Published in final edited form as:

Stat Med. 2015 July 20; 34(16): 2417–2426. doi:10.1002/sim.6490.

Hypothesis testing for two-stage designs with over or under enrollment

Donglin Zeng[a], **Fei Gao**[a], **Kuolung Hu**[b], **Catherine Jia**[b], and **Joseph G. Ibrahim**[a,*,†]

[a]Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599, U.S.A

[b]Global Biostatistical Science, Amgen Inc., One Amgen Center Drive, Thousand Oaks, CA 91320, U.S.A

Abstract

Simon's two-stage designs are widely used in cancer phase II clinical trials for assessing the efficacy of a new treatment. However in practice, the actual sample size for the second stage is often different from the planned sample size, and the original inference procedure is no longer valid. Previous work on this problem has certain limitations in computation. In this paper, we attempt to maximize the unconditional power while controlling for the type I error for the modified second stage sample size. A normal approximation is used for computing the power, and the numerical results show that the approximation is accurate even under small sample sizes. The corresponding confidence intervals for the response rate are constructed by inverting the hypothesis test, and they have reasonable coverage while preserving the type I error.

Keywords

clinical trials; adaptive design; sample size modification

1. Introduction

Clinical trials involving new treatments are commonly classified into four development phases. A treatment could be a drug, medical device, or biologic, such as a vaccine, blood product, or gene therapy. Each phase could include many separate clinical trials in order to properly build up the safety and efficacy profile of the treatment. It typically takes many years to advance a therapeutic treatment through all four phases. Therefore, the use of efficient trial designs in the early treatment development phase, such as phase I or phase II, is highly desirable in order to quickly and accurately identify promising treatments while also identifying treatments for which all further development should be stopped. Traditional oncology phase II trial designs typically use the endpoint of clinical response for single arm trials. Simon's two-stage design [1] is widely used in cancer phase II clinical trials for assessing the efficacy of a new treatment. However, based on this design, appropriate computation of a *p*-value or confidence interval is not readily available, and several different approaches have been proposed for these goals [2].

*Correspondence to: Joseph G. Ibrahim, Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599, U.S.A.
†ibrahim@bios.unc.edu

A common scenario in Simon's two-stage design setting is that a trial may enroll additional subjects beyond what is specified in the design once the efficacy criteria have been met. The reason for these additional enrollments could be for the purpose of accumulating more safety and/or efficacy data. This is attractive from a cost perspective as compared with launching a new trial. There is also a possibility of enrolling less patients due to practical reasons. However, over-enrollment or under-enrollment in this setting poses many statistical inference challenges. First, extending (or shortening) the sample size for stage II implies that the final study inference can no longer be based on the original Simon's two-stage design. Inference based on existing methods that followed the two-stage feature are no longer valid[3]. Other multi-stage methods may also not be appropriate for this situation because 'the third stage' of this trial is directly related to the hypothesis testing feature of the primary endpoint. The data analyst could consider providing the estimated response rate by maximum likelihood with an exact confidence interval [4] ignoring the additional enrollment. However, this approach does not account for the original design features. Therefore, an appropriate estimation method is needed for this common scenario.

In this paper, we propose a novel methodology for carrying out inferences under Simon's phase II design when there is over-enrollment or under-enrollment in the second stage of the design after claiming success of the trial. We propose a method to directly calculate the stage II critical value in a hypothesis testing framework. There has been some previous work performed in this over-enrollment problem. Green and Dahlberg [3] extended the standard phase II approach used by the Southwest Oncology Group to accommodate a modified sample size in both stages. However, the method only works for a type I error rate of 0.05, and the choice of cut points in their two stage hypothesis testing procedure is arbitrary and lacks a theoretical justification. Chen and Ng [5] considered a range of possible stage I and total sample sizes, for which the stage I and the total sample size would occur with equal probability, and they searched for the 'optimal' and 'minimax' designs with a desired type I error and power. Masaki *et al.* [6] extended their work by allowing unequal probabilities on the sample sizes. These approaches would properly control the type I error, but they required a prespecified difference between the planned and modified sample sizes. Koyama and Chen [7] controlled the conditional type I error for the modified stage II sample size, but the corresponding overall type I error could be highly deflated, and the power would be lower than desired. Li *et al*. [8] formulated the two-stage design in a Bayesian setting and applied a Bayesian criterion to the observed outcome with a modified sample size. The method of Li *et al.* obtains desirable frequentist properties under certain types of priors.

In this paper, we attempt to maximize the unconditional power while controlling for the type I error for the modified stage 2 sample size. Because enumerating all possible scenarios in the power calculation is computationally intensive, we propose a normal approximation in the evaluation of the power, and our numerical results show that the proposed approximation is very accurate even under small sample sizes. Finally, we construct confidence intervals for the response rate by inverting the hypothesis test. The rest of this paper is organized as follows. In section 2, we describe the proposed method to account for sample size change in Simon's two-stage design. Our method includes an explicit formula for the power calculation and an analytic derivation of the confidence intervals. Extensive simulations are

conducted in section 3 to demonstrate the finite-sample performance of the proposed method. Some concluding remarks are given in the final section.

2. Method

2.1. Hypothesis testing in Simon's two-stage design

Suppose that Simon's two-stage design is implemented to test the null hypothesis that the response rate (π) $H_0 : \pi = \pi_0$ versus $H_1 : \pi = \pi_1$ ($\pi_1 > \pi_0$) with a desired power $1 - \beta$ and type I error α. Let n_1 and n_2 be the stage I and stage II sample sizes. Also, R_1 and R_t denote the critical values for rejecting the null hypothesis, in specific, we precede to stage II if we observe $x_1 \geq R_1$ and reject the null hypothesis if $x_2 \geq R_t - x_1 =: R_2(x_1)$. In practice, the stage II sample size may not be the same as n_2 but can be a larger number due to study extension (over-enrollment) or a smaller number due to early stopping (under-enrollment) of the trial. Therefore, when the stage II sample size changes to n_2', the critical value R_t, which was pre-specified at the design stage, can no longer be used to determine the validity of the alternative hypothesis, and the desired type I error may not be preserved. In the following development, we propose a revised critical value depending on the number of successes at stage I so as to preserve the type I error α while maximizing the power.

Specifically, we let $R_2'(x_1)$ denote the critical value for the modified stage II sample size when we observe $x_1 \geq R_1$ in the first stage, and $R_t'(x_1) \equiv R_2'(x_1) + x_1$. The rejection region for the null hypothesis is $\{X_1 \geq R_1, X_2 \geq R_2'(X_1)\}$, where X_1 and X_2 are the numbers of successes in stage I and II, respectively. Then we find the $R_2'(x_1)$ to maximize the power of the test with the overall type I error controlled. This is equivalent to finding the $R_2'(x_1)$ such that

$$\text{Power} = \sum_{x_1=R_1}^{n_1} \binom{n_1}{x_1} \pi_1^{x_1}(1-\pi_1)^{n_1-x_1} \sum_{x_1=R_2'(x_1)}^{n_2'} \binom{n_2'}{x_2} \pi_1^{x_2}(1-\pi_1)^{n_2'-x_2}$$

is maximized subject to

$$\text{Type I error} = \sum_{x_1=R_1}^{n_1} \binom{n_1}{x_1} \pi_0^{x_1}(1-\pi_0)^{n_1-x_1} \sum_{x_2=R_2'(x_1)}^{n_2'} \binom{n_2'}{x_2} \pi_0^{x_2}(1-\pi_0)^{n_2'-x_2} \leq \alpha. \quad (1)$$

Unfortunately, this maximization problem does not have a closed form solution. Although it is theoretically possible to find $R_2'(x_1)$ by searching among all the possible combinations of $(R_2'(R_1), R_2'(R_1+1), \ldots, R_2'(n_1))$ in evaluating the corresponding power, the computation is very intensive. Instead, we consider the following approximation: the cumulative distribution function of the binomial random variable in the previous expression will be approximated by the cumulative distribution function of a normal random variable, that is,

$$\sum_{x_2=R_2^{'}(x_1)}^{n_2^{'}} \left(\begin{array}{c} n_2^{'} \\ x_2 \end{array} \right) \pi^{x_2}(1-\pi)^{n_2^{'}-x_2} \approx 1-\Phi\left(\frac{R_2^{'}(x_1)-n_2^{'}\pi}{\sqrt{n_2^{'}\pi(1-\pi)}}\right),$$

where $\Phi(.)$ denotes the cumulative distribution function of standard normal distribution. Under this approximation, we thus maximize

$$\sum_{x_1=R_1}^{n_1} b(x_1)\left[1-\Phi\left(\frac{R_2^{'}(x_1)-n_2^{'}\pi_1}{\sqrt{n_2^{'}\pi_1(1-\pi_1)}}\right)\right]$$

subject to

$$\sum_{x_1=R_1}^{n_1} a(x_1)\left[1-\Phi\left(\frac{R_2^{'}(x_1)-n_2^{'}\pi_0}{\sqrt{n_2^{'}\pi_0(1-\pi_0)}}\right)\right] \leq \alpha,$$

where $a(x_1)=\left(\begin{array}{c} n_1 \\ x_1 \end{array} \right) \pi_0^{x_1}(1-\pi_0)^{n_1-x_1}$ and $b(x_1)=\left(\begin{array}{c} n_1 \\ x_1 \end{array} \right) \pi_1^{x_1}(1-\pi_1)^{n_1-x_1}$. Using Lagrange multipliers and differentiating with respect to $R_2^{'}(x_1)$, we obtain

$$b(x_1)\phi\left(\frac{R_2^{'}(x_1)-n_2^{'}\pi_1}{\sqrt{n_2^{'}\pi_1(1-\pi_1)}}\right)=\lambda a(x_1)\phi\left(\frac{R_2^{'}(x_1)-n_2^{'}\pi_0}{\sqrt{n_2^{'}\pi_0(1-\pi_0)}}\right)$$

where λ is the Lagrange multiplier and $\phi(.)$ is the density function for standard normal distribution. Because here we can only take discrete value for $R_2^{'}(X_1)$ from 0 to $n_2^{'}$, one could find all possible values for $\phi\left(\frac{R_2^{'}(x_1)-n_2^{'}\pi_1}{\sqrt{n_2^{'}\pi_1(1-\pi_1)}}\right)$ and $\phi\left(\frac{R_2^{'}(x_1)-n_2^{'}\pi_0}{\sqrt{n_2^{'}\pi_0(1-\pi_0)}}\right)$ to get a reasonable range of λ, and search within. Then the problem is equivalent to solving the equation

$$\left(\frac{1}{\pi_0(1-\pi_0)}-\frac{1}{\pi_1(1-\pi_1)}\right) R_2^{'}(x_1)^2-\frac{2n_2^{'}(\pi_0-\pi_1)}{(1-\pi_0)(1-\pi_1)}R_2^{'}(x_1)+\frac{n_2^{'2}(\pi_0-\pi_1)}{(1-\pi_0)(1-\pi_1)}-2n_2^{'}\log\frac{\lambda a(x_1)}{b(x_1)}=0.$$

We redefine $R_2^{'}(x_1)$ as

$$R_2^{'}(x_1)=\max\left(0,\min\left(\left[R_2^{'}(x_1)\right],n_2^{'}\right)\right).$$

Now we search over a grid of λ to find the λ such that the type I error defined in equation (1) is as close to α as possible. The corresponding $\{ R_2'(x_1), x_1 = R_1, \ldots, n_1\}$ is the optimal critical value for the modified second stage.

2.2. Confidence interval for the response rate

With a modified stage II sample size, the confidence interval for the response rate could be calculated by inverting the hypothesis test. Specifically, we consider the hypothesis H_0 : response rate ≤ π, versus H_1 : response rate > π. Following the development in the previous section, we define the rejection region as

$$\{X_1 \geq R_1, X_1+X_2 \geq C(\pi)\},$$

where $C(\pi)$ is the critical value for rejection. To construct a $(1 - \alpha) \times 100\%$– confidence interval for π, we require

$$P_\pi\ (X_1 \geq R_1, X_1+X_2 \geq C(\pi)) = \alpha. \quad (2)$$

The solution of $C(\pi)$ has no closed form, but it could be calculated numerically: for each π, we simulate $\{X_1, X_2\}$ under the response rate π then determine $C(\pi)$ satisfying equation (2). Thus, for any given observations (x_1, x_2), the confidence interval is chosen as $[\pi_1, \pi_2]$ where π_1 and π_2 are the corresponding lower bound and upper bound of the π's satisfying $x_1 + x_2 \geq C(\pi)$.

3. Numerical studies

3.1. Comparison of our method with Koyama and Chen

To examine the performance of our method for hypothesis testing (named AG), we compare our method with the conditional type I error method of Koyama and Chen[7], denoted by (KC), in which $R_2'(x_1)$ is chosen so that the conditional rejection probability $P(X_2 \geq R_2'(x_1)|X_1=x_1) \leq \alpha$. Table I gives the critical value $R_2'(x_1)$ for one example scenario (n_1 = 15, n_2 = 31, R_1 = 6, R_2 = 13, which is the optimal design for testing π = 0.3 versus π = 0.5, and we changed the stage II sample size n_2 to 1.5n_2). After adjustment, the sum of critical value $R_t'(x_1)=R_2'(x_1)+x_1$ are not necessarily the same for different x_1. In general, $R_2'(x_1)$ decreases with x_1 increases.

For comparison, several scenarios of Simon's two-stage design are considered. We consider that the sample size in stage II is extended to 1.5 or 2 times, remains the same or is reduced to two-third of the originally designed sample size. The results are shown in Tables II and III, where in the first panel, π_0 and π_1 are the response rates under the null and alternative hypotheses, respectively, and (n_1, R_1, n_2, R_t) are the design parameters including the enrollment number in the first stage, the minimal number of responses in the first stage to move to the second stage, the enrollment number in the second stage, and the total number of responses at the end of the two stages to achieve a designed power. Furthermore, 'Min'

denotes the minimax design and 'Opt' refers to the optimal design. With overall enrollment n_2' in the second stage, we report the corresponding type I error and power based on our method (AG) and the method of Koyama and Chen.

Table II and III show that both AG and KC have protected type I error rate, while in almost all scenarios, AG has more power than KC. Bold numbers in the tables indicate the scenarios for which AG's power is at least 0.03 higher than KC's power. Because KC tries to control the conditional type I error rate for each stage I sample path, the overall unconditional type I error rate could be possibly much less than α, especially in small original sample size scenarios. With a deflated type I error rate, the power for detecting a treatment effect will decrease. There is only one scenario (optimal design for testing $\pi = 0.3$ versus $\pi = 0.5$, stage II sample size doubles) where KC has larger protected type I error rate and a slightly larger power than AG.

In calculating the critical value $R_2'(x_1)$ for the modified stage II sample size, we adopt a normal approximation to simplify the computation. However, for small sample sizes, an exhaustive grid search could also be used to search exactly among all the possible combinations of $\{ R_2'(R_1), R_2'(R_1+1), \ldots, R_2'(n_1)\}$. It is of interest to examine the power differences between the grid search method and the normal approximation. Towards this goal, we conducted an additional numerical study to compare the results with the normal approximation and the exhaustive grid search based on our method. The settings and results are given in Table IV. It shows that even when the total sample size for the original Simon's two-stage design is smaller than 20, there is not much difference in power (less than 0.02) between the normal approximation and the grid search, especially for those cases with a relatively large extended stage II sample size. However, the computation time gain using the normal approximation can be enormous when n_2 is not small.

3.2. Simulation studies for obtaining a confidence interval

To examine the performance of our method (AG) in constructing a two-sided 90% confidence interval for π, we simulate data to calculate the 90% coverage rate and width of the computed confidence interval. A desired confidence interval would have relatively narrow width while preserving the 90% nominal coverage. For comparison, we consider two other methods:

(A1) The sample path with a larger X_1 and a larger $X_1 + X_2$ is considered as more extreme. Then the $(1 - 2\alpha) \times 100\%$ two-sided confidence interval is constructed as

$$\{\pi: \alpha < P_\pi (X_1 \geq x_1, X_1 + X_2 \geq x_1 + x_2) < 1 - \alpha\}.$$

(A2) The confidence interval proposed by Koyama and Chen [7]. If we observed the sample path (x_1, x_2), then if $x_1 < R_t$, for each π, we want to find $\pi^*(\pi)$ such that the probability of observing x_2 for the modified stage II sample size n_2' with true response rate π is equal to the probability of observing the critical value R_2 for

the original stage II sample size n_2 with response rate π_*, which yields the equation

$$\sum_{y=R_t-x_1}^{n_2} \binom{n_2}{y} \pi_*(\pi)^y(1-\pi_*(\pi))^{n_2-y} = \sum_{y=x_2}^{n_2'} \binom{n_2'}{y} \pi^y(1-\pi)^{n_2'-y}. \quad (3)$$

Their method provides the criterion in comparing different sample paths by transferring them to the original stage design. Then, the $(1 - 2\alpha) \times 100\%$ confidence interval is computed as

$$\alpha < \sum_{x=R_1}^{n_1} \binom{n_1}{x} \pi^x(1-\pi)^{n_1-x} \sum_{y=R_t-x}^{n_2} \binom{n_2}{y} \pi_*(\pi)^y(1-\pi_*(\pi))^{n_2-y} < 1-\alpha.$$

The computation is very intensive to compute π^* in simulation studies with large number of replicates. In order to implement and compare our method with that of Koyama and Chen [7], we use the normal approximation similar to that in our proposed method in section 2.1 to calculate π_* for each π.

In order to better illustrate the construction of the confidence intervals and rejection regions for those methods, we give an example for the minimax two-stage design for testing $H_0 : \pi_0 = 0.3$ versus $H_1 : \pi_1 = 0.5$, with $\alpha = 0.05$, $1 - \beta = 0.8$. $n_1 = 19$, $n_2 = 20$, $R_1 = 7$, $R_t = 17$. If we observe $x_1 = 10$, proceed to stage II, and have an extended stage II sample size $n_2' = 23$ and observe $x_2 = 10$, then the rejection region could be specified as the upper right portion in Figure 1. The rejection region of our method (AG) is quite similar with (A2), which is the method of Koyama and Chen. However, (A2) can only handle the situation with $x_1 < R_t$. If we observe $X_1 \geq R_t$, the left-hand side of equation (3) would be one regardless of the $\pi_*(\pi)$ value, and therefore, no solution of $\pi_*(\pi)$ can be found. For this particular example, we obtain the two-sided 90% confidence intervals for the three respective methods (AG), (A1), and (A2), as (0.3436, 0.5947), (0.3681, 0.6804), and (0.3444, 0.5947).

Table V displays the simulation results in two minimax two-stage designs for testing (i) $H_0 : \pi_0 = 0.05$ versus $H_1 : \pi_1 = 0.15$ and (ii) $H_0 : \pi_0 = 0.1$ versus $H_1 : \pi_1 = 0.2$, with $\alpha = 0.05$, $1 - \beta = 0.8$, and a modified stage II sample size. We consider three different possibilities for the true underlying π: i) π is equal to the null response rate π_0, ii) π is equal to the alternative rate π_1, and iii) π is larger than π_1. For each scenario, 1000 two-stage studies are simulated, and the two-sided 90% confidence intervals are calculated using the three methods. To compare the methods in Table V, we report the average width and the corresponding coverage probabilities. Compared with the methods (A1) and (A2), our method has the narrowest width while preserving the nominal 90% coverage. For (A2), the confidence interval width is small when the true underlying response rate is close to the null response rate in the original design and the stage II sample size is not extended much. When the true response rate gets larger, the probability that $X_1 \geq R_t$ gets larger. In this case, (A2) fails to provide a valid confidence interval for the scenario of $X_1 \geq R_t$.

4. Conclusion

We have proposed a computationally simple method to modify the rejection rule in a Simon's two-stage design when the sample size in Stage II is changed in the trial. The proposed method guarantees a preservation of the type I error and leads to superior power compared with the existing methods. We also proposed a method for calculating the confidence interval by inverting the rejection region of the corresponding hypothesis test. The latter is shown to yield confidence intervals with proper coverage and smaller width compared with existing methods. The proposed method can be potentially generalized to multiple stage designs with varied sample sizes.

Furthermore, in the constructing of the confidence interval, out rejection region is of the form $\{X_1 \geq R_1, X_1 + X_2 \geq C(\pi)\}$. We may generally allow $C(\pi)$ to depend on X_1. In this case, the derived confidence interval may be even narrower but at a price of increased computation because we need to examine each sample path $\{(X_1, C(\pi, X_1)), X_1 = R_1, \ldots, n_1\}$ that preserves Type I error. The proposed method is computationally simpler and superior, although not optimal.

The proposed method focused on the improvement of inference reporting while controlling type I error properly where there is over-enrollment or under-enrollment in the second stage. However it may not be used as a generalization approach for sample size re-adjustment, which often needs re-adjusting the study hypotheses and the changes of R_1 and R_2. Under that circumstance, the trials should have already met the R_1 and R_2 criteria of Simon's original design, and the study null hypothesis was already rejected.

References

1. Simon R. Optimal two-stage designs for phase II clinical trials. Controlled Clinical Trials. 1989; 10(1):1–10. [PubMed: 2702835]
2. Porcher R, Desseaux K. What inference for two-stage phase II trials? BMC Medical Research Methodology. 2012; 12(1):117. [PubMed: 22867439]
3. Green SJ, Dahlberg S. Planned versus attained design in phase II clinical trials. Statistics in Medicine. 1992; 11(7):853–862. [PubMed: 1604065]
4. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika. 1934; 26(4):404–413.
5. Chen TT, Ng TH. Optimal flexible designs in phase II clinical trials. Statistics in Medicine. 1998; 17(20):2301–2312. [PubMed: 9819829]
6. Masaki N, Koyama T, Yoshimura I, Hamada C. Optimal two-stage designs allowing flexibility in number of subjects for phase II clinical trials. Journal of Biopharmaceutical Statistics. 2009; 19(4): 721–731. [PubMed: 20183436]
7. Koyama T, Chen H. Proper inference from Simon's two-stage designs. Statistics in Medicine. 2008; 27(16):3145–3154. [PubMed: 17960777]
8. Li Y, Mick R, Heitjan DF. A Bayesian approach for unplanned sample sizes in phase II cancer clinical trials. Clinical Trials. 2012; 9(3):293–302. [PubMed: 22523304]

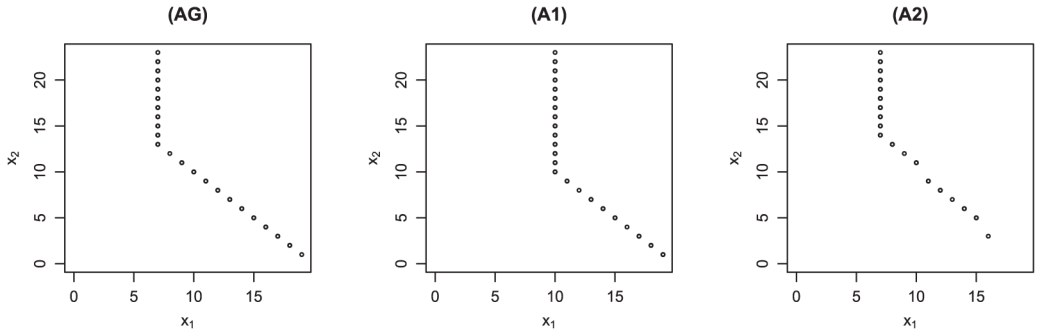

Figure 1.

Rejection region for three methods. Note: x_1 and x_2 are the numbers of responses in the first and second stage, respectively. The circles are the boundaries of the rejection regions.

Table I

One example of $R_2^{'}(x_1)$.

	AG		KC	
$X_1 = x_1$	$R_2^{'}(x_1)$	$R_t^{'}(x_1)=R_2^{'}(x_1)+x_1$	$R_2^{'}(x_1)$	$R_t^{'}(x_1)=R_2^{'}(x_1)+x_1$
6	18	24	19	25
7	18	25	18	25
8	18	26	18	26
9	18	27	18	27
10	18	28	18	28
11	17	28	18	29
12	17	29	17	29
13	17	30	17	30
14	17	31	17	31
15	17	32	17	32

Table II

Calculation of type I error and power for extended stage 2 sample size.

Simon's two-stage design							$n_2'=1.5n_2$					$n_2'=2n_2$				
One sided α = 0.05, 1 − β = 0.8								**AG**		**KC**			**AG**		**KC**	
π_0	π_1		n_1	R_1	n_2	R_t	n_2'	**Type I**	**Power**	**Type I**	**Power**	n_2'	**Type I**	**Power**	**Type I**	**Power**
0.05	0.15	Min	30	2	22	6	33	0.040	0.841	0.031	0.817	44	0.045	0.881	0.033	0.866
		Opt	23	2	33	6	49	0.043	0.825	0.028	0.803	66	0.045	0.854	0.031	0.843
0.1	0.2	Min	45	5	33	13	49	**0.047**	**0.858**	**0.033**	**0.821**	66	0.048	0.897	0.035	0.871
		Opt	30	4	59	14	88	0.044	0.838	0.037	0.832	118	0.048	0.863	0.041	0.859
0.15	0.25	Min	55	9	42	21	63	0.049	0.853	0.040	0.832	84	0.045	0.884	0.041	0.875
		Opt	38	7	78	24	117	0.049	0.843	0.042	0.838	156	0.050	0.861	0.042	0.857
0.2	0.3	Min	66	14	50	31	75	0.049	0.858	0.041	0.839	100	0.047	0.892	0.040	0.876
		Opt	46	11	95	36	142	0.046	0.832	0.043	0.830	190	0.047	0.848	0.043	0.846
0.05	0.2	Min	13	1	14	4	21	**0.048**	**0.851**	**0.025**	**0.817**	28	**0.047**	**0.896**	**0.021**	**0.842**
		Opt	10	1	19	4	28	0.039	0.831	0.033	0.828	38	0.029	0.854	0.029	0.854
0.1	0.25	Min	22	3	18	8	27	**0.048**	**0.867**	**0.024**	**0.808**	36	0.046	0.890	0.028	0.864
		Opt	18	3	25	8	37	0.050	0.832	0.034	0.818	50	0.050	0.851	0.038	0.843
0.2	0.35	Min	31	7	22	16	33	0.047	0.844	0.038	0.826	44	0.045	0.879	0.039	0.865
		Opt	22	6	50	20	75	0.047	0.824	0.042	0.822	100	0.050	0.833	0.040	0.831
0.1	0.3	Min	15	2	10	6	15	0.040	0.858	0.025	0.832	20	**0.043**	**0.900**	**0.023**	**0.861**
		Opt	10	2	19	6	28	0.036	0.819	0.026	0.811	38	0.042	0.841	0.033	0.838
0.2	0.4	Min	18	5	15	11	22	0.040	0.831	0.037	0.827	30	0.044	0.866	0.033	0.847
		Opt	13	4	30	13	45	0.048	0.820	0.038	0.816	60	0.048	0.828	0.037	0.826
0.3	0.5	Min	19	7	20	17	30	0.049	0.857	0.035	0.834	40	0.044	0.881	0.034	0.865
		Opt	15	6	31	19	46	0.050	0.832	0.042	0.826	62	0.038	0.840	0.042	0.841
0.05	0.35	Min	6	1	6	7	9	**0.049**	**0.895**	**0.006**	**0.803**	12	**0.044**	**0.912**	**0.010**	**0.874**
		Opt	4	1	12	7	18	0.047	0.820	0.014	0.811	24	0.030	0.820	0.025	0.820
0.1	0.4	Min	8	2	5	3	7	**0.045**	**0.853**	**0.014**	**0.762**	10	0.037	0.857	0.024	0.846
		Opt	4	1	11	3	16	0.047	0.847	0.032	0.841	22	0.039	0.861	0.029	0.859

Simon's two-stage design							$n_2'=1.5n_2$					$n_2'=2n_2$				
One sided α = 0.05, 1 − β = 0.8								**AG**		**KC**			**AG**		**KC**	
π_0	π_1		n_1	R_1	n_2	R_t	n_2'	**Type I**	**Power**	**Type I**	**Power**	n_2'	**Type I**	**Power**	**Type I**	**Power**
0.2	0.5	Min	9	3	8	4	12	**0.041**	**0.860**	**0.018**	**0.794**	16	0.041	0.883	0.027	0.860
		Opt	8	3	10	4	15	**0.049**	**0.838**	**0.023**	**0.805**	20	0.036	0.841	0.028	0.837
One sided α = 0.05, 1 − β=0.9								AG		KC			AG		KC	
π_0	π_1		n_1	R_1	n_2	R_t	n_2'	Type I	Power	Type I	Power	n_2'	Type I	Power	Type I	Power
0.2	0.3	Min	92	19	68	41	102	0.048	0.936	0.041	0.928	136	0.050	0.960	0.043	0.953
		Opt	71	16	113	46	169	0.044	0.923	0.042	0.922	226	0.048	0.933	0.042	0.932
0.3	0.5	Min	24	8	29	22	43	0.049	0.939	0.036	0.926	58	0.045	0.956	0.038	0.951
		Opt	24	9	39	25	58	0.049	0.918	0.040	0.916	78	0.045	0.922	0.042	0.922

Table III

Calculation of type I error and power for the same or reduced stage 2 sample size.

Simon's two-stage design							$n_2^{'}=n_2$					$n_2^{'}=2/3n_2$				
One sided α = 0.05, 1 − β = 0.8								**AG**		**KC**			**AG**		**KC**	
π_0	π_1		n_1	R_1	n_2	R_t	$n_2^{'}$	**Type I**	**Power**	**Type I**	**Power**	$n_2^{'}$	**Type I**	**Power**	**Type I**	**Power**
0.05	0.15	Min	30	2	22	6	22	0.047	0.807	0.043	0.802	14	**0.050**	**0.752**	**0.021**	**0.662**
		Opt	23	2	33	6	33	0.039	0.768	0.050	0.800	22	**0.047**	**0.732**	**0.022**	**0.662**
0.1	0.2	Min	45	5	33	13	33	0.046	0.805	0.044	0.802	22	**0.049**	**0.757**	**0.030**	**0.691**
		Opt	30	4	59	14	59	0.049	0.803	0.048	0.802	39	**0.047**	**0.735**	**0.034**	**0.704**
0.15	0.25	Min	55	9	42	21	42	0.049	0.802	0.049	0.801	28	**0.049**	**0.745**	**0.036**	**0.700**
		Opt	38	7	78	24	78	0.049	0.804	0.049	0.803	52	0.047	0.740	0.037	0.715
0.2	0.3	Min	66	14	50	31	50	0.048	0.802	0.047	0.801	33	**0.050**	**0.754**	**0.037**	**0.706**
		Opt	46	11	95	36	95	0.050	0.801	0.050	0.801	63	0.050	0.750	0.042	0.728
0.05	0.2	Min	13	1	14	4	14	0.042	0.801	0.042	0.801	9	0.032	0.649	0.022	0.663
		Opt	10	1	19	4	19	0.047	0.801	0.047	0.801	12	0.027	0.660	0.021	0.646
0.1	0.25	Min	22	3	18	8	18	0.044	0.810	0.040	0.803	12	**0.047**	**0.774**	**0.021**	**0.660**
		Opt	18	3	25	8	25	0.050	0.801	0.048	0.800	16	**0.043**	**0.741**	**0.034**	**0.705**
0.2	0.35	Min	31	7	22	16	22	0.050	0.802	0.050	0.802	14	**0.045**	**0.725**	**0.037**	**0.693**
		Opt	22	6	50	20	50	0.050	0.801	0.049	0.800	33	0.041	0.738	0.038	0.730
0.1	0.3	Min	15	2	10	6	10	0.036	0.808	0.033	0.802	6	**0.047**	**0.776**	**0.014**	**0.637**
		Opt	10	2	19	6	19	0.047	0.805	0.047	0.805	12	0.037	0.721	0.033	0.706
0.2	0.4	Min	18	5	15	11	15	0.049	0.806	0.046	0.801	10	**0.041**	**0.737**	**0.033**	**0.703**
		Opt	13	4	30	13	30	0.050	0.800	0.050	0.800	20	0.046	0.754	0.038	0.740
0.3	0.5	Min	19	7	20	17	20	0.047	0.806	0.045	0.804	13	**0.041**	**0.720**	**0.031**	**0.690**
		Opt	15	6	31	19	31	0.050	0.803	0.050	0.803	20	**0.043**	**0.732**	**0.031**	**0.699**
0.05	0.35	Min	6	1	6	7	6	0.040	0.847	0.018	0.822	4	0.011	0.729	0.011	0.729
		Opt	4	1	12	7	12	0.034	0.805	0.027	0.803	8	0.024	0.756	0.015	0.747
0.1	0.4	Min	8	2	5	3	5	0.031	0.802	0.031	0.802	3	**0.042**	**0.758**	**0.018**	**0.698**
		Opt	4	1	11	3	11	0.045	0.819	0.043	0.818	7	0.018	0.670	0.017	0.666

Simon's two-stage design							$n_2^{'}=n_2$					$n_2^{'}=2/3n_2$				
One sided α = 0.05, 1 − β = 0.8								**AG**		**KC**			**AG**		**KC**	
π_0	π_1		n_1	R_1	n_2	R_t	$n_2^{'}$	**Type I**	**Power**	**Type I**	**Power**	$n_2^{'}$	**Type I**	**Power**	**Type I**	**Power**
0.2	0.5	Min	9	3	8	4	8	0.040	0.814	0.034	0.806	5	**0.047**	**0.782**	**0.012**	**0.603**
		Opt	8	3	10	4	10	0.042	0.803	0.039	0.800	6	**0.041**	**0.750**	**0.023**	**0.661**
One sided α = 0.05, 1 − β = 0.9								AG		KC			AG		KC	
π_0	π_1		n_1	R_1	n_2	R_t	$n_2^{'}$	Type I	Power	Type I	Power	$n_2^{'}$	Type I	Power	Type I	Power
0.2	0.3	Min	92	19	68	41	68	0.049	0.901	0.049	0.900	45	0.048	0.856	0.038	0.830
		Opt	71	16	113	46	113	0.050	0.901	0.048	0.900	75	0.048	0.858	0.039	0.839
0.3	0.5	Min	24	8	29	22	29	0.047	0.902	0.047	0.902	19	**0.045**	**0.841**	**0.031**	**0.804**
		Opt	24	9	39	25	39	0.050	0.903	0.050	0.903	26	0.043	0.864	0.040	0.857

Table IV

Comparison of exact and normal approximation calculation.

Simon's two-stage design							$n_2'=n_2$				$n_2'=1.5n_2$				$n_2'=2n_2$			
One sided α = 0.05, 1 − β = 0.8							**Normal approx**		**Grid search**		**Normal approx**		**Grid search**		**Normal approx**		**Grid search**	
π_0	π_1		n_1	R_1	n_2	R_t	**Type I**	**Power**	**Type I**	**Power**	**Type I**	**Power**	**Type I**	**power**	**Type I**	**Power**	**Type I**	**Power**
0.05	0.35	Min	6	1	6	3	0.040	0.847	0.040	0.847	0.049	0.895	0.049	0.895	0.044	0.912	0.044	0.912
		Opt	4	1	12	3	0.034	0.805	0.034	0.805	0.047	0.820	0.048	0.820	0.030	0.820	0.034	0.820
0.1	0.4	Min	8	2	5	4	0.031	0.802	0.050	0.814	0.045	0.853	0.045	0.853	0.037	0.857	0.050	0.860
		Opt	4	1	11	4	0.045	0.819	0.045	0.819	0.047	0.847	0.047	0.847	0.039	0.861	0.040	0.861
0.2	0.5	Min	9	3	8	7	0.040	0.814	0.043	0.815	0.041	0.860	0.046	0.860	0.041	0.883	0.050	0.887
		Opt	8	3	10	7	0.042	0.803	0.043	0.803	0.049	0.838	0.049	0.838	0.036	0.841	0.050	0.848

Table V

Width and coverage for 90% two-sided confidence intervals from the three methods.

Design	π	n_2'	(AG) Width	(AG) Coverage	(A1) Width	(A1) Coverage	(A2) Width	(A2) Coverage
(i) $n_1 = 30$ $n_2 = 22$ $R_1 = 2$ $R_t = 6$	0.05	22	0.1343	0.902	0.1503	0.832	0.1355	0.919
		33	0.1288	0.905	0.1470	0.850	0.1295	0.933
		44	0.1252	0.925	0.1439	0.880	0.1256	0.938
	0.15	22	0.1584	0.960	0.1868	0.930	0.1553	0.997
		33	0.1451	0.956	0.1811	0.921	0.1460	0.994
		44	0.1358	0.953	0.1746	0.932	0.1401	0.997
	0.3	22	0.2020	0.909	0.2316	0.887	0.1929	0.205
		33	0.1843	0.902	0.2228	0.884	0.1931	0.248
		44	0.1707	0.901	0.2129	0.897	0.1929	0.230
(ii) $n_1 = 45$ $n_2 = 33$ $R_1 = 5$ $R_t = 13$	0.1	33	0.1356	0.895	0.1513	0.864	0.1360	0.891
		49	0.1292	0.917	0.1475	0.850	0.1300	0.920
		66	0.1245	0.886	0.1463	0.824	0.1250	0.933
	0.2	33	0.1470	0.946	0.1714	0.917	0.1479	0.972
		49	0.1348	0.933	0.1653	0.906	0.1374	0.979
		66	0.1246	0.931	0.1586	0.900	0.1298	0.982
	0.3	33	0.1662	0.874	0.1914	0.905	0.1731	0.849
		49	0.1529	0.907	0.1833	0.899	0.1657	0.811
		66	0.1408	0.919	0.1784	0.887	0.1626	0.818

Note: The coverage rate is calculated as the percentage of datasets where the confidence interval includes the true response rate. The column labeled ‘width’ displays the mean width based on the simulations.