

12-2019

STUDYING THE PROPERTIES OF SEAMLESS PHASE II/PHASE III CLINICAL TRIALS UNDER COVARIATE ADAPTIVE RANDOMIZATION

GARRETT MILLER

Follow this and additional works at: https://digitalcommons.library.tmc.edu/uthsph_dissertsopen



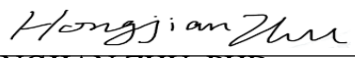
Part of the [Community Psychology Commons](#), [Health Psychology Commons](#), and the [Public Health Commons](#)

STUDYING THE PROPERTIES OF SEAMLESS PHASE II/PHASE III CLINICAL
TRIALS UNDER COVARIATE ADAPTIVE RANDOMIZATION

by

GARRETT MILLER, BS

APPROVED:


HONGJIAN ZHU, PHD


HAN CHEN, PHD

Copyright
by
Garrett Miller, BS, MS
2019

STUDYING THE EFFECTS OF SEAMLESS PHASE II/PHASE III CLINICAL TRIALS
UNDER COVARIATE ADAPTIVE RANDOMIZATION

by

GARRETT MILLER
BS MATHEMATICS, North Carolina State University, 2017

Presented to the Faculty of The University of Texas

School of Public Health

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

THE UNIVERSITY OF TEXAS
SCHOOL OF PUBLIC HEALTH
Houston, Texas
December, 2019

ACKNOWLEDGEMENTS

Thank you to my thesis supervisor Dr. Zhu, and my minor advisor Dr. Chen for guidance on this thesis. Thank you to Dr. Susan Todd and Dr. Nigel Stallard for their help on understanding the seamless Phase II/Phase III clinical trial designs. Thank you to my family and friends for their support.

STUDYING THE PROPERTIES OF SEAMLESS PHASE II/PHASE III CLINICAL TRIALS WITH COVARIATE ADAPTIVE RANDOMIZATION

Garrett Miller, MS
The University of Texas
School of Public Health, 2019

Thesis Chair: Hongjian Zhu, PhD

Covariate adaptive randomization is an allocation procedure used in clinical trials that seeks to balance treatment groups. While this method has been shown to reduce bias due to imbalanced treatment groups, the effects of covariate adaptive randomization have not been studied under seamless Phase II/Phase III clinical trials that incorporate short-term endpoint information and treatment selection. Therefore, this analysis sought to determine whether these adaptive randomization methods can be applied to seamless Phase II/Phase III trials while preserving Type I error. In addition, this analysis sought to create R-packages that employ seamless Phase II/Phase III techniques to provide analysis tools for future research. Two covariate adaptive randomization schemes, Pocock and Simon's procedure and stratified permuted block randomization were applied to simulated datasets to determine treatment groups. Seamless Phase II/Phase III clinical trials with two interim analyses were then simulated with 10,000 repetitions to determine overall Type I error rates. Only the most promising treatment group was selected to continue to the second interim analysis. Both randomization procedures were compared against trials that used simple randomization to

allocate treatment groups. Ultimately, Type I error rates under the two adaptive techniques were not preserved. Pocock and Simon's randomization saw an inflated level of Type I error, while stratified permuted block randomization resulted in lower levels of Type I error compared to simple randomization. In addition, power calculations revealed that both allocation methods resulted in lower levels of power compared to simple randomization.

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES	iv
BACKGROUND	1
Literature Review	1
PUBLIC HEALTH SIGNIFICANCE.....	4
SPECIFIC AIMS	5
METHODS	6
Study Design	6
Data Collection.....	7
Study Sample.....	7
Data Analysis	9
R-package for A New Clinical Trial Design Combining Phases 2 and 3: Sequential Designs with Treatment Selection and a Change of Endpoint.	9
R-package for A Confirmatory Seamless Phase II/Phase III Clinical Trial Design Incorporating Short-Term Endpoint Information	11
R-package for Statistical Analysis for Two-Stage Seamless Design with Different Study Endpoints.....	14
Covariate Adaptive Randomization in a Seamless Phase II/Phase III Clinical Trial.	16
RESULTS	17
DISCUSSION	19
CONCLUSION.....	20
ETHICAL AND SAFETY CONSIDERATIONS	21
REFERENCES	22

LIST OF TABLES

Table 1: Type I and Type II Error.....	4
Table 2: Type I Error Rates under Pocock & Simon's and Stratified Permuted Block Randomization	18
Table 3: Power Under Pocock and Simon's and Stratified Permuted Block Randomization	18

LIST OF FIGURES

Figure (1): Distribution of Bernoulli Covariates	7
Figure (2): Probability of Selecting Treatment A in Pocock Design.....	7
Figure (3): Distribution of Endpoints in Stallard (2010) Design.....	8
Figure (4): Distribution of Error Term in Stallard (2010) Design.....	8
Figure (5): Treatment Effect at First Interim Analysis	9
Figure (6): Distribution of Efficient Score at First Interim Analysis.....	10
Figure (7): Correlation Between $Z_{S,2}^{(2)}$ And $Z_{S,1}^{(1)}$	10
Figure (8): Correlation Between $Z_{S,1}^{(1)}$ And $Z_{S,1}^{(2)}$	10
Figure (9): Covariance Between $Z_{S,1}^{(1)}$ And $Z_{S,1}^{(2)}$	11
Figure (10): Estimated $\hat{\mathbf{B}}$ Parameter in Double Regression Method	12
Figure (11): Estimated $\hat{\mathbf{A}}$ Parameter in Double Regression Method	12
Figure (12) Estimated Variance in Double Regression Method	12
Figure (13): Variance of $\hat{\mathbf{B}}$ in Double Regression Method.....	12
Figure (14): Distribution of Secondary Endpoint in Stallard (2010).....	12
Figure (15): Conditional Distribution of Primary Endpoint Given Secondary Endpoint in Stallard (2010).....	13
Figure (16): Test Statistic for Stallard (2010).....	13
Figure (17): Density of Selected Treatment in Stallard (2010)	13
Figure (18): Spending Function Requirements in Stallard (2010).....	13
Figure (19): Expectation and Variance of Endpoints in Chow, Lu, And Tse (2007)	14
Figure (20): Linear Relationship Between Endpoints in Chow, Lu, And Tse (2007)	15
Figure (21): Weighted Mean Between Estimated and True Long-Term Endpoint	15
Figure (22): Weight in Weighted Mean Estimator	15
Figure (23): Weighted Variance Between Estimated and True Long-Term Endpoint.....	15
Figure (24): Test Statistic For Chow, Lu, and Tse (2007).....	16
Figure (25): Interval For Test Statistic For Chow, Lu, and Tse (2007).....	16

BACKGROUND

Clinical trials are an essential function of the healthcare system as they are the basis for developing new therapies and treatments. Once a therapy has been developed, it must undergo a multi-stage process to determine whether the therapy is safe, effective, and ready for public use. Phase I of a clinical trial is used to administer the drug to a small set of patients in order to determine that the drug is safe, and Phase II is used to determine an effective dosage and treatment schedule from a variety of different treatment policies (Whitehead, 1991). Phase III compares the chosen treatment policy to a control using a large number of patients, and Phase IV is used to evaluate the performance of the drug after it has reached public market (Whitehead, 1991). To reduce the large costs and sample sizes associated with clinical trials, many methods have been proposed that combine Phase II and Phase III into a single seamless trial that selects a most promising treatment and compares that treatment to a control (Stallard & Todd, 2005). While these methods have been shown to reduce overall sample size, further analysis needs to be done to test these designs under different allocation schemes such as covariate adaptive randomization. Since balancing covariates is an essential way to balance treatment groups with respect to their key covariates, it is important to ensure that seamless Phase II/Phase III methods are still valid under these allocation schemes (Ma, Hu, & Zhang, 2015). As there is always a need to reduce overall costs associated with clinical trials, more analysis needs to be done to develop the most ethical and efficient statistical methods used in clinical trials analysis.

Literature Review

In clinical trials, the Phase II trial, often described as an exploratory phase, is designed to determine the most promising treatment out of several treatments or to determine the most

promising dosage of a treatment (Whitehead, 1991). Once a most promising treatment has been determined, the trial proceeds to Phase III, where more long-term, primary effects are examined to determine if the treatment is effective. Since patients are often recruited as the trial goes on, it is possible to conduct analysis at interim points during the clinical trial, rather than waiting until all the patients have been recruited. Such designs that use interim inspections are known as sequential designs. These designs can lead to more ethical and efficient studies that stop analysis before all patients are recruited.

Seamless Phase II/Phase III clinical trials seek to eliminate some of the information lost between trials, as well as time lost to a recruitment by combining the phases into a seamless trial design. These methods use both phases in a single trial that selects the most promising treatment, while comparing the most promising treatment with a control to test for efficacy (Stallard & Todd, 2005). There are many methods that have been proposed to conduct this combined design. The seamless designs typically assume that there is a secondary endpoint available at the Phase II portion for the trial, and a primary endpoint available during the Phase III portion of the trial. Chow, Lu, and Tse (2007) developed a method that assumes there is a well-defined relationship between the two endpoints, which allows for estimation of the primary endpoint using a linear relationship. While this method was shown to reduce overall sample size, it does not incorporate the sequential design that is often used in the confirmatory phase of a clinical trial as there are no interim analyses. In addition, it is highly dependent on there being a well-defined relationship between two normally distributed endpoints. Furthermore, it does include treatment selection after the Phase II portion of the trial, which is a common element of Phase II/Phase III trials. To allow for sequential design in combination with a Phase II/Phase III clinical trial, Stallard (2010) considered a method that incorporates short-term endpoint information to select one treatment

after the exploratory phase and incorporates the short-term endpoint information to develop the stopping boundaries of the Phase III portion of the trial. This method was shown as a promising way to incorporate short-term information while preserving Type I error rates; however, this method was restricted to normally distributed endpoints, with a relatively high correlation between the two endpoints. To address the possibility of differently distributed endpoints at the separate analysis stages, Stallard and Todd (2003) proposed a seamless Phase II/Phase III design that incorporates short-term endpoint information while considering a change in endpoint in addition to treatment selection. This method considers a normally distributed short-term endpoint, used to select one treatment to continue to the Phase III portion of the trial, and a binomially distributed long-term endpoint, used to compare the most promising treatment to a control in a sequential design. This design was able to reduce sample size while preserving Type I error compared to a Phase II trial followed by a multi-stage Phase III trial.

Clinical trials allocate patients to treatment groups using randomization, to reduce possible biases in the treatment arms. With simple randomization, there may be a bias associated with clusters of patients receiving the same treatment. One method to reduce this bias is covariate adaptive randomization. Covariate adaptive randomization assigns patients to treatment groups while balancing possible covariates and preserving randomization. Typical methods in covariate adaptive randomization change the allocation probabilities at each step, while adjusting for previous assignments and the covariates of the subjects in order to reduce the possibility of bias. (Kahan, Morris, 2001). While there are many techniques for performing this randomization, only Pocock-Simons and stratified permuted block Design (SPB) were explored in this analysis. Pocock and Simon (1975) use a generalized version of a minimization procedure that utilizes a weighted sum to determine the differences between the number of patients over the separate

treatment groups. In stratified permuted block randomization, patients are assigned to strata that are defined by characteristics and randomization is performed on each stratum in equal-sized blocks. This is done to ensure that those baseline characteristic variables are balanced in each block. (Kahan, Morris, 2001). These methods intend to minimize imbalance between treatment groups, while reducing selection and accidental bias. Further analysis needs to be done to determine how these covariate adaptive randomization techniques affect the results of a seamless Phase II/Phase III trial design.

In any trial design, investigators seek to test a set of hypotheses: the null and alternate hypothesis. The null assumes that a new treatment has no effect on an outcome of interest, while the alternative assumes that there is an association. The null hypothesis enables investigators to conduct statistical tests that assume a reasonable measure of doubt set by the Type I and Type II error rates. Type I error, also known as false positive rate, occurs when the null hypothesis is rejected when then the hypothesis is actually true in the population. The Type II error rate occurs when the trial does not reject a null hypothesis that is actually false in the population. Ultimately, a test can result in the 4 situations shown in Table 1.

Table 1: Type I and Type II Error

Truth	Association	No Association
Reject Null Hypothesis	Correct Decision	Type I Error
Fail to reject Null Hypothesis	Type II error	Correct

PUBLIC HEALTH SIGNIFICANCE

Clinical trials are an essential pathway for developing new and effective therapies. In these trials, it is important to reduce possible harm caused to patients by giving them an ineffective treatment. Using a sequential design allows investigators to stop recruitment if at any

point during the interim analyses, the treatment is determined to be effective, or futile. Costs of clinical trials collected from seven major biopharma companies from 2010 to 2015 found that the median cost of Phase I, Phase II, and Phase III trials was \$3.4 million, \$8.6 million, and \$21.4 million respectively (Martin, Hutches, Hawkins, & Radnov, 2017). Methods such Stallard and Todd (2005), Stallard (2010) and Chow, Lu and Tse (2007) use study designs that attempt to reduce overall sample size, while still achieving a desired level of power. Therefore, these methods can be employed to attempt to reduce some of the overall costs of clinical trials. In addition, if it is determined that covariate adaptive randomization preserves Type I error rates in these trials, further cost reduction could be achieved by reducing biases that would prevent effective treatments from being discovered. With the vast number of clinical trials per year, it is important to develop statistical methods that efficiently and ethically identify new treatments.

SPECIFIC AIMS

While previous studies were able to reduce overall sample size and preserve Type I error rates in seamless Phase II/Phase III clinical trials, these methods have not been explored over different sets of data. In addition, these methods have not been adapted to include different allocation schemes such as covariate adaptive randomization. Based on these limitations, this paper had two specific aims:

Aim 1: Develop R-packages for the three seamless Phase II/Phase III clinical trial designs described by Stallard and Todd (2005), Stallard (2010), and Chow, Lu and Tse (2007) to disseminate these innovative approaches to clinical trial practitioners.

Aim 2: Study whether Type I error rates in confirmatory seamless Phase II/Phase III clinical trial incorporating short-term endpoint information are controlled under covariate adaptive randomization.

In order to examine the results of the three seamless Phase II/Phase III clinical trial designs, R-packages were created that followed the methodology described by each of the three papers Stallard and Todd (2005), Stallard (2010), and Chow, Lu and Tse (2007).

In order to determine whether Type I error rates in a seamless Phase II/Phase III design are controllable under covariate adaptive randomization, multiple factors were analyzed in order to determine their influence on the Type I error. These factors included: randomization scheme, correlation between study endpoints, and incorporation of Bernoulli covariates.

METHODS

Study Design

The designs described in Stallard and Todd (2005), Stallard (2010), and Chow, Lu, and Tse (2007) were recreated in R-packages to allow for application of these methods for innovative analysis. These packages were developed by following the methodology of each of the three designs, while allowing for flexible options for trial practitioners.

The design in Stallard (2010) was adapted to incorporate covariate adaptive randomization using both the Pocock-Simons and stratified permuted block design for treatment allocation. Multiple parameter settings were evaluated with 10,000 simulated seamless Phase II/Phase III trials under the Stallard (2010) design. Estimates for the overall Type I error rate were taken from the 10,000 simulations to determine whether Type I error was conserved under covariate adaptive randomization.

Data Collection

In the creation of the R-packages, no data was collected. All data was simulated in R to ensure that the results align with those described in Stallard and Todd (2005), Stallard (2010), and Chow, Lu, and Tse (2007). No data was collected from the three packages.

Study Sample

To adapt the design in Stallard (2010) study populations were simulated in R and generated with the following steps: (1) generate Bernoulli covariates for each observation, (2) assign observations to a treatment group based on covariate adaptive randomization design (3) generate treatment responses.

For the design in Stallard (2010), the primary and secondary endpoints came from a bivariate normal distribution. In each clinical trial simulation, two Bernoulli covariates were created. The covariates are independent and follow the distribution in Equation (1).

$$r_i = \begin{cases} 1 & \text{with probability } p_i \\ 0 & \text{with probability } 1 - p_i \end{cases} \quad (1)$$

After generating the Bernoulli covariates for each observation, the two covariate adaptive randomization techniques, Pocock-Simons and stratified permuted block design, were used to allocate observations to treatment groups. In Pocock-Simons design, were assigned to treatment group A with the probability given in Equation (3).

$$Pr_{n+1,(j,l)}(A) = \begin{cases} p & \min\{G_k\} = G_A \\ \frac{1}{k} & G_1 = \dots = G_k \\ \frac{1-p}{k-1} & \min\{G_k\} \neq G_A \end{cases} \quad (2)$$

In Equation (2), k is the number of treatment groups, and $p = 3/4$ as suggested by the original authors (Pocock & Simons, 1975). In Equation (2) G_k is defined as the “total amount of imbalance” which is taken as a weighted sum of the range of a set of non-negative integers $\{z_l\}_{l=1}^N$. Therefore, the treatment assignment that results in the smallest “total amount of imbalance” was assigned with probability p .

For trials simulated using the stratified permuted block design, this analysis used four levels of discretization with a block size of twenty. Details for the stratified permuted block design can be found in (Broglia, 2018).

The primary and secondary responses were assigned based on the distribution shown in Equation (3).

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim N \left(\begin{pmatrix} a + b'z_i + \tau_1 r_i + \tau_2 r_i \\ A + B'z_i + \tau_1 r_i + \tau_2 r_i \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \rho\sigma_0\sigma \\ \rho\sigma_0\sigma & \sigma^2 \end{pmatrix} \right) \quad (3)$$

In Equation (3), r_i and r_i are the Bernoulli covariates for the i^{th} patient defined in Equation (1). In addition, b and B are the treatment effects for the short-term and long-term endpoints, respectively. ρ is the partial correlation between Y_i and X_i , σ_0^2 is the variance of the error term for X_i , σ^2 is the variance of the error term $Y_i|X$. The remaining parameter values in this distribution are shown in Table 3. τ_1 and τ_2 were included to account for the covariate values in the distribution. A and a are the intercept terms. These terms were later excluded from this analysis. Error was applied to each observation and came from the distribution in Equation (4).

$$\varepsilon \sim N(0,1) \quad (4)$$

The primary and secondary responses, with their error terms, were generated for each of the 10,000 simulated clinical trials to give our study sample.

Data Analysis

To create R-packages for each of the three seamless Phase II/Phase III clinical trial designs, the methodology for each paper was recreated in R.

R-package for A New Clinical Trial Design Combining Phases 2 and 3: Sequential Designs with Treatment Selection and a Change of Endpoint.

To create an R-package that follows the design in Stallard and Todd (2005), R code was written that followed the steps of the seamless Phase II/Phase III trial design defined in Stallard and Todd (2005). The design takes any K number of treatments and observes two endpoints on each patient. In order to make inferences about the different treatments, this method uses two test statistics, corresponding to the Efficient Score and Fisher's information for a given treatment. The steps for this methodology are described in steps 1-6, with additional details available in (Stallard & Todd, 2005).

- (1) The two test statistics, efficient score and Fisher's information, are given in detail by Whitehead (1997). The statistics depend on the distribution of the outcome of interest. In the case of the first interim analysis, the outcome is normally distributed. The treatment effect at each analysis is shown in Equation (5).

$$\hat{\theta}_i^{(1)} = \frac{Z_{i,1}^{(1)}}{V_{i,1}^{(1)}} \quad (5)$$

In Equation (5), $V_{i,1}^{(1)}$ corresponds to the Fisher's information for secondary endpoint at the first interim analysis and $Z_{i,1}^{(1)}$ is the Efficient Score for the secondary endpoint at the first interim analysis

- (2) Out of the possible treatments, the most promising treatment is determined by using the largest value for the treatment effect. Once the most promising treatment has been chosen, the Efficient Score for this treatment is compared against a lower boundary. If does not cross the lower boundary, then the trial proceeds to the next interim analysis. The rationale for choosing a lower stopping boundary at this analysis can be found in Stallard and Todd (2005)
- (3) At the next analysis, the outcome, which now comes from a binomial distribution, is used to calculate the same test statistics as in the first analysis. In this case however, the formulae for the test statistics depend on the binomial distribution (Whitehead, 1997). The treatment effect is calculated in the same way as at the first interim analysis. Stopping boundaries are calculated at the second interim analysis are based on the Fisher's information at this point, and the Fisher's information obtained at the first interim analysis. The distribution of the Efficient Score at this point is shown in Equation (6)

$$Z_{S,2}^{(2)} \sim N(\theta_S^{(2)} V_2^{(2)} - \rho_2 \sqrt{\frac{V_2^{(2)}}{V_1^{(1)}}} \theta_S^{(1)} V_1^{(1)} - z, V_2^{(2)} (1 - \rho_2^2)) \quad (6)$$

$$\text{where } \rho_2 = \rho \sqrt{\frac{V_1^{(2)}}{V_2^{(2)}}} \quad (7)$$

$$\hat{\rho} = \frac{\widehat{cov}(Z_{S,1}^{(1)}, Z_{S,1}^{(2)})}{\sqrt{V_{S,1}^{(1)} V_{S,1}^{(2)}}} \quad (8)$$

$$\widehat{cov}(Z_{S,1}^{(1)}, Z_{S,1}^{(2)}) = \frac{n_S n_C}{n^2 \hat{\sigma}} \{n_C \hat{p}_S (\bar{x}_S^s - \bar{X}_S) + n_E \hat{p}_C (\bar{x}_C^s - \bar{x}_C)\} \quad (9)$$

Equation (7) gives the correlation between $Z_{S,2}^{(2)}$ and $Z_{S,1}^{(1)}$. Equation (8) is the correlation between $Z_{S,1}^{(1)}$ and $Z_{S,1}^{(2)}$. In Equation (9), the covariance between $Z_{S,1}^{(1)}$ and $Z_{S,1}^{(2)}$, $\hat{\sigma}^2$ is the MLE of σ for normally distributed endpoint under assumption that $\mu_S = \mu_C$, n_S is the number of individuals from the selected treatment group, n_C is the number of individuals from control group, \hat{p}_S is the probability of success for binary endpoint in treatment group, \hat{p}_C is the probability of success for binary endpoint in control group, \bar{x}_S^s is the treatment mean from normally distributed endpoint, and \bar{x}_C^s is the treatment mean value of short-term endpoint among those classified as a success in the binary outcome. Additional specifics for calculating these boundaries can be found in Stallard and Todd (2005).

- (4) If the observed efficient Score at the current analysis crosses the upper boundary, the trial stops and the treatment is determined to be superior to the control. If the test statistic crosses below the lower boundary, the trial stops for futility.
- (5) If the observed efficient Score at the current analysis crosses the upper boundary, the trial stops and the treatment is determined to be superior to the control. If the test statistic crosses below the lower boundary, the trial stops for futility.
- (6) Steps (3), (4), and (5) continue until the maximum number of analyses is reached, which is set before the trial begins.

R-package for A Confirmatory Seamless Phase II/Phase III Clinical Trial Design

Incorporating Short-Term Endpoint Information

To recreate the confirmatory seamless Phase II/Phase III clinical trial design described in Stallard (2010), the design was rewritten in R to better test the method against new sets of data.

The design takes any number of treatments, provided that the treatments responses are normally distributed. The steps for performing the design are described in steps 1-4.

- (1) A double regression method is used to estimate the parameters of the model in the case where secondary endpoint responses are available for all patients, and primary responses are available for a subset of patients (Engel & Walstra, 1991). The parameter estimates for the secondary endpoint are obtained from the regression of X_i on z_i where z_i is the vector that indicates which treatment group allocation. The estimates for α, β, γ , and can be obtained from regressing the primary outcome on z_i and X_i . In addition, the parameter σ_1^2 , the variance of the error term for Y_i , can be obtained. These estimates can then be used to obtain the parameters used in the calculation of the test statistic by the Equations (10), (11), and (12).

$$\hat{B} = \hat{\beta} + \gamma \hat{b} \quad (10)$$

$$\hat{A} = \hat{\alpha} + \hat{\gamma} \hat{\alpha} \quad (11)$$

$$s^2 = s_1^2 + \hat{\gamma} s_0^2 \quad (12)$$

The variance of the parameter \hat{B} is given by Equation (13).

$$var(\hat{\beta}) + \gamma^2 var(\hat{b}) + b' b var(\hat{\gamma}) + 2b' cov(\hat{\beta}, \hat{\gamma}) \quad (13)$$

The distribution of the secondary endpoint X_i , and the conditional distribution of $Y_i|X_i$ are shown in Equation (14) and Equation (15).

$$X_i \sim N(a + b' z_i, \sigma_0^2) \quad (14)$$

$$(Y_i|X_i = x_i) \sim N(\alpha + \beta' z_i + \gamma x_i, \sigma_1^2) \quad (15)$$

Once the parameter estimates are obtained, the test statistic at the first analysis, S_{k1} , is calculated from Equation (16).

$$S_{k1} = \frac{\hat{B}_k}{\text{var}(\hat{B}_k)} \quad (16)$$

In Equation (16), k is the corresponds to the treatment being used in the calculation of the test statistic. After calculating the test statistics for each treatment, one treatment will be selected to continue based on the largest value of the test statistic at the first analysis

- (2) At this point in the analysis, stopping boundaries are calculated based on the information from the selected treatment, while taking into account the total number of treatments used at the start of the trial. Stopping boundaries are obtained so that they satisfy Equation (17) and Equation (18).

$$f_{S_{[1]1}} = \int_{-\infty}^{\infty} \frac{K}{\frac{\varphi_1}{2}} \phi\left(\frac{x-s}{\sqrt{\frac{\varphi_2}{2}}}\right) \phi\left(\frac{x}{\sqrt{\frac{\varphi_1}{2}}}\right) \left[\Phi\left(\frac{x}{\sqrt{\frac{\varphi_1}{2}}}\right) \right]^{K-1} dx \quad (17)$$

$$Pr(\text{Stop and reject some } H_{0k} \text{ in favor of } B_k > 0 \text{ by look } j; H_0) = \alpha_U^*(j) \quad (18)$$

$$Pr(\text{Stop and reject some } H_{0k} \text{ in favor of } B_k < 0 \text{ by look } j; H_0) = \alpha_L^*(j)$$

In Equation (17), K is the number of treatments, φ_1 is the variance of the selected treatment's test statistic at the first interim analysis given by $\frac{1}{\text{var}\hat{B}_k}$. If the test statistic at the current

analysis crosses the upper boundary, the trial stops and the treatment is determined to be effective. If the test statistic crosses below the lower boundary, the trial stops for futility.

- (3) At the next analysis, recalculate the test statistics and stopping boundary values and compare the test statistic against these stopping boundaries.
- (4) Steps 1, 2, 3 and 4 continues until one of these situations occurs, or the trial reaches the maximum number of interim analyses.

These general steps were re-created in an R-package. The package will be designed to take any number of treatments, as well as different values for the known parameters in the design. Further details for the methodology in this method can be found in Stallard (2010).

R-package for Statistical Analysis for Two-Stage Seamless Design with Different Study Endpoints

Just as for Stallard and Todd (2005) and Stallard (2010), an R-package will be created to replicate the design in Chow, Lu, and Tse (2007). To create this package, R code will be written that follows the steps for the Two-Stage Seamless Design with Different Study Endpoints in Chow, Lu, and Tse (2007). This package will take two treatments, and test for equality of the two treatments. In addition, this design assumes that secondary endpoint information is available for all patients, and that primary endpoint information is available for a subset of those patients. The methodology for Chow, Lu and Tse (2007) is described in steps 1-3.

- (1) The secondary and primary endpoints, denoted by x_i and y_i , follow the relationship defined in Equation (19).

$$E(x_i) = v \text{ and } Var(x_i) = \tau^2 \tag{19}$$

$$E(y_i) = \mu \text{ and } Var(y_i) = \sigma^2$$

Since long-term endpoint data are only available for a subset of the patients, the patients with only short-term endpoint data have their long-term endpoint estimated by Equation (20).

$$y = \beta_0 + \beta_1 x + \epsilon \quad (20)$$

In Equation (20), β_0 and β_1 are assumed to come from a well-defined relationship and are known beforehand.

- (2) Once the estimated points are obtained, an estimate of the weighted mean and variance are obtained for each treatment group with Equation (21) and Equation (23).

$$\hat{\mu}_{GDi} = \hat{\omega}_i \bar{\hat{y}}_i + (1 - \hat{\omega}_i) \bar{y}_i \quad (21)$$

In Equation (21), $\bar{\hat{y}}$ is the weighted estimated mean, \bar{y} is the long-term endpoint mean, $\hat{\omega}$ is the weight given by Equation (22).

$$\hat{\omega}_i = \frac{\frac{n_i}{S_{1i}^2}}{\frac{n_i}{S_{1i}^2} + \frac{m_i}{S_{2i}^2}} \quad (22)$$

$$\hat{V}(\hat{\mu}_{GDi}) = \frac{1}{\frac{n_i}{S_{1i}^2} + \frac{m_i}{S_{2i}^2}} \left[1 + 4\hat{\omega}_i(1 - \hat{\omega}_i) \left(\frac{1}{n_i - 1} + \frac{1}{m_i - 1} \right) \right] \quad (23)$$

In Equation (22) and Equation (23), n is the number of patients in the i th treatment group at the secondary endpoint and m_i is the number of patients in the i th treatment group at the primary clinical endpoint. Additionally, S_{1i}^2 and S_{2i}^2 are the sample variances from the estimated primary endpoint and the true primary endpoints, respectively.

(3) Once the weighted mean and variance are calculated, an estimate for the test statistic can be found using Equation (24).

$$T_1 = \frac{\hat{\mu}_{GD1} - \hat{\mu}_{GD2}}{\sqrt{\hat{V}(\hat{\mu}_{GD1}) + \hat{V}(\hat{\mu}_{GD2})}} \quad (24)$$

The test statistic follows a standard normal distribution and the significance of the test can be found using the interval in Equation (25) where $V_t = \hat{V}(\hat{\mu}_{GD1}) + \hat{V}(\hat{\mu}_{GD2})$ and $z_{\frac{\alpha}{2}}$ is the z quantile a prespecified alpha level.

$$\left(\hat{\mu}_{GD1} - \hat{\mu}_{GD2} - z_{\frac{\alpha}{2}}\sqrt{V_T}, \hat{\mu}_{GD1} - \hat{\mu}_{GD2} + z_{\frac{\alpha}{2}}\sqrt{V_T} \right) \quad (25)$$

These steps will be coded in R in order to develop an R-package that gives practitioners the ability to implement these statistical methods. More details behind this method can be found in Chow, Lu, and Tse (2007).

Covariate Adaptive Randomization in a Seamless Phase II/Phase III Clinical Trial.

In order to determine if Type I error rates are conserved in a confirmatory seamless Phase II/Phase III clinical trial that incorporates short-term endpoint information and uses covariate adaptive randomization for allocation, study sample populations were simulated in R for 10,000 trials for each parameter setting, and for each covariate adaptive randomization scheme described above. Using the Bernoulli covariate values defined above, the steps for this analysis followed the same methodology as described in the confirmatory seamless Phase II/Phase III clinical trial design in Stallard (2010), but the treatment responses followed the distribution in Equation (3). The treatment groups were assigned using either the Pocock-Simons or the stratified permuted block design. Each simulation tested for overall significance, and the Type I

error rate was taken as the proportion of trials that rejected the null hypothesis. The null hypothesis in this case was that all treatments are equally effective.

For the stratified permuted block design, a block size of 20 was used with 4 stratification levels.

RESULTS

Simulations were performed with $\mathbf{b} = \{(0,0,0)', (0,0,0.333)', (0,0, -0.333)'\}$ and $\mathbf{B} = \{(0,0,0)'\}$ to test for Type I error. The overall Type I error rate was calculated as the number of trials that the treatment effect exceeded the stopping boundary and therefore rejected the global null hypothesis. Under simple randomization, the Type I Error Rate was 0.025 with a standard error of 0.001. In the simulations performed under stratified permuted block randomization, the Type I Error Rate was 0.016 with a standard error of 0.0002. In the simulations performed under Pocock and Simon's randomization, the Type I Error Rate was 0.027 with a standard error of 0.008.

Simulations were performed with $\mathbf{b} = \{(0,0,0)', (0,0,0.333)', (0,0, -0.333)'\}$ and $\mathbf{B} = \{(0,0,0.333)'\}$ to test for power. Power was calculated as the proportion of simulations that rejected the global null hypothesis and selected the third treatment as the selected treatment. Under simple randomization, the power was 0.763 with a standard error of 0.009. Those simulations performed under Pocock and Simon's randomization resulted in a power of .598 with a standard error of 0.0064. stratified permuted block randomization resulted in a power of 0.621 with a standard error of 0.007

Table 2: Type I Error Rates under Pocock & Simon's and Stratified Permuted Block Randomization

ρ	b_1	Type I Error Rates		
		Pocock and Simon's Randomization	Stratified Permuted Block Randomization	Simple Randomization
0.5	0.000	0.028	0.016	0.025
	0.333	0.028	0.015	0.025
	-0.333	0.026	0.016	0.024
0.6	0.000	0.026	0.015	0.023
	0.333	0.023	0.014	0.023
	-0.333	0.028	0.016	0.026
0.7	0.000	0.023	0.016	0.024
	0.333	0.031	0.017	0.022
	-0.333	0.024	0.014	0.025
0.8	0.000	0.033	0.015	0.024
	0.333	0.030	0.014	0.026
	-0.333	0.028	0.016	0.028
0.9	0.000	0.034	0.017	0.028
	0.333	0.022	0.017	0.029
	-0.333	0.020	0.018	0.026

Table 3: Power Under Pocock and Simon's and Stratified Permuted Block Randomization

ρ	b_1	Power		
		Pocock and Simon's Randomization	Stratified Permuted Block Randomization	Simple Randomization
0.5	0.000	0.564	0.587	0.720
	0.333	0.569	0.585	0.718
	-0.333	0.574	0.594	0.722
0.6	0.000	0.584	0.601	0.735
	0.333	0.577	0.594	0.741
	-0.333	0.580	0.599	0.733
0.7	0.000	0.594	0.620	0.754
	0.333	0.590	0.622	0.747
	-0.333	0.589	0.628	0.763
0.8	0.000	0.615	0.629	0.788
	0.333	0.617	0.642	0.785
	-0.333	0.619	0.642	0.785
0.9	0.000	0.627	0.666	0.821
	0.333	0.629	0.656	0.814
	-0.333	0.640	0.656	0.820

DISCUSSION

Using Pocock and Simon's randomization to balance treatment assignments between three treatment groups with two covariates, Type I error was not equal to Type I error under simple randomization. This adaptive method approximately inflated the Type I error in this simulation. This conclusion is consistent with previous research on Pocock and Simon's randomization procedure, where Type I error rates were over-estimated in clinical trial simulations. If Pocock and Simon's randomization is used to balance treatment groups, there is a higher chance of exceeding the stopping boundaries in the Phase III portion of the trial. This is concerning as trials under this randomization scheme might have more false positives, and ultimately might falsely conclude that a treatment is determined to be effective. The Type I error rate did not drastically change with variations in correlation between the long-term and short-term endpoints. The power under Pocock and Simon's randomization procedure did not reach the same level as with simple randomization. However, the power did increase with the correlation between the primary and secondary clinical endpoints.

Using stratified permuted block randomization with a block size of 20 to balance treatment assignments between three treatment groups, Type I error rates were not held at the same level as with simple randomization. Type I error rates under stratified permuted block randomization underestimated the Type I error rate compared to simple randomization. One possible explanation for this result is that this randomization procedure results in different numbers of treatment assignments. As a result, the block size isn't always a multiple of the number of treatment groups; therefore, there are some blocks that won't be filled. Another possible explanation is that the different number of patients per treatment group results in largely different test statistics for this clinical trial design. This trial design uses an estimate for Fisher's

information that is a function of the progress in the trial. Therefore, if there are different numbers of patients per treatment group, then the Fisher's information will be different for each treatment group. This contrasts with the original design under (Stallard, 2009) that uses equal treatment assignments between treatment groups. Similar to the results from Pocock and Simon's randomization, the power under stratified permuted block randomization did not reach the same level as with simple randomization. However, the power under this method was slightly higher compared to Pocock and Simon's method.

CONCLUSION

As shown in above, Type I error rates do not appear to be conserved when randomization is performed according to Pocock and Simon's randomization or stratified permuted block randomization. Further analysis needs to be performed that incorporates these two methods to understand how to control Type I error rates with covariate adaptive randomization.

Future analysis of Pocock and Simon's method could be to analyze the randomization scheme under different covariate weights, or to include Bootstrap sampling to estimate the variance of the test statistic. If this method controls control Type I error rates at the prespecified level, then this method would be useful to minimize the imbalance between treatment groups in a trial that considers more than two treatment groups, over multiple phases.

Using Bootstrap sampling to estimate the variance of the long-term endpoint test statistic could be a way to better estimate the Type I error at the prespecified level of 0.025. Although there is a closed-form solution for the variance of the long-term endpoint test statistic Z_S , the formula assumes that treatment group sizes equal across all treatments. Therefore, using a bootstrap estimate of variance might result in better construction of stopping boundaries and test statistics.

There are many other opportunities for future studies. Three treatment groups and a control group were tested in this analysis; however, different numbers of treatment groups could be tested in future analysis to determine the effect on Type I error. In addition, this method assumes that only one treatment continues past the first interim analysis. In future studies, more than one treatment group could continue to the next phase of the trial. Future analysis could test the results of this method under a wider range of correlations, parameter settings, and covariates. In Pocock and Simon's randomization, there are multiple ways to calculate the amount of variation and the total amount of imbalance between treatment groups. Although a range, and non-weighted sum were chosen in this analysis, other methods could be used to determine the amount of imbalance and determine the next treatment group. In stratified permuted block randomization, a random block size could be used to determine the effect on Type I error.

R-packages that follow the designs in Stallard and Todd (2005), Stallard (2010), and Chow, Lu, and Tse (2007) are available from the author.

ETHICAL AND SAFETY CONSIDERATIONS

There are no ethical concerns and no concerns about keeping records confidential. No data will be collected from human subjects since this is a simulated study.

REFERENCES

- Chow, S., Lu, Q., & Tse, S. (2007). Statistical analysis for two-stage seamless design with different study endpoints. *Journal of Biopharmaceutical Statistics*, 17(6), 1163-1176.
- Engel, B., & Walstra, P. (1991). Increasing precision or reducing expense in regression experiments by using information from a concomitant variable. *Biometrics*, 47(1), 13-20.
- Kahan, B. C., & Morris, T. P. (2012). Improper analysis of trials randomised using stratified blocks or minimisation. *Statist. Med.*, 31, 328-340.
- Pocock, S. J., & Simon, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*, 31(1), 103-115.
- Stallard, N. (2010). A confirmatory seamless phase II/III clinical trial design incorporating short-term endpoint information. *Statist. Med.*, 29, 959-971.
- Stallard, N., & Todd, S. (2003). Sequential designs for phase III clinical trials incorporating treatment selection. *Statist. Med*, 22, 689-703.
- Susan, T., & Nigel, S. (2005). A new clinical trial design combining phases 2 and 3: Sequential designs with treatment selection and a change of endpoint. *Drug Information Journal*, 39(2), 109-118.
- Wei, M., Feifang, H., & Lixin, Z. (2015). Testing hypotheses of covariate-adaptive randomized clinical trials. *Journal of the American Statistical Association*, 110(510), 669-680.

Whitehead, J. (1997). *The design and analysis of sequential clinical trials*. (Second ed.)
Chichester, England Wiley.