by Tsai et al. (Tsai et al., 2008), they showed that $P(X_1 \geq R) \times (1 - p) + P(X_1 \geq R - 1) \times p$ is an increasing function of p by using the relationship between binomial distribution function and incomplete beta function. Therefore, $P(X_1 \geq R_1, (X_1 + B_1) \geq R)$ is an increasing function of p .

In the next step of mathematical induction, we assume that $P(X_1 \geq R_1, (X_1 + B_1 + B_2 + \dots + B_k) \geq R)$ is an increasing function of p , where $R \leq R_1$, and we need to check the monotonic property of $P(X_1 \geq R_1, (X_1 + B_1 + B_2 + \dots + B_k + B_{k+1}) \geq R)$. It is easy to show that

$$\begin{aligned} P(X_1 \geq R_1, (X_1 + B_1 + \dots + B_k + B_{k+1}) \geq R) \\ = P(X_1 \geq R_1, (X_1 + B_1 + \dots + B_k) \geq R) \times (1 - p) + P(X_1 \geq R_1, (X_1 + B_1 + \dots + B_k) \geq R - 1) \times p. \end{aligned}$$

Its derivative is

$$\begin{aligned}
& \frac{\partial P(X_1 \geq R_1, (X_1 + B_1 + \dots + B_k + B_{k+1}) \geq R)}{\partial p} \\
&= \frac{\partial P(X_1 \geq R_1, (X_1 + B_1 + \dots + B_k) \geq R)}{\partial p} (1-p) - P(X_1 \geq R_1, (X_1 + B_1 + \dots + B_k) \geq R) \\
&+ \frac{\partial P(X_1 \geq R_1, (X_1 + B_1 + \dots + B_k) \geq R-1)}{\partial p} + P(X_1 \geq R_1, (X_1 + B_1 + \dots + B_k) \geq R-1)
\end{aligned}$$

It is obvious that $P(X_1 \geq R_1, (X_1 + B_1 + \dots + B_k) \geq R-1) - P(X_1 \geq R_1, (X_1 + B_1 + \dots + B_k) \geq R) = P(X_1 \geq R_1, (X_1 + B_1 + \dots + B_k) \geq R-1) - 0$. By the assumption that $P(X_1 \geq R_1, (X_1 + B_1 + \dots + B_k) \geq R)$ is an increasing function of p , we know that

$$\frac{\partial P(X_1 \geq R_1, (X_1 + B_1 + \dots + B_k) \geq R)}{\partial p} > 0. \text{ For } \frac{\partial P(X_1 \geq R_1, (X_1 + B_1 + \dots + B_k) \geq R-1)}{\partial p},$$

we need to consider two cases: $R-1 \geq R_1$ and $R = R_1$. In the first case with $R-1 \geq R_1$, based on the assumption when $n_2 = k$, this partial derivative is positive. In the second case with $R = R_1$, we have

$$P(X_1 \geq R_1, (X_1 + B_1 + \dots + B_k) \geq R-1) = P(X_1 \geq R_1, X_1 \geq R-1 - (B_1 + \dots + B_k)) = P(X_1 \geq R).$$

Similar to the proof in the first step, we have shown that $P(X_1 \geq R)$ is an increasing function of p . Thus, $P(X_1 \geq R_1, (X_1 + B_1 + B_2 + \dots + B_k + B_{k+1}) \geq R)$ is an increasing function of p .

Hence by mathematical induction, $P(R_1, n_1, R, n_2|p)$ is an increasing function of p when $R_1 \geq 1$.

It has been proved that the tail probability $P(R_1, n_1, R, n_2|p)$ is an increasing function of p when the trial goes to the second stage with the tail area as $\Phi = \{(X_1, X_2) : X_1 \geq R_1, (X_1 + X_2) \geq R\}$. When a study is stopped in the first stage, the following remark is used to show that the monotonic condition is still satisfied.

Remark 2.1

In the case that a study has a very small number of responses in the first stage, the study could be stopped earlier. In this case, the tail area is

$$\Phi = \{(X_1, X_2) : X_1 \geq R_1\}.$$

It follows that the tail probability is

$$P(R_1, n_1|p) = P(X_1 \geq R_1).$$

It has been shown in Theorem 2.1 that $P(X_1 \geq R_1)$ is an increasing function of p when $R_1 \geq 1$. When $R_1 = 0$, the tail probability is 1, which is independent of the parameter p .

Based on the results from Theorem 2.1 and Remark 2.1, the tail probability is an increasing function of p . This monotonic condition of the tail probability is very important to guarantee

that the actual type I and II error rates occur at the boundary of their associated parameter spaces.

3 Discussion

In this note, we prove the monotonic condition of the tail probability in a two-stage setting. Then, p-value and power of a study occur at the boundary of the null parameter space and the alternative parameter space, respectively. This proof is very important to reduce the computational intensity to find an optimal two-stage design.

Recently, adaptive two-stage designs are being increasingly developed and used in practice to make traditionally used study designs flexible and effective. In these adaptive designs, the second stage sample size depends on the number of responses observed from the first stage. The sample sizes in the first stage and the second stage are pre-determined to guarantee the overall type I and II error rates. The study design is presented as

$$(n_1, R_1, n_2(X_1), R_t(X_1)), \text{ where } X_1 = R_1, R_1 + 2, \dots, n_1.$$

If $X_1 < R_1$, then the study is stopped in the first stage. When $X_1 \geq R_1$, the trial proceeds to the second stage with the sample size $n_2(X_1)$. It is reasonable to assume that $n_2(X_1)$ is a non-increasing function of X_1 when $X_1 \geq R_1$. This could be considered as a natural constraint in a study design because if there are more responders in the first stage, then less participants are needed in the second stage.

Currently, most researchers assume that the tail probability function of adaptive designs (Shan, Wilding, Hutson, & Gerstenberger, 2016; Kang & Tian, 2013; Kang, Xiong, Crane, & Tian, 2013) is an increasing function of p during the design search. Although the monotonic condition of the tail probability for the final optimal design is often checked, it is very difficult and time consuming to check the monotonic condition for every possible design during the design search. The extension of proving the monotonic condition from Simon's two-stage design to adaptive two-stage designs is not trivial as more design parameters are involved. Therefore, we consider this as future work to further investigate the monotonic condition of the tail probability in adaptive two-stage designs.

Tsou et al. (Tsou, Hsiao, Chow, & Liu, 2008) proposed a two-stage design with continuous endpoints. They developed two optimal designs for use in practice: an optimal design with the smallest expected sample size under the null hypothesis, and a minimax design with the smallest possible maximum sample size. The hypothesis is one-sided, and the null hypothesis is rejected when a large test statistic is observed from both the first stage and the two stages combined. The type I and II error rates were only evaluated at the boundary of their associated sample spaces. It would be interesting to investigate the monotonic condition of the tail probability for a two-stage design with continuous endpoints.

Acknowledgments

The authors are very grateful to the Associate Editor and a referee for their insightful comments that help improve the manuscript. Shan's research is partially supported by grants from the National Institute of General Medical

Sciences from the National Institutes of Health: P20GM109025, P20GM103440, and 5U54GM104944. Zhang's work was supported by the Zhejiang Provincial Natural Science Foundation of China (grant no. LY15F020001) and the National Natural Science Foundation of China (grant no. 61170099). We also thank Dr. Huichi Huang from University of Muenster for the discussion of this research.

References

- Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*. 1934 Dec 1; 26(4):404–413. Retrieved from <http://dx.doi.org/10.1093/biomet/26.4.404>.
- Jung S-HH, Kim KMM. On the estimation of the binomial probability in multistage clinical trials. *Statistics in medicine*. 2004 Mar 30; 23(6):881–896. Retrieved from <http://dx.doi.org/10.1002/sim.1653>. [PubMed: 15027078]
- Kang L, Tian L. Estimation of the volume under the ROC surface with three ordinal diagnostic categories. *Computational Statistics & Data Analysis*. 2013 Jun; 62:39–51. Retrieved from <http://dx.doi.org/10.1016/j.csda.2013.01.004>.
- Kang L, Xiong C, Crane P, Tian L. Linear combinations of biomarkers to improve diagnostic accuracy with three ordinal diagnostic categories. *Statistics in medicine*. 2013 Feb 20; 32(4):631–643. Retrieved from <http://dx.doi.org/10.1002/sim.5542>. [PubMed: 22865796]
- Katz PO, Gerson LB, Vela MF. Guidelines for the diagnosis and management of gastroesophageal reflux disease. *The American journal of gastroenterology*. 2013 Mar; 108(3) Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/23419381>.
- Röhmel J. Problems with existing procedures to calculate exact unconditional P-values for non-inferiority/superiority and confidence intervals for two binomials and how to resolve them. *Biometrical Journal*. 2005 Feb; 47(1):37–47. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/16395995>. [PubMed: 16395995]
- Röhmel J, Mansmann U. Unconditional Non-Asymptotic One-Sided Tests for Independent Binomial Proportions When the Interest Lies in Showing Non-Inferiority and/or Superiority. *Biometrical Journal*. 1999 May 1; 41(2):149–170. Retrieved from [http://dx.doi.org/10.1002/\(sici\)1521-4036\(199905\)41:2%3C149::aid-bimj149%3E3.0.co;2-e](http://dx.doi.org/10.1002/(sici)1521-4036(199905)41:2%3C149::aid-bimj149%3E3.0.co;2-e).
- Shan G, Chen JJ, Ma C. Boundary problem in Simon's two-stage clinical trial designs. *Journal of biopharmaceutical statistics*. 2016 Feb 16. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/26881325>.
- Shan G, Ma C. Unconditional tests for comparing two ordered multinomials. *Statistical methods in medical research*. 2016 Feb 01; 25(1):241–254. Retrieved from <http://dx.doi.org/10.1177/0962280212450957>. [PubMed: 22700600]
- Shan G, Wilding GE, Hutson AD, Gerstenberger S. Optimal adaptive two-stage designs for early phase II clinical trials. *Statist. Med*. 2016 Apr 15; 35(8):1257–1266. Retrieved from <http://dx.doi.org/10.1002/sim.6794>.
- Siefker-Radtke AO, Dinney CP, Shen Y, Williams DL, Kamat AM, Grossman HB, Millikan RE. A phase 2 clinical trial of sequential neoadjuvant chemotherapy with ifosfamide, doxorubicin, and gemcitabine followed by cisplatin, gemcitabine, and ifosfamide in locally advanced urothelial cancer. *Cancer*. 2013 Feb 1; 119(3):540–547. Retrieved from <http://dx.doi.org/10.1002/cncr.27751>. [PubMed: 22914978]
- Simon R. Optimal two-stage designs for phase II clinical trials. *Controlled clinical trials*. 1989 Mar; 10(1):1–10. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/2702835>. [PubMed: 2702835]
- Tsai W-YY, Chi Y, Chen C-MM. Interval estimation of binomial proportion in clinical trials with a two-stage design. *Statistics in medicine*. 2008 Jan 15; 27(1):15–35. Retrieved from <http://view.ncbi.nlm.nih.gov/pubmed/17566141>. [PubMed: 17566141]
- Tsou H-H, Hsiao C-F, Chow S-C, Liu J-p. A Two-Stage Design for Drug Screening Trials Based on Continuous Endpoints. *Drug Information Journal*. 2008 May 01; 42(3):253–262. Retrieved from <http://dx.doi.org/10.1177/009286150804200307>.
- Zheng L, Rosenkranz SL, Taiwo B, Para MF, Eron JJ, Hughes MD. The Design of Single-Arm Clinical Trials of Combination Antiretroviral Regimens for Treatment-Naive HIV-Infected Patients. *AIDS Research and Human Retroviruses*. 2012 Dec 10; 29(4):652–657. Retrieved from <http://dx.doi.org/10.1089/aid.2012.0180>.