Published in final edited form as:

J Biopharm Stat. 2014 ; 24(4): 802–816. doi:10.1080/10543406.2014.901343.

Randomized Phase II Cancer Clinical Trials

Sin-Ho Jung and

Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA

Daniel J. Sargent

Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA

SUMMARY

Traditionally, phase II trials have been conducted as single-arm trials to compare the response probabilities between an experimental therapy and a historical control. Historical control data, however, often have a small sample size, are collected from a different patient population, or use a different response assessment method, so that a direct comparison between a historical control and an experimental therapy may be severely biased. Randomized phase II trials entering patients prospectively to both experimental and control arms have been proposed to avoid any bias in such cases. The small sample sizes for typical phase II clinical trials imply that the use of exact statistical methods for their design and analysis is appropriate. In this paper, we propose two-stage randomized phase II trials based on Fisher's exact test which does not require specification of the response probability of the control arm for testing. Through numerical studies, we observe that the proposed method controls the type I error accurately and maintains a high power. If we specify the response probabilities of the two arms under the alternative hypothesis, we can identify good randomized phase II trial designs by adopting the Simon's minimax and optimal design concepts that were developed for single-arm phase II trials.

Keywords

Fisher's exact test; Minimax design; Optimal design; Two-stage design; Unbalanced allocation

1 Introduction

The purpose of a phase II cancer clinical trial is to investigate if an experimental therapy has promising efficacy and this is worth further investigation. For phase II clinical trials of cytotoxic drugs, the most popular primary outcome is overall response, meaning the therapy shrinks the tumor(s), of the experimental therapy. As an effort to speed the assessment of new therapies, a phase II clinical trial usually recruits a small number of patients only to the experimental therapy arm to be compared to a historical control. This implies that the traditional single-arm phase II trials are appropriate only when reliable and valid data for an existing standard therapy are available for the same patient population. Furthermore, the

response assessment method in the historical control data should be identical to the one that will be used for a new study.

If no historical control data satisfying these conditions exist or the existing data are too small to represent the whole patient population, we have to consider a randomized phase II clinical trial with a prospective control to be compared with the experimental therapy under investigation. Cannistra (2009) recommends a randomized phase II trial if a single-arm design is subject to any of the above issues. Readers may refer to Gan et al. (2010) about more issues associated with which design to choose between a single-arm phase II trial and a randomized phase II trial.

Let p_x and p_y denote the response probabilities of an experimental and a control arms, respectively. In a randomized phase II trial, we want to test $H_0 : p_x = p_y$ against $H_1 : p_x > p_y$. The null distribution of the binomial test statistic depends on the common response probability $p_x = p_y$, see Jung (2008). Consequently, if the true response probabilities are different from the specified ones, the testing based on binomial distributions may not maintain the type I error close to the specified design value. In order to avoid this issue, Jung (2008) proposes to control the type I error rate at $p_x = p_y = 1/2$. This results in a strong conservativeness when the true response probability is different from 50%. Asymptotic tests avoid specification of $p_x = p_y$ by replacing them with their consistent estimators, but the sample sizes of phase II trials usually are not large enough for a good large sample approximation.

Fisher's (1935) exact test has been a popular testing method for comparing two sample binomial proportions with small sample sizes. In a randomized phase II trial setting, Fisher's exact test is based on the distribution of the number of responders on one arm conditioning on the total number of responders which is a sufficient statistic of $p_x = p_y$ under H_0 . Hence, the rejection value of Fisher's exact test does not require specification of the common response probabilities $p_x = p_y$ under H_0 . In this paper, we propose two-stage randomized phase II trial designs based on Fisher's exact test. Using some example designs, we show that Fisher's exact test accurately controls the type I error over the wide range of true response values, and is more powerful than Jung's binomial test if the true response probabilities are different from 50%. If we can project the true response probabilities accurately at the design stage, we can identify efficient designs by adopting the Simon's (1989) optimal and minimax design concepts that were proposed for single-arm phase II trials. We provide tables of minimax and optimal two-stage designs under various practical design settings.

In this paper, we limit our focus on randomized phase II trials for evaluating the efficacy of an experimental therapy compared to a prospective control. Although we demonstrate the proposed method using response as the endpoint, it can be applied to any binomial endpoint, e.g. the proportion of patients progression-free at a fixed time point (say 6 months). Other types of randomized phase II trial designs have been proposed by many investigators including Simon, Wittes and Ellenberg (1985), Sargent and Goldberg (2001), Thall, Simon and Ellenberg (1989), Palmer (1991), and Steinberg and Venzon (2002). Rubinstein et al. (2005) discuss the strengths and weaknesses of some of these methods, and propose a

method for randomized phase II screening designs based on the usual large-sample approximation. As we will demonstrate in Section 3.2, we observed from numerical studies that the designs based on the large-sample theory may not control the type I error accurately with small sample sizes, motivating the present work.

2 Single-Stage Design

If patient accrual is fast or it takes a lengthy time (say, longer than 6 months) for response assessment, we may consider using a single-stage design. Suppose that n patients are randomized to each arm, and let X and Y denote the number of responders in arms x (experimental) and y (control), respectively. Let $q_k = 1 - p_k$ for arm $k (= x, y)$. Then the frequencies (and response probabilities in the parentheses) can be summarized as in Table 1.

At the design stage, n is prespecified. Fisher's exact test is based on the conditional distribution of X given the total number of responders $Z = X + Y$ with a probability mass function

$$f(x|z, \theta) = \frac{\binom{n}{x} \binom{n}{z-x} \theta^x}{\sum_{i=m_-}^{m_+} \binom{n}{i} \binom{n}{z-i} \theta^i}$$

for $m_- \leq x \leq m_+$, where $m_- = \max(0, z - n)$, $m_+ = \min(z, n)$, and $\theta = p_x q_y / (p_y q_x)$ denotes the odds ratio.

Suppose that we want to limit the type I error rate to be below α^* . Given $X + Y = z$, we reject $H_0 : p_x = p_y$ (i.e. $\theta = 1$) in favor of $H_1 : p_x > p_y$ (i.e. $\theta > 1$) if $X - Y \geq a$, where a is the smallest integer satisfying

$$\alpha(z) \equiv P(X - Y \geq a | z, H_0) = \sum_{x=\langle (z+a)/2 \rangle}^{m_+} f(x|z, \theta=1) \leq \alpha^*,$$

where $\langle c \rangle$ is the round-up integer of c . Hence, the critical value a depends on the total number of responders z . Under $H_1 : \theta = \theta_1 (> 1)$, the power conditional on $X + Y = z$ is given by

$$1 - \beta(z) \equiv P(X - Y \geq a | z, H_1) = \sum_{x=\langle (z+a)/2 \rangle}^{m_+} f(x|z, \theta_1).$$

We propose to choose n so that the marginal power is no smaller than a specified power level $1 - \beta^*$, i.e.

$$E\{1 - \beta(Z)\} = \sum_{z=0}^{2n} \{1 - \beta(z)\} g(z) \geq 1 - \beta^*$$

where $g(z)$ is the probability mass function of $Z = X + Y$ under $H_1 : p_x > p_y$ that is given as

$$g(z) = \sum_{x=m_-}^{m_+} \binom{n}{x} p_x^x q_x^{n-x} \binom{n}{z-x} p_y^{z-x} q_y^{n-z+x}$$

for $z = 0, 1, \dots, 2n$. Note that the marginal type I error rate is controlled below α^* since the conditional type I error rate is controlled below α^* for any z value.

Given a type I error rate and a power ($\alpha^*, 1 - \beta^*$) and a specific alternative hypothesis $H_1 : (p_x, p_y)$, we find a sample size n as follows.

Algorithm for Single-Stage Design:

1. For $n = 1, 2, \dots$,

a. For $z = 0, 1, \dots, 2n$, find the smallest $a = a(z)$ such that

$$\alpha(z) = P(X - Y \geq a | z, \theta = 1) \leq \alpha^*$$

and calculate the power conditional on $X + Y = z$ for the chosen $a = a(z)$

$$1 - \beta(z) = P(X - Y \geq a | z, \theta_1).$$

b. Calculate the marginal power $1 - \beta = E\{1 - \beta(Z)\}$.

2. Find the smallest n such that $1 - \beta \geq 1 - \beta^*$.

Given a fixed n , Fisher's test that is based on the conditional distribution is valid under $\theta = 1$ (i.e. controls the type I error rate exactly), and its power conditional on the total number of responders depends only on the odds ratio θ_1 under H_1 . However, the marginal power, and hence the sample size n , depends on (p_1, p_2) , so that we need to specify (p_1, p_2) at the design stage. If (p_1, p_2) are mis-specified, the trial may be over- or under-powered but the type I error in data analysis will always be appropriately controlled.

3 Two-Stage Design

For ethical and economical reasons, clinical trials are often conducted using multiple stages. Phase II trials usually enter small number of patients, so that practically the number of stages is two at the most. We consider designs with the same features as popular two-stage phase II trial designs, with an early stopping rule when the experimental therapy has a low probability achieving additional benefits to the patients.

Suppose that n_l ($l = 1, 2$) patients are randomized to each arm during stage l ($l = 1, 2$). Let $n_1 + n_2 = n$ denote the maximal sample size for each arm, X_l and Y_l denote the number of responders during stage l in arms x and y , respectively, $X = X_1 + X_2$ and $Y = Y_1 + Y_2$.

At the design stage, n_l are prespecified. Note that X_1 and X_2 are independent, and, given $X_l + Y_l = z_l$, X_l has the conditional probability mass function

$$f_l(x_l|z_l, \theta) = \frac{\binom{n_l}{x_l} \binom{n_l}{z_l - x_l} \theta^{x_l}}{\sum_{i=m_{l-}}^{m_{l+}} \binom{n_l}{i} \binom{n_l}{z_l - i} \theta^i} \quad (1)$$

for $m_{l-} \leq x_l \leq m_{l+}$, where $m_{l-} = \max(0, z_l - n_l)$ and $m_{l+} = \min(z_l, n_l)$.

We consider a two-stage randomized phase II trial whose rejection values are chosen conditional on z_1 and z_2 as follows.

Stage 1: Randomize n_1 patients to each arm; observe x_1 and y_1 .

- a. Given $z_1 (= x_1 + y_1)$, find a stopping value $a_1 = a_1(z_1)$.
- b. If $x_1 - y_1 \leq a_1$, proceed to stage 2.
- c. Otherwise, stop the trial.

Stage 2: Randomize n_2 patients to each arm; observe x_2 and y_2 ($z_2 = x_2 + y_2$).

- a. Given (z_1, z_2) , find a rejection value $a = a(z_1, z_2)$.
- b. Accept the experimental arm if $x - y \geq a$.

Now, the question is how to choose rejection values (a_1, a) conditioning on (z_1, z_2) .

3.1 Choice of a_1 and a

In this section, we assume that n_1 and n_2 are given. We consider different options of choosing a_1 . For example,

- We may wish to stop the trial if the experimental arm is worse than the control. In this case, we choose $a_1 = 0$. This a_1 is constant with respect to z_1 .
- We may choose a_1 so that the conditional probability of early termination (PET) given z_1 is no smaller than a level γ_0 ($= 0.6$ to 0.8) under $H_0 : \theta = 1$, i.e.

$$\text{PET}_0(z_1) = P(X_1 - Y_1 < a_1 \mid z_1, H_0) = \sum_{x_1=m_{1-}}^{[(a_1+z_1)/2]-1} f_1(x_1 \mid z_1, \theta=1) \geq \gamma_0,$$

where $[c]$ denotes the largest integer not exceeding c .

- We may choose a_1 so that the conditional probability of early termination given z_1 is no larger than a level γ_1 ($= 0.02$ to 0.1) under $H_1 : \theta = \theta_1$, i.e.

$$\text{PET}_1(z_1) = P(X_1 - Y_1 < a_1 \mid z_1, H_1) = \sum_{x_1=m_{1-}}^{[(a_1+z_1)/2]-1} f_1(x_1 \mid z_1, \theta_1) \leq \gamma_1.$$

Among these options, we propose to use $a_1 = 0$. Most standard optimal two-stage phase II trials also stop early when the observed response probability from stage 1 is no larger than the specified response probability under H_0 , refer to Simon (1989) and Jung *et al.* (2004) for single-arm trial cases and Jung (2008) for randomized trial cases.

With a_1 fixed at 0, we choose the second stage rejection value a conditioning on (z_1, z_2) . Given type I error rate α^* , a is chosen as the smallest integer satisfying

$$\alpha(z_1, z_2) \equiv P(X_1 - Y_1 \geq a_1, X - Y \geq a \mid z_1, z_2, \theta=1) \leq \alpha^*.$$

We calculate $\alpha(z_1, z_2)$ by

$$\begin{aligned} &P(X_1 \geq (a_1 + z_1)/2, X_1 + X_2 \geq (a + z_1 + z_2)/2 \mid z_1, z_2, \theta) \\ &= 1 - \sum_{x_1=m_{1-}}^{m_{1+}} \sum_{x_2=m_{2-}}^{m_{2+}} I\{x_1 \geq (a_1 + z_1)/2, x_1 + x_2 \geq (a + z_1 + z_2)/2\} f_1(x_1 \mid z_1, 1) f_2(x_2 \mid z_2, 1), \end{aligned}$$

where $I(\cdot)$ is the indicator function.

Under $H_1 : \theta = \theta_1$, the power conditional on (z_1, z_2) is obtained by

$$\begin{aligned} &1 - \beta(z_1, z_2) \\ &= P(X_1 - Y_1 \geq a_1, X - Y \geq a \mid z_1, z_2, \theta_1) \\ &= \sum_{x_1=m_{1-}}^{m_{1+}} \sum_{x_2=m_{2-}}^{m_{2+}} I\{x_1 \geq (a_1 + z_1)/2, x_1 + x_2 \geq (a + z_1 + z_2)/2\} f_1(x_1 \mid z_1, \theta_1) f_2(x_2 \mid z_2, \theta_1). \end{aligned}$$

Note that, as in the single-stage case, the calculation of type I error rate $\alpha(z_1, z_2)$ and rejection values (a_1, a) does not require specification of the common response probability $p_x = p_y$ under H_0 , and that the conditional power $1 - \beta(z_1, z_2)$ requires specification of the odds ratio θ_1 under H_1 , but not the response probabilities for the two arms, p_x and p_y .

3.2 Choice of n_1 and n_2

In this section we discuss how to choose sample sizes n_1 and n_2 at the design stage based on various optimality criteria.

Given (α^*, β^*) , we propose to choose n_1 and n_2 so that the marginal power is maintained above $1 - \beta^*$ while controlling the conditional type I error rates for any (z_1, z_2) below α^* as described in Section 3.1. For stage $l (= 1, 2)$, the marginal distribution of $Z_l = X_l + Y_l$ has a probability mass function

$$g_l(z_l) = \sum_{x_l=m_{l-}}^{m_{l+}} \binom{n_l}{x_l} p_x^{x_l} q_x^{n_l-x_l} \binom{n_l}{z_l-x_l} p_y^{z_l-x_l} q_y^{n_l-z_l+x_l} \quad (2)$$

for $z_l = 0, \dots, 2n_l$. Under $H_0 : p_x = p_y = p_0$, this is expressed as

$$g_{0l}(z_l) = p_0^{z_l} q_0^{2n_l-z_l} \sum_{x_l=m_{l-}}^{m_{l+}} \binom{n_l}{x_l} \binom{n_l}{z_l-x_l}.$$

Z_1 and Z_2 are independent. Hence, we choose n_1 and n_2 so that the marginal power is no smaller than a specified level $1 - \beta^*$, i.e.

$$1 - \beta \equiv \sum_{z_1=0}^{2n_1} \sum_{z_2=0}^{2n_2} \{1 - \beta(z_1, z_2)\} g_1(z_1) g_2(z_2) \geq 1 - \beta^*.$$

The marginal type I error is calculated by

$$\alpha \equiv \sum_{z_1=0}^{2n_1} \sum_{z_2=0}^{2n_2} \alpha(z_1, z_2) g_{01}(z_1) g_{02}(z_2).$$

Since the conditional type I error rate is controlled below α^* for any (z_1, z_2) , the marginal type I error rate is no larger than α^* .

Although we do not have to specify the response probabilities for testing, we need to do so when choosing (n_1, n_2) at the design stage. If the specified response probabilities are different from the true ones, then the marginal power may be different from that expected. But in this case, our proposed test is still valid in the sense that it always controls both the conditional and marginal type I error rates below the specified level. Let

$PET_0 \equiv E\{PET_0(Z_1)|H_0\} = \sum_{z_1=0}^{2n_1} PET_0(z_1) g_{01}(z_1)$ denote the marginal probability of early termination under H_0 . Then, among those (n_1, n_2) satisfying the $(\alpha^*, 1 - \beta^*)$ -condition, the Simon-type (1989) minimax and the optimal designs can be chosen as follows.

- *Minimax design* chooses (n_1, n_2) with the smallest maximal sample size $n (= n_1 + n_2)$.
- *Optimal design* chooses (n_1, n_2) with the smallest marginal expected sample size EN under H_0 , where

$$EN = n_1 \times PET_0 + n \times (1 - PET_0).$$

Tables 2 to 5 report the sample sizes (n, n_1) of the minimax and optimal two-stage designs for $\alpha^* = 0.15$ or 0.2 , $1 - \beta^* = 0.8$ or 0.85 , and various combinations of (p_x, p_y) under H_1 . For comparison, we also list the sample size n of the single-stage design under each setting. Note that the maximal sample size of the minimax is slightly smaller than or equal to the sample size of the single-stage design. If the experimental therapy is inefficacious, however, the expected sample sizes of minimax and optimal designs are much smaller than the sample size of the single-stage design. We also observe from Tables 2 and 5 that the sample sizes under $(\alpha^*, 1 - \beta^*) = (0.15, 0.8)$ are similar to those under $(\alpha^*, 1 - \beta^*) = (0.2, 0.85)$.

One of popular approaches for randomized phase II trials is to use the asymptotic method. Given $(\alpha^*, p_y, n_1, n_2)$, we find c satisfying

$$\alpha = P(X_1 - Y_1 \geq 0, \frac{X - Y}{\sqrt{2n\hat{p}\hat{q}}} \geq c | p_x = p_y)$$

using the normal approximation to binomial distributions, where $\hat{p} = (X + Y)/2n$ and $\hat{q} = 1 - \hat{p}$. For an approximate critical value c , the exact type I error rate is calculated by using the true binomial distribution. For a specified p_x (p_y), the exact power is calculated similarly. From Table 2, the minimax design under $(\alpha^*, 1 - \beta^*, p_x, p_y) = (0.15, 0.8, 0.5, 0.35)$ has $(n, n_1) = (86, 66)$, for which the asymptotic method has $\alpha = 0.157$ and $1 - \beta = 0.840$. Since the sample size is relatively large in this case, the asymptotic method controls the power close to the nominal $\alpha^* = 0.15$.

Now, we consider the minimax design $(n, n_1) = (29, 11)$ under $(\alpha^*, 1 - \beta^*, p_x, p_y) = (0.15, 0.8, 0.35, 0.05)$. In this case, the asymptotic method has $\alpha = 0.244$, which is far larger than the nominal $\alpha^* = 0.15$ because of the small sample size.

4 Numerical Studies

Jung (2008) proposed a randomized phase II design method based on the binomial test, called MaxTest in this paper, by controlling the type I error rate at $p_x = p_y = 50\%$. Since the type I error rate of the two-sample binomial test is maximized at $p_x = p_y = 50\%$, this test will be conservative if the true response probability under H_0 is different from 50%.

Here we compare the performance of our Fisher's test with that of MaxTest. All the calculations in this section are based on exact distributions, not on simulations. Figure 1 displays the type I error rate and power in the range of $0 < p_y < 1 - p_x$ for single-stage designs with $n = 60$ per arm, $p_x - p_y = 0.15$ or 0.2 under H_1 and $\alpha = 0.1, 0.15$ or 0.2 under $H_0 : p_x = p_y$. The solid lines are for Fisher's test and the dotted lines are for MaxTest; the lower two lines represent type I error rate and the upper two lines represent power. As is well known, Fisher's test controls the type I error conservatively over the range of p_y . The conservativeness gets slightly stronger with small p_y values close to 0. MaxTest controls the type I error accurately around $p_y = 0.5$, but becomes more conservative for p_y values far from 0.5, especially with small p_y values. For $\alpha = 0.1$, Fisher's test and MaxTest have similar power around $0.2 < p_y < 0.4$ except that MaxTest is slightly more powerful for $p_y \approx 0.4$. Otherwise, Fisher's test is more powerful. The difference in power between the two

methods becomes larger with $\alpha = 0.15$. We observe similar trends overall, but the difference in power becomes smaller with $\alpha = 0.2$, especially when combined with a large $\alpha (= 0.2)$.

Figure 2 displays the type I error rate and power of two-stage designs with $n_1 = n_2 = 30$ per arm. We observe that, compared to MaxTest, Fisher's test controls the type I error more accurately in most range of p_y values. If $\alpha = 0.1$, Fisher's test is more powerful than MaxTest over the range of $p_y < 0.2$ or $p_y > 0.6$. But with a larger α , such as 0.15 or 0.2, MaxTest is more powerful in the range of p_y between 0.2 and 0.6. As in the single-stage design case, the difference in power diminishes as α and α increase. While Fisher's exact test is more powerful than the binomial test in single-stage designs, their performance is comparable in two-stage designs. With a_1 value fixed at 0, the two-stage designs based on Fisher's exact test are not fully optimal. A fully optimal choice of a_1 value will depend on z_1 and the two-stage design using the optimal a_1 value is believed to outperform the two-stage designs based on binomial test in as single-stage design case. We will develop an efficient algorithm to identify fully optimal two-stage designs using Fisher's exact test in our future study.

5 Unbalanced Two-Stage Randomized Trials

One may want to accrue more patients to one arm than the other for some reasons, e.g. to treat more patients by an experimental therapy than a control. In this case, the test statistic based on the difference in number of responders between two arms that has been considered so far is not appropriate. Let m_l and n_l denote the sample sizes at stage $l (= 1, 2)$ of arms x and y, respectively ($m = m_1 + m_2$, $n = n_1 + n_2$). Also, let X_l and Y_l denote the number of responders among stage l patients of arms x and y, respectively ($X = X_1 + X_2$, $Y = Y_1 + Y_2$). If we want to assign γ times larger number of patients to arm x than to arm y, then we have $m_l = \gamma \times n_l$ and $m = \gamma \times n$. Note that a choice of $\gamma = 1$ corresponds to the balanced two-stage designs considered in the previous section. When $\gamma \neq 1$, it does not make sense to directly compare the numbers of responders between arms at each stage. For the odds ratio, $\theta = p_x q_y / (q_x p_y)$, we want to design a study for $H_0 : \theta = 1$ vs. $H_0 : \theta = \theta_a (> 1)$, where $q_k = 1 - p_k$.

We propose following two-stage design.

- **Stage 1:** Accrue m_1 patients to arm x and n_1 patients to arm y, and observe X_1 and Y_1 . For $\hat{p}_{x,1} = X_1/m_1$, $\hat{p}_{y,1} = Y_1/n_1$, $\hat{q}_{k,1} = 1 - \hat{p}_{k,1}$, and $\hat{\theta}_1 = (\hat{p}_{x,1}\hat{q}_{y,1})/(\hat{q}_{x,1}\hat{p}_{y,1})$,
 - a. If $\hat{\theta}_1 < 1$, reject arm x and stop the trial.
 - b. Otherwise, proceed to the second stage.
- **Stage 2:** Accrue an additional m_2 patients to arm x and n_2 patients to arm y, and observe X_2 and Y_2 . For $\hat{p}_x = X/m$, $\hat{p}_y = Y/n$, $\hat{q}_k = 1 - \hat{p}_k$, and $\hat{\theta} = (\hat{p}_x\hat{q}_y)/(\hat{q}_x\hat{p}_y)$,
 - a. Accept arm x for further investigation if $\hat{\theta} \geq a$.
 - b. Otherwise, reject arm x.

Given $Z_l = z_l$, X_l has probability mass function

$$f_l(x_l|z_l, \theta) = \frac{\binom{m_l}{x_l} \binom{n_l}{z_l - x_l} \theta^{x_l}}{\sum_{i=m_{l-}}^{m_{l+}} \binom{m_l}{i} \binom{n_l}{z_l - i} \theta^i}$$

for $m_{l-} \leq x_l \leq m_{l+}$, where $m_{l-} = \max(0, z_l - n_l)$ and $m_{l+} = \min(z_l, m_l)$. Let $\hat{\theta}_1 = x_1(n_1 - y_1)/\{y_1(m_1 - x_1)\}$ and $\hat{\theta} = \{x(n - y)\}/\{y(m - x)\}$ denote the estimates of θ after stage 1 and 2, respectively. Note that $\hat{\theta}_1 = \hat{\theta}_1(x_1)$ is a function of x_1 given z_1 , and $\hat{\theta} = \hat{\theta}(x_1, x_2)$ is a function of (x_1, x_2) given (z_1, z_2) . Given type I error rate α^* and (m_1, m_2, z_1, z_2) , we find a satisfying $\alpha(z_1, z_2) = \alpha^*$, where

$$\alpha(z_1, z_2) = P\{\hat{\theta}_1(X_1) \geq 1, \hat{\theta}(X_1, X_2) \geq a | z_1, z_2, H_0\} = \sum_{x_1=m_{1-}}^{m_{1+}} \sum_{x_2=m_{2-}}^{m_{2+}} I\{\hat{\theta}_1(x_1) \geq 1, \hat{\theta}(x_1, x_2) \geq a\} f_1(x_1|z_1, 1) f_2(x_2|z_2, 1).$$

The conditional power of the two-stage design is calculated by

$$1 - \beta(z_1, z_2) = P\{\hat{\theta}_1(X_1) \geq 1, \hat{\theta}(X_1, X_2) \geq a | z_1, z_2, H_a\} = \sum_{x_1=m_{1-}}^{m_{1+}} \sum_{x_2=m_{2-}}^{m_{2+}} I\{\hat{\theta}_1(x_1) \geq 1, \hat{\theta}(x_1, x_2) \geq a\} f_1(x_1|z_1, \theta_a) f_2(x_2|z_2, \theta_a).$$

Noting that Z_l is sum of two independent binomial random variables, X_l and Y_l , its probability mass function is given as

$$g_l(z_l) = \sum_{x_l=m_{l-}}^{m_{l+}} \binom{m_l}{x_l} p_x^{x_l} q_x^{m_l - x_l} \binom{n_l}{z_l - x_l} p_y^{z_l - x_l} q_y^{n_l - z_l + x_l}$$

for $z_l = 0, \dots, m_l + n_l$. Under $H_0 : \theta = 1$, $g_l(z_l)$ is expressed as

$$g_{0l}(z_l) = \sum_{x_l=m_{l-}}^{m_{l+}} \binom{m_l}{x_l} \binom{n_l}{z_l - x_l} p_y^{z_l} q_y^{m_l + n_l - z_l}.$$

Hence, the marginal type I error rate and power of above two-stage design are calculated by

$$\alpha = E\{\alpha(Z_1, Z_2)\} = \sum_{z_1=0}^{m_1+n_1} \sum_{z_2=0}^{m_2+n_2} \alpha(z_1, z_2) g_{01}(z_1) g_{02}(z_2)$$

and

$$1 - \beta = E\{1 - \beta(Z_1, Z_2)\} = \sum_{z_1=0}^{m_1+n_1} \sum_{z_2=0}^{m_2+n_2} \{1 - \beta(z_1, z_2)\} g_1(z_1) g_2(z_2),$$

respectively. Since $\alpha(z_1, z_2) = \alpha^*$ for all (z_1, z_2) , we have $\alpha = \alpha^*$.

For specified a type I error rate α^* and a power $1 - \beta^*$, we want to select a two-stage design satisfying $\alpha = \alpha^*$ and $1 - \beta = 1 - \beta^*$. Under H_0 , the probability of early termination and the expected sample size for arm x are calculated as

$$PET_0 = P\{\hat{\theta}_1(X_1) < 1 | z_1, H_0\}$$

and

$$EN = m_1 \times PET_0 + m \times (1 - PET_0),$$

, respectively. Among the two-stage designs satisfying the $(\alpha^*, 1 - \beta^*)$ -restriction, the optimal design is defined as the one with the smallest EN. The minimax design is defined as the one with the smallest m (or $m+n$) among the two-stage designs satisfying the $(\alpha^*, 1 - \beta^*)$ -restriction.

6 Discussions

We have proposed design and analysis methods for two-stage randomized phase II clinical trials based on Fisher's exact test. While the binomial test by Jung (2008) requires specification of the response probability of the control arm p_y or conservatively controls the type I error rate at $p_y = 0.5$, Fisher's exact test does not require specification of p_y . If p_x and p_y under H_1 can be accurately specified at the design stage, we can calculate the expected sample size under H_0 and the sample sizes (n_1, n_2) for the minimax and optimal two-stage designs. Even if p_x and p_y are mis-specified at the design, Fisher's test accurately controls the type I error rate and maintains a higher power than the binomial test, especially if p_x and p_y are different from 50%.

Jung's (2008) designs based on the binomial test are specified by the sample sizes and rejection values (n_1, n_2, a_1, a) . For two-stage Fisher's exact test, however, the rejection value of the second stage a is chosen depending on the total numbers of responders through two stages, (z_1, z_2) , so that the protocol of a randomized phase II trial design based on Fisher's exact test may include a list of a values for all possible outcomes of (z_1, z_2) .

Even though the sample sizes (n_1, n_2) are determined at the design stage, the realized sizes when the study is completed may be slightly different from the pre-specified ones. This kind of discrepancy in sample sizes is no issue for our method by performing a two-stage Fisher's exact test conditioning on the realized sample sizes as well as the total number of responders.

The computing time of our sample size calculation procedure depends on how large the target sample size is, but it takes only a few seconds for most of the cases reported in Tables 2 to 5. The Fortran program to find minimax and optimal designs are available from the first author.

Acknowledgments

This research was supported by a grant from the National Cancer Institute (CA142538-01).

References

1. Cannistra SA. Phase II trials in *Journal of Clinical Oncology*. *Journal of Clinical Oncology*. 2009; 27(19):3073–3076. [PubMed: 19451415]
2. Fisher RA. The logic of inductive inference (with discussion). *Journal of Royal Statistical Society*. 1935; 98:39–82.
3. Gan HK, Grothey A, Pond GP, Moore MJ, Siu LL, Sargent DJ. Randomized Phase II Trials: Inevitable or Inadvisable? *Journal of Clinical Oncology*. 2010; 28(15):2641–2647. [PubMed: 20406933]
4. Jung SH. Randomized phase II trials with a prospective control. *Statistics in Medicine*. 2008; 27:568–583. [PubMed: 17573688]
5. Jung SH, Lee TY, Kim KM, George S. Admissible two-stage designs for phase II cancer clinical trials. *Statistics in Medicine*. 2004; 23:561–569. [PubMed: 14755389]
6. Palmer CR. A comparative phase II clinical trials procedure for choosing the best of three treatments. *Statistics in Medicine*. 1991; 10:1327–1340. [PubMed: 1925164]
7. Rubinstein LV, Korn EL, Freidlin B, Hunsberger S, Ivy SP, Smith MA. Design issues of randomized phase II trials and a proposal for phase II screening trials. *Journal of Clinical Oncology*. 2005; 23(28):7199–7206. [PubMed: 16192604]
8. Sargent DJ, Goldberg RMA. exible design for multiple armed screening trials. *Statistics in Medicine*. 2001; 20:1051–1060. [PubMed: 11276035]
9. Simon R. Optimal two-stage designs for phase II clinical trials. *Controlled Clinical Trials*. 1989; 10:1–10. [PubMed: 2702835]
10. Simon R, Wittes RE, Ellenberg SS. Randomized phase II clinical trials. *Cancer Treatment Reports*. 1985; 69:1375–1381. [PubMed: 4075313]
11. Steinberg SM, Venzon DJ. Early selection in a randomized phase II clinical trial. *Statistics in Medicine*. 2002; 21:1711–1726. [PubMed: 12111907]
12. Thall PF, Simon R, Ellenberg SS. A two-stage design for choosing among several experimental treatments and a control in clinical trials. *Biometrics*. 1989; 45:537–547. [PubMed: 2765637]

Figure 1.
Single-stage designs with $n = 60$ per arm: Type I error rate and power for Fisher's test (solid lines) and MaxTest (dotted lines)

Figure 2. Two-stage designs with $n_1 = n_2 = 30$ per arm: Type I error rate and power for Fisher's test (solid lines) and MaxTest (dotted lines)

Table 1

Frequencies (and response probabilities in the parentheses) of a single-stage randomized phase II trial

		Arm 1	Arm 2	Total
Response	Yes	$x (p_x)$	$y (p_y)$	z
	No	$n - x (q_x)$	$n - y (q_y)$	$2n - z$
Total		n	n	

Table 2

Single-stage designs, and minimax and optimal two-stage Fisher designs for $(\alpha^*, 1 - \beta^*) = (.15, .8)$

p_y	p_x	θ	Single-Stage Design			Minimax Two-Stage Design			Optimal Two-Stage Design				
			n	α	$1 - \beta$	(n, n_1)	α	$1 - \beta$	EN	(n, n_1)	α	$1 - \beta$	EN
.05	.15	3.353	79	.0827	.8005	(78, 40)	.0823	.8001	63.00	(81, 26)	.0836	.8008	60.83
	.2	4.750	45	.0631	.8075	(44, 17)	.0620	.8033	35.11	(44, 17)	.0620	.8033	35.11
	.25	6.333	29	.0450	.8109	(29, 11)	.0448	.8014	23.96	(29, 11)	.0448	.8014	23.96
.1	.25	3.000	56	.0884	.8033	(56, 25)	.0896	.8003	43.46	(58, 19)	.0925	.8016	42.80
	.3	3.857	36	.0747	.8016	(36, 16)	.0783	.8009	28.41	(37, 12)	.0800	.8024	28.03
	.15	.3	2.429	65	.1036	.8023	(65, 36)	.1036	.8004	52.42	(69, 22)	.1060	.8009
.2	.35	3.051	41	.0879	.8025	(41, 19)	.0919	.8010	32.01	(42, 14)	.0925	.8016	30.99
	.35	2.154	74	.1076	.8054	(74, 42)	.1097	.8001	59.74	(79, 26)	.1133	.8003	56.17
	.4	2.667	46	.0953	.8074	(46, 23)	.1016	.8007	36.19	(49, 14)	.1061	.8019	34.81
.25	.4	2.000	83	.1149	.8022	(81, 37)	.1147	.8005	61.35	(84, 30)	.1155	.8006	60.21
	.45	2.455	47	.0986	.8056	(47, 27)	.1007	.8005	38.25	(50, 17)	.1053	.8003	36.10
	.3	.45	1.909	85	.1112	.8005	(85, 65)	.1115	.8000	75.76	(95, 27)	.1210	.8004
.35	.5	2.333	53	.1121	.8015	(49, 26)	.1100	.8007	38.88	(52, 19)	.1113	.8014	37.82
	.5	1.857	86	.1155	.8006	(86, 66)	.1153	.8000	76.73	(95, 32)	.1186	.8002	66.78
	.55	2.270	54	.1009	.8026	(54, 21)	.1169	.8016	39.62	(55, 19)	.1168	.8012	39.43
.4	.55	1.833	87	.1228	.8032	(87, 59)	.1216	.8004	74.05	(94, 35)	.1205	.8000	67.36
	.6	2.250	54	.1015	.8013	(54, 22)	.1150	.8017	39.95	(56, 18)	.1147	.8006	39.56
	.45	.6	1.833	87	.1265	.8032	(87, 59)	.1252	.8004	74.03	(94, 35)	.1234	.8000
.5	.65	2.270	54	.1043	.8026	(54, 21)	.1151	.8016	39.53	(55, 19)	.1148	.8012	39.33
	.65	1.857	86	.1263	.8006	(86, 66)	.1260	.8000	76.69	(95, 32)	.1235	.8002	66.63
	.7	2.333	53	.1033	.8015	(49, 26)	.1147	.8007	38.77	(52, 19)	.1126	.8014	37.62
.55	.7	1.909	85	.1237	.8005	(85, 65)	.1234	.8000	75.70	(96, 26)	.1203	.8010	64.87
	.75	2.455	47	.1114	.8056	(47, 27)	.1240	.8005	38.09	(50, 17)	.1221	.8003	35.75
	.6	.75	2.000	83	.1173	.8022	(81, 37)	.1134	.8005	61.08	(84, 30)	.1193	.8006
.65	.8	2.667	46	.1208	.8074	(46, 23)	.1163	.8007	35.87	(50, 12)	.1145	.8011	34.13
	.8	2.154	74	.1150	.8054	(74, 42)	.1214	.8001	59.46	(81, 23)	.1213	.8006	55.56
	.85	3.051	41	.1020	.8025	(41, 19)	.1004	.8010	31.48	(43, 12)	.1060	.8024	30.12

	Single-Stage Design					Minimax Two-Stage Design					Optimal Two-Stage Design				
	p_y	p_x	θ	n	α	$1 - \beta$	(n, n_1)	α	$1 - \beta$	EN	(n, n_1)	α	$1 - \beta$	EN	
.7		.85	2.429	65	.1088	.8023	(65, 36)	.1083	.8004	51.98	(69, 22)	.1139	.8009	48.57	
		.9	3.857	36	.1023	.8016	(36, 16)	.1068	.8009	27.53	(38, 9)	.1075	.8004	26.45	
.75		.9	3.000	56	.1098	.8033	(56, 25)	.1099	.8003	42.52	(59, 17)	.1110	.8016	41.30	
		.95	6.333	29	.0929	.8109	(29, 11)	.0919	.8014	21.76	(30, 7)	.0949	.8022	21.32	
.8		.95	4.750	45	.0948	.8075	(44, 17)	.1036	.8033	32.81	(46, 11)	.1041	.8007	32.23	
		.85	3.353	79	.1065	.8005	(78, 40)	.1071	.8001	61.38	(83, 22)	.1098	.8011	57.67	

Table 3

Single-stage designs, and minimax and optimal two-stage Fisher designs for $(\alpha^*, 1 - \beta^*) = (.15, .85)$

Single-Stage Design			Minimax Two-Stage Design			Optimal Two-Stage Design								
p_y	p_x	θ	n	α	$1 - \beta$	(n, n_1)	α	$1 - \beta$	EN	(n, n_1)	α	$1 - \beta$	EN	
.05	.15	3.353	92	.0868	.8502	(92, 48)	.0870	.8500	74.20	(94, 35)	.0881	.8502	71.17	
	.2	4.750	51	.0677	.8526	(51, 24)	.0676	.8506	41.27	(52, 18)	.0678	.8505	40.61	
	.25	6.333	35	.0531	.8581	(34, 11)	.0516	.8506	27.56	(34, 11)	.0516	.8506	27.56	
.1	.25	3.000	65	.0952	.8529	(65, 37)	.0950	.8502	53.18	(68, 24)	.0963	.8501	50.29	
	.3	3.857	41	.0785	.8515	(41, 21)	.0823	.8500	33.09	(42, 16)	.0847	.8502	32.14	
	.15	.3	2.429	78	.1058	.8521	(78, 48)	.1067	.8502	64.71	(82, 29)	.1091	.8506	59.40
.2	.35	3.051	49	.0948	.8541	(49, 23)	.0976	.8506	38.15	(51, 17)	.1004	.8527	37.28	
	.3	2.154	88	.1113	.8518	(88, 43)	.1120	.8502	67.92	(93, 32)	.1147	.8500	66.49	
	.4	2.667	52	.0990	.8506	(52, 29)	.0992	.8506	42.01	(56, 18)	.1062	.8506	40.16	
.25	.4	2.000	94	.1132	.8511	(94, 68)	.1133	.8502	82.03	(102, 37)	.1203	.8503	72.98	
	.45	2.455	59	.1067	.8535	(59, 30)	.1122	.8503	46.22	(62, 19)	.1129	.8506	43.71	
	.3	.45	1.909	104	.1210	.8504	(100, 55)	.1200	.8501	79.37	(107, 39)	.1209	.8503	76.34
.35	.5	2.333	60	.1034	.8524	(60, 39)	.1062	.8501	50.53	(65, 22)	.1138	.8510	46.31	
	.5	1.857	106	.1148	.8508	(106, 79)	.1148	.8501	93.40	(114, 39)	.1252	.8503	80.04	
	.55	2.270	61	.1089	.8551	(61, 37)	.1082	.8507	50.16	(65, 24)	.1112	.8504	46.96	
.4	.55	1.833	107	.1178	.8520	(107, 74)	.1170	.8500	91.60	(115, 45)	.1205	.8500	83.00	
	.6	2.250	61	.1147	.8538	(61, 38)	.1132	.8500	50.57	(66, 23)	.1138	.8502	47.07	
	.45	.6	1.833	107	.1214	.8520	(107, 74)	.1203	.8500	91.59	(115, 45)	.1196	.8500	82.95
.45	.65	2.270	61	.1184	.8551	(61, 37)	.1163	.8507	50.11	(65, 24)	.1156	.8504	46.86	
	.65	1.857	106	.1215	.8508	(106, 79)	.1210	.8501	93.36	(114, 39)	.1240	.8503	79.88	
	.5	.7	2.333	60	.1176	.8524	(60, 39)	.1163	.8501	50.45	(65, 22)	.1151	.8510	45.07
.55	.7	1.909	104	.1180	.8504	(100, 55)	.1159	.8501	79.22	(107, 39)	.1234	.8503	76.08	
	.75	2.455	59	.1145	.8535	(59, 30)	.1104	.8503	46.00	(62, 19)	.1114	.8506	43.28	
	.6	.75	2.000	94	.1205	.8511	(94, 68)	.1256	.8502	81.91	(102, 37)	.1277	.8503	72.57
.6	.8	2.667	52	.0999	.8506	(52, 29)	.1071	.8506	41.72	(56, 18)	.1147	.8506	39.56	
	.65	.8	2.154	88	.1180	.8518	(88, 43)	.1146	.8502	67.52	(93, 32)	.1178	.8500	65.68
	.85	3.051	49	.1158	.8541	(49, 23)	.1160	.8506	37.60	(52, 15)	.1155	.8519	36.31	

p_y	Single-Stage Design			Minimax Two-Stage Design			Optimal Two-Stage Design						
	p_x	θ	n	α	$1 - \beta$	(n, n_1)	α	$1 - \beta$	EN	(n, n_1)	α	$1 - \beta$	EN
.7	.85	2.429	78	.1141	.8521	(78, 48)	.1190	.8502	64.33	(84, 26)	.1201	.8503	58.49
	.9	3.857	41	.0980	.8515	(41, 21)	.1000	.8500	32.34	(43, 14)	.1050	.8514	30.87
.75	.9	3.000	65	.1082	.8529	(65, 37)	.1102	.8502	52.50	(69, 22)	.1150	.8500	48.76
	.95	6.333	35	.1019	.8581	(34, 11)	.0997	.8506	24.75	(35, 7)	.1016	.8505	24.43
.8	.95	4.750	51	.0983	.8526	(51, 24)	.1021	.8506	39.45	(53, 15)	.1073	.8518	37.47
	.85	.95	3.353	.92	.1065	.8502	.1101	.8500	72.52	(98, 27)	.1140	.8503	67.92

Table 4

Single-stage designs, and minimax and optimal two-stage Fisher designs for $(\alpha^*, 1 - \beta^*) = (.2, .8)$

p_y	p_x	θ	Single-Stage Design			Minimax Two-Stage Design			Optimal Two-Stage Design					
			n	α	$1 - \beta$	(n, n_1)	α	$1 - \beta$	EN	(n, n_1)	α	$1 - \beta$	EN	
.05	.15	3.353	65	.1078	.8031	(65, 36)	.1121	.8005	53.73	(68, 24)	.1152	.8004	52.14	
	.2	4.750	38	.0799	.8040	(38, 15)	.0813	.8007	30.73	(38, 15)	.0813	.8007	30.73	
	.25	6.333	26	.0509	.8050	(25, 10)	.0481	.8005	20.98	(25, 10)	.0481	.8005	20.98	
.1	.25	3.000	48	.1200	.8043	(47, 19)	.1222	.8002	36.08	(47, 19)	.1222	.8002	36.08	
	.3	3.857	30	.0991	.8012	(30, 19)	.1022	.8005	25.71	(31, 13)	.1066	.8024	24.43	
	.15	.3	2.429	54	.1343	.8005	(54, 40)	.1349	.8003	47.88	(60, 18)	.1445	.8015	42.94
.2	.35	3.051	35	.1220	.8057	(34, 15)	.1209	.8022	26.46	(35, 13)	.1221	.8020	26.44	
	.2	.35	2.154	63	.1512	.8003	(62, 29)	.1491	.8011	47.66	(65, 23)	.1507	.8002	47.09
	.4	2.667	39	.1264	.8005	(39, 26)	.1342	.8005	33.40	(40, 14)	.1408	.8001	29.46	
.25	.4	2.000	67	.1424	.8028	(67, 47)	.1467	.8001	54.95	(73, 24)	.1582	.8003	51.75	
	.45	2.455	41	.1295	.8101	(40, 35)	.1292	.8000	37.78	(45, 12)	.1477	.8001	31.59	
	.3	.45	1.909	68	.1518	.8006	(68, 55)	.1517	.8001	62.04	(75, 30)	.1570	.8003	55.02
.35	.5	2.333	41	.1392	.8034	(41, 28)	.1389	.8002	35.25	(45, 16)	.1445	.8003	32.72	
	.5	1.857	69	.1631	.8012	(69, 54)	.1625	.8001	62.10	(76, 32)	.1602	.8010	56.29	
	.55	2.270	42	.1515	.8092	(42, 25)	.1479	.8004	34.50	(46, 16)	.1491	.8003	33.20	
.4	.55	1.833	70	.1713	.8043	(70, 50)	.1692	.8005	60.81	(77, 32)	.1656	.8010	56.78	
	.6	2.250	42	.1581	.8079	(42, 26)	.1545	.8007	34.90	(45, 18)	.1528	.8007	33.32	
	.45	.6	1.833	70	.1751	.8043	(70, 50)	.1727	.8005	60.80	(77, 32)	.1685	.8010	56.75
.5	.65	2.270	42	.1618	.8092	(42, 25)	.1572	.8004	34.46	(46, 16)	.1549	.8003	33.11	
	.65	1.857	69	.1746	.8012	(69, 54)	.1737	.8001	62.07	(76, 32)	.1686	.8010	56.19	
	.7	2.333	41	.1601	.8034	(41, 28)	.1581	.8002	35.19	(45, 16)	.1542	.8003	32.53	
.55	.7	1.909	68	.1716	.8006	(68, 55)	.1711	.8001	62.00	(77, 28)	.1651	.8022	55.11	
	.75	2.455	41	.1589	.8101	(41, 23)	.1559	.8000	37.74	(45, 12)	.1505	.8001	31.17	
	.6	.75	2.000	67	.1660	.8028	(67, 47)	.1638	.8001	57.84	(73, 24)	.1601	.8003	51.37
.65	.8	2.667	39	.1491	.8005	(39, 26)	.1472	.8005	33.23	(40, 14)	.1430	.8001	28.98	
	.8	2.154	63	.1522	.8003	(62, 29)	.1531	.8011	47.31	(65, 23)	.1599	.8002	46.58	
	.85	3.051	35	.1311	.8057	(34, 15)	.1425	.8022	25.94	(36, 11)	.1458	.8003	25.71	

	p_y	p_x	θ	Single-Stage Design			Minimax Two-Stage Design			Optimal Two-Stage Design				
				n	α	$1 - \beta$	(n, n_1)	α	$1 - \beta$	EN	(n, n_1)	α	$1 - \beta$	EN
	.7	.85	2.429	54	.1466	.8005	(54, 40)	.1518	.8003	47.68	(60, 18)	.1590	.8015	42.03
		.9	3.857	30	.1386	.8012	(30, 19)	.1427	.8005	25.27	(33, 7)	.1428	.8008	22.99
	.75	.9	3.000	48	.1446	.8043	(47, 19)	.1427	.8002	35.09	(48, 17)	.1458	.8009	35.09
		.95	6.333	26	.1305	.8050	(25, 10)	.1249	.8005	19.04	(26, 6)	.1267	.8007	18.65
	.8	.95	4.750	38	.1392	.8040	(38, 15)	.1406	.8007	28.60	(40, 9)	.1419	.8017	28.16
		.85	3.353	65	.1411	.8031	(65, 36)	.1410	.8005	52.42	(70, 20)	.1483	.8010	49.45

Table 5

Single-stage designs, and minimax and optimal two-stage Fisher designs for $(\alpha^*, 1 - \beta^*) = (.2, .85)$

p_y	p_x	θ	Single-Stage Design			Minimax Two-Stage Design			Optimal Two-Stage Design		
			n	α	$1 - \beta$	(n, n_1)	α	$1 - \beta$	(n, n_1)	α	$1 - \beta$
.05	.15	3.353	81	.1148	.8535	(78, 37)	.1197	.8503	(81, 29)	.1214	.8501
	.2	4.750	44	.0897	.8459	(44, 19)	.0924	.8504	(44, 19)	.0924	.8504
	.25	6.333	30	.0618	.8542	(30, 11)	.0621	.8519	(30, 11)	.0621	.8519
.1	.25	3.000	56	.1292	.8511	(56, 36)	.1301	.8501	(59, 21)	.1321	.8503
	.3	3.857	35	.1046	.8545	(35, 19)	.1097	.8521	(37, 12)	.1132	.8502
.15	.3	2.429	65	.1411	.8523	(65, 35)	.1411	.8506	(69, 26)	.1453	.8508
	.35	3.051	43	.1294	.8577	(42, 19)	.1344	.8508	(43, 16)	.1345	.8511
.2	.35	2.154	74	.1469	.8528	(74, 51)	.1497	.8502	(80, 30)	.1575	.8507
	.4	2.667	45	.1298	.8559	(45, 27)	.1349	.8501	(48, 18)	.1443	.8505
.25	.4	2.000	80	.1575	.8515	(78, 50)	.1530	.8500	(84, 36)	.1567	.8515
	.45	2.455	46	.1402	.8514	(46, 32)	.1398	.8501	(50, 20)	.1453	.8508
.3	.45	1.909	87	.1531	.8506	(87, 68)	.1561	.8501	(92, 36)	.1670	.8503
	.5	2.333	53	.1468	.8501	(50, 29)	.1568	.8510	(52, 21)	.1547	.8500
.35	.5	1.857	89	.1541	.8533	(89, 63)	.1537	.8502	(93, 40)	.1660	.8501
	.55	2.270	54	.1392	.8523	(53, 27)	.1589	.8503	(55, 23)	.1599	.8506
.4	.55	1.833	89	.1600	.8508	(89, 71)	.1595	.8501	(96, 39)	.1686	.8516
	.6	2.250	54	.1405	.8510	(53, 28)	.1569	.8501	(57, 21)	.1563	.8502
.45	.6	1.833	89	.1637	.8508	(89, 71)	.1632	.8501	(96, 39)	.1713	.8516
	.65	2.270	54	.1437	.8523	(53, 27)	.1564	.8503	(55, 23)	.1559	.8506
.5	.65	1.857	89	.1649	.8533	(89, 63)	.1627	.8502	(92, 50)	.1604	.8502
	.7	2.333	53	.1426	.8501	(50, 29)	.1548	.8510	(52, 21)	.1559	.8500
.55	.7	1.909	87	.1610	.8506	(87, 68)	.1603	.8501	(92, 36)	.1664	.8503
	.75	2.455	46	.1437	.8514	(46, 32)	.1616	.8501	(50, 20)	.1659	.8508
.6	.75	2.000	80	.1485	.8515	(78, 50)	.1621	.8500	(85, 34)	.1692	.8502
	.8	2.667	45	.1664	.8559	(45, 27)	.1615	.8501	(48, 18)	.1576	.8505
.65	.8	2.154	74	.1672	.8528	(74, 51)	.1679	.8502	(82, 27)	.1627	.8501
	.85	3.051	43	.1544	.8577	(42, 19)	.1439	.8508	(44, 14)	.1473	.8505

Single-Stage Design					Minimax Two-Stage Design					Optimal Two-Stage Design				
p_y	p_x	θ	n	α	$1 - \beta$	(n, n_1)	α	$1 - \beta$	EN	(n, n_1)	α	$1 - \beta$	EN	
.7	.85	2.429	65	.1480	.8523	(65, 35)	.1503	.8506	51.56	(69, 26)	.1591	.8508	50.09	
	.9	3.857	35	.1286	.8545	(35, 19)	.1362	.8521	28.13	(37, 12)	.1461	.8502	26.71	
.75	.9	3.000	56	.1536	.8511	(56, 36)	.1545	.8501	47.08	(59, 21)	.1538	.8503	42.70	
	.95	6.333	30	.1408	.8542	(30, 11)	.1373	.8519	22.36	(31, 7)	.1386	.8519	21.94	
.8	.95	4.750	44	.1290	.8549	(44, 19)	.1429	.8504	33.53	(46, 14)	.1451	.8520	33.02	
	.85	3.353	81	.1506	.8535	(78, 37)	.1496	.8503	60.17	(83, 26)	.1516	.8506	58.94	