Simulation Controlled Seamless Phase II/III Clinical Trials

by

Lindsay Berry

Department of Mathematics
University of Texas at Austin

Date: _____

Approved:

_____
Dr. Peter Mueller, Supervisor

_____
Dr. James Scott

_____
Dr. David Rusin

Thesis submitted in partial fulfillment of graduation within the
Dean's Scholars Honors Degree in the Department of Mathematics,
The University of Texas at Austin
2015

Abstract

Clinical trials have become a standard part of modern medicine, creating great interest in the field of adaptive clinical trials. The goal of adaptive trials is greater efficiency in the use of patients and time, higher probability of demonstrating treatment effect, and more informative trials. This paper focuses on seamless phase II/III trials, an adaptive trial consisting of two stages: the first stage compares multiple treatment arms to a control and determines the appropriate arm, and the second stage undertakes a more traditional comparison of this arm to a control arm. The data from both stages is used in analysis of the treatment, with the goal of more power than two separate trials where data is not shared. This paper begins with a discussion of an established method of analyzing seamless trials, referred to as the Posch method. Proofs are presented that conclude that the Posch method has analytical control over type one error. This paper also introduces a new method of analyzing seamless trials called the simulation method. This simulation method relies on simulation of clinical trials under the null hypothesis to find critical values that control type one error. This paper includes a model of the simulation method as a first step in a potential proof that the simulation method has analytical control over type one error. Power comparisons between the Posch method, the simulation method, and two separate trials conclude that seamless trials are more powerful than

separate trials, and the simulation method is slightly more powerful than the Posch method when analyzing seamless trials. Power comparisons of separate trials and the simulation method with a fixed sample size reveal that implementation of the simulation method would result in larger phase II trials than when separate trials are run. This paper serves as an initial step in establishing the simulation method as a method of analyzing seamless phase II/III clinical trials.

I grant the Dean's Scholars Program permission to post a copy of my thesis on the University of Texas Digital Repository. For more information, visit http://repositories.lib.utexas.edu/about.

Simulation Controlled Seamless Phase II/III Clinical Trials

Department of Mathematics

_____

(Your name, printed)

_____ _____

Signature Date

_____

(Supervisor name, printed)

_____ _____

Signature Date

Contents

Acknowledgements

I would like to thank my research advisor, Dr. Peter Mueller, for his valuable suggestions and encouragement throughout this research project. I would also like to thank Dr. James Scott for his support and motivating teaching. I would also like to thank Dr. David Rusin, the Dean's Scholars Honors Program, and the University of Texas for the opportunity to complete this work. Many thanks to John Murray, Oksana Hays, and my fellow cohort of DS students for serving as indispensable distractions from this work. Finally, I would like to thank my parents for their unwavering support and guidance throughout the past twenty-two years.

1

Background

Clinical trials study whether an intervention is effective in humans under certain conditions. Many clinical trials test if a drug or medical device is effective in treating a medical condition. Clinical trials have become a standard part of modern medicine and the definitive tool for evaluating the effect of treatments. The importance of clinical trials and ever increasing ethical and efficiency standards generates increasing interest in the field of *apaptive* design or analyses guided by examination of the accumulated data at an interim point in the trial. These adaptive features may make clinical trials more efficient (fewer patients, shorter duration), more likely to demonstrate the effect of treatments, or more informative (FDA (2010)).

Clinical trials of drugs are traditionally divided into four phases. In many cases, in particular for chemotherapy, phase I trials aim to find the maximum safe dose of a treatment. Phase II trials seek to establish efficacy (effectiveness) while protecting against toxicity and futility (continuing a trial that is unlikely to produce a

significant result). Phase III trials, often called confirmatory trials, are usually large patient trials (300-3,000 patients depending on disease) aimed at being the definitive assessment of the efficacy of a drug. Phase IV studies are post-market studies that continue studying a drug after it is placed on the market. (Berry et al. (2011))

In drug development, phase II trials are typically run with the goal of determining the next step in the development process. The data from dose-finding phase II trials is used to make decisions about whether a phase III trial should be run and which dose(s) should be included. The length of time between phase II and phase III trials can be quite lengthy; filled with regulatory meetings to design a phase III trial. If a phase III trial is conducted, the data collected in phase II is ignored in evaluating the same treatment in phase III. These issues lead to the desire for more adaptive phase II/III trials. Seamless phase II/III trials are composed of two stages: the first stage begins with a dose-finding phase II trial which leads 'seamlessly' to a phase III trial which undertakes a more traditional comparison of a dose with a control arm. The seamless design removes the time between trials, typically leading to a shorter development time for the drug. Seamless phase II/III trials combine the data from stage one and stage two to analyze the trial as a whole. The hope for this seamless design is that combining the data from both trials will result in a more powerful and efficient design than two separate trials. Although power is not the major concern for regulatory agencies, trial sponsors find power very important. A more powerful design lessens the risk of running a clinical trial.

The following notation will be used throughout the paper to refer to the design of seamless phase II/II trials. The trial aims to study the elementary null hypotheses $H_i : \theta_i = \theta_0$ where θ_i is the treatment effect of treatment i. Stage one begins with

k treatment arms, $T_i, i \in \{1, \ldots, k\} = \mathscr{T}_1$, compared to the placebo, T_0. After stage one, one treatment $T_m, m \in \mathscr{T}_1$ is chosen to move on to stage two where it is compared to T_0. Data from stage one and stage two are combined to analyze the elementary null hypothesis H_m. In this paper, our hypothesis tests are one-sided at level 0.025 because in a phase III trial there is scientific evidence that the treatment has an effect greater than or equal to zero.

Seamless phase II/III trials are considered a type of confirmatory trial. Confirmatory trials are overseen by regulatory agencies with predetermined statistical thresholds that the trial design must meet in order to be approved for public use. The statistical hurdle that confirmatory trials must meet is to get a statistically significant result at a specified type-I error level. In the first stage of seamless phase II/III trial, multiple treatment arms are compared to a placebo. These multiple comparisons may cause an inflated type-I error rate.

This paper discusses two methods of analyzing seamless phase II/III clinical trials. Chapter 2 discusses the method of Posch et. al. (2005), an established method which has analytical control over type-I error. Chapter 3 introduces a new method of analyzing seamless phase II/III trials with simulation control of type-I error.

2

Method of Posch, et al

2.1 Introduction

This section discusses the method of Posch et. al. (2005) of analyzing seamless phase II/III clinical trials. For simplicity, we will refer to this method as the "Posch method". The seamless trial design is more complicated than a traditional single phase trial. The Posch method develops procedures to address each of the involved challenges. The three methods that make up the Posch method include the (i) combination test method, (ii) the closed testing principle, and (iii) the Simes test. The combination test is a procedure to combine the data from stage one and stage two in order to assess the trial as a whole. The closed testing principle is a method that allows investigators to control family-wise type-I error rate across a family of hypotheses. The Simes test defines a p-value for the intersection of multiple (ideally) independent hypotheses. That is, the Simes test allows us to control the overall

type-I error rate for multiple independent hypotheses. Each of these methods will be fully explained and partially proven in the following sections. In section 2.6, we will combine all of these steps in an example clinical trial that uses the Posch method. Finally, section 2.7 presents a discussion of the benefits and limitations of the Posch method. The notation described in the Introduction will be used heavily throughout this chapter.

2.2 Hypothesis tests and p-values

A combination test combines the evidence from stage one and two in a way to assess the trial as a whole. Ordinarily, clinical trials are assessed with p-values. Consider a two-stage test of a null hypothesis H. Let p be a p-value for H based only on data from stage one, and q be a p-value for H based only on data from stage two. Section 2.3 defines a combination function, $C(p, q)$, that takes the p-values from each stage and combines them into a p-value to evaluate the entire trial. Therefore, $C(p, q)$ must satisfy the definition of a p-value, so we review p-values in this section.

Assume we are testing a hypothesis, denoted H_0 for the null hypothesis. We begin with our data, $y_1, \ldots, y_n$, which are realizations of the i.i.d. random variables $Y_1, \ldots, Y_n$. These random variables often follow a distribution indexed by a parameter θ. A statistical test is based on a test statistic, $T(y)$, with some known distribution under H_0. Here T is a function of the data $y = (y_1, \ldots, y_n)$. A test is then constructed by defining a rejection region C, such that for $T \in C$ the conclusion is to reject H_0; and for $T \notin C$ we fail to reject H_0. Usually C is defined by a threshold on T.

Let p denote our p-value for this test. One definition of a p-value is

$$p = P(T \geqslant t|H_0) \tag{2.1}$$

Here $P(A|H_0)$ is a slight abuse of notation to indicate a probability under H_0. In words, the p-value is the probability that we have a test statistic equal to or more extreme than the data we observed, assuming the null hypothesis is true. The following Lemma will be useful in explaining an alternate definition of p-values used in this paper.

Lemma 1. *Let $u_1 \sim \mathcal{U}(0,1)$ (u_1 has a standard uniform distribution). Then $1-u_1 \sim \mathcal{U}(0,1)$.*

Proof. By definition of a standard uniform distribution, $P(u_1 \leqslant a) = a$ for $a \in [0,1]$. We can rewrite this as $1 - P(u_1 > a) = a \Rightarrow 1 - P(1-u_1 > 1-a) = a \Rightarrow P(1-u_1 \leqslant 1-a) = 1-a$. Let $b = 1-a$. We know that $P(1-u_1 \leqslant b) = b$ for $b \in [0,1]$. Therefore, $1-u_1 \sim \mathcal{U}(0,1)$. □

Proposition 2. *Assume the distribution function $F(t)$ for the test statistic is continuous. Then the p-value p is a uniform random variable, $p \sim U(0,1)$, under the null hypothesis.*

For the upcoming discussion this property is a convenient alternative definition of a p-value. We use it in place of Equation 2.1 as the defining property. A short 'proof' of this definition follows. We previously defined a p-value as $p = P(T \geqslant t|H_0)$. We can rewrite this as $p = 1 - P(T < t|H_0)$. A cumulative distribution function (cdf) F of a random variable X is a function given by $F_X(x) = P(X \leqslant x)$. Let F_0 be the cdf of T

under H_0. It follows that $p = 1-F_0(t)$. All cdf's are monotonic, increasing, and right-continuous, therefore $p = P(T \geqslant t|H_0) = P(F_0(T) \geqslant F_0(t)) = 1 - P(F_0(T) < F_0(t))$

It follows that $p = 1 - F_0(t) = 1 - P(F_0(T) < F_0(t))$ and we conclude that

$$P(F_0(T) < F_0(t)) = F_0(t)$$

By definition, $F_0(T)$ follows a uniform distribution. Since a cdf results in values from 0 to 1, we known $F_0(T)$ has a standard uniform distribution. By Lemma 1, $1 - F_0(T) \sim \mathcal{U}(0,1)$, and therefore $p \sim \mathcal{U}(0,1)$.

2.3 Combination tests

$C(p,q)$ must satisfy the definition of a p-value and be a function of the data that is $\mathcal{U}(0,1)$ under the null hypothesis.

The result of $C(p,q)$ is used to determine the following decision function:

$$\phi_C(p,q) = \begin{cases} 1 & C(p,q) \leqslant \alpha \\ 0 & \text{otherwise} \end{cases}$$

If $\phi_C = 1$, then we reject the null hypothesis. If $\phi_C = 0$, we fail to reject the null hypothesis. $C(p,q)$ is used to combine the data from both stages, and then then ϕ_C evaluates the entire trial. The combination function is the weighted inverse normal combination function, defined as follows:

$$C(p,q) := 1 - \Phi[v\Phi^{-1}(1-p) + w\Phi^{-1}(1-q)]$$

where v, w denote pre-defined weights such that $v^2 + w^2 = 1$, and Φ is the cumulative distribution function of the standard normal distribution and Φ^{-1} is its quantile. We will use the following lemma to prove that $C(p,q)$ satisfies the definition of a p-value.

Lemma 3. *Let $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ be independent random variables. Then $Y = aX_1 + bX_2 \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$.*

Proof. We will use the moment generating functions (mgf) to find the mean and variance of Y. The moment generating function for a variable with $N(\mu, \sigma^2)$ distribution is $M(t) = \mathrm{e}^{t\mu + \frac{1}{2}\sigma^2 t^2}$. Therefore, $M_{X_1}(t) = \mathrm{e}^{t\mu_1 + \frac{1}{2}\sigma_1^2 t^2}$ and $M_{X_2}(t) = \mathrm{e}^{t\mu_2 + \frac{1}{2}\sigma_2^2 t^2}$. One important quality of mgf's is that for independent random variables X and Y with mgf M_x and M_y, respectively,

$$M_{aX+bY}(t) = M_X(at) \times M_Y(bt)$$

We can use this fact to find the moment generating function of Y,

$$M_Y(t) = M_{aX_1+bX_2}(t) = M_{X_1}(at)M_{X_2}(bt) =$$
$$= \mathrm{e}^{(at)\mu_1 + \frac{1}{2}\sigma_1^2(at)^2}\mathrm{e}^{(bt)\mu_2 + \frac{1}{2}\sigma_2^2(bt)^2} = \mathrm{e}^{t(a\mu_1+b\mu_2) + \frac{1}{2}t^2(a^2\sigma_1^2+b^2\sigma_2^2)}$$

This result is recognizable as the moment generating function for a normally distributed variable with mean $a\mu_1 + b\mu_2$ and variance $a^2\sigma_1^2 + b^2\sigma_2^2$. Since distributions have unique moment generating functions, we conclude that $Y \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$. □

Theorem 4. *$C(p, q)$ is a p-value.*

Proof. We must show that $C(p, q) \sim U(0, 1)$ when the null hypothesis is true. Assume that the null hypothesis is true, and definition 2 tells us that that p-values p and q have standard uniform distributions. Therefore, by Lemma 1, $1 - p$ and $1 - q$ also have standard uniform distributions. We know that $\Phi^{-1} : \mathcal{U}(0, 1) \to N(0, 1)$, so $\Phi^{-1}(1-p)$ takes the standard uniform $1-p$ and results in a $N(0, 1)$ random variable,

as does $\Phi^{-1}(1-q)$. Therefore, $v\Phi^{-1}(1-p)+w\Phi^{-1}(1-q)$ is the linear combination of two $N(0,1)$ random variables.

By Lemma 3, $v\Phi^{-1}(1-p)+w\Phi^{-1}(1-q) \sim N(0, v^2+w^2) = N(0,1)$. (Recall that we defined v, w so that $v^2+w^2=1$). Let $W = v\Phi^{-1}(1-p)+v\Phi^{-1}(1-q)$ (so $W \sim N(0,1)$). We know that $\Phi : N(0,1) \to \mathcal{U}(0,1)$, so Φ sends a $N(0,1)$ variable to $\mathcal{U}(0,1)$. Therefore $\Phi(W) \sim \mathcal{U}(0,1)$ and $1-\Phi(W) \sim \mathcal{U}(0,1)$ by Lemma 1. Therefore, $C(p,q) = 1-\Phi(W) \sim \mathcal{U}(0,1)$. □

2.4 Closed testing principle

The closed testing principle is a general method to control family wise type-I error rate in the strong sense. The goal of this seamless trial is to test a set of elementary null hypothesis denoted by $H_i : \theta_i = 0, i \in \mathscr{T}_1 = \{1, \ldots, k\}$. In this paper, our tests of hypotheses will be a one-sided test with $\alpha = 0.025$. If we test each H_j at significance level .025, we risk inflating the type-I error rate above our pre-specified 0.025 level. This inflated type-I error rate creates the need for closed testing procedures which strongly control the familywise error rate (FWER). Strong FWER is defined as the maximum probability of rejecting at least one true null hypothesis regardless of the configuration of true or false hypotheses (Posch et al. (2005)).

Definition 1. *An intersection hypothesis is defined as* $H_{\mathscr{S}} = \bigcap_{i \in \mathscr{S}} H_i, \mathscr{S} \subset \mathscr{T}_1$.

Definition 2. *Consider a family* HH *of hypotheses that are closed under intersection. That is, for any two* H_i, $H_j \in HH$ *also the intersection* $H_i \cap H_j \in HH$. *Let* $\phi_i(X) \in \{0,1\}$ *denote a nominal level alpha test for each hypothesis. That is,* $P(\phi_i(X) = 1 | H_i) = \alpha$. *Instead of using* ϕ_i *to decide rejection, the closed testing*

principle specifies the following rule:

Reject H_i if and only if $\phi_j(X) = 1$ for all $H_j \subset H_i$.

Note that ϕ_i controls type-I error rate for individual tests, but not family-wise error rate. We do not use ϕ_i itself to define the desired test. Instead, the closed testing principle allows us to use these individual ϕ_i to construct a test that controls the overall family-wise error rate. For example, consider stage one has 3 treatment arms, and T_1 moves on to stage two while the other arms are dropped. After stage two, we want to test $H_1 : \theta_1 = 0$. The intersection hypotheses are all hypotheses that also include treatment 1 such as $H_1, H_{\{1,2\}} : \{\theta_1 = \theta_2 = 0\}, H_{\{1,3\}} : \{\theta_1 = \theta_3 = 0\}$, and $H_{\{1,2,3\}} : \{\theta_1 = \theta_2 = \theta_3 = 0\}$. In order to reject the elementary null hypothesis, H_1, with strong control at level 0.025, each of the intersection hypotheses has to be rejected at 0.025.

Theorem 5. *The closed testing principle controls the probability of committing any type-1 error at α.*

Proof. Let A denote the event that any true H_i is rejected under the closed testing method. Next, let H_r denote the intersection of all true H_i. By assumption H_r is true, since it is the intersection of all true H_i. Let $B = \{\phi_r = 1\}$. By assumption, $P(B) = \alpha$, since H_r is true. Also $A \subset B$ since the closed testing principle only rejects H_i if all $H_j \subset H_i$ are rejected, and by construction $H_r \subset H_i$ for all true H_i. Thus $A \cap B = A$ and we have

$$P(A) = P(A \cap B) \leqslant P(B)p(A|B) \leqslant \alpha$$

□

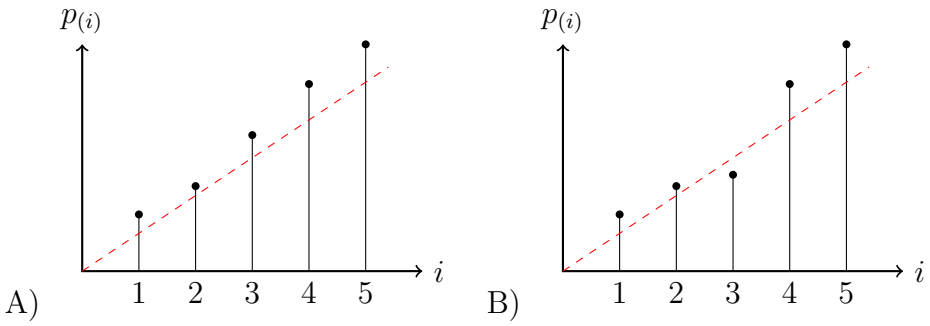


FIGURE 2.1: Illustration of Simes test

To understand the closed testing principle, consider the following example. In stage one, we test 3 treatments and decide to move on to stage two with T_1. We want to test $H_1 : \theta_1 = 0$. In order to reject H_1, we have to first look at the intersection hypotheses. The intersection hypotheses of H_1 are $\{H_1, H_{(1,2)}, H_{(1,3)}, H_{(1,2,3)}\}$ because each of these hypotheses also tests whether $\theta_1 = 0$. We test $\phi_{(1,2,3)}(X)$, $\phi_{(1,2)}(X)$, and $\phi_{(1,3)}(X)$. If any of these functions return a 0, we fail to reject H_1, and we did not commit a type-I error. If all of these functions return a 1, then we can finally test $\phi_1(X)$. If $\phi_1(X) = 1$, we reject H_1. If $\phi_1(X) = 0$, we fail to reject H_1. A type-I error can only be committed if and only if H_1 is true and $\phi_1(X) = 1$.

When we apply the closed testing principle, we define combination tests for all intersection hypotheses. Recall that in the second stage, no data is available for dropped treatment arms. When we are testing an intersection hypothesis $H_{\mathscr{S}}$, some of the treatments in $\mathscr{S}$ may have no data from $\mathscr{T}_2$. Let $q_{\mathscr{S}}$ denote the second stage p-value for $H_{\mathscr{S}}$. We define $q_{\mathscr{S}}$ as follows:

$$q_{\mathscr{S}} = q_{\mathscr{S} \bigcap \mathscr{T}_2}$$

where $q_{\varnothing} = 1$. These are conservative p-values for $H_{\mathscr{S}}$

2.5 p-values for intersection hypothesis

As seen in section 2.4, the closed testing principle uses α-level tests ϕ_β to control the FWER in testing a family of hypotheses. The scheme required level-alpha tests for all intersections H_i. There are many possible ways to define such level-alpha tests. This paper uses the Simes test (Simes (1986)).

During a statistical test on a hypothesis, H, a rejection region, R, is defined. If our data lies in R, then we reject H, and if our data is outside of R, then we fail to reject H. The Simes test defines a rejection region such that the probability the data lies in R under the null hypothesis is less than or equal to α.

According to the the closed testing principle, in order to reject an elementary hypothesis, $H_j, j \in \mathscr{T}_1$ at level α, for all subsets $\mathscr{S} \subset \mathscr{T}_1$ that contain j, the intersection hypothesis $H_\mathscr{S}$ must be rejected at level α. The Simes test evaluates an intersection hypothesis $H_\mathscr{S}$. Denote the number of treatments in $\mathscr{S}$ by s.

The Simes test defines R as follows:

$$R := \{\text{any } p_{(i)} \leqslant \frac{i}{s}\alpha | i \in \mathscr{S}\} = \{\text{any } \frac{s}{i}p_{(i)} \leqslant \alpha | i \in \mathscr{S}\} = \{\min_{i \in \mathscr{S}} \frac{s}{i}p_{(i)} \leqslant \alpha\}$$

Two illustrations of the Simes test is shown in Figure 2.1. Each figure shows five ordered p-values. The red dotted line is the Simes cutoff which is shown as a line with slope $\alpha/s = \alpha/5$. In Figure 2.1A, every p-value is above the line. Therefore each $p_{(i)} > \alpha i/5$, so our data does not lie in the rejection region and we fail to reject our null hypothesis. An alternate situation is depicted in Figure 2.1B, where $p_{(3)} < 3/5\alpha$. Therefore, the data does lie in the rejection region, and we reject our null hypothesis.

Let $P_{(1)}, \ldots, P_{(n)}$ be the ordered p-values for testing hypotheses $H_0 = \{H_{(1)}, \ldots, H_{(n)}\}$.

The Simes test says that H_0 is rejected if $p_{(j)} \leqslant j\alpha/n$ for any $j = 1, \ldots, n$.

Theorem 6. *Let $P_{(1)}, \ldots, P_{(n)}$ be order statistics of n Uniform$(0,1)$ random variables, and let $A_n(\alpha) = P\{P_{(j)} > j\alpha/n | j = 1, \ldots, n\}$. Then $A_n(\alpha) = 1 - \alpha$.*

Proof. We will complete this proof by induction. We begin with the case that $n = 1$. Thus, $A_1(\alpha) = P(P_{(1)} > \alpha) = 1 - \alpha$ because $P_{(1)}$ is a standard uniform random variable.

Assume the theorem is true for $n-1$, and we will show that the theorem follows for n. Our assumption is that $A_{n-1}(\alpha) = 1-\alpha$, and we must show that $A_n(\alpha) = 1-\alpha$. First, note the joint pdf of the $P_{(i)}$ is $f_{P_{(1)},\ldots,P_{(n)}}(p_{(1)}, \ldots, p_{(n)}) = 1_{(p_{(1)} \leqslant \ldots \leqslant p_{(n)})}$. We define random variables $W_1, \ldots, W_n$ so that $W_i = P_{(i)}/P_{(n)}, i \in (1, n-1)$ and $W_n = P_{(n)}$. *Claim 1*: $W_1, \ldots, W_{n-1}$ are $n-1$ independent and identically distributed $\mathcal{U}(0,1)$ random variables that are independent from W_n. To show this, we use the following property of functions of random variables found in Ross (2010). Let $Y_1, \ldots, Y_k$ be functions of the random variables $X_1, \ldots, X_k$ which have joint pdf $f_{X_1,\ldots,X_k}$. Say $Y_1 = g_1(X_1, \ldots, X_k), \ldots, Y_k = g_k(X_1, \ldots, X_k)$, and assume $g_1, \ldots, g_k$ are invertible functions. Then, the joint pdf of $Y_1, \ldots, Y_k$ is $f_{Y_1,\ldots,Y_k}(y_1, \ldots, y_k) = f_{X_1,\ldots,X_k}(x_1, \ldots, x_k)|J(x_1, \ldots, x_k)|^{-1}$ where J is the Jacobian and $x_i = g_i^{-1}(y_1, \ldots, y_k)$ (Ross (2010)). We use this fact to find the joint pdf of the W_i's which is

$$f_{W_1,\ldots,W_n}(w_1, \ldots, w_n) = f_{P_{(1)},\ldots,P_{(n)}}(p_{(1)}, \ldots, p_{(n)})|J(p_{(1)}, \ldots, p_{(n)}|^{-1}$$

where $p_{(i)} = T^{-1}(w_i)$.

The Jacobian is equal to:

$$J(p_{(1)},\ldots,p_{(n)}) = det\begin{pmatrix} 1/p_{(n)} & 0 & \cdots & 0 \\ 0 & 1/p_{(n)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = \left(\tfrac{1}{p_{(n)}}\right)^{n-1}$$

Therefore, $|J(p_{(1)},\ldots,p_{(n)}|^{-1} = p_{(n)}^{n-1}$, and

$$f_{W_1,\ldots,W_n}(w_1,\ldots,w_n) = 1_{(w_1 w_n \leqslant \ldots \leqslant w_{n-1} w_n \leqslant w_n)} w_n^{n-1} = 1_{(w_1 \leqslant \ldots \leqslant w_{n-1})} w_n^{n-1}$$

Because we can factor the joint pdf, this proves that W_n is independent from $W_1,\ldots,W_{n-1}$. The joint pdf for $W_1,\ldots,W_{n-1}$ is $1_{(w_1 \leqslant \ldots \leqslant w_{n-1})}$, which proves that these are $n-1$ iid $\mathcal{U}(0,1)$ random variables. So, we have shown Claim 1 is true. We know by definition that,

$$A_n(\alpha) = Pr(P_{(j)} > j\alpha/n, j = 1,\ldots,n) \tag{2.2}$$

$$= Pr(P_{(1)} > \frac{\alpha}{n},\ldots,P_{(n-1)} > \frac{(n-1)\alpha}{n}, P_{(n)} > \alpha) \tag{2.3}$$

$$= Pr\Big(\frac{P_{(1)}}{P_{(n)}} > \frac{\alpha}{nP_{(n)}},\ldots,\frac{P_{(n-1)}}{P_{(n)}} > \frac{(n-1)\alpha}{nP_{(n)}}, 1 > \frac{\alpha}{P_{(n)}}\Big) \tag{2.4}$$

$$= Pr(W_1 > \frac{\alpha}{nP_{(n)}},\ldots,W_{n-1} > \frac{(n-1)\alpha}{nP_{(n)}}, W_n > \alpha) \tag{2.5}$$

$$= Pr\Big(W_1 > \frac{\alpha}{nP_{(n)}},\ldots,W_{n-1} > \frac{(n-1)\alpha}{nP_{(n)}}\Big) \times Pr(W_n > \alpha) \tag{2.6}$$

$$= A_{n-1}\Big(\frac{(n-1)\alpha}{nP_{(n)}}\Big) \times Pr(W_n > \alpha) \tag{2.7}$$

Step 6 follows because W_n is independent of $W_1,\ldots,W_{n-1}$. According to Ross (2010), the density function of a uniform order statistic $X_{(j)}$ is

$$f_{X_{(j)}}(x) = \frac{n!}{(n-1)!(n-j)!}x^{j-1}(1-x)^{n-j}$$

Using this equation, $f_{W_{(n)}}(x) = nx^{n-1}$. Therefore, we know that

$$A_n(\alpha) = \int_\alpha^1 A_{n-1}\Big(\frac{\alpha(n-1)}{pn}\Big) np^{n-1} dp$$

We assumed, by induction that $A_{n-1}(\alpha) = 1 - \alpha$, so we can simplify this to

$$\begin{aligned} A_n(\alpha) = \int_\alpha^1 \Big(1 - \frac{\alpha(n-1)}{pn}\Big) np^{n-1} dp &= \int_\alpha^1 np^{n-1} - \alpha(n-1)p^{n-2} dp \\ &= \Big[p^n - \alpha p^{n-1}\Big]_\alpha^1 \\ &= (1-\alpha) - (\alpha^n - \alpha\alpha^{n-1}) = (1-\alpha) \end{aligned}$$

Therefore, $A_n(\alpha) = 1 - \alpha$ for all n. □

2.6 Example Study Using the Posch Method

This section puts all of the steps together to illustrate use of the Posch method in a seamless phase II/III clinical study. This example is loosely based on the ADVENT trial of the drug Crofelemer designed by Cytel (Chaturvedi and Mehta (2014)). In this example, we will test our hypotheses with a one-sided test at $\alpha = .025$ level. Stage one of the study compares 3 dosages of a drug, $\{T_1, T_2, T_3\}$, to the placebo, T_0. After stage one, we compute a p-value for each treatment, and the p-value for treatment T_i is denoted as p_i. The stage one p-values are $p_1 = .0019$, $p_2 = .0563$, and $p_3 = .0024$. Because p_1 is the smallest p-value, we will continue to stage two with T_1 and a placebo and the rest of the arms will be dropped. In stage two, we randomize patients to T_1 and T_0, and again calculate a p-value for T_1 based on the stage two data. This study results in a second stage p-value for T_1 of $q_1 = .1690$.

Now we have all of our information for the study, and we will begin using the Posch method to analyze the data. Since T_1 moved to stage 2, this analysis is testing the hypothesis $H_1 : \theta_1 > \theta_0$ at the overall .025 level. Recall that the closed testing principle says that we must reject the intersection hypotheses at $\alpha = 0.025$ before we can reject H_1.

To test every intersection hypothesis, we will calculate first and second stage p-values for each intersection hypothesis, then use the combination function to get the overall p-value for that intersection hypothesis. The intersection hypotheses we must test are $\{H_{\{1,2\}}, H_{\{1,3\}}, H_{\{1,2,3\}}\}$.

We will reject $H_{\{1,2\}}$ only if $C(p^{(1,2)}, q^{(1,2)}) < .025$. Similarly we will only reject $H_{\{1,3\}}$ only if $C(p^{(1,3)}, q^{(1,3)}) < .025$, and we will reject $H_{\{1,2,3\}}$ only if $C(p^{(1,2,3)}, q^{(1,2,3)}) < .025$. We will use the Simes test to compute $p^{(1,2)}, p^{(1,2)}, p^{(1,2,3)}$, the adjusted stage one p-values.

$$p^{(1,2)} = \min\{2\min(p_1, p_2), \max(p_1, p_2)\} = 0.0038$$

$$p^{(1,3)} = \min\{2\min(p_1, p_3), \max(p_1, p_3)\} = 0.0024$$

$$p^{(1,2,3)} = \min\{3\min(p_1, p_2, p_3), 2\text{med}(p_1, p_2, p_3), \max(p_1, p_2, p_3)\} = 0.0036$$

As shown in Section 2.4, our adjusted stage two p-values are $q_1 = q^{(1,2)} = q^{(1,3)} = q^{(1,2,3)} = 0.1690$.

Now, we use the combination function to find the p-value for both stages for each intersection hypothesis.

$$C(p^{(1,2)}, q^{(1,2)}) = 1 - \Phi[\sqrt{1/2}\Phi^{-1}(1-.0038) + \sqrt{1/2}\Phi^{-1}(1-.1690)] = 0.00514 < 0.025$$

$$C(p^{(1,3)}, q^{(1,3)}) = 1 - \Phi[\sqrt{1/2}\Phi^{-1}(1-.0024) + \sqrt{1/2}\Phi^{-1}(1-.1690)] = 0.00382 < 0.025$$

$$C(p^{(1,2,3)}, q^{(1,2,3)}) = 1-\Phi[\sqrt{1/2}\Phi^{-1}(1-.0036)+\sqrt{1/2}\Phi^{-1}(1-.1690)] = 0.00503 < 0.025$$

Since all of these p-values are less than 0.025, we can reject $H_{(1,2)}, H_{(1,3)}, H_{(1,2,3)}$. Because we have rejected all of the intersection hypotheses, we can now test H_1.

We test hypothesis H_1 at the local 0.025 level using the combination function with p_1 and q_1.

$$C(.0019, .1690) = 1 - \Phi[\sqrt{1/2}\Phi^{-1}(1 - .0019) + \sqrt{1/2}\Phi^{-1}(1 - .1690)] = 0.0032$$

Since $.0032 < .025$, we can reject H_1 at the overall 0.025 level, based on the closed testing principle.

2.7 Discussion

In this section, we have proven that the Posch method has analytical control over type-I error. This type-I control is achieved through the use of the conservative tests (Simes, Bonferroni test). These conservative methods may sacrifice power in order to guarantee type-I error is controlled. There is criticism of the Posch method for violating the Likelihood Principle. The **Likelihood Principle** states that in an inference about θ, after x is observed, all relevant experimental information should be contained in the likelihood function for x (Berry et al. (2011)). Furthermore, two likelihood functions contain the same information if they are proportion to one another. The Posch method violates the Likelihood Principle because the same data may result in different conclusions. For example, consider scenario 1 where patient A is in a stage 1 of size n_1 and patient B is in a stage 2 of size n_2. Then consider scenario 2 where patient A and patient B switch stages. Our inferences from the data would be different, even though our data is the same but in a different order.

This difference results form the weighting of each stage in the combination function $C(p, q)$. If stage 1 has many more patients than stage 2, the weighting process will give more importance to a patient from stage 1. In the following section, we introduce an alternate method for analyzing seamless trials.

3

Simulation Controlled Adaptive Trials

3.1 Introduction

In this section, I propose a new method of analyzing seamless phase II/III clinical trials. This method, referred to as the simulation method, simulates clinical trials under the null hypothesis in order to find a value that controls type-I error for these simulated trials. As discussed earlier, regulatory agencies expect confirmatory trials to control type-I error at a specified level. In section 2, we proved that the Posch method has analytical control over type-I error. Currently, we cannot prove that the simulation method has analytical control over type-I error. Up to this point, simulated control of type-I error has not been as trusted in the industry as analytical control. Instead of using simulation methods to analyze seamless trials, analytical methods, or, more likely, separate trials, are relied upon. If a general formula to describe the result of this simulation method were found, it may lead to greater

acceptance of simulation methods in analyzing clinical trials. In this chapter, I introduce the simulation method in section 3.2. Then, section 3.3 describes the general results of the simulation method. Next, section 3.3.1 attempts to find a model of the simulation method. Section 3.3.2 includes power comparisons between the simulation method, the Posch method, and two separate trials. Finally, section 3.3.3 discusses implications of usage of the simulation method in the drug development process.

3.2 Method

The null hypothesis tested during a seamless phase II/III clinical trial is $H_m : \theta_m = 0$ with alternative hypothesis $H_a : \theta_m > 0$. We are performing a one-sided test with $\alpha = 0.025$. Our data comes from stage one with k arms with n_1 patients per arm, and stage two with T_m and T_0 with n_2 patients each. T_m is the arm chosen to move on from stage one to stage two. From this data, we compute a test statistic with the formula

$$Z_m = \frac{Y_{comb} - Y_0}{\frac{1}{n_1+n_2} + \frac{1}{n_1+n_2}}$$

where Y_{comb} and Y_0 are summaries of the data. We will define details later. We use the simulation method to find the critical value to compare Z_m to in order to control the type-I error rate at α. We denote this critical Z value obtained through simulation as Z_{sim}. If $Z_m > Z_{sim}$, we reject H_m. If $Z_m < Z_{sim}$, then we fail to reject H_m. In this section, we will explain the method used to find Z_{sim}. The goal of this simulation method is to find a critical value Z_{sim} such that

$$P(Z_m > Z_{sim}|H_m) = \alpha$$

Z_{sim} is found by simulating clinical trials under H_m and finding the test statistic at the $(1-\alpha)$th percentile. Let $\hat{\theta}_i(data)$ denote an estimate of θ_i.

The simulation method is as follows:

1. Simulate stage one values of $\hat{\theta}_i(data)$ for $i \in 0, 1, \ldots, k$ under the null hypothesis. Each $\theta_i \sim N(0, 1/n_1)$. Let $n_1 = 1$ for ease of computation. Let $Y_1[j]$ denote the stage one value of θ_j.

2. Identify $T_m = \max_{i \in \mathscr{T}_1}(Y_1)$. This is the best performing treatment arm from stage one, and the dose which will move to stage two.

3. Simulate stage two values of $\hat{\theta}_m$ and $\hat{\theta}_0$ under the null hypothesis. Each $\theta_i \sim N(0, 1/n_2)$ for $i = 0, \ldots, m$. Let $Y_2[j]$ denote the stage two values of θ_j.

4. Combine the data for T_m from stage one and two by finding a weighted average of $Y_1[m]$ and $Y_2[m]$.

$$Y_{comb} = \frac{n_1 Y_1[m] + n_2 Y_2[m]}{n_1 + n_2}$$

   Combine the data for the control, T_0, from stage one and two by finding the weighted average of $Y_1[0]$ and $Y_2[0]$.

$$Y_0 = \frac{n_1 Y_1[0] + n_2 Y_2[0]}{n_1 + n_2}$$

5. Calculate the Z statistic for T_m:

$$Z_m = \frac{Y_{comb} - Y_0}{\sqrt{\frac{1}{n_1+n_2} + \frac{1}{n_1+n_2}}}$$

6. Repeat Step (1)-(5) N times, and let M_z denote the N resulting Z_m values. Order M_z from smallest to largest. Define Z_{sim} as the $(1-\alpha)N$-th percentile value of Z_m in M_z.

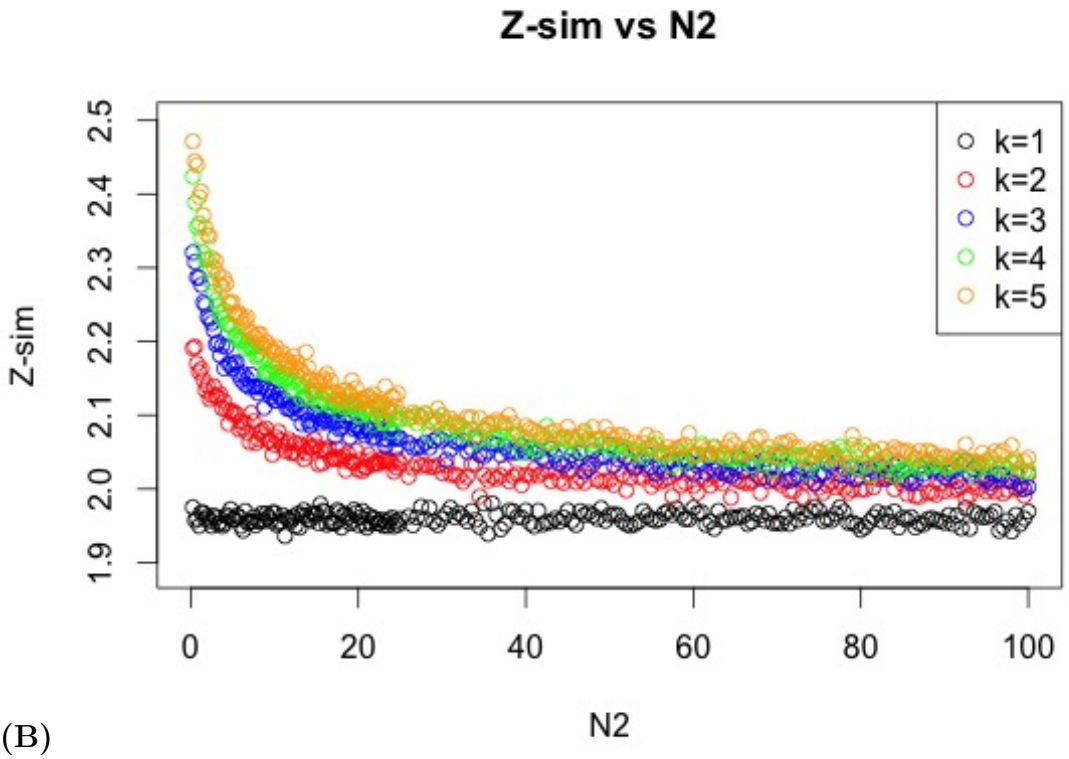
The R code for this simulation can be found in Appendix A. For all simulations in this paper, I set $\alpha = 0.025$. This method returns a value of Z_{sim} such that $P(Z_m > Z_{sim}|H_m) \leqslant 0.025$.

3.3 Results

This section discusses the results of the simulation method. First, we start with a simple example to show how to interpret Z_{sim}. In this example, let $k = 4$ and $n_2 = 1$. Because $n_1 = 1$ in our simulation method, n_2 can be thought of as a multiple of n_1. In this example we are testing the situation where arms in stage one and stage two have the same number of patients because $n_2 = n_1$. When I run the simulation 5,000 times with $k = 4$ and $n_2 = 1$, the value of $Z_{sim} = 2.3301$. A success in the seamless trial would occur if $Z_m > 2.3301$. If the null hypothesis is true, the probability of a type-I error when using 2.3301 as the critical Z value should be less than 0.025.

Figure 3.1A shows the results of the simulation method for on $n_2 = 0, \ldots, 100$ for different values of k. Figure 3.1B shows the results of the simulation method for on $n_2 = 0, \ldots, 25$ for different values of k. Every simulation of Z_{sim} was done with 10,000 simulations runs. If $n_2 = .5$, that means that arms in stage two have half as many patients as treatment arms in stage one. Similarly, $n_2 = 10$ means that stage two arms have 10 times as many patients as stage one treatment arms. In practice, values of n_2 will usually be between 0 and 25, but we include higher values to completely illustrate the relationship between Z_{sim} and n_2. In these figures, the

(A)



(B)

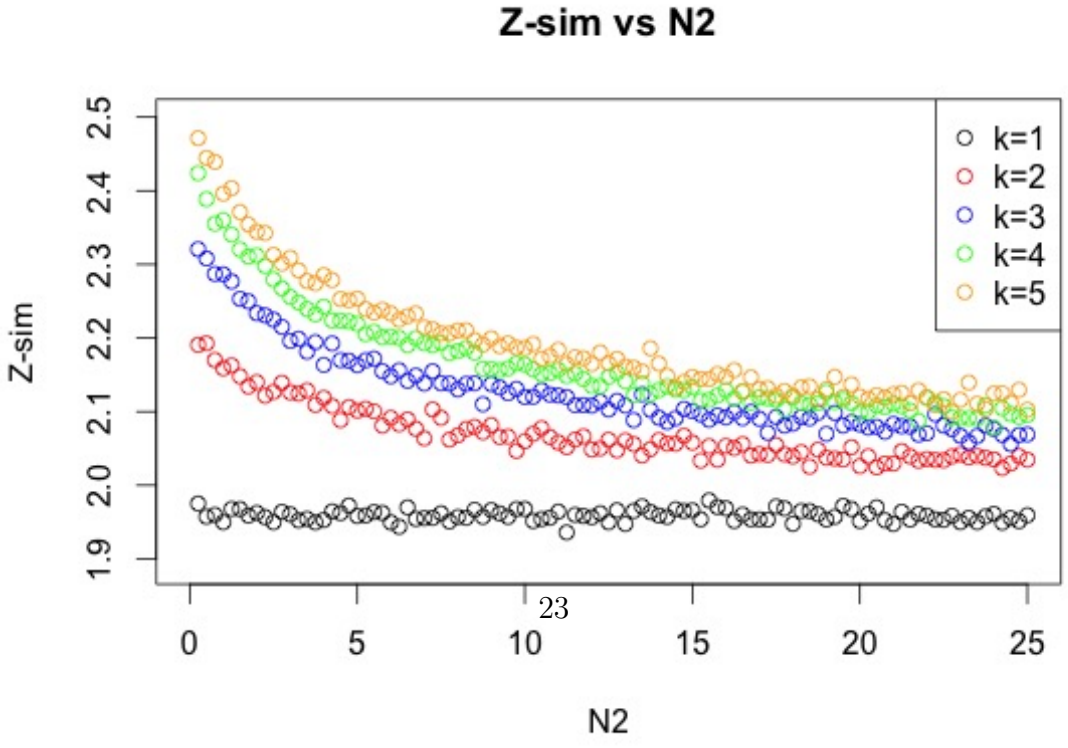


FIGURE 3.1: Results of the simulation method. (A) $n_2 = [0, 100]$ (B) $n_2 = [0, 25]$

different colors of dots correspond to different values of k. As the graph shows, when $k = 1$, no multiplicity correction is needed, so we can use the cutoff $Z^* = 1.96$ for $\alpha = .025$. As k increases, the number of choices in stage one increases, so the critical value needed to control type-I error, Z_{sim}, increases.

3.3.1 Modeling Z_{sim}

In this section, we attempt to model Z_{sim} based on n_2 and k. A model of Z_{sim} could serve several purposes. First of all, a model is much simpler than a large simulation. Rather than running a lengthy simulation, we can plug n_2 into a model and find a value of Z_{sim} immediately. Second, fitting a curve to Z_{sim} formalizes borrowing strength across n_2 and k. We end up pooling information from simulations for similar values of n_2 and k. Furthermore, a model of the critical Z^* needed to control type-I error based on n_2 may give more credibility to the simulation method possibly leading to greater acceptance of the simulation method. A model may also motivate a mathematical proof that the simulation method analytically controls type-I error.

Modeling Z_{sim} with $\log(n_2)$

The results of our simulation, shown in Figure 3.3, appear to show an approximately logarithmic relationship between Z_{sim} and n_2. For each $k = 2, 3, 4, 5$, we create a linear model of Z_{sim} based on $\log(n_2)$. The results of each linear model is shown in Table 3.1 and Figure 3.2. In Figure 3.2, the points are the actual values of Z_{sim} from our simulation method, and the lines show the linear model we fit to these results. The model fits reasonably well, and the values of R^2 are high, but the model underestimates values of Z_{sim} when $\log(n_2)$ is close to zero and large. In actual

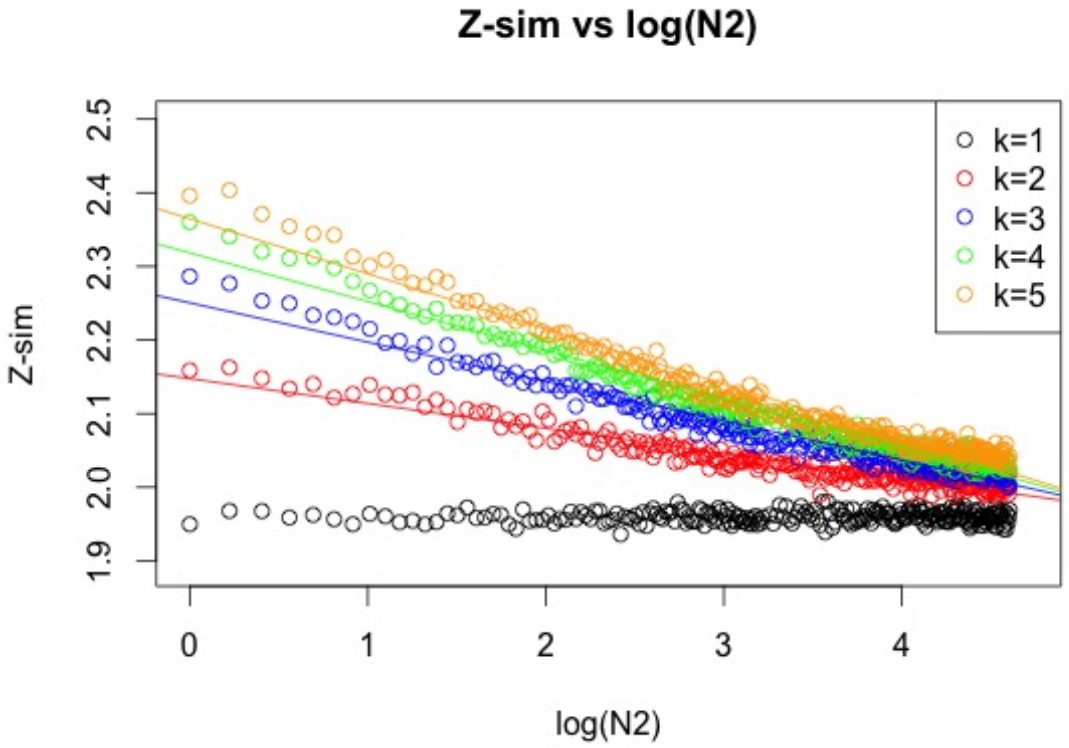


FIGURE 3.2: Relationship between Z_{sim} and $\log(n_2)$.

clinical trials, our value of n_2 will usually be between zero and 5, so it is important that our model fits at small values of n_2. Underestimating these values of Z_{sim} at low n_2 would lead to a loss of control over type-I error. In the next section, we attempt to find a model with a better fit of Z_{sim} at low values of n_2.

Modeling Z_{sim} as a weighted average of 1.96 and Bonferroni critical Z-value

For this model, I try to provide a better fit for the low values of n_2. The simulation method produces values of Z_{sim} that are bounded between 1.96 and the Bonferroni

Model	β_0	β_1	R^2
$k = 2$	2.1475	-0.0338	0.9233
$k = 3$	2.2505	-0.0535	0.9607
$k = 4$	2.3183	-0.0662	0.97
$k = 5$	2.3646	-0.0749	0.9653

Table 3.1: Summary of models $Z_{sim} = \beta_0 + \beta_1 log(n_2)$ for $k = 2, 3, 4, 5$

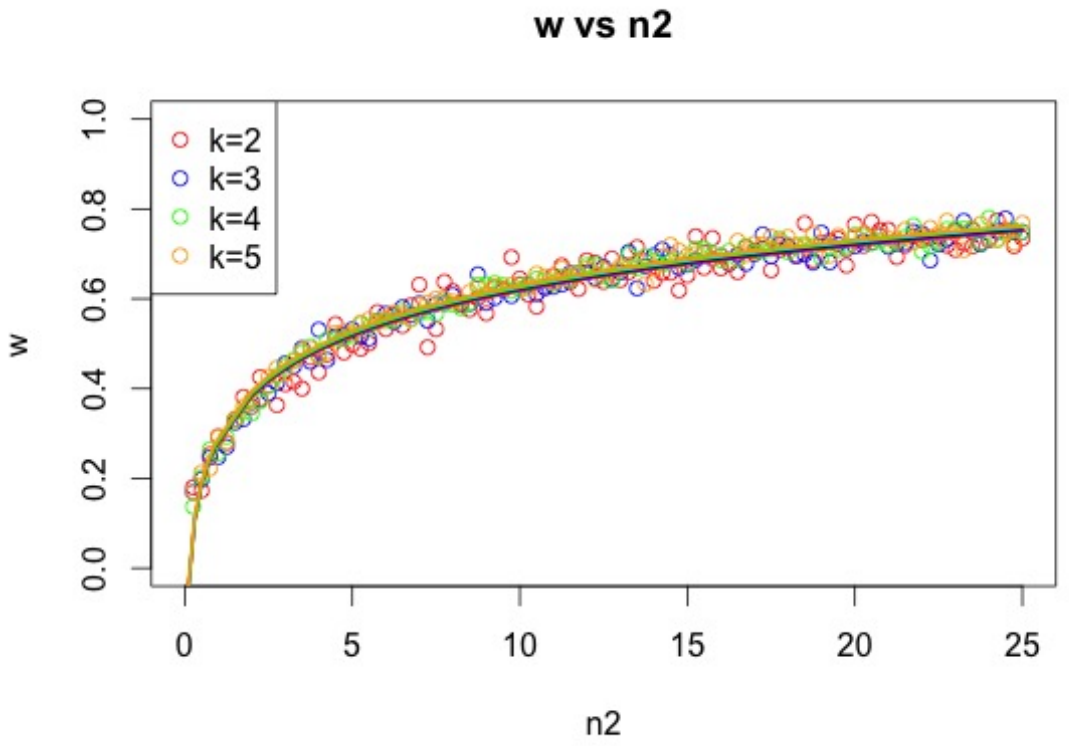


FIGURE 3.3: Relationship between w and n_2.

Model	β_0	β_1	R^2
$k = 2$	0.2824699	0.1455681	0.9445
$k = 3$	0.2861188	0.1455137	0.9744
$k = 4$	0.2874319	0.1466722	0.9858
$k = 5$	0.2945057	0.1456999	0.9826

Table 3.2: Summary of models $w = \beta_0 + \beta_1 \log(n_2)$ for $k = 2, 3, 4, 5$

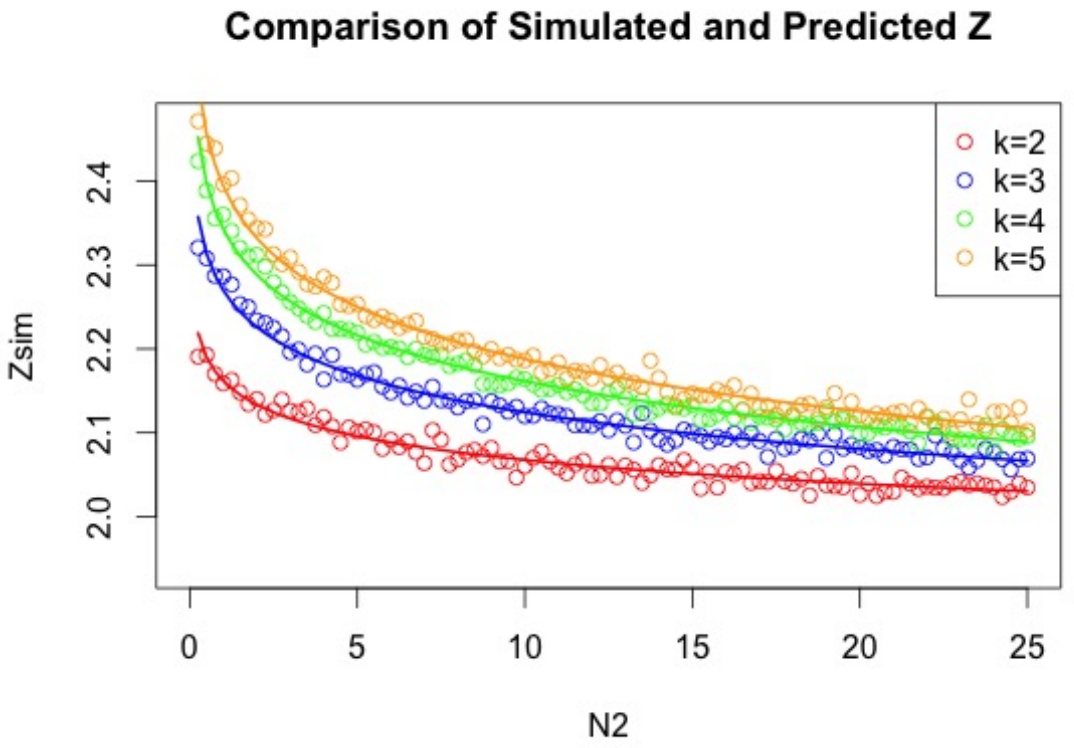


FIGURE 3.4: Comparison of simulated Z_{sim} (points) and predicted Z_{sim} (lines)

critical Z-value for k arms. As n_2 increases, Z_{sim} gets closer to 1.96. As n_2 gets closer to zero, the value of Z_{sim} approaches the Bonferroni corrected Z value.

Recall that the Bonferroni correction with k arms in stage one would compare each p-value to α/k rather than α. The critical Bonferroni Z value would then be $\Phi^{-1}(1-.025/k)$. The Bonferroni correction is more conservative than the simulation method, so the Bonferroni Z value will always be larger than Z_{sim} for the same values of k and n_2. For example, for 3 arms, the Bonferroni Z value is $\Phi^{-1}(1-.025/3) = 2.39398$. After 100,000 simulations with 3 arms, we find that $Z_{sim} = 2.340666$ when n_2 is very close to zero. When $n_2 = 10000000$, then $Z_{sim} = 1.971483$. This relationship holds for all values of k.

Since Z_{sim} is bounded by these two numbers, I can model Z_{sim} as a weighted average of 1.96 and Z_{bonf} (the Bonferroni corrected Z value). The weight given to 1.96 would be a function of n_2. I will use the following function to estimate Z_{sim}.

$$Z_{sim} = w1.96 + (1-w)Z_{bonf} \text{ for } w \in [0,1] \tag{3.1}$$

Here w is the weight placed on 1.96, and $1-w$ is the weight placed on Z_{bonf}. After running the simulation method, and plugging in values of Z_{sim} into this formula, we can solve for an estimate of w from the data. This estimate of $\hat{w}$ is

$$\hat{w} = \frac{Z_{sim} - Z_{bonf}}{1.96 - Z_{bonf}}$$

Figure 3.3 shows a plot of $\hat{w}$ against n_2. The color of each point corresponds to the number of arms in the trial, but as the plot shows, the relationship between $\hat{w}$ and n_2 seems to be the same regardless of the number of arms.

Our goal is to find a model for Z_{sim} based on n_2. We can model w with n_2, and

then plug this value of w into Equation 3.1 to model Z_{sim}. First, I created a linear model of w based on $\log(n_2)$ for $k = 2, 3, 4, 5$. The result of each of these models is approximately

$$w = .29 + .146 \log(n_2)$$

The exact values for $k = 2, 3, 4, 5$ are shown in Table 3.2. The fitted lines can be seen in Figure 3.3. The model of Z_{sim} becomes

$$Z_{sim} = 1.96(.29 + .146 \log(n_2)) + Z_{bonf}(1 - (.29 + .146 \log(n_2)))$$

This model can be used to predict Z_{sim} based on k and n_2. The predicted Z_{sim} are shown with simulated Z_{sim} in Figure 3.4. The fit is very good for all values of k. In future work, this model could be the basis of a proof that the simulation method controls type-I error rate at α.

3.3.2 Power Comparisons

In this section, I compare the power of a seamless trial using the Posch method versus a seamless trial using the simulation method, and separate phase II and III trials. As discussed earlier, the motivation for seamless phase II/III trials is to gain more power than separate trials.

The power of each method is calculated by simulation in R. The code for each method can be found in Appendix A. Every power calculation is based on a scenario that represents the 'truth'. Before calculating the power, we must choose the true treatment effect size of each k arms that we would like to test. In reality, we do not know the true effect size of each arm, so we test many different scenarios to get an idea of the power of each method under different circumstances. The power is

defined as the probability of rejecting the null hypothesis under an assumed effect (in H_A). We estimate the power using Monte Carlo simulations as the proportion of times that the method correctly rejects the null hypothesis given a true scenario S.

I now consider several different scenarios, and show the resulting power of each method in the given scenario. Every computation that follows is based on 5,000 simulations. Notation for scenarios will be $S = (\theta_1, \theta_2, \ldots, \theta_k)$. Scenario S denotes a trial with k arms where θ_i is the treatment effect of T_i. For example, $S_1 = (0, 0, 1)$ represents a trial with 3 treatment arms. The treatment effect of two of the arms is zero, and the treatment effect of the third arm is 1. Note that the control arm is not included in S, and the placebo always has a treatment effect of 0.

The figures in Table 3.3.1 show scenarios where only one arm is non-zero. Each figure shows power comparisons for a different number of arms. In each graph, the simulation method is slightly better than the other two methods. Between the other two methods, the Posch method has higher power for $k = 2, 3$, both methods perform about the same for $k = 4$, and then the separate trials overtakes Posch for $k = 5, 6$. Although there are slight differences between each method, no method performs drastically better than another.

The figures in Table 3.3.1 show scenarios where each arm has linearly increasing treatment effect. This would be a common scenario in a clinical trial where the different arms indicate increasing dosages of a drug. Each figure shows a different number of arms, beginning with 2 arms and increasing to 6 arms. For each figure, the Posch method and simulation method have higher power than separate trials. The simulation method performs very slightly better than the Posch method, but not significantly better in any scenario.

Detailed tables of exact power for the above scenarios can be found at the end of the chapter. For each scenario, these tables show values of $n_2 = .5, 1, 2, 5$, which are values of n_2 that would be commonly seen in a clinical trial. These tables also include the probability that each arm is selected as the 'best' treatment in stage one for each arm. Also included for each scenario is the probability that the trial results in a success for each arm. A success means that the arm was chosen as the best in stage one, and then null hypothesis was rejected after stage two. Based on these tables, it seems like the probability that each method selects an arm is the same for each method. For the Posch method and the seamless method, the probability of success is around the same for each arm besides the arm with the maximum treatment effect. The simulation method's slightly superior power seems to come from higher power in rejecting the arm with the maximum treatment effect.

3.3.3 Implications for Drug Development

In this section, I consider the implications of using seamless trials with the simulation method rather than separate trials. I will set a fixed sample size for the entire seamless trial and both separate trials. The sample size of patients in stage one is $(k+1)n_1$ because each arm (including control) has n_1 patients. The sample size of patients in stage two is $2n_2$ because both arms T_m and T_0 have n_2 patients on them. Therefore, the entire sample size for both trials is $(k+1)n_1 + 2n_2$. For a fixed sample size, I set this equal to a fixed value, in this case I have used 10.

$$(k+1)n_1 + 2n_2 = 10$$

Note that different values of total sample size will affect the meaning of the effect sizes in power comparisons. The power charts in this section cannot be compared to

power charts in section 3.3.2.

For this fixed sample size, the goal is to find the most powerful division of patients into stage one and stage two. I simulate power for different values of W_1 such that

$$W_1[(k+1)n_1] + (1 - W_1)[2n_2] = 10$$

$$W_1\Big[\frac{(k+1)n_2}{10}\Big] + (1 - W_1)\Big[\frac{2n_2}{10}\Big] = 1$$

Now, W_1 can be interpreted as the percentage of the fixed samples size in stage one (or phase II for separate trials).

The goal of the following simulations is to find the optimal weight to give to stage one to produce the maximum power for both methods. The results of these simulations are shown in Table 3.5. The figures in Table 3.3.3 show scenarios where only one treatment arm is effective, and the different numbers of k. The figures in Table 3.3.3 show power results for scenarios with linearly increasing treatment effects, and different numbers of k. For each scenario, the power of a seamless trial with simulation method is higher than the power of separate trials regardless of W_1. When W_1 is close to zero, the power is about the same for both methods. The power of separate trials drops off significantly as W_1 increases, approaching zero as W_1 approaches 1. The power of the simulation method is much more stable with a slight drop in power as W_1 approaches 1. The power of seamless trials drops slightly more as k increases because the number of choices in stage one increases. For all k, the value of W_1 that results in the highest power for seamless trials is slightly larger than the value of W_1 that results in a maximum power for separate trials. If clinical trials begin applying the seamless trial with simulation method, then these clinical trials would benefit from slightly larger stage one trials than are traditionally used.

One reason for this result is that the simulation method does a better job of choosing a correct arm in stage one than the separate trials.

3.4 Conclusions

In Chapter 2, I showed that the method of Posch has analytical control over type-1 error. Currently, this is an advantage of the Posch method over the simulation method. The simulation method has not been proven to have analytical control over type-1 error. The models included in this paper may serve as an initial step in future work of proving the simulation method has analytical control. Simulated control of type-1 error is not strong enough control because we cannot simulate every possible 'true' scenario. If the truth violated our assumptions made in the simulation method, then the type-1 error rate may be inflated. The results of the power analyses showed that the simulation method and Posch method have very similar results when arms have linearly increasing treatment effects. Both methods perform better than two separate trials. When only one arm has a non-zero treatment effect, this increase in power is diminished and each method performs about the same in terms of power. When considering the maximum power of simulation method and separate trials with a fixed sample size, I conclude that the simulation method would result in designs with a slightly larger phase II (stage one) trial. This paper serves as the first stepping stone for the simulation method for seamless trials in the drug development world.

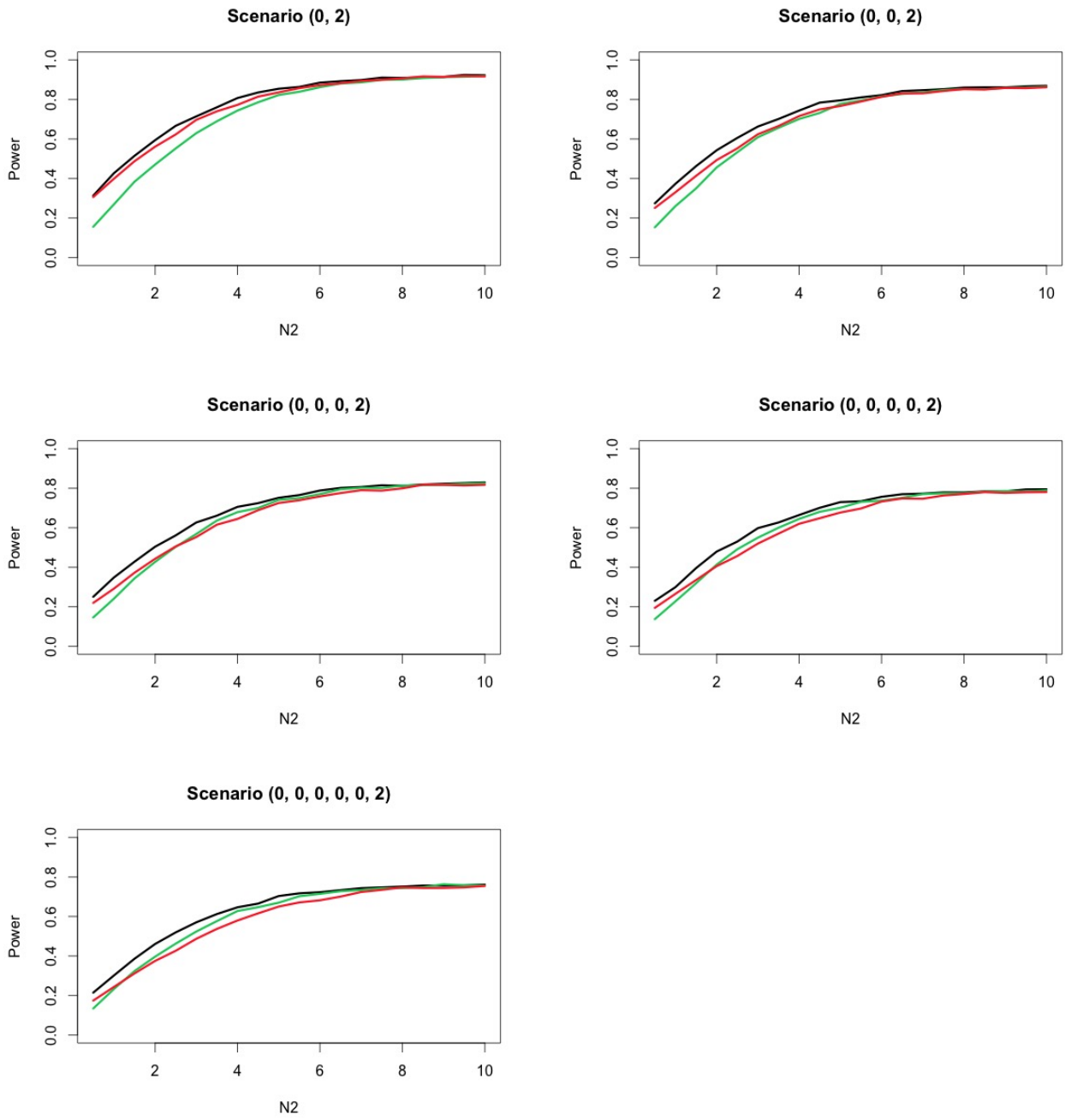


Table 3.3: Power comparisons with one effective treatment arm of simulation method (black), Posch method (red), separate trials (green)

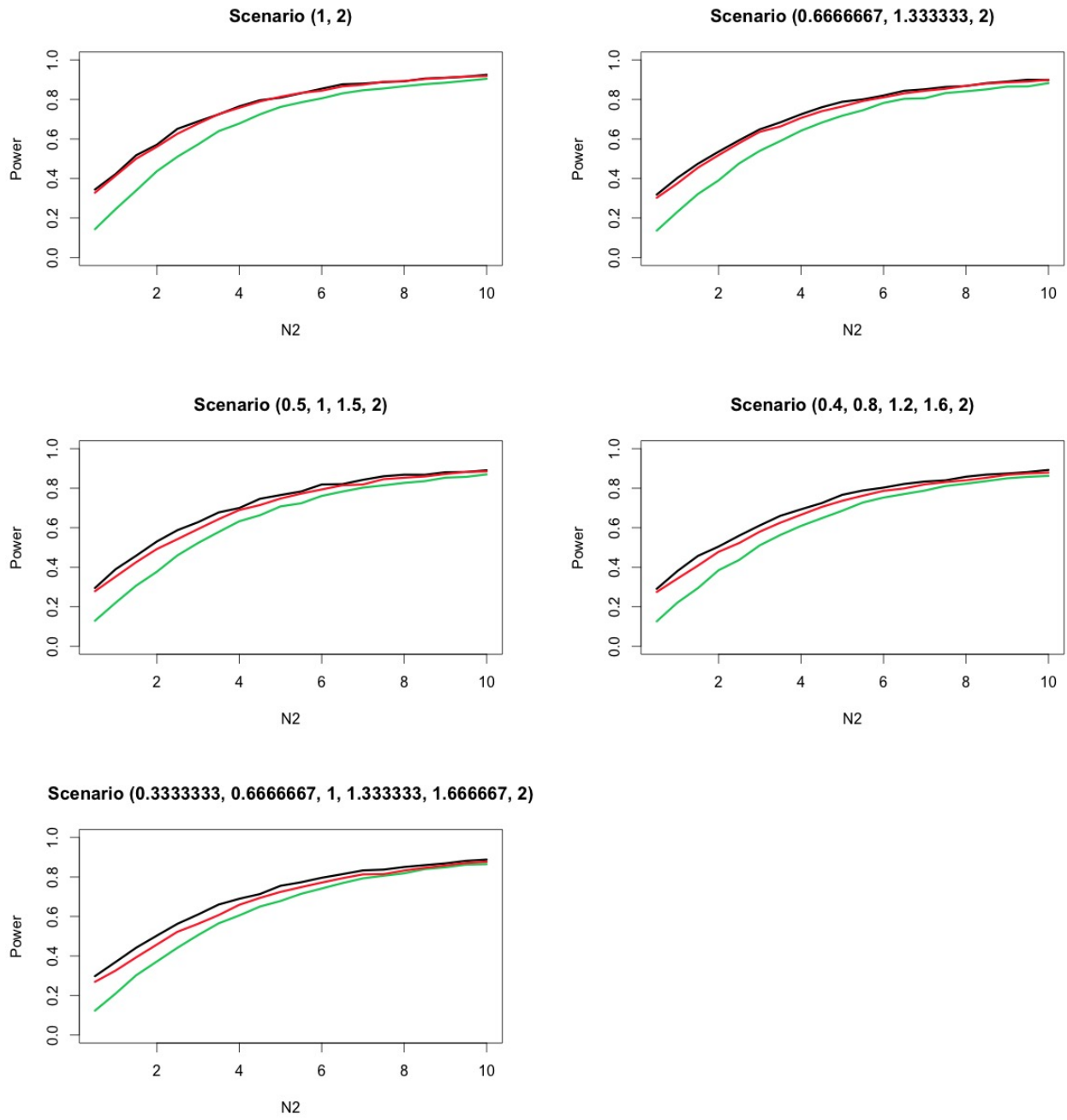


Table 3.4: Power comparisons with linearly increasing treatment effects of simulation method (black), Posch method (red), separate trials (green)

Scenario	Max Power Seamless	Seamless W_1	Max Power Separate	Separate W_1
(0, 2)	0.7738	.3	0.7176	.18
(0, 0, 2)	0.6724	.42	0.6232	.24
(0, 0, 0, 2)	0.5872	.40	0.5426	.35
(0, 0, 0, 0, 2)	0.5086	.42	0.4958	.36
(0, 0, 0, 0, 0, 2)	0.4610	.42	0.4428	.35
(1, 2)	0.7392	.375	0.6806	.12
(.67, 1.33 , 2)	0.6584	.34	0.6120	.14
(.5, 1, 1.5, 2)	0.6062	.15	0.5872	.125
(.4, .8, 1.2, 1.6, 2)	0.5814	.21	0.5436	.06
(.33, .67, 1, 1.33, 1.67, 2)	0.5574	.175	0.5292	.105

Table 3.5: The maximum power for each scenario for the seamless trial and separate trials and their respective weight W_1 which produces the maximum power.

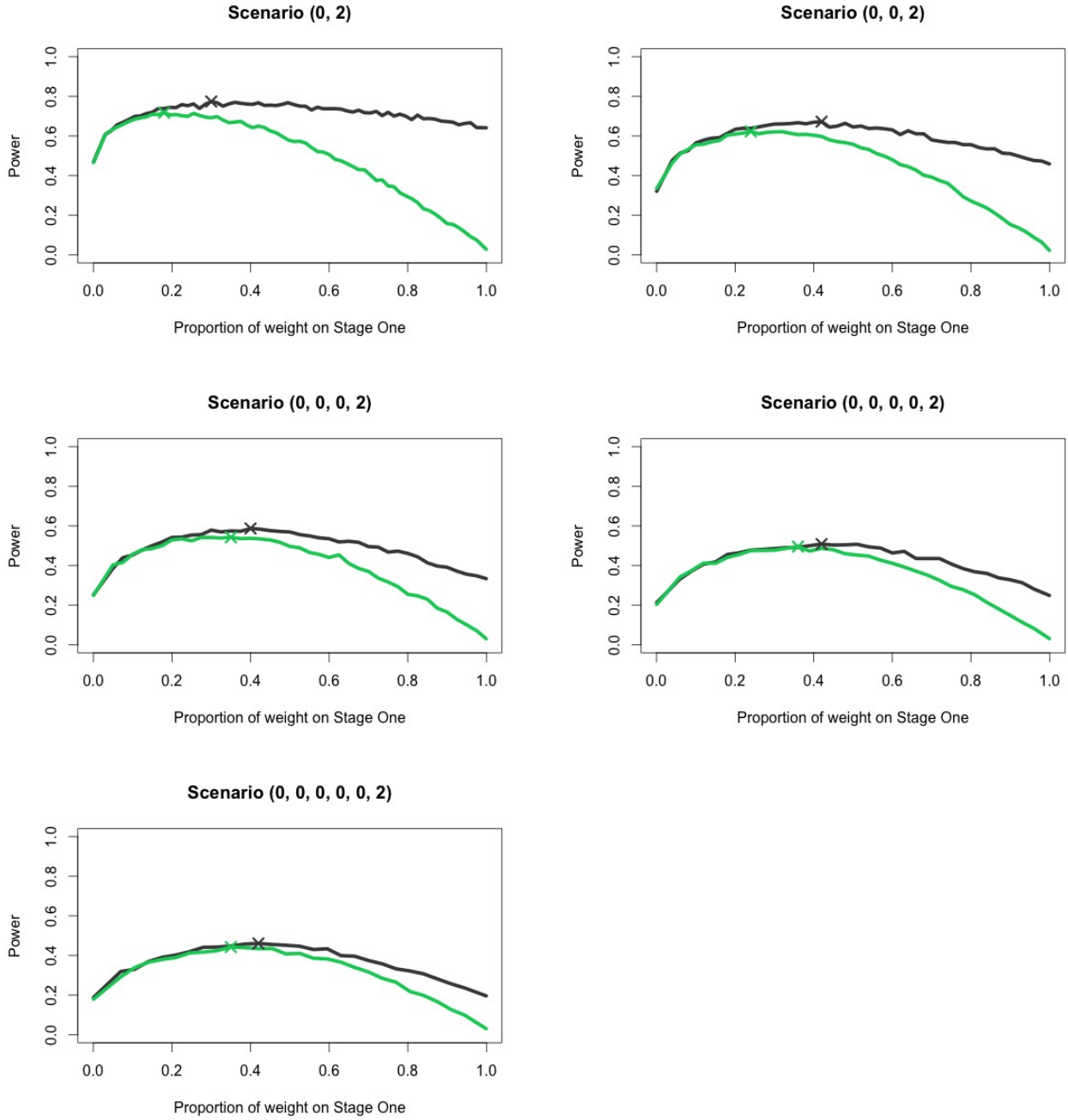


Table 3.6: Power comparisons with one effective treatment arm of the simulation method (gray) and separate trials (green) based on the proportion of weight on stage one. The 'X' indicates maximum power of each method.

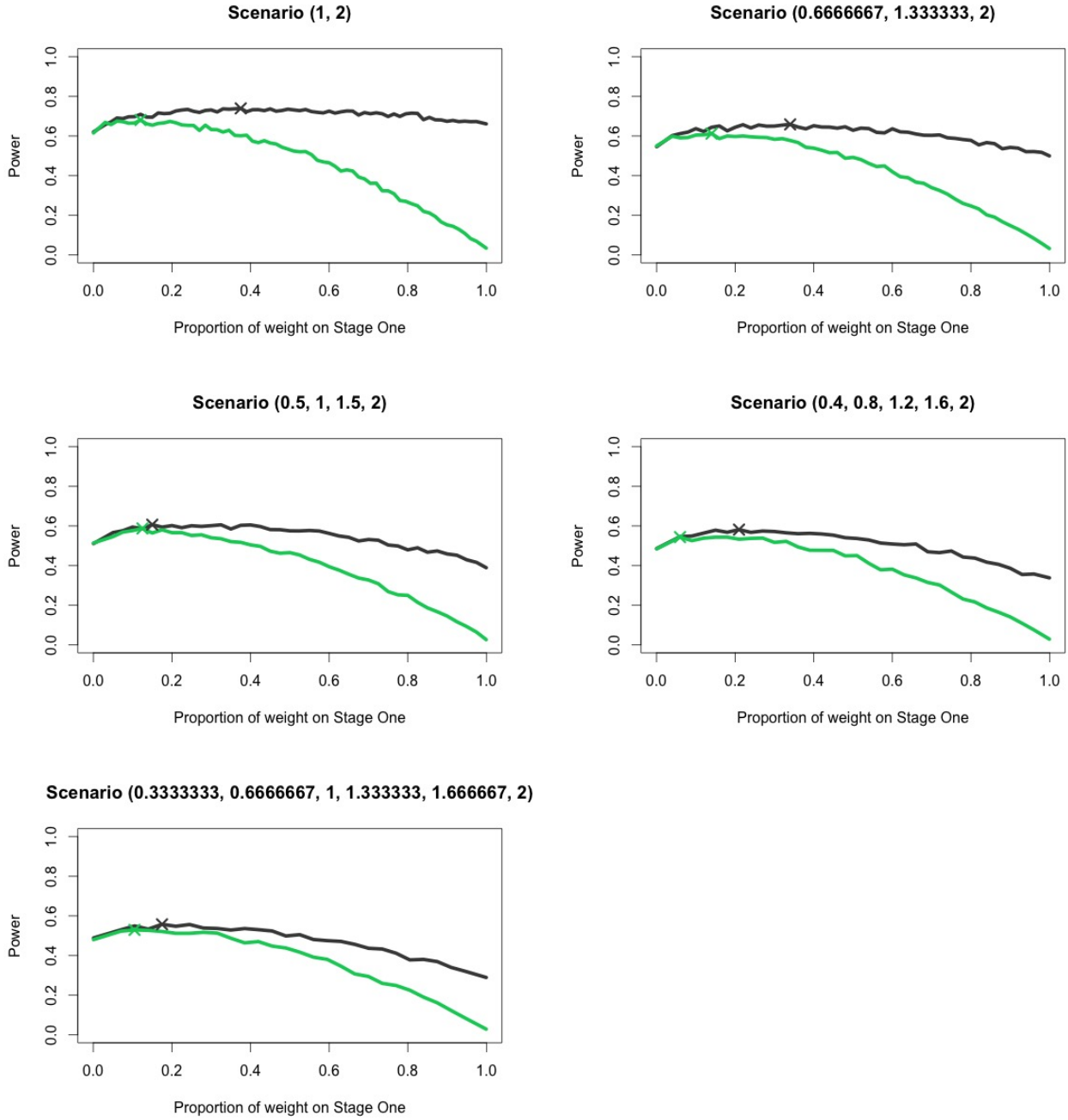


Table 3.7: Power comparisons with linearly increasing treatment effects of simulation method (gray) and separate trials (green) based on the proportion of weight on stage one. The 'X' indicates maximum power of each method.

Appendix A

Detailed Power Tables

Table A.1: Power Per Arm for Scenario (0, 2)

Method	Power	Select 1	Select 2	Success 1	Success 2
$n_2 = .5$					
Posch	0.3046	0.0834	0.9166	0.0065	0.2981
Simulation	0.3219	0.0797	0.9203	0.0036	0.3183
Separate	0.155	0.0732	0.9268	0.0022	0.1528
$n_2 = 1$					
Posch	0.3984	0.0792	0.9208	0.0067	0.3917
Simulation	0.4266	0.0871	0.9129	0.0057	0.4209
Separate	0.2725	0.0819	0.9181	0.0021	0.2704
$n_2 = 2$					
Posch	0.5731	0.0790	0.9210	0.0046	0.5685
Simulation	0.5908	0.0776	0.9224	0.0038	0.5870
Separate	0.476	0.0790	0.9210	0.0024	0.4746
$n_2 = 5$					
Posch	0.8388	0.0751	0.9249	0.0029	0.8359
Simulation	0.8509	0.0802	0.9198	0.0027	0.8475
Separate	0.8118	0.0816	0.9184	0.0017	0.8101

Table A.2: Power Per Arm for Scenario (0, 0, 2)

Method	Power	Select 1	Select 2	Select 3	Success 1	Success 2	Success 3
$n_2 = .5$							
Posch	0.2499	0.0663	0.0704	0.8633	0.0024	0.0040	0.2435
Simulation	0.2706	0.0700	0.0688	0.8612	0.0034	0.0023	0.2649
Separate	0.1499	0.0668	0.0675	0.8657	0.0021	0.0012	0.1466
$n_2 = 1$							
Posch	0.3299	0.0647	0.0658	0.8695	0.0031	0.0024	0.3244
Simulation	0.361	0.0658	0.0672	0.8670	0.0037	0.0029	0.3544
Separate	0.2561	0.0699	0.0648	0.8653	0.0015	0.0018	0.2528
$n_2 = 2$							
Posch	0.4974	0.0631	0.0696	0.8673	0.0031	0.0023	0.4920
Simulation	0.5386	0.0681	0.0677	0.8642	0.0020	0.0028	0.5338
Separate	0.4487	0.0665	0.0643	0.8692	0.0019	0.0016	0.4452
$n_2 = 5$							
Posch	0.7639	0.0678	0.0701	0.8621	0.0021	0.0027	0.7591
Simulation	0.7965	0.0697	0.0695	0.8608	0.0018	0.0032	0.7915
Separate	0.7735	0.0703	0.0611	0.8686	0.0016	0.0018	0.7701

Table A.3: Power Per Arm for Scenario (0, 0, 0, 2)

Method	Power	Select 1	Select 2	Select 3	Select 4	Success 1	Success 2	Success 3	Success 4
$n_2 = .5$									
Posch	0.2201	0.0569	0.0559	0.0626	0.8246	0.0026	0.0028	0.0034	0.2113
Simulation	0.2412	0.0594	0.0595	0.0623	0.8188	0.0028	0.0024	0.0029	0.2331
Separate	0.1417	0.0615	0.0546	0.0566	0.8273	0.0017	0.0017	0.0016	0.1373
$n_2 = 1$									
Posch	0.2965	0.061	0.0568	0.0579	0.8243	0.0032	0.0026	0.0022	0.2885
Simulation	0.3364	0.0592	0.0635	0.0559	0.8214	0.0023	0.0031	0.0024	0.3286
Separate	0.2413	0.064	0.0618	0.0589	0.8153	0.0016	0.0016	0.0005	0.2376
$n_2 = 2$									
Posch	0.4403	0.055	0.0601	0.0622	0.8227	0.0021	0.0021	0.0025	0.4336
Simulation	0.5049	0.0579	0.0595	0.058	0.8246	0.0022	0.0022	0.0022	0.4983
Separate	0.4265	0.0595	0.0559	0.0554	0.8292	0.0022	0.0009	0.0017	0.4217
$n_2 = 5$									
Posch	0.7258	0.0527	0.0617	0.0575	0.8281	0.0019	0.0023	0.0025	0.7191
Simulation	0.7501	0.0586	0.0584	0.0637	0.8193	0.0023	0.0014	0.0029	0.7435
Separate	0.738	0.0563	0.0587	0.06	0.825	0.0021	0.001	0.0024	0.7325

Table A.4: Power Per Arm for Scenario (0, 0, 0, 0, 2)

Method	Power	Select 1	Select 2	Select 3	Select 4	Select 5	Success 1	Success 2	Success 3	Success 4	Success 5
$n_2 = .5$											
Posch	0.2008	0.0513	0.0493	0.0531	0.0535	0.7928	0.0025	0.0014	0.0019	0.0017	0.1933
Simulation	0.2316	0.051	0.0536	0.0518	0.0496	0.794	0.0027	0.0035	0.003	0.002	0.2204
Separate	0.1436	0.055	0.0495	0.0555	0.0502	0.7898	0.0014	0.0011	0.0015	0.0015	0.1381
$n_2 = 1$											
Posch	0.2666	0.0567	0.0547	0.0543	0.0501	0.7842	0.002	0.0021	0.0022	0.0023	0.258
Simulation	0.3203	0.0524	0.0503	0.0534	0.0535	0.7904	0.0026	0.0021	0.0029	0.0023	0.3104
Separate	0.241	0.048	0.0532	0.0545	0.0566	0.7877	0.0012	0.0014	0.0013	0.0016	0.2355
$n_2 = 2$											
Posch	0.403	0.0517	0.0546	0.0534	0.0533	0.787	0.0022	0.0018	0.0016	0.0019	0.3955
Simulation	0.4622	0.0536	0.0531	0.0525	0.0544	0.7864	0.002	0.0019	0.002	0.0014	0.4549
Separate	0.4092	0.054	0.0559	0.0556	0.0542	0.7803	0.0015	0.0017	0.0012	0.0008	0.404
$n_2 = 5$											
Posch	0.6786	0.0518	0.0553	0.0559	0.0497	0.7873	0.0011	0.0019	0.0013	0.0013	0.673
Simulation	0.7262	0.0506	0.0533	0.0519	0.0513	0.7929	0.0017	0.0016	0.0019	0.0013	0.7197
Separate	0.7036	0.053	0.0522	0.0513	0.0529	0.7906	0.0009	0.0014	0.0017	0.001	0.6986

Table A.5: Power Per Arm for Scenario (0, 0, 0, 0, 0, 2)

Method	Power	Select 1	Select 2	Select 3	Select 4	Select 5	Select 6	Success 1	Success 2	Success 3	Success 4	Success 5	Success 6
$n_2 = .5$													
Posch	0.1783	0.0454	0.0506	0.0469	0.0483	0.0475	0.7613	0.0013	0.0015	0.0022	0.001	0.0017	0.1706
Simulation	0.2213	0.0496	0.0444	0.0483	0.0485	0.0475	0.7617	0.002	0.0023	0.0026	0.002	0.0022	0.2102
Separate	0.1327	0.0462	0.0526	0.0508	0.0461	0.0462	0.7581	0.0014	0.0013	0.0016	0.0007	0.0012	0.1265
$n_2 = 1$													
Posch	0.2448	0.0499	0.049	0.0474	0.0457	0.0487	0.7593	0.0015	0.0013	0.0019	0.0015	0.0018	0.2368
Simulation	0.3	0.0515	0.0444	0.0507	0.0492	0.0483	0.7559	0.0015	0.0014	0.0015	0.0014	0.0027	0.2915
Separate	0.2352	0.0501	0.0501	0.0508	0.0483	0.0461	0.7546	0.0009	0.0015	0.001	0.0012	0.0007	0.2299
$n_2 = 2$													
Posch	0.3748	0.0493	0.047	0.0491	0.0452	0.0472	0.7622	0.0013	0.0014	0.0019	0.0012	0.0016	0.3674
Sim	0.4477	0.0467	0.046	0.0472	0.0556	0.0483	0.7562	0.0016	0.0019	0.0014	0.0021	0.0027	0.438
Separate	0.4026	0.0442	0.0488	0.0438	0.0498	0.0534	0.76	0.001	0.0022	0.0011	0.0012	0.0015	0.3956
$n_2 = 5$													
Posch	0.6381	0.0493	0.0509	0.0451	0.05	0.0492	0.7555	0.0009	0.0016	0.0021	0.0016	0.0012	0.6307
Sim	0.6917	0.048	0.0487	0.0483	0.0481	0.0549	0.752	0.0014	0.0011	0.0017	0.0024	0.0017	0.6834
Separate	0.6812	0.0456	0.0481	0.0486	0.0512	0.0476	0.7589	0.0009	0.0009	0.0013	0.0008	0.0008	0.6765

Table A.6: Power Per Arm for Scenario (1, 2)

Method	Power	Select 1	Select 2	Success 1	Success 2
$n_2 = .5$					
Posch	0.3303	0.2378	0.7622	0.0447	0.2856
Sim	0.3408	0.2328	0.7672	0.0448	0.296
Separate	0.1388	0.2389	0.7611	0.0171	0.1217
$n_2 = 1$					
Posch	0.4141	0.2454	0.7546	0.0567	0.3574
Sim	0.4294	0.2407	0.7593	0.0515	0.3779
Separate	0.2525	0.2383	0.7617	0.0242	0.2283
$n_2 = 2$					
Posch	0.5613	0.2415	0.7583	0.0727	0.4886
Sim	0.5825	0.2407	0.7593	0.072	0.5105
Separate	0.434	0.2416	0.7584	0.0424	0.3916
$n_2 = 5$					
Posch	0.8107	0.2354	0.7646	0.1055	0.7052
Sim	0.8204	0.2397	0.7603	0.112	0.7084
Separate	0.7643	0.2371	0.7629	0.0836	0.6807

Table A.7: Power Per Arm for Scenario (.67, 1.33, 2)

Method	Power	Select 1	Select 2	Select 3	Success 1	Success 2	Success 3
$n_2 = .5$							
Posch	0.3037	0.1001	0.2829	0.617	0.0157	0.0717	0.2163
Sim	0.3104	0.1074	0.2724	0.6202	0.0148	0.066	0.2296
Separate	0.1328	0.1041	0.2704	0.6255	0.0049	0.0248	0.1031
$n_2 = 1$							
Posch	0.3871	0.1089	0.275	0.6161	0.0178	0.0842	0.2851
Sim	0.4014	0.1048	0.2766	0.6186	0.0171	0.0844	0.2999
Separate	0.2334	0.1069	0.2836	0.6095	0.008	0.0459	0.1795
$n_2 = 2$							
Posch	0.511	0.1046	0.2856	0.6098	0.0179	0.1106	0.3825
Sim	0.5417	0.1088	0.275	0.6162	0.0221	0.1109	0.4087
Separate	0.4023	0.1077	0.2841	0.6082	0.0123	0.0741	0.3159
$n_2 = 5$							
Posch	0.7684	0.1062	0.2828	0.611	0.0291	0.1811	0.5582
Sim	0.7792	0.105	0.2839	0.6111	0.0268	0.1859	0.5665
Separate	0.7172	0.1056	0.2788	0.6156	0.0184	0.1554	0.5434

Table A.8: Power Per Arm for Scenario (.5, 1, 1.5, 2)

Method	Power	Select 1	Select 2	Select 3	Select 4	Success 1	Success 2	Success 3	Success 4
$n_2 = .5$									
Posch	0.2888	0.0609	0.1408	0.2802	0.5181	0.0077	0.0265	0.0719	0.1827
Sim	0.3039	0.0595	0.1366	0.2854	0.5185	0.0078	0.0269	0.0845	0.1847
Separate	0.136	0.0573	0.14	0.2791	0.5236	0.0029	0.0113	0.0348	0.087
$n_2 = 1$									
Posch	0.357	0.0636	0.1384	0.2773	0.5207	0.0075	0.0284	0.0908	0.2303
Sim	0.378	0.0576	0.1354	0.2946	0.5124	0.0065	0.0284	0.0967	0.2464
Separate	0.2201	0.0574	0.1373	0.2799	0.5254	0.003	0.0132	0.0538	0.1501
$n_2 = 2$									
Posch	0.4889	0.0656	0.1388	0.2777	0.5179	0.0097	0.0369	0.1237	0.3186
Sim	0.5042	0.0627	0.1373	0.2847	0.5153	0.0084	0.035	0.1271	0.3337
Separate	0.3827	0.0608	0.1375	0.2768	0.5249	0.004	0.0257	0.0893	0.2637
$n_2 = 5$									
Posch	0.753	0.0583	0.1347	0.2791	0.5279	0.0096	0.0609	0.2035	0.479
Sim	0.7657	0.061	0.1356	0.2845	0.5189	0.0111	0.0609	0.2119	0.4818
Separate	0.7028	0.0624	0.1411	0.2825	0.514	0.0066	0.0494	0.1878	0.459

Table A.9: Power Per Arm for Scenario (.4, .8, 1.2, 1.6, 2)

Method	Power	Select 1	Select 2	Select 3	Select 4	Select 5	Success 1	Success 2	Success 3	Success 4	Success 5
$n_2 = .5$											
Posch	0.2722	0.0398	0.0849	0.1537	0.273	0.4486	0.0038	0.0126	0.0325	0.0698	0.1535
Sim	0.3057	0.0406	0.0787	0.1578	0.2689	0.454	0.0053	0.014	0.0353	0.0799	0.1712
Separate	0.1202	0.0399	0.0826	0.1516	0.2782	0.4477	0.001	0.005	0.0138	0.0316	0.0688
$n_2 = 1$											
Posch	0.3343	0.039	0.0855	0.1481	0.2775	0.4499	0.0032	0.0161	0.0369	0.0938	0.1843
Sim	0.3831	0.039	0.0835	0.1495	0.2711	0.4569	0.0046	0.014	0.0394	0.1059	0.2192
Separate	0.2143	0.0358	0.0835	0.1564	0.2713	0.453	0.0012	0.0069	0.019	0.0567	0.1305
$n_2 = 2$											
Posch	0.4736	0.0374	0.0807	0.1597	0.2685	0.4537	0.0043	0.0154	0.0509	0.1297	0.2733
Sim	0.5091	0.0356	0.0786	0.1566	0.275	0.4542	0.0039	0.0172	0.0549	0.1361	0.297
Separate	0.3789	0.0411	0.0832	0.1548	0.2658	0.4551	0.0026	0.0107	0.0364	0.0974	0.2318
$n_2 = 5$											
Posch	0.7286	0.0368	0.0805	0.1564	0.2721	0.4542	0.0049	0.0251	0.0841	0.2071	0.4074
Sim	0.7637	0.0389	0.0808	0.153	0.2751	0.4522	0.007	0.0277	0.0892	0.2168	0.423
Separate	0.6901	0.0408	0.0783	0.1593	0.2736	0.448	0.0037	0.019	0.0777	0.1961	0.3936

Table A.10: Power Per Arm for Scenario (.33, .67, 1, 1.33, 1.67, 2)

Method	Power	Select 1	Select 2	Select 3	Select 4	Select 5	Select 6	Success 1	Success 2	Success 3	Success 4	Success 5	Success 6
$n_2 = .5$													
Posch	0.2618	0.0284	0.055	0.0923	0.1657	0.2666	0.392	0.003	0.0081	0.0163	0.0357	0.0722	0.1265
Sim	0.2935	0.0288	0.0531	0.0993	0.1671	0.2544	0.3973	0.0048	0.0084	0.0204	0.0402	0.0784	0.1413
Separate	0.1268	0.0287	0.0516	0.0975	0.1632	0.2597	0.3993	0.0008	0.0031	0.0077	0.0175	0.0296	0.0681
$n_2 = 1$													
Posch	0.3335	0.0279	0.0552	0.0965	0.1587	0.2643	0.3974	0.0028	0.0092	0.0176	0.0439	0.0898	0.1702
Sim	0.3687	0.0258	0.0537	0.0929	0.1611	0.2532	0.4133	0.0024	0.0076	0.0213	0.0456	0.0957	0.1961
Separate	0.214	0.0292	0.0523	0.1038	0.1615	0.249	0.4042	0.0017	0.0034	0.0096	0.0291	0.0548	0.1154
$n_2 = 2$													
Posch	0.4584	0.0291	0.0503	0.0928	0.1631	0.2622	0.4025	0.0024	0.0078	0.0205	0.0605	0.1216	0.2456
Sim	0.51	0.0267	0.0472	0.0933	0.1605	0.2566	0.4157	0.0027	0.0096	0.025	0.0627	0.1403	0.2697
Separate	0.3732	0.0273	0.052	0.0984	0.1598	0.2542	0.4083	0.001	0.0045	0.0154	0.0438	0.0965	0.212
$n_2 = 5$													
Posch	0.7195	0.0292	0.0529	0.0956	0.1651	0.2644	0.3927	0.0035	0.0122	0.04	0.0996	0.2115	0.3527
Sim	0.7531	0.0279	0.0507	0.0961	0.1536	0.2699	0.4018	0.0039	0.0141	0.0425	0.0989	0.221	0.3727
Separate	0.6876	0.0267	0.0579	0.0923	0.159	0.267	0.3971	0.0022	0.0095	0.0348	0.0893	0.2017	0.3501

Appendix B

R code

The following R code creates a function 'fTypeOne' which returns Z_{sim} based on the following input:

- $n2$ = the number of patients per arm in stage two
- $nsim$ = the number of simulations to run
- k = the number of arms tested in stage one

```
fTypeOne = function(n2, nsim, k){
  y=numeric(k+1)
  ts = numeric(nsim)
  for(i in 1:nsim){
    y[1:(k+1)] = rnorm(k+1, 0, 1)
    y2 = rnorm(k+1, 0, sd=1/sqrt(n2))
    yy=(y + n2*y2)/(1 + n2)

    zz=(yy[2:(k+1)] - yy[1])/sqrt(1/(1+n2)+1/(1+n2))

    ts[i]=zz[which.max(y[2:(k+1)])]
  }
  return(quantile(ts, .975))
}
```

The following code creates a function 'fPower' that returns the power of the simulation method based on the following input:

- x is a vector of the true treatment effects for each arm. If phase two has three arms, and example of x is $(.5, 1, 1.5)$
- $n2$ is the number of patients per arm in stage two
- ns is the number of simulations to run

```
fPower = function(x, n2, ns){
  k=length(x)
  y = numeric(k+1)
  y2 = numeric(k+1)
  rejTrue = numeric(ns)
  pMat = numeric(ns)
  zVal = fTypeOne(n2, 100000, k)
  for (sims in 1:ns){
    y[2:(k+1)] = rnorm(k, x, sd=1)
    y[1]=rnorm(1, 0, sd=1)

    y2[2:(k+1)]=rnorm(k, x, sd=1/sqrt(n2))
    y2[1]=rnorm(1, 0, sd=1/sqrt(n2))

    yy = (y + n2*y2)/(1+n2)
    zz = (yy[2:(k+1)] - yy[1])/sqrt(1/(1+n2) + 1/(1+n2))
    best = which.max(y[2:(k+1)])

    rejTrue[sims]= (zz[best] > zVal)
  }
  p = sum(rejTrue)/ns
  return(p)
}
```

The following code creates a function 'simesPower' which simulates the Posch method using the Simes test, and returns the power of the method based on the following input:

- x is a vector of the treatment effects for each arm in stage one
- $n2$ is the number of patients per arm in stage two

- ns is the number of simulations to run

```
simesPower = function(x, n2, ns){
  k=length(x)
  y1 = numeric(k+1)
  y2 = numeric(k+1)
  ts = numeric(ns)
  rejTrue = numeric(ns)

  for (sims in 1:ns){
    y1[2:(k+1)] = rnorm(k, x, sd=1)
    y1[1] = rnorm(1, 0, sd=1)

    y2[2:(k+1)]=rnorm(k, x, sd=1/sqrt(n2))
    y2[1]=rnorm(1, 0, sd=1/sqrt(n2))

    z1 = (y1[2:(k+1)]-y1[1])/sqrt(2)
    z2 = (y2[2:(k+1)]-y2[1])/sqrt((2/n2))
    pval1 = 1 - pnorm(z1)
    pval2 = 1 - pnorm(z2)

    w1 = sqrt(1/(1+n2))
    w2 = sqrt(n2/(1+n2))
    best = which.max(y1[2:(k+1)])
    pbest = min(pval1)
    sortedP = sort(pval1, index.return = TRUE)

    stage2 = pval2[best]

    combFunction = function(p1, p2){ return(1 - pnorm(w1*qnorm(1-p1)
       +w2*qnorm(1-p2))) }

    pvalVector = numeric(k)

    pvalVector[1] = combFunction(pbest, stage2)
    if(k > 1){
      for(j in 2:k){
        pList=numeric(j)
        pList[1] = j * pbest
        for(i in 2:length(pList))
        {
          pList[i] = (j / i) * sortedP$x[k-j+i]
        }
        pvalVector[j] = combFunction(min(pList), stage2)
      }
    }

```

```
    rejTrue[sims]= (max(pvalVector) < .025)
  }
  p = sum(rejTrue)/ns
  return(p)
}
```

The following code creates a function 'sepPower' which simulates a trial with separate phase II/III, and returns the power of the method based on the following input:

- x is a vector of the treatment effects for each arm in stage one
- $n2$ is the number of patients per arm in stage two
- ns is the number of simulations to run

```
sepPower = function(x, n2, ns){
  k=length(x)
  y1 = numeric(k+1)
  y2 = numeric(k+1)
  ts = numeric(ns)
  rejTrue = numeric(ns)
  pMat = numeric(ns)
  for (sims in 1:ns){
    y1[2:(k+1)] = rnorm(k, x, sd=1)
    y1[1] = rnorm(1, 0, sd=1)

    y2[2:(k+1)]=rnorm(k, x, sd=1/sqrt(n2))
    y2[1]=rnorm(1, 0, sd=1/sqrt(n2))

    z2 = (y2[2:(k+1)] - y2[1])/sqrt((2/n2))
    best = which.max(y1[2:(k+1)])

    rejTrue[sims] = (z2[best] > 1.96)
  }
  p = sum(rejTrue) / ns
  return(p)
}
```

Bibliography

Berry, S., Carlin, B., Lee, J., and Mueller, P. (2011), *Bayesian Adaptive Methods for Clinical Trials*, Chapman and Hall, CRC Biostatistics Series, Boca Raton, Fl.

Chaturvedi, P. and Mehta, C. (2014), "ADVENT: The First Execution of an Adaptive Phase 3 Trial Resulting in FDA Approval of Crofelemer," Presentation for DIA KOL Series.

FDA (2010), "Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics," .

Posch, M., Koenig, F., Branson, M., Brannath, W., Dunger-Baldauf, C., and Bauer, P. (2005), "Testing and estimation in flexible group sequential designs with adaptive treatment selection," *Statistics in Medicine*, 24, 3697–3714.

Ross, S. (2010), *A First Course in Probability*, Pearson Prentice Hall, Upper Saddle River, NJ.

Simes, R. J. (1986), "An improved Bonferroni procedure for multiple tests of significance," *Biometrika*, 23, 751–754.