Approximation For Response Adaptive Designs Using Stein's Method

by

© **Anita Shahrokhian**

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of Master of Science.

Department of Mathematics and Statistics

Memorial University

September 2020

St. John's, Newfoundland and Labrador, Canada

Abstract

Stein's method introduced by Charles Stein (1972) is a powerful tool in distributional approximation, especially in classes of random variables that are stochastically dependent. In recent years, researchers have concentrated more on adaptive designs. For the response adaptive randomization procedures, the patient's allocation depends on the aggregated information that is acquired from the responses of the previously treated patients. This design uses the information of patients' responses to modify treatment allocation in order to assign more patients to a successful treatment, thus introduce dependent structure in the data. In this thesis we investigate the use of Stein's method in statistical inference for response adaptive design. We have acquired asymptotic normality of the maximum likelihood estimators for treatment effects by deriving an upper bound for these estimators using Stein's method. We examine the performance of three types of response adaptive designs under various success probabilities through simulation studies. Since adaptive designs generate a dependent sequence of random variables that are not exchangeable, we present the advantage of using bootstrap re-sampling in adaptive designs and the efficiency of this method. We compare bootstrap confidence intervals with the asymptotic confidence interval under different success rates of three allocation methods. Also, we discuss the normal approximation based on the Wald's statistic in the numerical studies.

I dedicate this work to the victims of Ukraine Flight PS752

Lay summary

Stein's method introduced by Charles Stein (1972) is a powerful tool in proving well-known Central Limit Theorems for complex dependent problems and distributional approximation. The method later expanded to other targeted distributions, but was first used on normal approximation.
In recent years, researchers have concentrated more on adaptive designs. For the response adaptive randomization procedures, the patient's allocation depends on the aggregated information that is acquired from the responses of the previously treated patients. This design uses the information of patients' responses to modify treatment allocation in order to assign more patients to a successful treatment, thus introduce dependent structure in the data. In this thesis we investigate the use of Stein's method in statistical inference for response adaptive design. We have acquired asymptotic normality of the maximum likelihood estimators for treatment effects by deriving an upper bound for these estimators using Stein's method.

There are various allocation procedures that can be used in adaptive design here, so we investigate the performance for three allocations of response adaptive designs under different success rates in our numerical studies for two treatments. Since adaptive design generates a dependent sequence of random variables that are not exchangeable, we present the advantage of using bootstrap re-sampling in adaptive design and the efficiency of this method. We compare bootstrap versus asymptotic design under different success rates for the three methods of allocations.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Dr. Yanqing Yi for her patience, enthusiasm, motivation and guidance. I would like to thank my committee member Dr. Hong Wang for his support and advice. I gratefully acknowledge the financial support from Dr. Yi's NSERC research grant and my graduate fellowship from the school of Graduate studies as well as the teaching assistantship provided from the Department of Mathematics and Statistics of the Memorial University of Newfoundland.

Table of contents

List of tables

List of figures

Chapter 1

Introduction

In recent years, several authors and researchers investigated Response Adaptive Designs (RAD) for clinical trials because of the advantages over the other methods of randomization. Such designs create a dependency structure on the collected data for RAD because allocation operates on the previously treated assignments and their success in responses. Designs of this type assign more patients to potentially better treatment. This thesis seeks to find approximation for response adaptive design using Stein's method. To understand the context motivating this objective, we start by describing what a clinical trial is and why we need randomization in the procedure. Subsequently, we describe adaptive design and response adaptive-design. Also, we introduce the likelihood function to estimate the parameters. Moreover, we introduce the aim of this study, which is using Stein's method for the dependent data from RAD. This distributional approximation method is applied to handle the difficulty caused by the dependence structure and complexity of the adaptive design.

1.1 Response Adaptive Designs For Randomized Clinical Trial

The advantages of using response adaptive designs in clinical trials are introduced in this section. Clinical trials are the prospective comparison of two or more treatments, one or more of which is a new innovation being tested, while the others are controls based on Rosenberger and Lanchin (2002) [13]. There exist several stages and phases for the approval of clinical trials. From Rosenberger and Lanchin (2002) [13], we learn that in a Phase III clinical trial, the treatment tests on many patients, for treatment approval. A randomized clinical trial is a clinical trial in which patients are randomly allocated to groups with different treatments. There are two critical properties for randomization. The first property is the comparability among the study groups. This comparability can only be used to mediate studies in which the covariates are observable and adjustable, however, there are no guarantees or assurances regarding other covariates. Also, randomization develops a high probability of comparability for unknown covariates. The second property is that the process causes a probabilistic basis for inference if considering all possible results. Overall, the purpose of randomization is to achieve comparability without guarantee and to prevent bias in allocating subjects to a treatment group or avoid predictability. Therefore, it is necessary to use randomization to prevent bias and covariate imbalances.

Randomization in RAD applies on the best allocation procedures based on the previously treated assignments and their success in responses. Hu and Rosenberger (2006)[8] studied asymptotic properties of allocation proportions and introduced five types of randomization for adaptive design processes: complete randomization, restricted randomization, response-adaptive randomization, covariate-adaptive randomization and CARA or covariate-adjusted response-adaptive randomization. There are two primary objectives, one is maximizing the power, and the other is minimizing the number of failure of treatments allocations. From Hu and Rosenberger

(2006) [8] " the response adaptive randomization procedures have three defining characteristics: (1) they are myopic (2) They are fully randomized (3) They require a fixed sample". Although the dependency structure exists in adaptive design and the traditional methods are used for independent data more frequently, it is not possible to apply them without modification. Moreover, response adaptive design was developed to assign more patients to a better treatment. Also, they should be viewed in context of sequential analysis; such randomization causes the dependent samples.

1.2 The Estimation for Response Adaptive Designs

The maximum likelihood estimator properties following a RAD presented and discussed in this section. Generally, if the allocation proportion converges to a constant when $n \to \infty$, the maximum likelihood estimator has asymptotic properties similar to those in i.i.d sequences. Hu and Rosenberger(2006) [8] introduced in a general theorem on the asymptotic properties of ML estimators. This theorem was first proved by Rosenberger, Flrounoy, and Durham (1997)[14] for K treatments, but their conditions were more restrictive and can not be applied to the different types of response adaptive randomization. However, the theory in Hu and Rosenberger(2006) [8] only requires that the allocation proportion converges to a constant. The Taylor expansion is one of the standard ways of proving the ML estimators' asymptotic normality. Yi and Wang(2007)[21] obtained the likelihood function for $K > 2$ using the transition probability of the stochastic process and applied the strong law of large numbers for the Martingale. They used the Taylor expansion to show how under some the regularity conditions, the ML estimators are strongly consistent and asymptotically normally distributed. Melfi et al. (2001)[12] worked on the same properties, but their allocation limitation was only limited by the target allocation proportion. However, the allocation rule provided by Yi and Wang (2007) [21] considered a wide class of adaptive designs such as randomized play the winner (RPW) and Melfi's optimal design, the generalized Polya's urn (GPU) model and finally doubly adaptive biased coin design in Rosenberger et al. (2001) [16].

Suppose that patients arrive sequentially to the trials and receive only treatment A or B. The patients responses X_{1l}, X_{2l},.... from treatment l are independent and identically distributed with probability distribution function $f_l(X, \mu_l)$, where $l = A, B$ and $\mu_l \in \Theta$,and Θ is an open subset of $\mathbb{R}^{\mathrm{k}}$, for a positive integer k. Denote $\mu = (\mu_A, \mu_B)$. Let $X_j = (X_{jA}\mathbb{1}_{jA}, X_{jB}(1 - \mathbb{1}_{jA}))$ be the corresponding response for the patient j. Now, $\mathbb{1}_{jA} = 1$ means that the j^{th} patient receives treatment A, and i.e.

$$\mathbb{1}_{jA} = \begin{cases} 1 & \text{if treatment } A \\ 0 & \text{if treatment } B \end{cases} \tag{1.1}$$

For each observed sequence $\{(\mathbb{1}_{1A}, X_1), ..., (\mathbb{1}_{(j-1)A}, X_{(j-1)})\}$ under $\pi = \{\pi_j, j = 1, 2, ..\}$, the allocation probability π_j for j^{th} patient in a response adaptive design depends on the previously treated patients responses $X_1, X_2, ..., X_{j-1}$, where $\pi_j = P(\mathbb{1}_{jA}|F_{j-1})$ for $j \geq 2$; F_{j-1} is the σ-algebra generated by the observed sequence; and $\pi_1 = P(\mathbb{1}_{1A} = 1)$ is 1/2 for the first two patients. From Hu and Rosenberger (2006) [8] and Yi and Wang (2007) [21], the likelihood function can be formed by,

$$L(\mu) = \prod_{j=1}^{n} \pi_j^{\mathbb{1}_{jA}} (1 - \pi_j)^{(1-\mathbb{1}_{jA})} \prod_{j=1}^{n} f_A(X_{jA}, \mu_A)^{\mathbb{1}_{jA}} f_B(X_{jB}, \mu_B)^{(1-\mathbb{1}_{jA})}, \tag{1.2}$$

The number of patients who are allocated to treatment A or B is $N_A(n)$ and $N_B(n) = n - N_A(n)$, respectively, and their randomization depends on the adaption process of treatment allocation. Assume that $\frac{N_A(n)}{n} \to \rho(\mu)$(a.s) and $\rho(\mu) \in (0, 1)$. Let $l(\mu) = lnL(\mu)$ and $\hat{\mu}_l(n)$ be the solution of $\frac{\partial l}{\partial \mu_l} = 0$, when n patients are allocated by the adaptive design. Assume the regular condition for $f_l(X, \mu_l), l = A, B$ and the eighth moments of responses from treatment A and B and the second moment of the likelihood function to exist and be finite. Rosenberger and Lachin (2002)[13], Reseberger and Hu (2006)[8] and, Yi and Wang (2007)[21] studied consistency and asymptotic normality of the ML estimators. Yi and Li (2018) [20] obtained the rate of convergence error probability of the confidence interval, and they proved that, when the density function is normal, the convergence error and type I error rate is at the order of n^{-1}.

1.3 Stein's Method

Stein's method for normal approximation first appeared in Charles Stein (1972) [18]. In that paper, Stein came up with a new idea to prove the normal approximation to the distribution of a sum of dependent random variables by bounding distance between two random variables, W and Z, where Z follows a standard normal distribution, even when the condition of independence does not hold. Subsequently, Chen (1975) [1] established this characterization for Poisson distribution. Stein (1986) [17] developed these approaches for exchangeable pairs using binomial and Poisson distribution and other probability distributions additional to normal distribution. Furthermore, these characterizations and ideas were later discussed in Chen, Goldstein, and Shao (2005) [3] and in their book (2010) [2].

Recently, Ley, Reinert, and Swan (2014)[10] worked on the canonical definition of the Stein operator and Stein class of distributions, and they presented the comparison of several pairs of distribution. Ley, Reinert, and Swan (2017) [11] worked on the same issue but with a new generalization on Stein's method for univariate distributions. They introduced a canonical definition of Stein's operator of a probability distribution based in a linear differential operator, and applied Stein's identity to both discrete and continuous distributions. They provided an application to compare several pairs of distributions: normal vs. normal, sum of independent Rademacher vs. normal, normal vs Student, and maximum of random variable vs. exponential, Frechet and Gumbel for comparison of the mentioned univariate distribution.

Stein's method can be formed in multivariate approximation using the chi-square distribution characteristics. Gaunt and Reinert (2016) [6] obtained the upper bounds for the rate of convergence of some asymptotical chi-square distributed statistics at the order n^{-1} for the smooth test function. They estimated the bound to Friedman's statistic for comparison of statistical power. Gaunt, Pickett, and Reinert (2017) [5] expanded their research for chi-square approximation by Stein's method with application to Pearson's statistics. They derived the bound from the solution of the gamma Stein equation. Furthermore, they worked on an approximation to estimate the distance between Pearson's statistics and limiting chi-square distribution for the smooth test function.

1.4 My Thesis

The main objective of this thesis is to find the approximation for the RAD inference using Stein's method. For the response-adaptive randomization procedures, the patients' allocation depends on the aggregated information that is acquired from the responses of the previously treated patients. This design uses the information of patients' responses to modify treatment allocation in order to assign more patients to a successful treatment. Due to the dependent structure, such designs are more complex and challenging. We describe Stein's method and some of the practical applications with this methodology, which is a useful tool in distributional approximation. In Chapter 2 we introduce the K function method in Stein's equation for independent random variables. In Chapter 3, using Stein's method, we show the asymptotic normality of the maximum likelihood estimators for the response adaptive design. In Chapter 4, we conduct a simulation study.

Chapter 2

Detailed Illustration of Stein's Method

The goal of this chapter is to introduce Stein's method. Here we begin with the derivation of Stein's identity and obtain the solutions to Stein's equation. At the end of this chapter, we discuss a typical application of this method using the K function approach.

2.1 Fundamental of Stein's Method

One of the most important theorems for large sample sizes is the Central Limit Theorem (CLT). A classic form of the CLT states that a normal approximation applies to the distribution of quantities that can be modeled as a sum of the many independent contributions. However, in response adaptive design, we have dependency in the collected data. Therefore, we need a method for dependent data. Stein's method helps us to obtain a limiting distribution for such data.

2.1.1 Stein's Method

The characterization of the normal distribution presented for the first time by Charles Stein [18] was based on the fact that, $Z \sim N(0, \sigma^2)$ if and only if,

$$E[Zf(Z)] = \sigma^2 E[f^{'}(Z)] \tag{2.1}$$

for all absolutely continuous functions $f : \mathbb{R} \to \mathbb{R}$ for which the expectation exists with $E|f^{'}(Z)| \leq \infty$.

To show a random variable W has a distribution close to a target distribution, say that of random variable Z, we can compare the value of expectation on some class function of W and Z. From Chen, Goldstein and Shao (2010) [2] the goal is to estimate closeness of the distribution of W and Z given by the evaluation of the difference between $Eh(W)$ and $Eh(Z)$ over some collection of measurable function of h. If the distribution of W is close to the distribution Z, then the difference $Eh(W) - Eh(Z)$ should be small for a collection of measurable function h.
As a special case when $\sigma = 1$ in (2.1), $Z \sim N(0, 1)$ if and only if

$$E[f^{'}(Z) - Zf(Z)] = 0 \tag{2.2}$$

for all absolutely continuous function $f : \mathbb{R} \to \mathbb{R}$ for which the expectation exist $E|f^{'}(Z)| \leq \infty$.

If the distribution of W is close to the distribution of Z, then evaluating the left-hand side of (2.2), when Z is replaced by W, deduces something small. Hence, putting these two differences together using the Stein's characterization (2.2) we have Stein's equation given by,

$$f^{'}(w) - wf(w) = h(w) - Eh(Z) \tag{2.3}$$

where Z has a standard normal distribution.

The results (2.1) and (2.3) were introduced and (2.2) proved in Lemma 2.1 from Chen, Goldstein and Shao (2010) [2].

2.1.2 Solution to Stein's Equation

Equation (2.3) is a general form of Stein's equation. From Lemma 2.1 in Chen, Goldstein and Shao (2010) [2], for fixed $z \in R$ and $\Phi(x) = P(Z \leq z)$, the unique bounded solution $f(w) := f_z(w)$ of equation

$$f^{'}(w) - wf(w) = \mathbb{1}_{\{w \leq z\}} - \Phi(z) \tag{2.4}$$

is given by

$$f_z(w) = \begin{cases} \sqrt{2\pi} e^{\frac{w^2}{2}} \Phi(w)[1 - \Phi(z)] & \text{if} \quad w \leq z \\ \sqrt{2\pi} e^{\frac{w^2}{2}} \Phi(z)[1 - \Phi(w)] & \text{if} \quad w > z. \end{cases} \tag{2.5}$$

The solution (2.5) to the equation (2.4) is a special case of Stein's characterization for $h(w) = \mathbb{1}_{\{w \leq z\}}$. By the same method of integrating, a general solution for (2.3) is given by,

$$\begin{aligned} f_h(w) &= e^{\frac{w^2}{2}} \int_{-\infty}^{w} (h(x) - Nh) e^{\frac{-x^2}{2}} dx \\ &= -e^{\frac{w^2}{2}} \int_{w}^{+\infty} (h(x) - Nh) e^{\frac{-x^2}{2}} dx. \end{aligned} \tag{2.6}$$

where $Nh = Eh(Z)$.

(2.5) and (2.6) are the special case and general solutions to the Stein's equations (2.4) and (2.3), respectively (Chen, Goldstein and Shao (2010) [2]).

2.1.3 Boundary Conditions For the Solution of Stein's Equation

We describe some of the results on boundaries for the solution (2.5) and (2.6) to the Stein's equation (2.4) and (2.3), respectively, in this section.

The following results are from Lemma 2.3 in Chen, Goldstein, and Shao (2010) [2]. First, $wf_z(w)$ is an increasing function of w.

Moreover, for all real w , u and v,

$$|wf_z(w)| \leq 1, \qquad |wf_z(w) - uf_z(u)| \leq 1 \tag{2.7}$$

$$|f_z^{'}(w)| \leq 1, \qquad |f_z^{'}w) - f_z^{'}(u)| \leq 1 \tag{2.8}$$

$$0 \leq f_z(w) \leq \min\left(\frac{\sqrt{2\pi}}{4}, \frac{1}{|z|}\right) \tag{2.9}$$

$$|(w+u)f_z(w+u) - (w+u)f_z(w+v)| \leq \min\left(|w| + \frac{\sqrt{2\pi}}{4}\right)|u| + |v| \tag{2.10}$$

In addition, from Lemma 2.4 in Chen, Goldstein, and Shao (2010) [2] for any real valued function h on $\mathbb{R}$, if h is bounded, then

$$||f_h|| \leq \sqrt{\frac{\pi}{2}}||h(.) - Nh|| \tag{2.11}$$

$$||f_h{'}|| \leq 2||h(.) - Nh|| \tag{2.12}$$

$$||f_h|| \leq ||h^{'}|| \tag{2.13}$$

$$||f_h^{'}|| \leq \sqrt{\frac{2}{\pi}}||h^{'}|| \tag{2.14}$$

$$||f^{''}|| \leq 2||h^{'}|| \tag{2.15}$$

where,

$$||h|| = sup_{x\in \mathbb{R}}|h(x)|. \tag{2.16}$$

For Lemma 2.4 in Chen, Goldstein, and Shao (2010) [2], the properties (2.13), (2.14) and (2.15) holds if h is absolutely continuous.

2.2 Illustration of The Use of Stein's Method

There are four different approaches presented for handling Stein's equation in Chen, Goldstein and Shao (2010)[2]. The first approach, which plays a key role in our study, is the K function method when W is a sum of independent random variables. The other well-known method called the exchangeable pair approach of Stein's, while W has a particular dependency structure. Additionally, they have discussed zero bias distribution (the associated transformation for arbitrary mean zero with finite variance), which is called the zero bias method. Finally, there is the size bias transformation. This method and zero bias transformations are close to each other, but size bias is defined on the class of non-negative random variables x with finite non-zero means.

As one of the ordinary applications of Stein's method from Chen, Goldstein and Shao (2010)[2] in the Section 2.3.1, the K function approach is to handle Stein's equation (2.3) for the sum of the independent random variables.

Let W be the sum of the independent random variables $\xi_1, ..., \xi_n$, where $E(\xi_i) = 0$, $Var(\xi_i) = \frac{1}{n}$, and $E|\xi_i|^3 < \infty$ for $i = 1, ..., n$.

Define,

$$K_i(t) = E\xi_i\{\mathbb{1}_{\{0\le t\le \xi_i\}} - \mathbb{1}_{\{\xi_i \le t<0\}}\} \tag{2.17}$$

From Chen, Goldstein and Shao (2010)[2], in the Section 2.3.1 (page 19), since $K_i(t) \geq 0$ for all real t it can be shown that,

$$\int_{-\infty}^{+\infty} K_i(t)dt = E\xi_i^2 \quad and \quad \int_{-\infty}^{+\infty} |t|K_i(t)dt = \frac{1}{2}E|\xi_i|^3. \tag{2.18}$$

Let h be a measurable function with $E|h(Z)| < \infty$, and $f = f_h$ be the corresponding solution of the Stein's equation (2.3). The goal is to estimate the left-hand side of the following equation,

$$E[f'(W) - Wf(W)] = Eh(W) - Nh. \tag{2.19}$$

This means if we consider the right-hand side of (2.19) for some large class of function h, the expectations differences should be small if the distribution of W approximates to that of Z. Since the right-hand side of (2.19) contains two random variables and working with this side is not easy to proceed, working with the left-hand side of the equation is recommended.

To use the K function (2.17) in Stein's method, first of all, estimate $E[Wf(W)]$. Considering the definition of W we have,

$$\begin{aligned} E[Wf(W)] &= \sum_{i=1}^{n} E\Big[\int_{-\infty}^{+\infty} f'(W^{(i)}+t)\ \xi_i(\mathbb{1}_{\{0\le t\le \xi_i\}} - \mathbb{1}_{\{\xi_i\le t<0\}})dt\Big] \\ &= \sum_{i=1}^{n}\int_{-\infty}^{+\infty} E\big[f'(W^{(i)}+t)\big]K_i(t)dt. \end{aligned} \tag{2.20}$$

Since,

$$\sum_{i=1}^{n}\int_{-\infty}^{+\infty} K_i(t)dt = \sum_{i=1}^{n} E\xi_i^2 = 1 \tag{2.21}$$

it follows that

$$Ef^{'}(W) = \sum_{i=1}^{n} \int_{-\infty}^{+\infty} E\big[f^{'}(W)\big] K_i(t)dt. \tag{2.22}$$

Therefore,

$$E\big[f^{'}(W) - Wf(W)\big] = \sum_{i=1}^{n} \int_{-\infty}^{+\infty} E\big[f^{'}(W) - f^{'}(W^{(i)} + t)\big] K_i(t)dt. \tag{2.23}$$

Equations (2.20), (2.23) play a vital role in finding the corresponding upper bound to prove normal approximation. Also, (2.20), (2.23) hold for all bounded absolutely continuous function f. The normal approximation can be obtained from the bounds on the solution f in Lemma 2.4 in Chen, Goldstein, and Shao (2010) [2], and using mean value theorem, it can be proven

$$\lim_{n\to\infty} E|f^{'}(W) - Wf(W)| \to 0. \tag{2.24}$$

As well as the right-hand side of (2.19),

$$\lim_{n\to\infty} E|h(W) - h(Z)| \to 0. \tag{2.25}$$

This leads us to deduce that $W \xrightarrow{d} Z$ by the Kolmogrov distance properties,

$$||L(X) - L(Y)||_H = \sup_{h\in H} |Eh(X) - h(Y)|. \tag{2.26}$$

where,

$$H = \{\mathbb{1}_{(x\leq z)}, z \in \mathbb{R}\}.$$

Chapter 3

Normal Approximation Using Stein's Method In Response Adaptive Designs

In this chapter, we apply Stein's method using the K function approach in the inference of RAD and discuss the asymptotic normality of the maximum likelihood estimators. In Section 3.1, we introduce the notation and in Sections 3.2 and 3.3 we thoroughly discuss the application of the K function approach in Stein's method for the RAD inference.

3.1 The Maximum Likelihood Estimator for RAD

In this section, we introduce the MLE for RAD. We assume similar assumptions as those in Yi and Li(2018) [20] and Rosenberger et al.(2001) [16]:

1. The parameter space Θ is an open subset of $\mathbb{R}^k$ and k is a positive integer for treatment A and treatment B.
2. The distribution of $f_l(X, \mu_l), l = A, B$ belongs to exponential family.

3. For limiting allocation $\rho(\mu) \in (0,1)$, $\frac{N_A}{n} \to \rho(\mu)$ almost surely (a.s), where $0 < \rho(\mu) < 1$.

4. The moments of responses from treatment A and B at the order of one to three and the second moment of the likelihood function exist and are finite.

For the response adaptive randomization procedures, the patients' allocation depends on the aggregated information that is acquired from the responses of the previously treated patients. This design uses the information of patients' responses to modify treatment allocation, in order to assign more patients to a successful treatment. Such design creates dependency in the collected data for RAD.

From Hu and Rosenberger (2006) [8], Rosenberger et al.(1997) [14], Yi and Wang (2007) [21] the likelihood estimators for the mean responses for treatment A and B are given by,

$$\hat{\mu}_A = \frac{\sum_{j=1}^{n} \mathbb{1}_{jA} X_{jA}}{\sum_{j=1}^{n} \mathbb{1}_{jA}} \tag{3.1}$$

$$\hat{\mu}_B = \frac{\sum_{j=1}^{n} (1 - \mathbb{1}_{jA}) X_{jB}}{\sum_{j=1}^{n} (1 - \mathbb{1}_{jA})}. \tag{3.2}$$

Define,

$$W_A = \sum_{j=1}^{n} \mathbb{1}_{jA} \xi_{jA}, \tag{3.3}$$

where $\xi_{jA} = \frac{(X_{jA} - \mu_A)}{\sqrt{n\rho}\sigma_A}, j = 1, ..., n$. It is clear that $E\xi_{jA} = 0$ and $Var(\xi_{jA}) = E\xi_{jA}^2 = \frac{1}{n\rho}$.

Similarly, define

$$W_B = \sum_{j=1}^{n} (1 - \mathbb{1}_{jA}) \xi_{jB}, \tag{3.4}$$

where $\xi_{jB} = \frac{(X_{jB} - \mu_B)}{\sqrt{n(1-\rho)}\sigma_B}, j = 1, ..., n$. Therefore, $E\xi_{jB} = 0$ and $Var(\xi_{jB}) = E\xi_{jB}^2 =$

$\frac{1}{n(1-\rho)}$.

In the next section, we apply the Stein's method to obtain normal approximation for W_A and W_B.

3.2 Normal Approximation of MLE for RAD

In this section, we use Stein's method to prove the asymptotic normality of MLE for RAD. We establish the asymptotic properties for W_A and W_B first.

3.2.1 Setup For the K Function Method

We define the K functions as follows:

$$K_{iA}(t) = E[\xi_{iA}\{\mathbb{1}_{\{0\le t\le \xi_{iA}\}} - \mathbb{1}_{\{\xi_{iA}\le t<0\}}\}]. \tag{3.5}$$

Following the same steps as those in Chen, Goldstein and Shao (2010)[2] (page 19), it can be proven that, $K_{iA}(t) \ge 0$ for all real t and

$$\int_{-\infty}^{+\infty} K_{iA}(t)dt = E\xi_{iA}^2 \quad \text{and} \quad \int_{-\infty}^{+\infty} |t|K_{iA}(t)dt = \frac{1}{2}E|\xi_{iA}|^3. \tag{3.6}$$

Define the corresponding K function for W_B as

$$K_{iB}(t) = E[\xi_{iB}\{\mathbb{1}_{\{0\le t\le \xi_{iB}\}} - \mathbb{1}_{\{\xi_{iB}\le t<0\}}\}]. \tag{3.7}$$

We have that $K_{iB}(t) \ge 0$ for all real t as well as

$$\int_{-\infty}^{+\infty} K_{iB}(t)dt = E\xi_{iB}^2 \quad and \quad \int_{-\infty}^{+\infty} |t|K_{iB}(t)dt = \frac{1}{2}E|\xi_{iB}|^3. \tag{3.8}$$

3.2.2 Normal Approximation Using Stein's Method for $\mathbf{W}_A$

Here, we prove asymptotic normality of the maximum likelihood estimators using Stein's method. Throughout the process, we assume the listed regularity conditions are satisfied; First, we obtain normal approximation for W_A using Stein's method.

Lemma 1. *W_A has mean zero and variance $\sigma^2_{W_A} = Var(W_A) = E[\frac{N_A}{n\rho}]$.*

Similarly,

Lemma 2. *W_B has mean zero and variance $\sigma^2_{W_B} = Var(W_B) = E\left[\frac{N_B}{n(1-\rho)}\right]$.*

The details of the proofs of the Lemma 1 and Lemma 2 can be found in the Appendix A.

From the Lemma 1 we can state the following theorem.

Theorem 3. *Let h be a measurable function with $||h^{'}|| < \infty$, and $\frac{N_A}{n} \to \rho$ a.s. as $n \to \infty$. If $E|\xi_{iA}|^3 < \infty$ for $i = 1, ..., n$, then*

$$(1)\ E|h(W_A) - h(Z)| \leq \sqrt{\frac{2}{\pi}}||h^{'}||E|1 - \frac{N_A}{n\rho}| + 3||h^{'}||E\big[\sum_{i=1}^{n} \mathbb{1}_{iA} E|\xi_{iA}|^3\big];$$

$$(2)\ W_A \xrightarrow{d} Z, \text{as } n \to \infty, \text{ where } Z \text{ has a standard normal distribution.}$$

The normal approximation in Theorem 3 can be obtained by verifying the following steps. First, we use the K function approach in Stein's equation by defining the leave-one-out in the summation of the random variables. Second, we estimate the left-hand side of (2.19). Then, we derive an upper bound for the estimation we get in left-hand side of (2.19), and the result is valid for the right-hand side of the equation. Finally, we show that the upper bound converge to zero as $n \to \infty$ and $W_A \xrightarrow{d} Z$. The details of the proof of the Theorem 3 can be found in the Appendix A.

Since,

$$\begin{aligned} \frac{n\rho\, W_A}{N_A} &= \frac{n\rho\ \sum_{j=1}^{n} \mathbb{1}_{jA}(X_{jA} - \mu_A)}{N_A\ \sqrt{n\rho}\sigma_A} \\ &= \frac{\sqrt{n}\ \sum_{j=1}^{n} \mathbb{1}_{jA}(X_{jA} - \mu_A)}{N_A\ \sqrt{\rho^{-1}}\sigma_A} \\ &= \frac{\sqrt{n}(\hat{\mu}_A - \mu_A)}{\sqrt{\rho^{-1}}\sigma_A}, \end{aligned} \tag{3.9}$$

and $\frac{N_A}{n} \to \rho$ a.s as $n \to \infty$, we have the following using the Slutsky's theorem.

Theorem 4. *Under regularity conditions 1-3,* $\sqrt{n}(\hat{\mu}_A - \mu_A) \to N(0, \rho^{-1}\sigma_A^2)$.

Similarly, we have the following results for W_B and $\hat{\mu}_B$.

Corollary 5. *Let h be a measurable function with $||h^{'}|| < \infty$, and $\frac{N_B}{n} \to (1-\rho)$ a.s. as $n \to \infty$. If $E|\xi_{iB}|^3 < \infty$ for $i = 1, ..., n$, then*

(1) $E|h(W_B) - h(Z)| \leq \sqrt{\frac{2}{\pi}}||h^{'}||E|1 - \frac{N_B}{n(1-\rho)}| + 3||h^{'}||E\big[\sum_{i=1}^{n}(1 - \mathbb{1}_{iA})E|\xi_{iB}|^3\big]$;

(2) $W_B \xrightarrow{d} Z$ *as* $n \to \infty$, *where Z has a standard normal distribution.*

Corollary 6. *If $\frac{N_B}{n} \to (1-\rho)$ a.s. as $n \to \infty$ then,* $\sqrt{n}(\hat{\mu}_B - \mu_B) \to N(0, (1-\rho)^{-1}\sigma_B^2)$.

3.3 Normal Approximation Using the Stein's method for $(\hat{\mu}_A - \hat{\mu}_B)$

In the previous section, we prove the asymptotic normality of the MLE for RAD, using the K function approach in Stein's method. Here we discuss the same process for $W_A - W_B$ and obtain the asymptotic properties for $\hat{\mu}_A - \hat{\mu}_B$.

Define,

$$\begin{aligned} W_{rad} &= W_A - W_B \\ &= \sum_{j=1}^{n} \mathbb{1}_{jA}\xi_{jA} - \sum_{j=1}^{n}(1-\mathbb{1}_{jA})\xi_{jB} \\ &= \sum_{j=1}^{n} \big\{\mathbb{1}_{jA}\xi_{jA} - (1-\mathbb{1}_{jA})\xi_{jB}\big\}. \end{aligned} \tag{3.10}$$

Lemma 7. *W_{rad} has mean zero and variance $\sigma^2_{W_{rad}} = Var(W_{rad}) = E[\frac{N_A}{n\rho}] + E[\frac{N_B}{n(1-\rho)}]$.*

The details of the proof of the Lemma 7 can be found in the Appendix A.

Now to discuss the normal approximation for W_{rad} using the following theorem, $\frac{N_B}{n} \to (1-\rho)$ a.s. as $n \to \infty$. If $E|\xi_{iB}|^3 < \infty$ for $i = 1, ..., n$, then, we have the following theorem.

Theorem 8. *Let h be a measurable function with $||h^{'}|| < \infty$, and $\frac{N_A}{n} \to \rho$ a.s. as $n \to \infty$. If $E|\xi_{iA}|^3 < \infty$ and $E|\xi_{iB}|^3 < \infty$ for $i = 1, ..., n$, then*

$$\begin{aligned} (1)\ E|h(W_{rad}) - h(Z)| \le &\sqrt{\frac{2}{\pi}}||h^{'}||E|1 - \frac{N_A}{n\rho}| + 3||h^{'}||E\big[\sum_{i=1}^{n}\mathbb{1}_{iA}\ E|\xi_{iA}|^3\big] \\ &+ \sqrt{\frac{2}{\pi}}||h^{'}||E|1 - \frac{N_B}{n(1-\rho)}| + 3||h^{'}||E\big[\sum_{i=1}^{n}(1-\mathbb{1}_{iA})\ E|\xi_{iB}|^3\big]; \end{aligned}$$

(2) $W_{rad} \xrightarrow{d} Z$ *as* $n \to \infty$, *where Z has a standard normal distribution.*

The approximation in Theorem 8 can be proved by verifying the similar steps as of W_A in the previous subsection. Hence, we first use the K function approach in Stein's method by defining the leave-one-out in the summation of the random variables. Second, we estimate equation (2.19) using the left-hand side. Third, we achieve an upper bound for estimation we get for the left-hand side of (2.19). The same result is valid for the right-hand side of the equation. Ultimately, we show that the upper bound converges to zero as $n \to \infty$ and $W_{rad} \xrightarrow{d} Z$. The details of the proof of the Theorem 8 can be found in the Appendix A.

Similar to the Theorem 4 we have,

Corollary 9. *If* $\frac{N_A}{n} \to \rho$ *a.s. as* $n \to \infty$, *then* $\sqrt{n}((\hat{\mu}_A - \hat{\mu}_B) - (\mu_A - \mu_B)) \to N(0, \rho^{-1}\sigma_A^2 + (1-\rho)^{-1}\sigma_B^2)$.

Chapter 4

Simulation Study

Throughout this chapter, we conduct an extensive numerical study to examine the normal approximation of MLE for RAD. To do so, we considered three adaptive allocation procedures with various success rates. Furthermore, to attain a comprehensive measure of the obtained estimates' accuracy, we utilized Bootstrap sampling under different allocation methods and success probabilities. This chapter is structured in the following order. In Section 4.1, we first explain three different target allocation techniques, and then we outline the simulation setup for capturing the distribution of $\frac{N_A}{n}$. In Section 4.2, we carry out another simulation study to find confidence intervals for the success rates under each of the introduced allocation procedures using the Bootstrap re-sampling method. Finally, in Section 4.3, the results and comparisons of the different techniques are discussed.

4.1 The Adaptive Allocation Methods

In this section, we introduce three adaptive allocation methods for RAD. We begin by introducing the essentials for these allocations. Then we discuss the normal approximation for the RAD inference by using numerical methods.

4.1.1 Randomized Play the Winner Allocation (RPW)

Randomized Play the Winner (RPW) was introduced by Wei and Durham (1978) [19]. Suppose there is a two-treatment clinical trial, and an urn includes one "A" ball and one "B" ball for the first patient. Thus, the first patient is equally likely to receive either of the two treatments A or B. Suppose this patient is assigned to treatment A through randomization. Then, if the treatment is successful, the original "A" ball is placed back to and an additional "A" ball is added to the urn. Otherwise, if the patient fails on treatment A, we put the original "A" ball and a "B" ball in the urn. Thus, the second patient has a probability of 2/3 or 1/3 of receiving treatment A, depending on whether treatment A was a success or a failure for the first patient. We continue this process for the total number of patients in the trial. Due to this allocation, it is possible that a higher proportion of patients will be assigned to the more successful treatment. Hence, we can formulate the probability of assigning patients to treatment A in each step of this allocation process by,

$$\rho_{(A)} = \frac{A_{ball}}{A_{ball} + B_{ball}}. \tag{4.1}$$

It is straightforward to see that the probability of patients being assigned to treatment B can be obtained by $\rho_{(B)} = 1 - \rho_{(A)}$.

4.1.2 RSIHR Allocation

The optimal allocation proposed by Rosenberger, Stallard, Ivanova, Harper, and Ricks (RSIHR) (2001) [7] aims to minimize the expected number of treatment failures concerning the conditional variance of the Wald test statistic at a fixed level. The RSIHR method derives an optimal allocation proportion for binary responses that is independent of the test's power. Following this technique, the target allocation proportion of patients allocated to treatment A is given by,

$$\rho_{(A)} = \frac{\sqrt{p_A}}{\sqrt{p_A} + \sqrt{p_B}}, \tag{4.2}$$

where p_A is success probability on treatment A, and p_B is success probability on treatment B.

4.1.3 YW Allocation

In [22], Yi and Wang (2009), proposed a RAD with enhanced ethical benefits to patients. Such design considers both of the average number of patients allocated to better treatments as well as the power of the statistical test. The target allocation for YW is given by,

$$\rho_{(A)} = \frac{q_B + \epsilon \; min\{q_A, q_B\} sign(q_B - q_A)}{q_A + q_B}, \tag{4.3}$$

where $q_l = 1 - p_l, l = A, B$.

4.1.4 Doubly Biased Coin Design (DBCD)

From Hu and Zhang (2004)[9], when the allocation proportions are unknown for treatments A and B, a biased coin design (e.g., general Eisele biased coin design (1994) [4]) can be used. Here, the allocation function $g(x, \rho)$ plays a dominant role, and is given by,

$$g(x, \rho) = \frac{\rho(\rho/x)^\gamma}{\rho(\rho/x)^\gamma + (1-\rho)((1-\rho)/(1-x))^\gamma}, \tag{4.4}$$

where $\gamma > 0$. It is straightforward to see that, if $x > \rho$, then $g(x, \rho) < \rho$.

Due to the results we obtained in Chapter 3, also from Yi and Li (2018) [20], a $(1-\alpha)100\%$ confidence interval for $(\mu_A - \mu_B)$ is given by,

$$(\hat{\mu}_A - \hat{\mu}_B) \pm z_{(1-\frac{\alpha}{2})} \sqrt{\frac{\hat{\sigma}_A^2}{N_A} + \frac{\hat{\sigma}_B^2}{N_B}}, \tag{4.5}$$

where α is the nominal coverage error probability and $z_{(1-\frac{\alpha}{2})}$ is the critical value for the standard normal distribution.

4.1.5 Simulation Setup

In this subsection, we outline the setup for our numerical study under the three allocation procedures and various success rates. Our goal is to monitor the goodness of the normal approximation discussed in Chapter 3 for different success rates in RAD.

- The adaptive allocation considered are the RPW (1978) [19] and the optimal allocation proportion RSIHR (2001) [16], and the YW (2009) [22] allocation.
- We use the Doubly Biased Coin Design (DBCD) Hu and Rosenberger (2006)[8] with the allocation function $g(x, \rho)$ to target $\rho_{(A)}$ where $\gamma = 100$ is fixed as suggested in Hu, Rosenberger and Zhang (2006) [23] in RSHIR and YW allocation methods.
- We generate data from a uniform distribution $X \sim Unif(0, 1)$ for allocation purposes.
- Here, we consider three different success rates from the set $\{0.5, 0.7, 0.9\}$ as the true value for p_A and the values of p_B are determined such that, $0 \leq |p_A - p_B| \leq 0.3$.
- The total number of patients is fixed at $n \in \{100, 200, 300\}$ and we generate $r = 10,000$ replications for each of the RADs.
- The first two patients are allocated to the treatments A and B, respectively. Then, the next patients assign to each of the treatments by following one of the three RAD procedures.
- In all the numerical studies, we fix epsilon at $\epsilon = 1/4$ for YW adaptive design and, the type I error rate is fixed to the nominal level of $\alpha = 0.05$.

4.2 Bootstrap Resampling for Response Adaptive Design

Bootstrap is a powerful statistical tool that can be used to quantify the uncertainty associated with a given estimator. It resamples from a set of data to create many simulated samples. This type of resampling allows us to estimate standard errors and find the simulated confidence intervals, and apply the hypothesis testing for various test statistics. This methodology can be considered as an alternative approach to traditional estimation; however, we must note some fundamental differences. There are advantages to applying bootstrap resampling since it does not make any assumptions about the distribution of the data and use a wider variety of distributions. Additionally, it is very beneficial to draw inference when the size of the data is not adequately large.

Moreover, the confidence interval deduced by bootstrap is asymptotically consistent. From Rosenberger and Hu (1999) [15], the sequence of responses inferred from adaptive designs is dependent, and it cannot be modified during the process. In their paper, they used bootstrap resampling to obtain confidence intervals. This process uses the output of an adaptive experiment as the input of the bootstrap resampling procedure. It can be utilized for any sample size due to the complex covariance structure on the data in adaptive design.

4.2.1 Simulation Setup for Bootstrap

In Subsection 4.1.5, we describe the simulation setup to compute coverage probabilities and obtain confidence intervals based on asymptotic normal distribution. Here, we outline the setup for the simulating of bootstrap samples in the setting of RAD.

- The bootstrap method is applied to the three allocation methods of RPW, RSIHR, and YW for the total sample size fixed at $n = 200$.

- We consider various settings of true values for the success rates $p_A = (0.9, 0.7, 0.5)$ vs. $p_B = (0.7, 0.5, 0.3)$ such that, $|p_A - p_B| = 0.2$.
- We then find the observed success probabilities and the sample size for $\hat{p} = (\hat{p}_A, \hat{p}_B)$ and $N_l = (N_A, N_B), l = A, B$ under the three allocation methods.
- Using the adaptive allocation rules, we replicate a total of T sequences of treatment allocations and their corresponding responses.
- We estimate $\hat{p}_1^*, ..., \hat{p}_T^*$ and $N_1^*, ..., N_T^*$ as bootstrap estimates of the response probabilities and sample sizes respectively.
- Ultimately, we order $\hat{p}_l^{*1}, ..., \hat{p}_l^{*T}, l = A, B,$ as $\hat{p}_l^{*(1)}, ..., \hat{p}_l^{*(T)}$ to compute the desired quantiles.

Using the outlined steps, we then simply attain the bootstrap confidence intervals with $100(1-\alpha)\%$ confidence level for p_l, $l = A, B$ as follows,

$$(\hat{p_l}^{*(T\alpha/2)}, \hat{p_l}^{*(T(1-\alpha)/2)}). \tag{4.6}$$

Rosenberger and Hu (1999) [15] introduced two other approximations for confidence interval. In the first method, they considered the measure $\hat{p_l}^* - \hat{p_l}$ where $\hat{p_l^*}$ is an individual bootstrap estimate for $l = A, B$. Accordingly, a second confidence interval can be approximated by,

$$(2\hat{p_l} - \hat{p_l}^{*(T\alpha/2)}, 2\hat{p_l} - \hat{p_l}^{*(T(1-\alpha)/2)}), \tag{4.7}$$

where $l = A, B$.

Finally, the third confidence interval approximation method that they introduce is defined by,

$$(\hat{p_l} - \hat{Z_l}^{*(T\alpha/2)}, \hat{p_l} - \hat{Z_l}^{*(T(1-\alpha)/2)}), \tag{4.8}$$

where $Z_l^* = \sqrt{\left(\frac{N_l^* \hat{p_l}\hat{q_l}}{N_l \hat{p_l^*}\hat{q_l^*}}\right)(\hat{p_l^*} - \hat{p_l})}$ and $\hat{Z_l}^{*(T\alpha/2)}$ is the estimates $Z_l^{*1}, ..., Z_l^{*T}$ in an increasing order to obtain $100(1-\alpha)$ per cent bootstrap confidence interval for p_l.

Rosenberger and Hu's (1999) [15] paper shows that the simplest confidence interval method given by (4.6) outperforms its two competitors. Hence, we use this approximation for the bootstrap confidence intervals in our simulation study.

The computation code in R programming can be found in the Appendix B for the allocation RPW, RSHIR and YW using RAD from Yi and Li(2018) [20] and the steps for bootstrap method in RAD from Hu and Rosenberger (1999) [15].

For any of the allocation techniques, we generate one pass of data, and $\hat{p}$ for these two treatments is calculated. Then, concerning the bootstrap process, $T = 500$ replications of the success rate estimates is obtained for each of the allocation procedures. To estimate the coverage probabilities and confidence intervals using the simplest bootstrap method, we use a total of 5,000 repetitions and calculate the number of times that the constructed confidence interval captures the true success rate. We should also note that the DBCD adjustment is employed in the allocation process.

4.3 Results

In this section, we compare the obtained results for the three allocation methods and under different population configurations.

We consider $|p_A - p_B| = 0.3$, $|p_A - p_B| = 0.2$, $|p_A - p_B| = 0.1$ and $|p_A - p_B| = 0$ for different range of p_A and p_B to better monitor the power of the test under various scenarios. Simulated power is computed for two-sided hypothesis testing using Wald's statistic from Yi and Li (2018) and consequently the confidence interval from (4.5). Table 4.1, Table 4.2 and Table 4.3 demonstrate the results for $n = 100$, $n = 200$ and $n = 300$ respectively. The entries regarding the expected value of $\frac{N_A}{n}$, Standard Deviation (SD), Coverage Probability (CP), and statistical power are rounded to two decimal places. We also assumed that treatment A has a higher survival rate relative to treatment B.

Overall, for all the three adaptive designs, the results are in accordance with the superiority of the assigned treatment for the majority of patients. Also , there is a positive association between the statistical power and the magnitude of (p_A-p_B) and its slope is considerably large. Generally, for a fixed sample size, all three methods have relatively the same simulated statistical powers, but as the sample sizes increase, we observe higher values for statistical power.

Additionally, when $n = 300$, statistical power converges to one for the more considerable difference between the two success rates. From the results, we can observe that the standard deviation corresponding to the allocation proportion of RPW is substantially higher than its counterparts, and its value culminates at $(p_A, p_B) = (0.9, 0.7)$.

In Figure.4.1, Figure.4.2 and Figure.4.3, we can see the distribution of $\frac{N_A}{n}$. The distributions of YW has a significant shift to the right under $\epsilon = 1/4$. The results show that under $|p_A - p_B| > 0.1$, we reach the demanded power and accuracy of the coverage probabilities using an adaptive clinical trial.

Considering the bootstrap results in Table 4.4, we notice that for both the bootstrap and observed data, the coverage probabilities are close to the nominal level. The advantage of the simplest bootstrap confidence interval based on repeated copies of the simulated data is more striking in clinical trials. Also, the bootstrap techniques consider mechanisms during the process that can be combined in the data analysis of the desired timateses' sampling distribution. This table presents the simulated coverage probabilities (CP) and the average length of the intervals (L) for RPW, RSHIR, and YW for different true values of $p_A = (0.9, 0.7, 0.5)$ and $p_B = (0.7, 0.5, 0.3)$ for observed data and bootstrap re-sampling from RAD respectively for a total sample size $n = 200$. As we can see under each of these target allocations, the difference between success probabilities p_A, and p_B is set to be 0.2. Table 4.4 shows that when $n = 200$, bootstrap re-sampling and observed data work approximately the same in both coverage probabilities and the confidence interval that shows bootstrap estimation is appropriate in this study as well. Moreover, $n = 200$ is a sufficient sample size to use the asymptotic confidence interval obtained from theoretical results using bootstrap method in RAD.

Additionally, the normal approximation for $(p_A - p_B)$ under RPW, RSHIR, and YW is shown in Figure 4.4 by setting the total size at $n = 200$. Consider the true values for success rates $p_A = (0.9, 0.7, 0.5)$ and $p_B = (0.7, 0.5, 0.3)$ such that $|p_A - p_B| = 0.2$, it is obvious from the results that the estimates are normally distributed but a relatively small number of data points in normally distributed data fall in the few highest and few lowest quantiles, when $(p_A, p_B) = (0.9, 0.7)$ in RPW and YW. Clearly, the figures show that the quantile points lie on the theoretical normal line. We use the histograms in Figure 4.5, to show the frequency of $(\hat{p}_A - \hat{p}_B)$ under RPW, RSHIR, and YW for $n = 200$ for the same success rates as in Figure 4.4.

Table 4.1: Simulation results of the $E(N_A/n)$, S.D, Coverage Probability (CP) and statistical power, n=100, r=10,000

(P_A, P_B)	**RPW**				**RSIHR**				**YW**			
	$E(N_A/n)$	$S.D$	CP	$Power$	$E(N_A/n)$	$S.D$	CP	$Power$	$E(N_A/n)$	$S.D$	CP	$Power$
(0.9,0.6)	0.71	0.12	0.94	0.93	0.55	0.02	0.97	0.97	0.83	0.05	0.95	0.91
(0.9,0.7)	0.66	0.15	0.95	0.74	0.53	0.02	0.95	0.82	0.80	0.08	0.96	0.68
(0.9,0.8)	0.59	0.18	0.96	0.33	0.52	0.01	0.96	0.37	0.71	0.14	0.96	0.29
(0.9,0.9)	0.50	0.21	0.97	0.03	0.50	0.01	0.97	0.04	0.50	0.21	0.98	0.04
(0.7,0.4)	0.65	0.08	0.95	0.89	0.57	0.02	0.95	0.92	0.74	0.04	0.95	0.88
(0.7,0.5)	0.61	0.09	0.98	0.62	0.54	0.02	0.96	0.66	0.70	0.07	0.96	0.60
(0.7,0.6)	0.56	0.10	0.95	0.25	0.52	0.02	0.96	0.28	0.63	0.13	0.95	0.24
(0.7,0.7)	0.50	0.12	0.95	0.05	0.50	0.02	0.95	0.05	0.50	0.16	0.95	0.05
(0.5,0.2)	0.61	0.06	0.95	0.93	0.61	0.04	0.94	0.94	0.70	0.03	0.96	0.91
(0.5,0.3)	0.58	0.07	0.95	0.62	0.56	0.03	0.95	0.64	0.67	0.06	0.95	0.60
(0.5,0.4)	0.54	0.07	0.95	0.25	0.53	0.03	0.95	0.25	0.61	0.11	0.95	0.23
(0.5,0.5)	0.50	0.08	0.95	0.04	0.50	0.02	0.95	0.05	0.50	0.14	0.95	0.05

Table 4.2: Simulation results of the $E(N_A/n)$, S.D, Coverage Probability (CP) and statistical power, n=200, r=10,000

(P_A, P_B)	**RPW**				**RSIHR**				**YW**			
	$E(N_A/n)$	*S.D*	*CP*	*Power*	$E(N_A/n)$	*S.D*	*CP*	*Power*	$E(N_A/n)$	*S.D*	*CP*	*Power*
(0.9,0.6)	0.74	0.09	0.95	0.99	0.55	0.01	0.95	1	0.84	0.03	0.95	0.99
(0.9,0.7)	0.68	0.12	0.94	0.95	0.53	0.01	0.96	0.97	0.81	0.05	0.95	0.94
(0.9,0.8)	0.60	0.15	0.96	0.57	0.51	0.01	0.96	0.62	0.74	0.09	0.94	0.55
(0.9,0.9)	0.50	0.18	0.95	0.04	0.50	0.01	0.96	0.04	0.50	0.20	0.96	0.04
(0.7,0.4)	0.65	0.06	0.95	0.99	0.57	0.02	0.95	0.99	0.75	0.04	0.95	0.99
(0.7,0.5)	0.61	0.07	0.95	0.88	0.54	0.02	0.95	0.89	0.72	0.04	0.94	0.86
(0.7,0.6)	0.56	0.08	0.95	0.41	0.52	0.01	0.95	0.43	0.65	0.09	0.95	0.39
(0.7,0.7)	0.50	0.09	0.95	0.05	0.50	0.01	0.95	0.05	0.50	0.15	0.95	0.05
(0.5,0.2)	0.61	0.04	0.95	0.99	0.61	0.03	0.96	0.99	0.71	0.02	0.95	0.99
(0.5,0.3)	0.58	0.05	0.95	0.89	0.56	0.02	0.95	0.89	0.68	0.03	0.95	0.87
(0.5,0.4)	0.54	0.05	0.95	0.40	0.53	0.02	0.95	0.41	0.63	0.09	0.95	0.37
(0.5,0.5)	0.50	0.06	0.95	0.05	0.50	0.02	0.95	0.05	0.50	0.14	0.95	0.05

Table 4.3: Simulation results of the $E(N_A/n)$, S.D, Coverage Probability (CP) and statistical power, n=300, r=10,000

(P_A, P_B)	**RPW**				**RSIHR**				**YW**			
	$E(N_A/n)$	$S.D$	CP	$Power$	$E(N_A/n)$	$S.D$	CP	$Power$	$E(N_A/n)$	$S.D$	CP	$Power$
(0.9,0.6)	0.75	0.09	0.95	1	0.55	0.01	0.95	1	0.84	0.03	0.95	1
(0.9,0.7)	0.69	0.11	0.95	1	0.53	0.01	0.95	0.99	0.81	0.04	0.95	0.99
(0.9,0.8)	0.61	0.14	0.95	0.73	0.52	0.01	0.96	0.77	0.74	0.07	0.96	0.71
(0.9,0.9)	0.50	0.18	0.96	0.04	0.50	0.01	0.96	0.05	0.50	0.17	0.96	0.05
(0.7,0.4)	0.66	0.05	0.94	1	0.57	0.01	0.95	1	0.75	0.02	0.95	1
(0.7,0.5)	0.62	0.06	0.95	0.97	0.54	0.01	0.95	0.97	0.72	0.03	0.95	0.95
(0.7,0.6)	0.56	0.07	0.95	0.55	0.52	0.01	0.95	0.55	0.66	0.07	0.95	0.51
(0.7,0.7)	0.50	0.08	0.95	0.05	0.50	0.01	0.95	0.05	0.50	0.15	0.95	0.05
(0.5,0.2)	0.61	0.03	0.95	1	0.61	0.02	0.95	1	0.71	0.02	0.95	1
(0.5,0.3)	0.58	0.04	0.95	0.97	0.56	0.02	0.95	0.97	0.68	0.02	0.95	0.96
(0.5,0.4)	0.54	0.04	0.95	0.52	0.53	0.02	0.95	0.53	0.64	0.07	0.95	0.51
(0.5,0.5)	0.50	0.05	0.96	0.05	0.50	0.02	0.95	0.05	0.50	0.13	0.95	0.05

Table 4.4: Simulation of coverage probabilities (CP) and Length of intervals (L) for $|p_A - p_B| = 0.2$ from bootstrapping for n=200, r=5,000

(P_A, P_B)	**RPW**				**RSIHR**				**YW**			
	CP	L	CP_{boot}	L_{boot}	CP	L	CP_{boot}	L_{boot}	CP	L	CP_{boot}	L_{boot}
(0.9,0.7)	0.95	0.27	0.94	0.27	0.96	0.22	0.95	0.21	0.95	0.31	0.94	0.31
(0.7,0.5)	0.95	0.28	0.95	0.27	0.95	0.27	0.95	0.26	0.95	0.30	0.95	0.29
(0.5,0.3)	0.95	0.27	0.95	0.26	0.95	0.26	0.95	0.26	0.95	0.28	0.95	0.27

Figure 4.1: Distribution of N_A/n allocated to treatment A under RPW(solid line) and RSIHR(dashed line) and YW(dotted line) for n=100, r=10,000

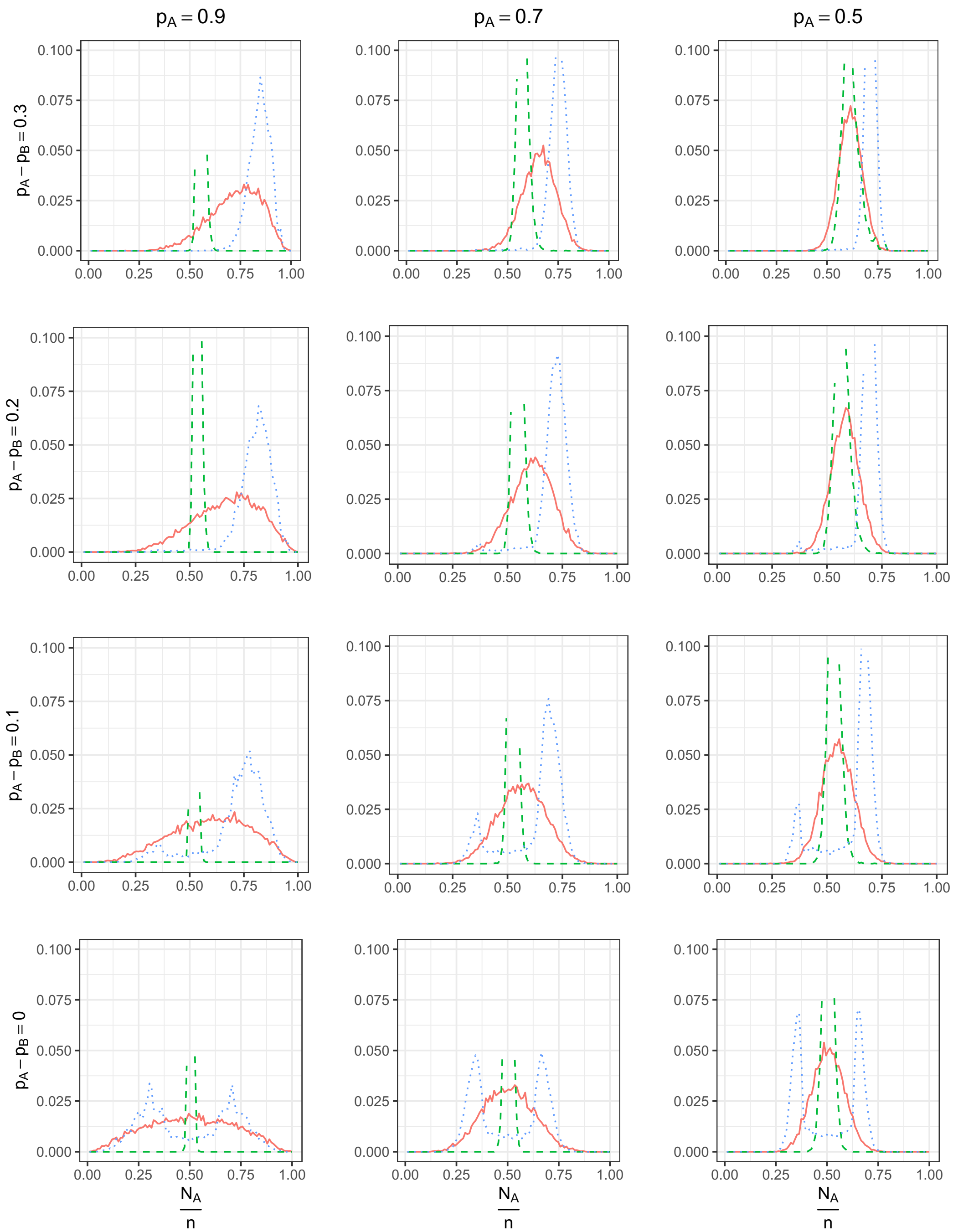


Figure 4.2: Distribution of N_A/n allocated to treatment A under RPW(solid line) and RSIHR(dashed line) and YW(dotted line) for n=200, r=10,000

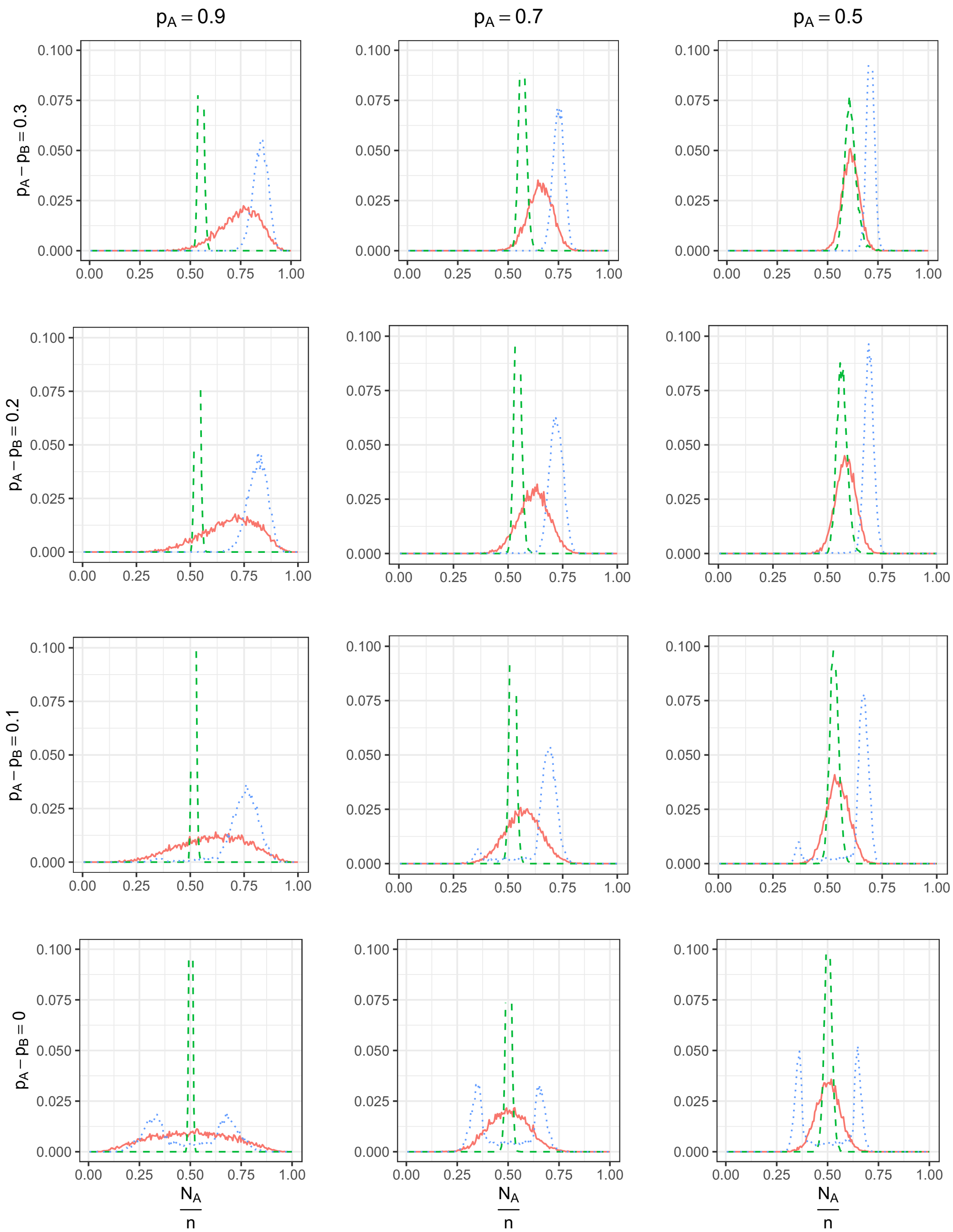


Figure 4.3: Distribution of N_A/n allocated to treatment A under RPW(solid line) and RSIHR(dashed line) and YW(dotted line) for n=300, r=10,000

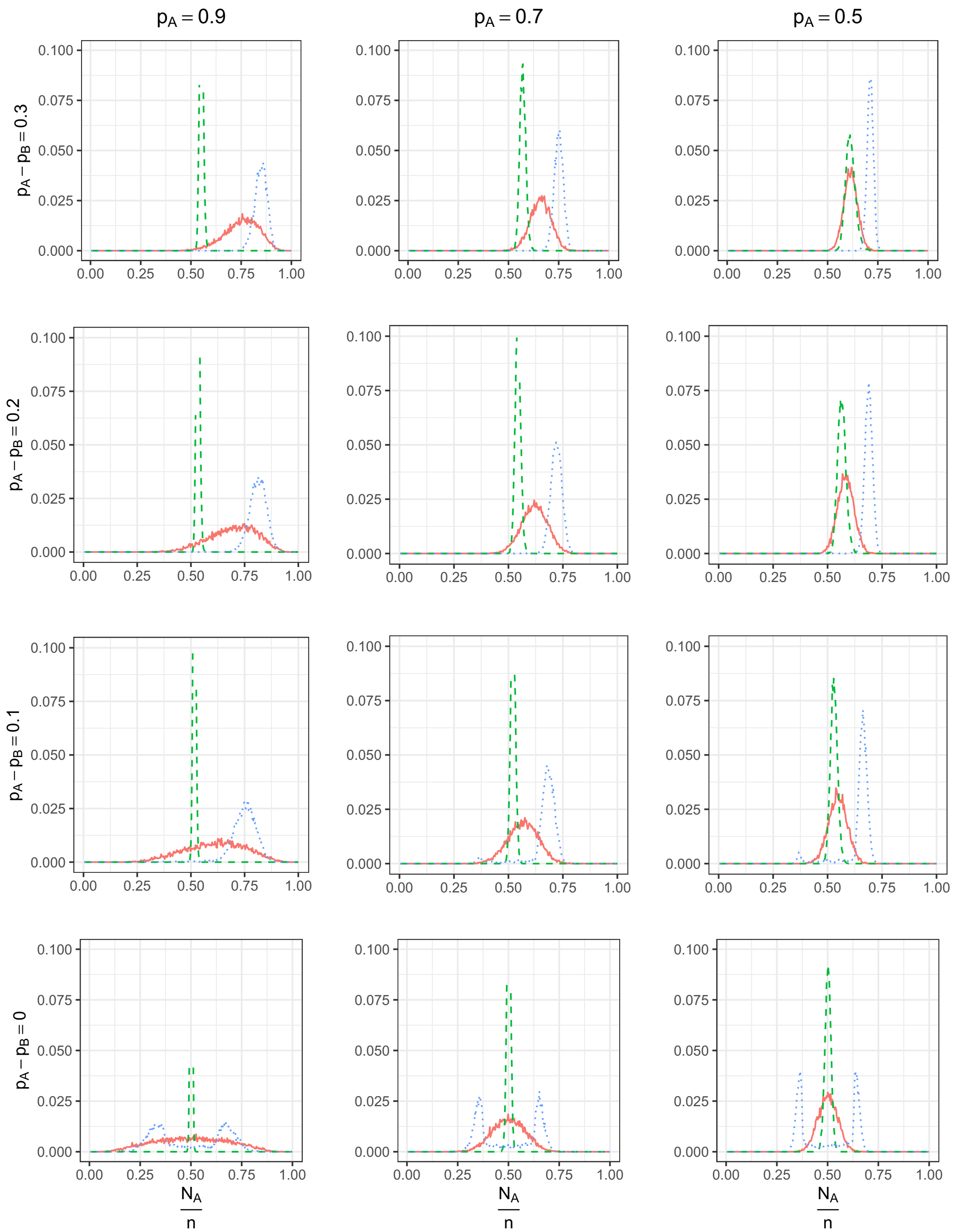


Figure 4.4: Normal Approximation of $(\hat{p}_A - \hat{p}_B)$ for RPW, RSIHR and YW, for n=200, r=5,000

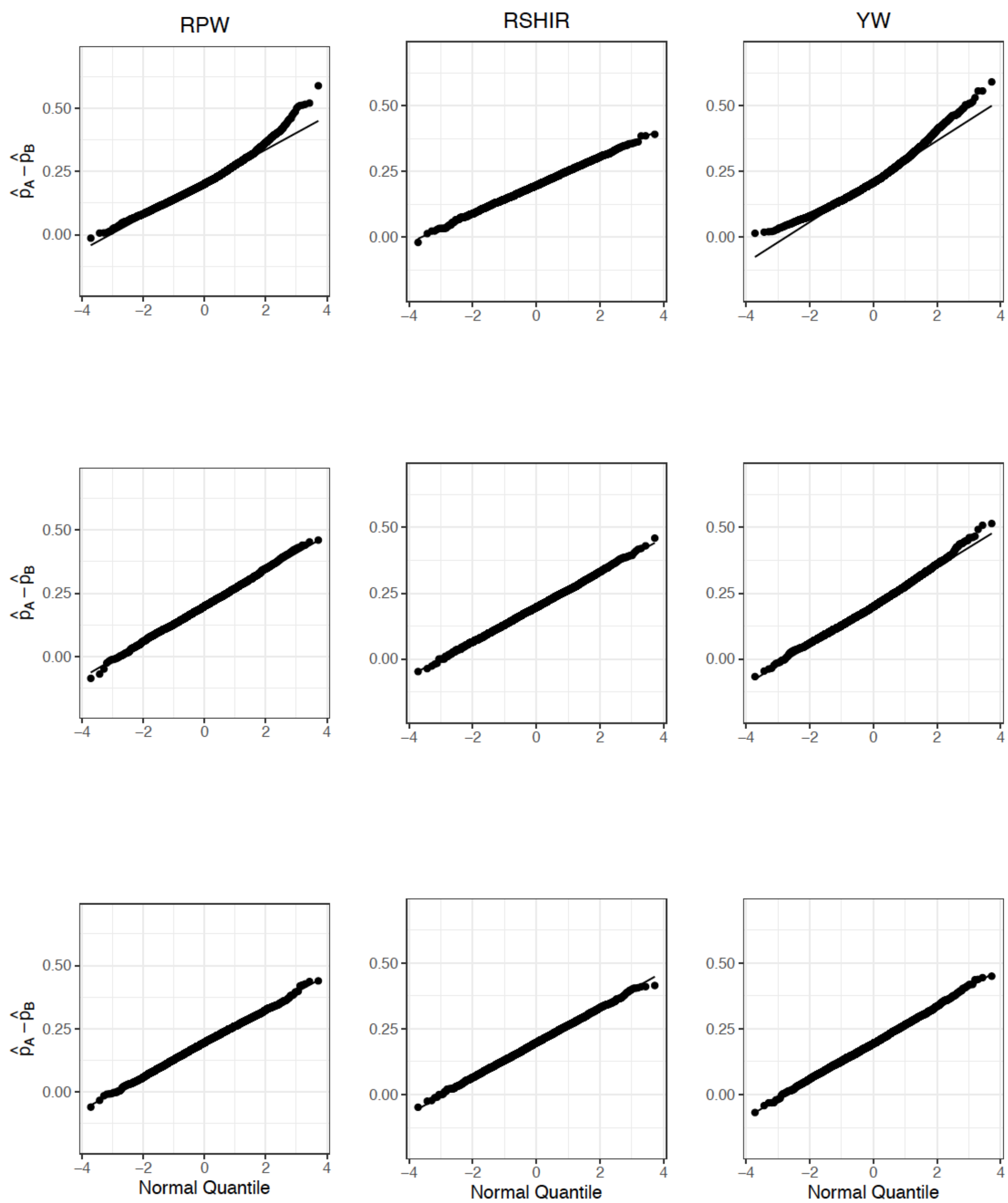
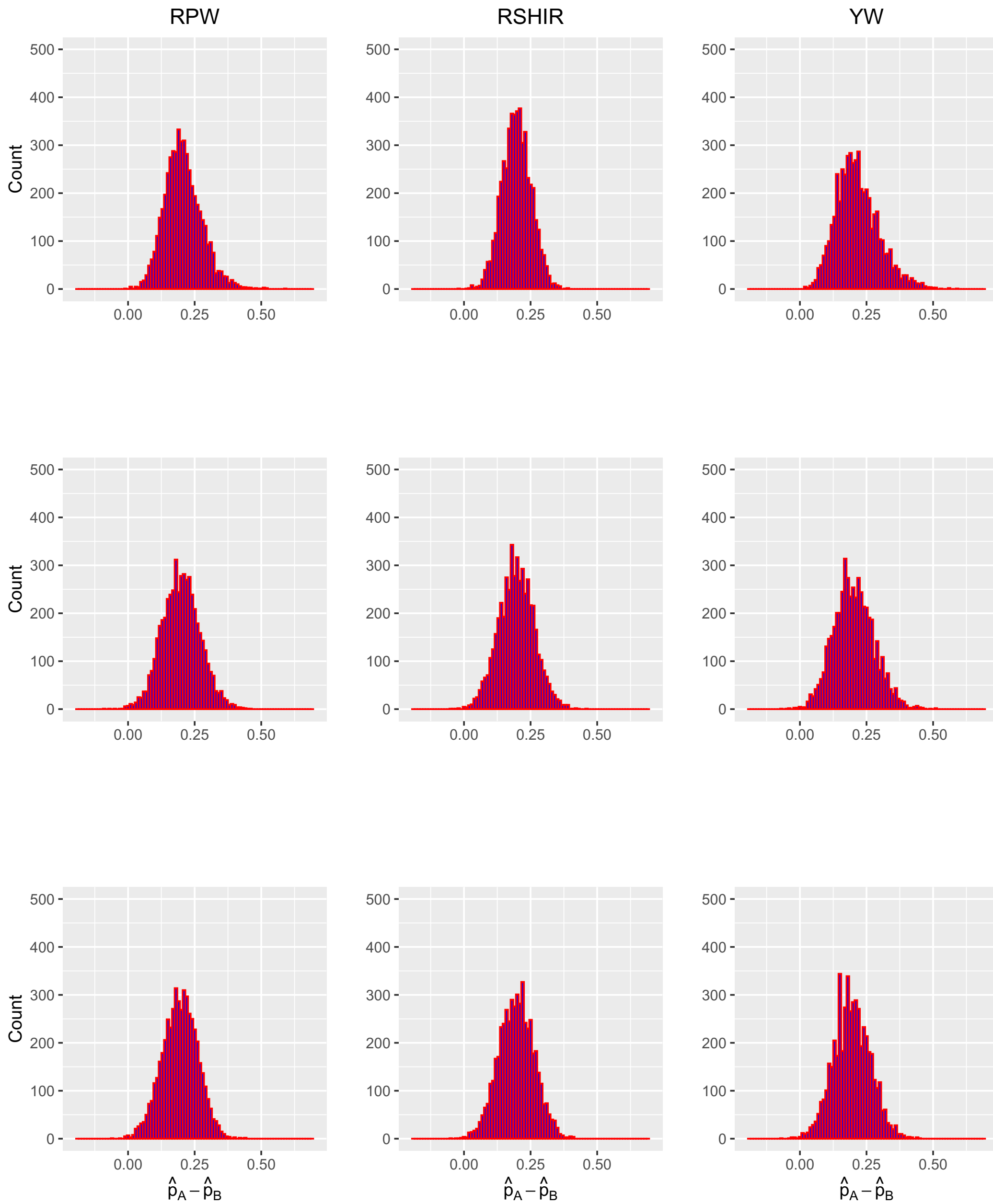


Figure 4.5: Histogram of $(\hat{p}_A - \hat{p}_B)$ for RPW, RSIHR and YW, for n=200, r=5,000



Chapter 5

Conclusion

5.1 Summary of Achievements

In this thesis, we study the asymptotic normality of MLE for response adaptive design (RAD). From the response adaptive randomization procedure, the patients' allocation depends on the aggregated information acquired from the previously treated patients' response, which creates the dependency structure in the data. Due to the dependency structure, we use Stein's method to obtain asymptotic normality of the ML estimators in various allocation procedures.

First, we consider the K function approach in Stein's equation by defining the leave-one-out in the summation of random variables. Then, we derive an upper bound using Stein's equation and obtain a normal approximation for W_A and W_B under a few regularity conditions. Furthermore, upon obtaining asymptotic normal distributions of W_A and W_B using the K function approach in Stein's method, we derive the MLE's asymptotic normality of the parameters of mean responses for RAD using the Slutsky's theorem under the regularity conditions provided.

Next, we prove the asymptotic normality of MLE for the likelihood estimators $(\hat{\mu}_A - \hat{\mu}_B)$ for the difference in mean responses for treatment A and B in the RAD setting and under some regularity conditions. Our main contribution is that by using Stein's method, we obtain the asymptotic normality for RAD with the dependency structure among the collected data, and accordingly, we have considerably more relaxed steps to obtain the results compared to the other references.

Ultimately, we conduct a numerical study to investigate the normal approximation accuracy for RAD under three different adaptive allocation methods –RPW, RSHIR, and YW– for various settings of success probabilities. Using different sample sizes, we compare the coverage probabilities and the power of the utilized statistical test. We also use the bootstrap method for the RAD in order to reduce the estimation bias in situations where the covariance structure is more complicated. Overall, the numerical results show that both estimators have a reasonable performance in the coverage probabilities and confidence intervals, for sample size of 200 and 300.

5.2 Future Reseach

While our research is focused on the asymptotic normality of MLE for RAD using Stein's method for two treatments A and B, these results can be generalized to more than two treatments. Therefore, in the future, we aim to utilize Stein's method in RAD's inference to find the joint distribution of the difference in mean responses for more than two treatments in the RAD setting.

Appendix A

Proofs

Proof of Lemma 1:

$$
\begin{aligned}
E[W_A] &= E\big[E\big[\sum_{j=1}^{n} \mathbb{1}_{jA}\xi_{jA}\big|\mathbb{1}_{jA,j=1,2,\ldots,n}\big]\big] \\
&= E\big[\sum_{j=1}^{n} E[\mathbb{1}_{jA}\xi_{jA}|\mathbb{1}_{jA,j=1,2,\ldots,n}]\big] \\
&= E\big[\sum_{j=1}^{n} \mathbb{1}_{jA}E[\xi_{jA}]\big] \\
&= E[N_A E[\xi_{jA}]] = 0 \qquad\qquad (from \quad (3.3) \quad E\xi_{jA} = 0)
\end{aligned}
\tag{A.1}
$$

$$\begin{aligned}
Var(W_A) &= E\big[Var\big(\sum_{j=1}^{n} \mathbb{1}_{jA}\xi_{jA}\big|\mathbb{1}_{jA,j=1,2,\ldots,n}\big)\big] \\
&\quad + Var\big(E\big[\sum_{j=1}^{n} \mathbb{1}_{jA}\xi_{jA}\big|\mathbb{1}_{jA,j=1,2,\ldots,n}\big]\big) \\
&= E\big[Var\big(\sum_{j=1}^{n} \mathbb{1}_{jA}\xi_{jA}\big|\mathbb{1}_{jA,j=1,2,\ldots,n}\big)\big] \qquad (from \quad (3.3) \quad E\xi_{jA} = 0) \\
&= E[\sum_{j=1}^{n} Var(\mathbb{1}_{jA}\xi_{jA}|\mathbb{1}_{jA,j=1,2,\ldots,n})] \\
&= E[\sum_{j=1}^{n} E(\mathbb{1}_{jA}^2\xi_{jA}^2|\mathbb{1}_{jA,j=1,2,\ldots,n})] \qquad (from \quad (3.3) \quad Var(\xi_{jA}) = E\xi_{jA}^2) \\
&= E[\sum_{j=1}^{n} E(\mathbb{1}_{jA}\xi_{jA}^2|\mathbb{1}_{jA,j=1,\ldots,n})] \\
&= E[\sum_{j=1}^{n} \mathbb{1}_{jA}E(\xi_{jA}^2)] \qquad (from \ (3.3) \quad E\xi_{jA}^2 = \frac{1}{n\rho}) \\
&= E[N_A \ \frac{1}{n\rho}] = E[\frac{N_A}{n\rho}].
\end{aligned} \tag{A.2}$$

Proof of Lemma 2:

$$\begin{aligned}
E[W_B] &= E\big[E\big[\sum_{j=1}^{n}(1-\mathbb{1}_{jA})\xi_{jB}\big|\mathbb{1}_{jA,j=1,2,\ldots,n}\big]\big] \\
&= E[\sum_{j=1}^{n} E[(1-\mathbb{1}_{jA})\xi_{jB}|\mathbb{1}_{jA,j=1,2,\ldots,n}]] \\
&= E[\sum_{j=1}^{n}(1-\mathbb{1}_{jA})E[\xi_{jB}]] \\
&= E[N_B E[\xi_{jB}]] = 0 \qquad (from \quad (3.4) \quad E\xi_{jB} = 0)
\end{aligned} \tag{A.3}$$

$$
\begin{aligned}
Var(W_B) &= E\big[Var\big(\sum_{j=1}^{n}(1-\mathbb{1}_{jA})\xi_{iB}\big|\mathbb{1}_{iA,i=1,2,\ldots,n}\big)\big] \\
&+ Var(E[\sum_{j=1}^{n}(1-\mathbb{1}_{jA})\xi_{jB}\big|\mathbb{1}_{jA,j=1,2,\ldots,n}\big]\big) \\
&= E[Var(\sum_{j=1}^{n}(1-\mathbb{1}_{jA})\xi_{jB}|\mathbb{1}_{jA,j=1,2,\ldots,n})] \qquad (from\ \ (3.4)\ \ E\xi_{jB}=0) \\
&= E[\sum_{j=1}^{n}Var((1-\mathbb{1}_{jA})\xi_{jB}|\mathbb{1}_{jA,j=1,2,\ldots,n})] \\
&= E[\sum_{j=1}^{n}E((1-\mathbb{1}_{jA})^2\xi_{jB}^2|\mathbb{1}_{jA,j=1,2,\ldots,n})] \qquad (from\ \ (3.4)\ \ \ Var(\xi_{jB})=E\xi_{jB}^2) \\
&= E[\sum_{j=1}^{n}E((1-\mathbb{1}_{jA})\xi_{jB}^2|\mathbb{1}_{jA,j=1,2,\ldots,n})] \\
&= E[\sum_{j=1}^{n}(1-\mathbb{1}_{jA})E(\xi_{jB}^2)] \qquad (from\ (3.4)\ \ \ E\xi_{jB}^2=\frac{1}{n(1-\rho)}) \\
&= E[N_B\ \frac{1}{n(1-\rho)}) = E(\frac{N_B}{n(1-\rho)}].
\end{aligned}
\tag{A.4}
$$

Proof of Theorem 3: We first investigate the K function approach for W_A

Define,

$$W_A^{(i)} = W_A - \xi_{iA}\mathbb{1}_{iA} \tag{A.5}$$

where W_A from (3.3) and $\mathbb{1}_{iA}$ from (1.1).

We have,

$$E[\xi_{iA}W_A^{(i)}|\mathbb{1}_{jA}, j = 1, ..., n] = E\big[\xi_{iA}(\sum_{j=1}^{n}\xi_{jA}\mathbb{1}_{jA} - \xi_{iA}\mathbb{1}_{iA}|\mathbb{1}_{jA}, j = 1, ..., n)\big]$$
$$= E[\xi_{iA}\sum_{\substack{j=1\\ j\neq i}}^{n}\mathbb{1}_{jA}\xi_{jA}] + \mathbb{1}_{iA}E[\xi_{iA}^2] - \mathbb{1}_{iA}E[\xi_{iA}^2]$$
$$= \sum_{\substack{j=1\\ j\neq i}}^{n}\mathbb{1}_{jA}E[\xi_{iA}\xi_{jA}] = 0 \ (Since\ i \neq j\ \ E[\xi_{iA}\xi_{jA}] = 0). \tag{A.6}$$

In the next step, we estimate the left-hand side of (2.19) and begin to calculate $E[W_Af(W_A)]$. Let h be a measurable function with $E|h(Z)| < \infty$, and $f = f_h$ be the corresponding solution to the Stein's equation (2.3).

$$E[W_Af(W_A)|\mathbb{1}_{jA}, j = 1, ..., n] = E\ \big[\sum_{i=1}^{n}\xi_{iA}\mathbb{1}_{iA}f(W_A)|\mathbb{1}_{jA}, j = 1, ..., n\big]$$
$$= \sum_{i=1}^{n}\mathbb{1}_{iA}E\big[\xi_{iA}f(W_A)\ \big|\mathbb{1}_{jA}, j = 1, ..., n\big]$$
$$= \sum_{i=1}^{n}\mathbb{1}_{iA}E\big[\xi_{iA}\big(f(W_A) - f(W_A^{(i)})\big)\big|\mathbb{1}_{jA}, j = 1, ..., n\big]. \tag{A.7}$$

From (A.6) and by rearranging the summation to the number of patients, we can conclude that,

$$E[\xi_{iA}f(W_A^{(i)})|\mathbb{1}_{jA}, j = 1, ..., n] = 0. \tag{A.8}$$

Due to this conclusion, the last equity in (A.7) conditionally holds for all $\{\mathbb{1}_{jA}, j =$

$1, ..., n\}$. Therefore,

$$
\begin{aligned}
&E[W_A f(W_A)|\mathbb{1}_{jA}, j = 1, ..., n] = \\
&\sum_{i=1}^{n} \mathbb{1}_{iA} E\big[\xi_{iA}\big(f(W_A) - f(W_A^{(i)})\big)\big|\mathbb{1}_{jA}, j = 1, ..., n\big] \\
&= \sum_{i=1}^{n} \mathbb{1}_{iA} E\big[\xi_{iA} \int_0^{\xi_{iA}} f^{'}(W_A^{(i)} + t)dt\big|\mathbb{1}_{jA}, j = 1, ..., n\big] \\
&= \sum_{i=1}^{n} \mathbb{1}_{iA} E\big[\int_{-\infty}^{0} \{f^{'}(W_A^{(i)} + t)\xi_{iA}(-\mathbb{1}_{\{\xi_{iA} \le t < 0\}})dt|\mathbb{1}_{jA}, j = 1, ..., n\} \\
&+ \int_0^{+\infty} \{f^{'}(W_A^{(i)} + t)\xi_{iA}(\mathbb{1}_{\{0 \le t \le \xi_{iA}\}})dt|\mathbb{1}_{jA}, j = 1, ..., n\}\big] \qquad (\textit{Property 1 for } K_i(t) \text{ (2.17)}).
\end{aligned}
\tag{A.9}
$$

$$
\begin{aligned}
&E[W_A f(W_A)|\mathbb{1}_{jA}, j = 1, ..., n] = \\
&= \sum_{i=1}^{n} \mathbb{1}_{iA} E\big[\int_{-\infty}^{+\infty} \{f^{'}(W_A^{(i)} + t)\xi_{iA}(\mathbb{1}_{\{0 \le t \le \xi_{iA}\}} - \mathbb{1}_{\{\xi_{iA} \le t < 0\}})dt|\mathbb{1}_{jA}, j = 1, ..., n\}\big] \\
&= \sum_{i=1}^{n} \mathbb{1}_{iA} \int_{-\infty}^{+\infty} E\big[f^{'}(W_A^{(i)} + t)|\mathbb{1}_{jA}, j = 1, ..., n\big]K_{iA}(t)dt \quad (\textit{Property 2 for } K_i(t) \text{ (2.17)}).
\end{aligned}
\tag{A.10}
$$

Moving froward to $Ef^{'}[W_A]$, we consider that, $\int_{-\infty}^{+\infty} K_{iA}(t)dt = E\xi_{iA}^2 = \frac{1}{n\rho}$, and $\sum_{i=1}^{n} \mathbb{1}_{iA} \int_{-\infty}^{+\infty} K_{iA}(t)dt = \frac{N_A}{n\rho}$. Then we have,

$$
\begin{aligned}
Ef^{'}(W_A|\mathbb{1}_{jA}, j = 1, ..., n) &= \frac{n\rho}{N_A\; n\rho} \sum_{i=1}^{n} \mathbb{1}_{iA} E\big[f^{'}(W_A)|\mathbb{1}_{jA}, j = 1, ..., n\big] \\
&= \frac{n\rho}{N_A} \sum_{i=1}^{n} \mathbb{1}_{iA} E\big[f^{'}(W_A)|\mathbb{1}_{jA}, j = 1, ..., n\big] E\xi_{iA}^2 \\
&= \frac{n\rho}{N_A} \sum_{i=1}^{n} \mathbb{1}_{iA} \int_{-\infty}^{+\infty} E\big[f^{'}(W_A)|\mathbb{1}_{jA}, j = 1, ..., n\big] K_{iA}(t)dt.
\end{aligned}
\tag{A.11}
$$

Denote the conditional variance of W_A as $\sigma^2_{\{W_A|\mathbb{1}_{jA},j=1,...,n\}}$, then,

$$\sigma^2_{\{W_A|\mathbb{1}_{jA},j=1,...,n\}} = Var(W_A|\mathbb{1}_{jA}, j = 1, ..., n) = \frac{N_A}{n\rho}. \tag{A.12}$$

Therefore, we have,

$$\begin{aligned} &E\big[\big(\sigma^2_{\{W_A|\mathbb{1}_{jA},j=1,...,n\}} f^{'}(W_A) - W_A f(W_A)\big)|\mathbb{1}_{jA}, j = 1, ..., n\big] \\ &\quad = E\big[\big(\frac{N_A}{n\rho} f^{'}(W_A) - W_A f(W_A)\big)|\mathbb{1}_{jA}, j = 1, ..., n\big] \\ &\quad = \sum_{i=1}^{n} \mathbb{1}_{iA} \int_{-\infty}^{+\infty} E\big[\frac{\frac{N_A}{n\rho}}{\frac{N_A}{n\rho}} f^{'}(W_A)|\mathbb{1}_{jA}, j = 1, ..., n\big] K_{iA}(t)dt \\ &\quad - \sum_{i=1}^{n} \mathbb{1}_{iA} \int_{-\infty}^{+\infty} E\big[f^{'}(W_A^{(i)} + t)|\mathbb{1}_{jA}, j = 1, ..., n\big] K_{iA}(t)dt. \end{aligned} \tag{A.13}$$

Or equivalently,

$$\begin{aligned} &E\big[\big(\sigma^2_{\{W_A|\mathbb{1}_{jA},j=1,...,n\}} f^{'}(W_A) - W_A f(W_A)\big)|\mathbb{1}_{jA}, j = 1, ..., n\big] \\ &= \sum_{i=1}^{n} \mathbb{1}_{iA} \int_{-\infty}^{+\infty} E\big[\big(f^{'}(W_A) - f^{'}(W_A^{(i)} + t)\big)|\mathbb{1}_{jA}, j = 1, ..., n\big] K_{iA}(t)dt. \end{aligned} \tag{A.14}$$

Now, we can write (A.14) such that,

$$\begin{aligned} E\big[\frac{N_A}{n\rho} f^{'}(W_A) - W_A f(W_A)\big] &= E\big[E\big[\big(\sigma^2_{\{W_A|\mathbb{1}_{jA},j=1,...,n\}} f^{'}(W_A) - W_A f(W_A)\big)|\mathbb{1}_{jA}, j = 1, ..., n\big]\big] \\ &= E\ \big[\sum_{i=1}^{n} \mathbb{1}_{iA} \int_{-\infty}^{+\infty} E\big[\big(f^{'}(W_A) - f^{'}(W_A^{(i)} + t)\big)|\mathbb{1}_{jA}, j = 1, ..., n\big] K_{iA}(t)dt\big]. \end{aligned} \tag{A.15}$$

The next step is to find the corresponding bound with respect to (2.16) to the estimation obtained for Stein's equation.

$$||h|| = \sup_{x \in R^p} |h(x)|.$$

From the mean value theorem we can obtain for (A.15) that,

$$f^{''}(c) = \frac{f^{'}(W_A^{(i)} + \xi_{iA}) - f^{'}(W_A^{(i)} + t)}{(\xi_{iA} - t)}, \tag{A.16}$$

and,

$$\begin{aligned} |f^{'}(W_A^{(i)} + \xi_{iA}) - f^{'}(W_A^{(i)} + t)| &\leq ||f^{''}|||\xi_{iA} - t| \qquad (from\ mean\ value\ theorem) \\ &\leq ||f^{''}||(|\xi_{iA}| + |t|) \quad (from\ Triangle\ inequality). \end{aligned} \tag{A.17}$$

Therefore we have,

$$\begin{aligned} E|E\big[\frac{N_A}{n\rho} f^{'}(W_A) - W_A f(W_A)|\mathbb{1}_{jA}, j = 1, ..., n\big]| \leq \\ ||f^{''}||E\big[\sum_{i=1}^{n} \mathbb{1}_{iA} \int_{-\infty}^{+\infty} E(|\xi_{iA}| + |t|)K_{iA}(t)dt\big] \\ \leq ||f^{''}|| \sum_{i=1}^{n} E\big[\mathbb{1}_{iA}\big(\int_{-\infty}^{+\infty} E|\xi_{iA}|K_{iA}(t)dt \\ + \int_{-\infty}^{+\infty} |t|K_{iA}(t)dt\big)\big] \qquad (using (3.6)) \\ = ||f^{''}|| \sum_{i=1}^{n} \mathbb{1}_{iA}\big[E|\xi_{iA}|E\xi_{iA}^2 + \frac{1}{2}E|\xi_{iA}^3|\big] \\ \leq ||f^{''}|| \ \frac{3}{2}E\big[\sum_{i=1}^{n} \mathbb{1}_{iA}E|\xi_{iA}^3|\big]. \qquad (from\ H\ddot{o}lder's\ inequality) \end{aligned} \tag{A.18}$$

Last inequality in (A.18) is valid from the following steps by the Hölder's inequality given by,

$$\sum_{i=1}^{n} |a_i \ b_i| \leq \big(\sum_{i=1}^{n} |a_i|^p\big)^{\frac{1}{p}} \big(\sum_{i=1}^{n} |b_i|^q\big)^{\frac{1}{q}} \tag{A.19}$$

when $\frac{1}{p} + \frac{1}{q} = 1$.
In the equality $\sum_{i=1}^{n} \mathbb{1}_{iA} E|\xi_{iA} E\xi_{iA}^2|$ from (A.18) consider $p = 3$ and $q = \frac{3}{2}$ such that,

$$\begin{aligned} \sum_{i=1}^{n} \mathbb{1}_{iA}|E\xi_{iA}E\xi_{iA}^2| &\leq \big(\sum_{i=1}^{n} \mathbb{1}_{iA}|E\xi_{iA}|^3\big)^{\frac{1}{3}} \big(\sum_{i=1}^{n} \mathbb{1}_{iA}|E\xi_{iA}^2|^{\frac{3}{2}}\big)^{\frac{2}{3}} \\ &\leq \big(\sum_{i=1}^{n} \mathbb{1}_{iA}|E\xi_{iA}|^3\big)^{\frac{1}{3}} \big(\sum_{i=1}^{n} \mathbb{1}_{iA}|E\xi_{iA}|^3\big)^{\frac{2}{3}} = \sum_{i=1}^{n} \mathbb{1}_{iA} E|\xi_{iA}^3|. \end{aligned} \tag{A.20}$$

This proves the last inequality in (A.18).

Now from Chapter 2 we know that $||f^{''}|| \leq 2||h^{'}||$ (2.15). For a given function $h : \mathbb{R} \to \mathbb{R}$, let f be the corresponding solution to the Stein's equation (2.3). If h is bounded, then we can rewrite (A.18) as,

$$E|E\big[\frac{N_A}{n\rho} f^{'}(W_A) - W_A f(W_A)|\mathbb{1}_{jA}, j = 1, ..., n\big]| \leq 3||h^{'}||\ E\big[\sum_{i=1}^{n} \mathbb{1}_{iA} E|\xi_{iA}|^3\big]. \tag{A.21}$$

Now, from the definition of $\xi_{iA} = \frac{X_{iA} - \mu_A}{\sqrt{n\rho}\sigma_A}$ in (3.3), we have,

$$E|E\big[\frac{N_A}{n\rho} f^{'}(W_A) - W_A f(W_A)|\mathbb{1}_{jA}, j = 1, ..., n\big]| \leq 3||h^{'}||E\big[\sum_{i=1}^{n} \mathbb{1}_{iA} E|\frac{X_{iA} - \mu_A}{\sqrt{n\rho}\sigma_A}|^3\big]. \tag{A.22}$$

Since in RAD, the patients' responses are i.i.d. Therefore, we have,

$$E|E\big[\frac{N_A}{n\rho} f^{'}(W_A) - W_A f(W_A)|\mathbb{1}_{jA}, j = 1, ..., n\big]| \leq 3||h^{'}||E\big[\frac{N_A}{n\sqrt{n}\rho\sqrt{\rho}} E|\frac{X_{1A} - \mu_A}{\sigma_A}|^3\big]. \tag{A.23}$$

From (2.19) we have,

$$\begin{aligned} E[h(W_A) - h(Z)] =& E[f^{'}(W_A) - W_A f(W_A)] \\ =& E[f^{'}(W_A) - \frac{N_A}{n\rho} f^{'}(W_A)] + E[\frac{N_A}{n\rho} f^{'}(W_A) - W_A f(W_A)]. \end{aligned} \tag{A.24}$$

From (2.14) we know that $||f^{'}|| \leq \sqrt{\frac{2}{\pi}}||h^{'}||$. Therefore, from (A.23) and (A.24) we have,

$$\begin{aligned} |E(h(W_A) - h(Z))| \leq& ||f^{'}||E|1 - \frac{N_A}{n\rho}| + 3||h^{'}||E\big[\frac{N_A}{n\sqrt{n}\rho\sqrt{\rho}} E|\frac{X_{1A} - \mu_A}{\sigma_A}|^3\big] \\ \leq& \sqrt{\frac{2}{\pi}}||h^{'}||E|1 - \frac{N_A}{n\rho}| + 3||h^{'}||E\big[\frac{N_A}{n\sqrt{n}\rho\sqrt{\rho}} E|\frac{X_{1A} - \mu_A}{\sigma_A}|^3\big]. \end{aligned} \tag{A.25}$$

Using Slutsky's theorem, proof of the theorem is completed because $\frac{N_A}{n} \to \rho$ almost surely(a.s) as $n \to \infty$.
Therefore, we have,

$$\lim_{n\to\infty} E|h(W_A) - h(Z))| \to 0. \tag{A.26}$$

Since Z follows a standard standard normal distribution, and $W_A = \sum_{j=1}^{n} \mathbb{1}_{jA}\xi_{jA}$ where $\xi_{jA}, j = 1, ...n$ are conditionally independent random variables on all $\{\mathbb{1}_{jA}, j = 1, ..., n\}$. From (2.26) and (A.26) we have,

$$W_A \xrightarrow{d} Z \quad as \quad n \to \infty. \tag{A.27}$$

■

Proof of Lemma 7:

$$
\begin{aligned}
E[W_{rad}] &= E[E[W_{rad}|\mathbb{1}_{jA,j=1,\dots,n}]] \\
&= E[E[W_A - W_B|\mathbb{1}_{jA,j=1,\dots,n}]] \\
&= E[E[\sum_{i=1}^{n}\mathbb{1}_{iA}\xi_{iA} - \sum_{i=1}^{n}(1-\mathbb{1}_{iA})\xi_{iB}|\mathbb{1}_{jA,j=1,\dots,n}]] \\
&= E[\sum_{i=1}^{n}\mathbb{1}_{iA}E(\xi_{iA}) - \sum_{i=1}^{n}(1-\mathbb{1}_{iA})E(\xi_{iB}]] \\
&= E[N_A\ E[\xi_{iA}] - N_B\ E[\xi_{iB}]] = 0
\end{aligned}
\tag{A.28}
$$

For the variance we only need to obtain the correlation between these two parameters, using Lemma 1 and Lemma 2.

$$
\begin{aligned}
Var(W_{rad}) = E\big[Var\big(W_{rad}\big|\mathbb{1}_{jA}, j=1,\dots,n\big)\big] + Var(E[W_{rad}|\mathbb{1}_{jA}, j=1,\dots,n\big]\big) \\
E\big[Var\big(\{\sum_{i=1}^{n}\mathbb{1}_{iA}\xi_{iA} - \sum_{i=1}^{n}(1-\mathbb{1}_{iA})\xi_{iB}\}|\mathbb{1}_{jA}, j=1,\dots,n\big)\big] \\
+ Var(E[\{\sum_{i=1}^{n}\mathbb{1}_{iA}\xi_{iA} - \sum_{i=1}^{n}(1-\mathbb{1}_{iA})\xi_{iB}\}\big|\mathbb{1}_{jA}, j=1,\dots,n\big]\big)
\end{aligned}
\tag{A.29}
$$

The second term of equality is zero because mean of W_{rad} is zero from (A.28).

$$\begin{aligned}
Var(W_{rad}) &= E[Var(\{\sum_{i=1}^{n} \mathbb{1}_{iA}\xi_{iA} - \sum_{i=1}^{n}(1-\mathbb{1}_{iA})\xi_{iB}\}|\mathbb{1}_{jA}, j=1,...,n)] \\
&= E[\sum_{i=1}^{n} Var(\mathbb{1}_{iA}\xi_{iA}|\mathbb{1}_{jA}, j=1,...,n) + \sum_{i=1}^{n} Var((1-\mathbb{1}_{iA})\xi_{iB}|\mathbb{1}_{jA}, j=1,...,n) \\
&- 2\sum_{i=1}^{n} Cov((\mathbb{1}_{iA}\xi_{iA}, (1-\mathbb{1}_{iA})\xi_{iB})|\mathbb{1}_{jA}, j=1,...,n)] \\
&= E[\sum_{i=1}^{n} E[\mathbb{1}_{iA}^2\xi_{iA}^2|\mathbb{1}_{jA}, j=1,...,n] + \sum_{i=1}^{n} E[(1-\mathbb{1}_{iA})^2\xi_{iB}^2|\mathbb{1}_{jA}, j=1,...,n] \\
&- 2\sum_{i=1}^{n}\{E[(\mathbb{1}_{iA}\xi_{iA}(1-\mathbb{1}_{iA})\xi_{iB})|\mathbb{1}_{jA}, j=1,...,n] \\
&- E[\mathbb{1}_{iA}\xi_{iA}|\mathbb{1}_{jA}, j=1,...,n)E((1-\mathbb{1}_{iA})\xi_{iB}|\mathbb{1}_{jA}, j=1,...,n]\}] \\
&= E(\sum_{i=1}^{n} E[\mathbb{1}_{iA}\xi_{iA}^2|\mathbb{1}_{jA}, j=1,...,n] + \sum_{i=1}^{n} E[(1-\mathbb{1}_{iA})\xi_{iB}^2|\mathbb{1}_{jA}, j=1,...,n] \\
&- 2\sum_{i=1}^{n}\{E[(\mathbb{1}_{iA}\xi_{iA}(1-\mathbb{1}_{iA})\xi_{iB})|\mathbb{1}_{jA}, j=1,...,n] \\
&- \mathbb{1}_{iA}E[\xi_{iA}](1-\mathbb{1}_{iA})E[\xi_{iB}]\}) \qquad (from \quad (3.3),(3.4) \quad E\xi_{iA}=0 \quad and \quad E\xi_{iB}=0) \\
&= E[\sum_{i=1}^{n} \mathbb{1}_{iA}E(\xi_{iA}^2) + \sum_{i=1}^{n}(1-\mathbb{1}_{iA})E(\xi_{iB}^2) \\
&- 2\sum_{i=1}^{n}\{(\mathbb{1}_{iA}(1-\mathbb{1}_{iA})E(\xi_{iA}\xi_{iB}))\}] \qquad (from \ (1.1) \ and \ E[\xi_{iA}\xi_{iB}]=0) \\
&= E[N_A \ E\xi_{iA}^2 + N_B \ E\xi_{iB}^2] \\
&= E[N_A \ \frac{1}{n\rho} + N_B \ \frac{1}{n(1-\rho)}] \\
&= E[\frac{N_A}{n\rho}] + E[\frac{N_B}{n(1-\rho)}].
\end{aligned} \tag{A.30}$$

Proof of Theorem 8: We first apply the K function approach to W_{rad}. Define,

$$W_{rad}^{(i)} = W_{rad} - (\xi_{iA}\mathbb{1}_{iA} + \xi_{iB}(1 - \mathbb{1}_{iA})) \tag{A.31}$$

where W_{rad} from (3.10).

We have,

$$E[W_{rad}^{(i)}\xi_{iA}\mathbb{1}_{iA}|\mathbb{1}_{jA,j=1,...,n}] + E[W_{rad}^{(i)}\xi_{iB}(1 - \mathbb{1}_{iA})|\mathbb{1}_{jA}, j = 1, ..., n] = 0. \tag{A.32}$$

Also,

$$E[f(W_{rad}^{(i)})\xi_{iA}\mathbb{1}_{iA}|\mathbb{1}_{jA,j=1,...,n}] + E[f(W_{rad}^{(i)})\xi_{iB}(1 - \mathbb{1}_{iA})|\mathbb{1}_{jA}, j = 1, ..., n] = 0. \tag{A.33}$$

Afterwards, we estimate the left-hand side of (2.19), and begin to calculate $E[W_{rad}f(W_{rad})]$. Let h is a measurable function with $E|h(Z)| < \infty$. Also, let $f = f_h$ be the corresponding solution to the Stein's equation (2.3).

$$\begin{aligned}
E[W_{rad}f(W_{rad})|\mathbb{1}_{jA}, j = 1, ..., n] &= E\ [\sum_{i=1}^{n}\big(\mathbb{1}_{iA}\xi_{iA} - (1 - \mathbb{1}_{iA})\xi_{iB}\big)f(W_{rad})|\mathbb{1}_{jA}, j = 1, ..., n] \\
&= \sum_{i=1}^{n} E\big[\big(\mathbb{1}_{iA}\xi_{iA} - (1 - \mathbb{1}_{iA})\xi_{iB}\big)f(W_{rad})\ \big|\mathbb{1}_{jA}, j = 1, ..., n\big] \\
&= \sum_{i=1}^{n}\big(\mathbb{1}_{iA}E\big[\xi_{iA}f(W_{rad})\ |\mathbb{1}_{jA}, j = 1, ..., n\big] \\
&- (1 - \mathbb{1}_{iA})E\ \big[\xi_{iB}f(W_{rad})|\mathbb{1}_{jA}, j = 1, ..., n\big]\big) \\
&= \sum_{i=1}^{n}\mathbb{1}_{iA}E[\xi_{iA}(f(W_{rad}) - f(W_{rad}^{(i)}))\big|\mathbb{1}_{jA}, j = 1, ..., n] \\
&- \sum_{i=1}^{n}(1 - \mathbb{1}_{iA})E\ \big[\xi_{iB}(f(W_{rad}) - f(W_{rad}^{(i)}))\big|\mathbb{1}_{jA}, j = 1, ..., n].
\end{aligned} \tag{A.34}$$

Due to conclusion in (A.32) and (A.33) the last equality of (A.34) conditionally holds on all $\{\mathbb{1}_{jA}, j = 1, ..., n\}$. Therefore,

$$
\begin{aligned}
&E[W_{rad} f(W_{rad}) | \mathbb{1}_{jA}, j = 1, ..., n] = \sum_{i=1}^{n} \mathbb{1}_{iA} E\big[\xi_{iA}\big(f(W_{rad}) - f(W_{rad}^{(i)})\big)\big| \mathbb{1}_{jA}, j = 1, ..., n\big] \\
&- \sum_{i=1}^{n} (1 - \mathbb{1}_{iA}) E\ \big[\xi_{iB}\big(f(W_{rad}) - f(W_{rad}^{(i)})\big)\big| \mathbb{1}_{jA}, j = 1, ..., n\big]\} \\
&= \sum_{i=1}^{n} \mathbb{1}_{iA} E\big[\xi_{iA} \int_{0}^{\xi_{iA}} f'(W_{rad}^{(i)} + t) dt | \mathbb{1}_{jA}, j = 1, ..., n\big] \\
&- \sum_{i=1}^{n} (1 - \mathbb{1}_{iA}) E\big[\xi_{iB} \int_{0}^{\xi_{iB}} f'(W_{rad}^{(i)} + t) dt | \mathbb{1}_{jA}, j = 1, ..., n\big] \\
&= \sum_{i=1}^{n} \mathbb{1}_{iA} E\big[\int_{-\infty}^{0} (f'(W_{rad}^{(i)} + t) \xi_{iA} (-\mathbb{1}_{\{\xi_{iA} \le t < 0\}}) dt\ | \mathbb{1}_{jA}, j = 1, ...n) \\
&+ \int_{0}^{+\infty} (f'(W_{rad}^{(i)} + t) \xi_{iA} (\mathbb{1}_{\{0 \le t \le \xi_{iA}\}}) dt\ | \mathbb{1}_{jA}, j = 1, ...n)\big] \\
&- \sum_{i=1}^{n} (1 - \mathbb{1}_{iA}) E[\int_{-\infty}^{0} (f'(W_{rad}^{(i)} + t) \xi_{iB} (-\mathbb{1}_{\{\xi_{iB} \le t < 0\}}) dt\ | \mathbb{1}_{jA}, j = 1, ...n) \\
&+ \int_{0}^{+\infty} (f'(W_{rad}^{(i)} + t) \xi_{iB} (\mathbb{1}_{\{0 \le t \le \xi_{iB}\}}) dt\ | \mathbb{1}_{jA}, j = 1, ...n)\big]\} \qquad (Property\ 1\ for\ K_i(t)\ (2.17)).
\end{aligned}
\tag{A.35}
$$

$$
\begin{aligned}
&E[W_{rad}f(W_{rad})|\mathbb{1}_{jA}, j=1,...,n] = \\
&\sum_{i=1}^{n}\mathbb{1}_{iA}E\Big[\int_{-\infty}^{+\infty}(f^{'}(W_{rad}^{(i)}+t)\xi_{iA}(\mathbb{1}_{\{0\le t\le \xi_{iA}\}}-\mathbb{1}_{\{\xi_{iA}\le t<0\}})\ dt|\mathbb{1}_{jA}, j=1,...n)\Big] \\
&-\sum_{i=1}^{n}(1-\mathbb{1}_{iA})E\Big[\int_{-\infty}^{+\infty}(f^{'}(W_{rad}^{(i)}+t)(\xi_{iB}(\mathbb{1}_{\{0\le t\le \xi_{iB}\}}-\mathbb{1}_{\{\xi_{iB}\le t<0\}}))dt|\mathbb{1}_{jA}, j=1,...,n)\Big] \\
&=\sum_{i=1}^{n}\mathbb{1}_{iA}\int_{-\infty}^{+\infty}E\big[f^{'}(W_{rad}^{(i)}+t)\ |\mathbb{1}_{jA}, j=1,...,n\big]K_{iA}(t)dt \\
&-\sum_{i=1}^{n}(1-\mathbb{1}_{iA})\int_{-\infty}^{+\infty}E\big[f^{'}(W_{rad}^{(i)}+t)|\mathbb{1}_{jA}, j=1,...,n\big]K_{iB}(t)dt \quad (Property\ 2\ for\ K_i(t)\ (2.17)).
\end{aligned} \tag{A.36}
$$

Moving froward to $Ef^{'}[W_{rad}]$ considering that $\int_{-\infty}^{+\infty}K_{iA}(t)dt = E\xi_{iA}^2 = \frac{1}{n\rho}$ also, $\int_{-\infty}^{+\infty}K_{iB}(t)dt = E\xi_{iB}^2 = \frac{1}{n(1-\rho)}$. Then, we have,

$$
\begin{aligned}
Ef^{'}[W_{rad}|\mathbb{1}_{jA}, j=1,...,n] &= E[f^{'}(W_{rad})|\mathbb{1}_{jA}, j=1,...,n] \\
&= E[\frac{(\frac{N_A}{n\rho}+\frac{N_B}{n(1-\rho)})}{(\frac{N_A}{n\rho}+\frac{N_B}{n(1-\rho)})}f^{'}(W_{rad})|\mathbb{1}_{jA}, j=1,...,n] \\
&= \frac{1}{(\frac{N_A}{n\rho}+\frac{N_B}{n(1-\rho)})}\sum_{i=1}^{n}\mathbb{1}_{iA}E[f^{'}(W_{rad})|\mathbb{1}_{jA}, j=1,...,n]E\xi_{iA}^2 \\
&+ \frac{1}{(\frac{N_A}{n\rho}+\frac{N_B}{n(1-\rho)})}\sum_{i=1}^{n}(1-\mathbb{1}_{iA})E[f^{'}(W_{rad})|\mathbb{1}_{jA}, j=1,...,n]E\xi_{iB}^2 \\
&= \frac{1}{(\frac{N_A}{n\rho}+\frac{N_B}{n(1-\rho)})}\sum_{i=1}^{n}\mathbb{1}_{iA}\int_{-\infty}^{+\infty}E[f^{'}(W_{rad})|\mathbb{1}_{jA}, j=1,...,n]K_{iA}(t)dt \\
&+ \frac{1}{(\frac{N_A}{n\rho}+\frac{N_B}{n(1-\rho)})}\sum_{i=1}^{n}(1-\mathbb{1}_{iA})\int_{-\infty}^{+\infty}E[f^{'}(W_{rad})|\mathbb{1}_{jA}, j=1,...,n]K_{iB}(t)dt.
\end{aligned} \tag{A.37}
$$

Denote conditional variance of W_{rad} as $\sigma^2_{\{W_{rad}|\mathbb{1}_{jA}, j=1,...,n\}}$, then,

$$
\sigma^2_{\{W_{rad}|\mathbb{1}_{jA}, j=1,...,n\}} = Var(W_{rad}|\mathbb{1}_{jA}, j=1,...,n) = \frac{N_A}{n\rho}+\frac{N_B}{n(1-\rho)}. \tag{A.38}
$$

Therefore we have,

$$
\begin{aligned}
&E\big[\big(\sigma^2_{\{W_{rad}|\mathbb{1}_{jA},j=1,...,n}f^{'}(W_{rad})-W_{rad}f(W_{rad})\big)|\mathbb{1}_{jA},j=1,...,n\big]\\
&=E\big[(\frac{N_A}{n\rho}+\frac{N_B}{n(1-\rho)})f^{'}(W_{rad})-W_{rad}f(W_{rad})|\mathbb{1}_{jA},j=1,...,n\big]\\
&=\sum_{i=1}^{n}\mathbb{1}_{iA}\int_{-\infty}^{+\infty}E[f^{'}(W_{rad})|\mathbb{1}_{jA},j=1,...,n]K_{iA}(t)dt\\
&+\sum_{i=1}^{n}(1-\mathbb{1}_{iA})\int_{-\infty}^{+\infty}E[f^{'}(W_{rad})|\mathbb{1}_{jA},j=1,...,n]K_{iB}(t)dt\\
&-\sum_{i=1}^{n}\mathbb{1}_{iA}\int_{-\infty}^{+\infty}E\big[f^{'}(W_{rad}^{(i)}+t)\ |\mathbb{1}_{jA},j=1,...n\big]K_{iA}(t)dt\\
&-\sum_{i=1}^{n}(1-\mathbb{1}_{iA})\int_{-\infty}^{+\infty}E\big[f^{'}(W_{rad}^{(i)}+t)|\mathbb{1}_{jA},j=1,2,...,n\big]K_{iB}(t)dt.
\end{aligned}
\tag{A.39}
$$

Equivalently, we can write,

$$
\begin{aligned}
&E\big[\big(\sigma^2_{\{W_{rad}|\mathbb{1}_{jA},j=1,...,n\}}f^{'}(W_{rad})-W_{rad}f(W_{rad})\big)|\mathbb{1}_{jA},j=1,...,n\big]\\
&=E\big[((\frac{N_A}{n\rho}+\frac{N_B}{n(1-\rho)})f^{'}(W_{rad})-W_{rad}f(W_{rad})\big)|\mathbb{1}_{jA},j=1,...,n\big]\\
&=\sum_{i=1}^{n}\mathbb{1}_{iA}\int_{-\infty}^{+\infty}E[f^{'}(W_{rad})-f^{'}(W_{rad}^{(i)}+t)|\mathbb{1}_{jA},j=1,...,n]K_{iA}(t)dt\\
&+\sum_{i=1}^{n}(1-\mathbb{1}_{iA})\int_{-\infty}^{+\infty}E[f^{'}(W_{rad})-f^{'}(W_{rad}^{(i)}+t)|\mathbb{1}_{jA},j=1,...,n]K_{iB}(t)dt.
\end{aligned}
\tag{A.40}
$$

Now, we can write (A.40) such that,

$$
\begin{aligned}
&E\big[\frac{N_A}{n\rho}+\frac{N_B}{n(1-\rho)}f^{'}(W_{rad})-W_{rad}f(W_{rad})\big]\\
&=E\big[E\big[\sigma^2_{\{W_{rad}|\mathbb{1}_{jA},j=1,\ldots,n\}}f^{'}(W_{rad})-W_{rad}f(W_{rad})|\mathbb{1}_{jA},j=1,...,n\big]\big]\\
&=E\big[\sum_{i=1}^{n}\mathbb{1}_{iA}\int_{-\infty}^{+\infty}E[f^{'}(W_{rad})-f^{'}(W^{(i)}_{rad}+t)|\mathbb{1}_{jA},j=1,...,n]K_{iA}(t)dt \qquad \text{(A.41)}\\
&+\sum_{i=1}^{n}(1-\mathbb{1}_{iA})\int_{-\infty}^{+\infty}E[f^{'}(W_{rad})-f^{'}(W^{(i)}_{rad}+t)|\mathbb{1}_{jA},j=1,...,n]K_{iB}(t)dt\big].
\end{aligned}
$$

The next step is to acquire the corresponding bound from (A.42). Note that the bound that we obtain for the left-hand side of (2.19) is valid for the right-hand side of the equation as well.

We can simplify (A.42) by using the mean value theorem, Triangle inequality, and Hölder's inequality. Then,

$$
f^{''}(c^{'})=\frac{f^{'}(W^{(i)}_{rad}+\xi_{iA}\mathbb{1}_{iA}+\xi_{iB}(1-\mathbb{1}_{iA}))-f^{'}(W^{(i)}_{rad}+t)}{((\xi_{iA}\mathbb{1}_{iA}+\xi_{iB}(1-\mathbb{1}_{iA}))-t)}. \qquad \text{(A.42)}
$$

Since we are using a summation on all $\{j=1,...,n\}$, we can consider them separately and the difference will depend on $\{\mathbb{1}_{jA},j=1,...,n\}$ outcomes. From triangle inequality we have,

$$
|f^{'}(W^{(i)}_{rad}+\xi_{iA})-f^{'}(W^{(i)}_{rad}+t)|\leq||f^{''}||\;|\xi_{iA}-t|
$$

when $\mathbb{1}_{iA}=1$,

$$
|f^{'}(W^{(i)}_{rad}+\xi_{iB})-f^{'}(W^{(i)}_{rad}+t)|\leq||f^{''}||\;|\xi_{iB}-t|
$$

when $\mathbb{1}_{iA}=0$.

By (3.6) and (3.8) and from Hölder's inequality, we have the following inequality for

W_{rad} using Stein's method.

$$E|E[(\frac{N_A}{n\rho}+\frac{N_B}{n(1-\rho)})f^{'}(W_{rad})-W_{rad}f(W_{rad})|\mathbb{1}_{jA},j=1,...,n]|\leq$$
$$||f^{''}||\ E\big[\sum_{i=1}^{n}\mathbb{1}_{iA}\ \big[E\xi_{iA}^2E|\xi_{iA}|+\frac{1}{2}E|\xi_{iA}^3|\big]$$
$$+\sum_{i=1}^{n}(1-\mathbb{1}_{iA})\ \big[E\xi_{iB}^2E|\xi_{iB}|+\frac{1}{2}E|\xi_{iB}^3|\big]\big]$$
$$\leq||f^{''}||\ E\big[\frac{3}{2}\sum_{i=1}^{n}\mathbb{1}_{iA}E|\xi_{iA}^3|$$
$$+\frac{3}{2}\sum_{i=1}^{n}(1-\mathbb{1}_{iA})E|\xi_{iB}^3|\big]$$
$$\leq\frac{3}{2}||f^{''}||\ E\big[\sum_{i=1}^{n}\mathbb{1}_{iA}E|\xi_{iA}^3|$$
$$+\sum_{i=1}^{n}(1-\mathbb{1}_{iA})E|\xi_{iB}^3|\big]. \tag{A.43}$$

From (2.15) $||f^{''}||\leq 2||h^{'}||$, If h is bounded, then we can rewrite (A.43)

$$E|E[(\frac{N_A}{n\rho}+\frac{N_B}{n(1-\rho)})f^{'}(W_{rad})-W_{rad}f(W_{rad})|\mathbb{1}_{jA},j=1,...,n]|\leq$$
$$3||h^{'}||\{\ E[\sum_{i=1}^{n}\mathbb{1}_{iA}\ E|\xi_{iA}|^3+\sum_{i=1}^{n}(1-\mathbb{1}_{iA})\ E|\xi_{iB}|^3\big]\}. \tag{A.44}$$

We know the definition for $E\xi_{iA}$, $E\xi_{iB}$ from (3.3) and (3.4), we have,

$$E|E[(\frac{N_A}{n\rho}+\frac{N_B}{n(1-\rho)})f^{'}(W_{rad})-W_{rad}f(W_{rad})|\mathbb{1}_{jA},j=1,...,n]|\leq$$
$$3||h^{'}||E\big[\sum_{i=1}^{n}\mathbb{1}_{iA}\;\frac{1}{n\sqrt{n}\rho\sqrt{\rho}}E|\frac{X_{iA}-\mu_A}{\sigma_A}|^3 + \sum_{i=1}^{n}(1-\mathbb{1}_{iA})\frac{1}{n\sqrt{n}(1-\rho)\sqrt{(1-\rho)}}E|\frac{X_{iB}-\mu_B}{\sigma_B}|^3\big]. \tag{A.45}$$

Since in RAD the patients' responses are i.i.d. Therefore, we have,

$$E|E[(\frac{N_A}{n\rho}+\frac{N_B}{n(1-\rho)})f^{'}(W_{rad})-W_{rad}f(W_{rad})|\mathbb{1}_{jA},j=1,...,n]|\leq$$
$$3||h^{'}||E\big[N_A\;\frac{1}{n\sqrt{n}\rho\sqrt{\rho}}E|\frac{X_{1A}-\mu_A}{\sigma_A}|^3 + N_B\frac{1}{n\sqrt{n}(1-\rho)\sqrt{(1-\rho)}}E|\frac{X_{1B}-\mu_B}{\sigma_B}|^3\big]. \tag{A.46}$$

Similar to (A.25) we have,

$$\begin{aligned}|E(h(W_{rad})-h(Z))| &\leq ||f^{'}||E|1-\frac{N_A}{n\rho}|+3||h^{'}||E\big[\frac{N_A}{n\sqrt{n}\rho\sqrt{\rho}}E|\frac{X_{1A}-\mu_A}{\sigma_A}|^3\big]\\ &+||f^{'}||E|1-\frac{N_B}{n(1-\rho)}|+3||h^{'}||E\big[\frac{N_B}{n\sqrt{n}(1-\rho)\sqrt{(1-\rho)}}E|\frac{X_{1B}-\mu_B}{\sigma_B}|^3\big]\\ &\leq\sqrt{\frac{2}{\pi}}||h^{'}||E|1-\frac{N_A}{n\rho}|+3||h^{'}||E\big[\frac{N_A}{n\sqrt{n}\rho\sqrt{\rho}}E|\frac{X_{1A}-\mu_A}{\sigma_A}|^3\big]\\ &+\sqrt{\frac{2}{\pi}}||h^{'}||E|1-\frac{N_B}{n(1-\rho)}|+3||h^{'}||E\big[\frac{N_B}{n\sqrt{n}(1-\rho)\sqrt{(1-\rho)}}E|\frac{X_{1B}-\mu_B}{\sigma_B}|^3\big].\end{aligned} \tag{A.47}$$

Using Slutsky's theorem, proof of the theorem is completed because $\frac{N_A}{n}\to\rho$ a.s. as

$n \to \infty$.

Therefore, we have,

$$\lim_{n\to\infty} E|h(W_{rad}) - E(h(Z))| \to 0. \tag{A.48}$$

Since Z follows a standard normal distribution. Similarly, from the definition of the L^1 mentioned before in (2.26) and (A.48) we can conclude that,

$$W_{rad} \xrightarrow{d} Z. \;\; as \;\; n \to \infty. \tag{A.49}$$

■

Appendix B

Simulation R Codes

This appendix provides the simulation program in R for the RPW design and the DBCD that is used to target the RSIHR and YW allocations from Yi and Li (2018) [20]. Also, we have provided the steps for bootstrap re-sampling from Rosenberger and Hu (1999) [15]. The steps to use the program is as follows:

- Input r - the repetition number; d - the total number of patients; cutZ - the critical value for the statistical power (one-sided); pa - the success probability for treatment A; pb - the success probability of treatment B.
- Run the R function for the RPW design or for the RSIHR and YW distrSRPW, distrSDBC2 and distrSDBC3 designs. Before run the function for RSIHR or YW the target allocation should be run in R firstly.
- From the return functions the results can be obtained.

```
###############
#Variables definitions
# r is the repetition number
# d is the total number of patients-2,
#Level of significant as an example alpha =0.05
```

```
#Success probability on treatment A is pa
#Success probability on treatment B is pb
#Difference of the success probabilities is called mu
###############
#RPW Using Response Adaptive Design
distrSRPW<-function(alpha,cutZ,pa,pb,r,d){
  p_cum=0
  p_sd=c()
  naCount<-rep(0,d+1)
  xNA<-1:(d+1)
  sCount<-rep(0,d+3)
  xS<-0:(d+2)
  zCount<-0
  zDistr<-rep(0,r)
  CI=c()
  CI_RPW=matrix(0,r,2)
  countCI=0
  SD=c()
#r for replicate simulated coverage probabilities
for (n in 1:r){ print(n)#Count the simulation run
    na<-1 #First two patients
    nb<-1
    sa<-0
    sb<-0
    u1<-runif(2,0,1)
    if (u1[1]<pa) {sa<-sa+1}
    if (u1[2]<pb) {sb<-sb+1}
    aBall<-1 #RPW(1,1,1)
    bBall<-1
    Add<-1
    p<-aBall/(aBall+bBall)
```

```
    for (i in 1:d){
      u<-runif(3,0,1)
      e<-c(0,0,0)
      if (u[1]<p){na<-na+1
      if (u[2]<pa){sa<-sa+1
      aBall<-aBall+Add
      e[2]<-1}
      if (e[2]==0){bBall<-bBall+Add}
      e[1]<-1}
      if (e[1]==0){nb<-nb+1
      if (u[3]<pb){sb<-sb+1
      bBall<-bBall+Add
      e[3]<-1}
      if (e[3]==0){aBall<-aBall+Add}
      }
      p<-aBall/(aBall+bBall)
    }
    s<-sa+sb
    naCount[na]<-naCount[na]+1
    sCount[s+1]<-sCount[s+1]+1
#SD and NA/n
    p_cum=p_cum+(na/(na+nb))
    p_sd=append(p_sd,(na/(na+nb)))
#Use an adjustment by Agresti and Caffo(DBDCD)
    paHat<-(sa+1)/(na+2)
    pbHat<-(sb+1)/(nb+2)
#Convergence probability and Power of the test
#Using CIs
    mu=pa-pb
    MuHat=paHat-pbHat
    SdHat=sqrt(paHat*(1-paHat)/(na)+pbHat*(1-pbHat)/(nb))
```

```
    E = qnorm(1-(alpha/2))*SdHat
    CI<-c(MuHat-E, MuHat + E)
    CI_RPW[n,]=CI
    if (CI_RPW[n,1]<=mu && mu<=CI_RPW[n,2]){countCI=countCI+1}
#With test statistics and cutZ
    z<-(paHat-pbHat)/sqrt(paHat*(1-paHat)/(na)+pbHat*
    (1-pbHat)/(nb))#power wald
    zDistr[n]<-z
    c1<-0
    if (z<cutZ){c1<-1} # for one-sided
    if (c1==0){zCount<-zCount+1}
  }
#Out comes
  mean_length=mean(CI_RPW[,2]-CI_RPW[,1])
  p_cum_r=p_cum/r #Mean (NA/n)
  p_sd_r=sd(p_sd)#SD(NA/n)
  probNA<-naCount/r #NA count
  probS<-sCount/r  # Success count
  powerZ<-zCount/r #simulated power
  CountCI_r=countCI/r #Simulated CI
  return(list(zDistr=zDistr,
    xNA=xNA,probNA=probNA,
    probS=probS,
    CI_RPW=CI_RPW,
    pa=pa,pb=pb,cutZ=cutZ,
    p_cum_r=p_cum_r,
    p_sd_r=p_sd_r,
    powerZ=powerZ,
    CountCI_r=CountCI_r,
    mean_length=mean_length
    ))}
```

```
###############
#RSIHR
#Using the doubly biased coin design allocation function g(x,rho)
#(DBCD) in Hu and Zhang (2007) allocating by p=1/2
TargProp2<-function(pa,pb){rho<-sqrt(pa)/(sqrt(pa)+sqrt(pb))
return(rho)}
# gamma is the parameter in allocation function g(x,rho);
# r is the repetition number.
# critical value for one-sided test
# d is the number of patients-2.
# success probability on treatment A pa
# success probability on treatment B pb
distrSDBC2<-function(alpha,pa,pb,gamma,d,r,cutZ){
  p_cum=0
  p_sd=c()
  naCount<-rep(0,d+1)
  xNA<-1:(d+1)
  sCount<-rep(0,d+3)
  xS<-0:(d+2)
  zCount<-0
  scCount<-0
  zDistr<-rep(0,r)
  CI=c()
  CI_RSIHR=matrix(0,r,2)
  countCI=0
  #SD=c()
  for (n in 1:r){ print(n)
    sa<-0
    sb<-0
    na<-1 # the first two patients
```

```
    nb<-1
    u1<-runif(2,0,1)
    if (u1[1]<pa) {sa<-sa+1}
    if (u1[2]<pb) {sb<-sb+1}
#Use an adjustment by Agresti and Caffo (the same as Rosenberger et al)
    paHat<-(sa+1)/(na+2)
    pbHat<-(sb+1)/(nb+2)
    EstiProp<-TargProp2(paHat,pbHat) #define target proportion
    x1<-1/2
    y<-EstiProp
    g1<-y*(y/x1)^gamma/(y*(y/x1)^gamma+(1-y)*((1-y)/(1-x1))^gamma)
    p<-g1
    for (i in 1:d){
      u<-runif(3,0,1)
      e<-0
      if (u[1]<p){na<-na+1
      if (u[2]<pa){sa<-sa+1}
      e<-1}
      if (e==0){nb<-nb+1
      if (u[3]<pb){sb<-sb+1}}
#Use an adjustment by Agresti and Caffo (the same as Rosenbergeret al)
      paHat<-(sa+1)/(na+2)
      pbHat<-(sb+1)/(nb+2)
      EstiProp<-TargProp2(paHat,pbHat)
      x1<-na/(i+2)
      y<-EstiProp
      g1<-y*(y/x1)^gamma/(y*(y/x1)^gamma+(1-y)*((1-y)/(1-x1))^gamma)
      p<-g1
    }
    s<-sa+sb
    naCount[na]<-naCount[na]+1
```

```
    sCount[s+1]<-sCount[s+1]+1
#SD and NA/n
    p_cum=p_cum+(na/(na+nb))
    p_sd=append(p_sd,(na/(na+nb)))
    paHat<-(sa+1)/(na+2)
    #print(paHat)
    pbHat<-(sb+1)/(nb+2)
    ComSuc<-(s+2)/(d+4)
#Using CIs
    mu=pa-pb
    MuHat=paHat-pbHat
    SdHat=sqrt(paHat*(1-paHat)/(na)+pbHat*(1-pbHat)/(nb))
    E = qnorm(1-(alpha/2))*SdHat
    CI<-c(MuHat-E, MuHat + E)
    CI_RSIHR[n,]=CI
    if (CI_RSIHR[n,1]<=mu && mu<=CI_RSIHR[n,2]){countCI=countCI+1}
#Using Zdist
    Def<-paHat-pbHat
    varDef<-sqrt(paHat*(1-paHat)/na+pbHat*(1-pbHat)/nb)
    z<-Def/varDef
    zDistr[n]<-z
    c2<-0
    if (z<cutZ){c2<-1} ##one-sided test
    if (c2==0){zCount<-zCount+1}
  }
  mean_length=mean(CI_RSIHR[,2]-CI_RSIHR[,1])
  p_cum_r=p_cum/r #mean(NA/n)
  p_sd_r=sd(p_sd) #SD(NA/n)
  probNA<-naCount/r
  probS<-sCount/r
  powerZ<-zCount/r
```

```
  CountCI_r=countCI/r
  return(list(zDistr=zDistr,
        xNA=xNA,probNA=probNA,
        probS=probS,
        CI_RSIHR=CI_RSIHR,
        pa=pa,pb=pb,cutZ=cutZ,gamma=gamma,
        p_cum_r=p_cum_r,p_sd_r=p_sd_r,
        powerZ=powerZ,
        CountCI_r=CountCI_r
        ,mean_length=mean_length))}

###############
#YW
#Using the (DBCD)
epsilin1<-0.25
#gamma is the parameter in allocation function g(x,rho)
TargProp3<-function(pa,pb){qa<-1-pa
qb<-1-pb
rho<-(qb+epsilin1*min(qa,qb)*sign(pa-pb))/(qb+qa)
return(rho)}
# gamma is the parameter in allocation function g(x,rho);
# r is the repetition number.
# critical value for one-sided test
# d is the number of patients-2.
# success probability on treatment A pa
# success probability on treatment B pb
distrSDBC3<-function(alpha,pa,pb,gamma,d,r,cutZ){
  p_cum=0
  p_sd=c()
  naCount<-rep(0,d+1)
  xNA<-1:(d+1)
```

```
sCount<-rep(0,d+3)
xS<-0:(d+2)
zCount<-0
scCount<-0
CI=c()
CI_YW=matrix(0,r,2)
countCI=0
zDistr<-rep(0,r)
for (n in 1:r){ print(n)
  sa<-0
  sb<-0
  na<-1 # the first two patients
  nb<-1
  u1<-runif(2,0,1)
  if (u1[1]<pa) {sa<-sa+1}
  if (u1[2]<pb) {sb<-sb+1}
  paHat<-(sa+1)/(na+2)
  pbHat<-(sb+1)/(nb+2)
  EstiProp<-TargProp3(paHat,pbHat)
  x1<-1/2
  y<-EstiProp
  g1<-y*(y/x1)^gamma/(y*(y/x1)^gamma+(1-y)*((1-y)/(1-x1))^gamma)
  p<-g1
  for (i in 1:d){
    u<-runif(3,0,1)
    e<-0
    if (u[1]<p){na<-na+1
    if (u[2]<pa){sa<-sa+1}
    e<-1}
    if (e==0){nb<-nb+1
    if (u[3]<pb){sb<-sb+1}}
```

```
      paHat<-(sa+1)/(na+2)
      pbHat<-(sb+1)/(nb+2)
      EstiProp<-TargProp3(paHat,pbHat)
      x1<-na/(i+2)
      y<-EstiProp
      g1<-y*(y/x1)^gamma/(y*(y/x1)^gamma+(1-y)*((1-y)/(1-x1))^gamma)
      p<-g1
    }
    s<-sa+sb
    naCount[na]<-naCount[na]+1
    sCount[s+1]<-sCount[s+1]+1
#E(NA/n) and SD
    p_cum=p_cum+(na/(na+nb))
    p_sd=append(p_sd,(na/(na+nb)))
    paHat<-(sa+1)/(na+2)
    pbHat<-(sb+1)/(nb+2)
    ComSuc<-(s+2)/(d+4)
#Using CIs
    mu=pa-pb
    MuHat=paHat-pbHat
    SdHat=sqrt(paHat*(1-paHat)/(na)+pbHat*(1-pbHat)/(nb))
    E = qnorm(1-(alpha/2))*SdHat
    CI<-c(MuHat-E, MuHat + E)
    CI_YW[n,]=CI
    if (CI_YW[n,1]<=mu && mu<=CI_YW[n,2]){countCI=countCI+1}
#Using Zdist
    Def<-paHat-pbHat
    varDef<-sqrt(paHat*(1-paHat)/na+pbHat*(1-pbHat)/nb)
    z<-Def/varDef
    zDistr[n]<-z
    c2<-0
```

```
    if (z<cutZ){c2<-1} #one-sided test
    if (c2==0){zCount<-zCount+1}
  }
#Out comes
  mean_length=mean(CI_YW[,2]-CI_YW[,1])
  p_cum_r=p_cum/r #mean(NA/n)
  p_sd_r=sd(p_sd) #sd(NA/n)
  probNA<-naCount/r
  probS<-sCount/r
  powerZ<-zCount/r
  CountCI_r=countCI/r
  return(list(zDistr=zDistr,
        xNA=xNA,probNA=probNA,
        probS=probS,
        CI_YW=CI_YW,
        pa=pa,pb=pb,cutZ=cutZ,gamma=gamma,
        p_cum_r=p_cum_r, p_sd_r=p_sd_r,
        powerZ=powerZ,
        CountCI_r=CountCI_r
        ,mean_length=mean_length))}
\end{lstlisting}

Now we move forward to Bootstrap method in RAD as follows:
\begin{lstlisting}
###############
#Estimate pa and pb
#RPW
distrSRPW_matix<-function(alpha,cutZ,pa,pb,r,d){
  p_cum=0
  p_sd=c()
  naCount<-rep(0,d+1)
```

```
xNA<-1:(d+1)
sCount<-rep(0,d+3)
xS<-0:(d+2)
zCount<-0
zDistr<-rep(0,r)
CI=c()
CI_RPW=matrix(0,r,2)
MuHat_RPW=c()
countCI=0
SD=c()
pHat_estimate=matrix(NA,r,2)
for (n in 1:r){ print(n)
na<-1
nb<-1
sa<-0
sb<-0
u1<-runif(2,0,1)
if (u1[1]<pa) {sa<-sa+1}
if (u1[2]<pb) {sb<-sb+1}
aBall<-1 # RPW(1,1,1)
bBall<-1
Add<-1
p<-aBall/(aBall+bBall)
for (i in 1:d){
u<-runif(3,0,1)
e<-c(0,0,0)
if (u[1]<p){na<-na+1
if (u[2]<pa){sa<-sa+1
aBall<-aBall+Add
e[2]<-1}
if (e[2]==0){bBall<-bBall+Add}
```

```
  e[1]<-1}
  if (e[1]==0){nb<-nb+1
  if (u[3]<pb){sb<-sb+1
  bBall<-bBall+Add
  e[3]<-1}
  if (e[3]==0){aBall<-aBall+Add}
   }
  p<-aBall/(aBall+bBall)}
  s<-sa+sb
  naCount[na]<-naCount[na]+1
   sCount[s+1]<-sCount[s+1]+1
#SD and NA/n
  p_cum=p_cum+(na/(na+nb))
  p_sd=append(p_sd,(na/(na+nb)))
#Use DCBD adjustment
# Use an adjustment by Agresti and Caffo
  paHat<-(sa+1)/(na+2)
  pbHat<-(sb+1)/(nb+2)
#Convergence probability and Power of the test
  pHat_estimate[n,1]=paHat
  pHat_estimate[n,2]=pbHat
#Using CIs
  mu=pa-pb
  MuHat=paHat-pbHat
  SdHat=sqrt(paHat*(1-paHat)/(na)+pbHat*(1-pbHat)/(nb))
  E = qnorm(1-(alpha/2))*SdHat
  CI<-c(MuHat-E, MuHat + E)
  CI_RPW[n,]=CI
  MuHat_RPW[n]=MuHat
  if (CI_RPW[n,1]<=mu && mu<=CI_RPW[n,2]){countCI=countCI+1}
#With test statistics and cutZ
```

```
    z<-(paHat-pbHat)/sqrt(paHat*(1-paHat)/(na)+pbHat*(1-pbHat)/(nb))#power wald
    zDistr[n]<-z
    c1<-0
    if (z<cutZ){c1<-1} #for one-sided
    if (c1==0){zCount<-zCount+1}
  }
#Out Comes
  mean_length=mean(CI_RPW[,2]-CI_RPW[,1])
  p_cum_r=p_cum/r #Mean (NA/n)
  p_sd_r=sd(p_sd) #SD(NA/n)
  probNA<-naCount/r  #NA count
  probS<-sCount/r  # Success count
  powerZ<-zCount/r #simulated power
  CountCI_r=countCI/r #Simulated CI
  return(list(zDistr=zDistr,
         xNA=xNA,probNA=probNA,
         probS=probS,
         powerZ=powerZ,
         pa=pa,pb=pb,cutZ=cutZ,
         CI_RPW=CI_RPW,
         p_cum_r=p_cum_r,
         p_sd_r=p_sd_r,
         CountCI_r=CountCI_r,pHat_estimate=pHat_estimat,
         mean_length=mean_length,MuHat_RPW=MuHat_RPW
  ))}
#Bootstrap
#RPW
distrSRPW_boot<-function(alpha,cutZ,B,BB,d,data){
  p_cum=0
  p_sd=c()
  naCount<-rep(0,d+1)
```

```
  xNA<-1:(d+1)
  sCount<-rep(0,d+3)
  xS<-0:(d+2)
  zCount<-0
  zDistr<-rep(0,B)
  CI=c()
  #CI_RPW=matrix(0,r,2)
  co=0
  SD=c()
  #mu_diff=c()
  CI_RPW=matrix(0,BB,2)
  #CI_RPWb=matrix(0,BB,2)
  boot_count=0
  #Cov_countb=0
  for(b in 1:BB){ print(b)
    pa=data[b,1]
    pb=data[b,2]
    pa_boot=pb_boot=c()
    theta1=as.vector(NULL)
    theta2=as.vector(NULL)
    theta=as.vector(NULL)
    #sort_theta=as.vector(NULL)
    p_theorya=as.vector(NULL)
    p_theory=as.vector(NULL)
    Quantile_Boot=as.vector(NULL)
    for (n in 1:B){#print(n)##Count the simulation run
#the first two patients are assigned to the two trs,one on each
    na<-1
    nb<-1
    sa<-0
    sb<-0
```

```
    u1<-runif(2,0,1)
    if (u1[1]<data[b,1]) {sa<-sa+1}
    if (u1[2]<data[b,2]) {sb<-sb+1}
    aBall<-1 # RPW(1,1,1)
    bBall<-1
    Add<-1
    p<-aBall/(aBall+bBall)
    #print(p)
    for (i in 1:d){
    u<-runif(3,0,1)
    e<-c(0,0,0)
    if (u[1]<p){na<-na+1
    if (u[2]<data[b,1]){sa<-sa+1
    aBall<-aBall+Add
    e[2]<-1}
    if (e[2]==0){bBall<-bBall+Add}
    e[1]<-1}
    if (e[1]==0){nb<-nb+1
    if (u[3]<data[b,2]){sb<-sb+1
    bBall<-bBall+Add
    e[3]<-1}
    if (e[3]==0){aBall<-aBall+Add}
    }
    p<-aBall/(aBall+bBall)
     }
    s<-sa+sb
    naCount[na]<-naCount[na]+1
    sCount[s+1]<-sCount[s+1]+1
#Use DCBD adjustment
    paHat<-(sa+1)/(na+2)
    pbHat<-(sb+1)/(nb+2)
```

```
    pa_boot=c(pa_boot,paHat)
    pb_boot=c(pb_boot,pbHat)
    }
    theta1=c(theta1,pa_boot)
    theta2=c(theta2,pb_boot)
    theta=sort(theta1-theta2)
    Quantile_Boot<- quantile(theta,probs =c(alpha/2,1-(alpha/2)))
    CI_RPW[b,]=Quantile_Boot
    if(CI_RPW[b,1]<=0.2 && 0.2<=CI_RPW[b,2]){boot_count=boot_count+1}
  }
  p_boot_count=boot_count/(BB)
  mean_length=mean(CI_RPW[,2]-CI_RPW[,1])
  return(list(CI_RPW=CI_RPW,
        p_boot_count=p_boot_count,
        mean_length=mean_length
        ))}
#Estimate pa and pb
#RSHIHR
distrSDBC2_matrix<-function(alpha,pa,pb,gamma,d,r,cutZ){
  pHat_estimate=matrix(NA,r,2)
  p_cum=0
  p_sd=c()
  naCount<-rep(0,d+1)
  xNA<-1:(d+1)
  sCount<-rep(0,d+3)
  xS<-0:(d+2)
  zCount<-0
  scCount<-0
  zDistr<-rep(0,r)
  CI=c()
  CI_RSIHR=matrix(0,r,2)
```

```
  MuHat_RSHIR=c()
  countCI=0
#SD=c()
  for (n in 1:r){ print(n)
    sa<-0
    sb<-0
    na<-1
    nb<-1
    u1<-runif(2,0,1)
    if (u1[1]<pa) {sa<-sa+1}
    if (u1[2]<pb) {sb<-sb+1}
    ## Use an adjustment by Agresti and Caffo (the same as Rosenberger et al)
    paHat<-(sa+1)/(na+2)
    pbHat<-(sb+1)/(nb+2)
    EstiProp<-TargProp2(paHat,pbHat)
    x1<-1/2
    y<-EstiProp
    g1<-y*(y/x1)^gamma/(y*(y/x1)^gamma+(1-y)*((1-y)/(1-x1))^gamma)
    p<-g1
    for (i in 1:d){
    u<-runif(3,0,1)
    e<-0
    if (u[1]<p){na<-na+1
    if (u[2]<pa){sa<-sa+1}
    e<-1}
    if (e==0){nb<-nb+1
    if (u[3]<pb){sb<-sb+1}}
    paHat<-(sa+1)/(na+2)
    pbHat<-(sb+1)/(nb+2)
    EstiProp<-TargProp2(paHat,pbHat)
    x1<-na/(i+2)
```

```
    y<-EstiProp
    g1<-y*(y/x1)^gamma/(y*(y/x1)^gamma+(1-y)*((1-y)/(1-x1))^gamma)
    p<-g1
    }
    s<-sa+sb
    naCount[na]<-naCount[na]+1
    sCount[s+1]<-sCount[s+1]+1
#SD and NA/n
    p_cum=p_cum+(na/(na+nb))
    p_sd=append(p_sd,(na/(na+nb)))
    paHat<-(sa+1)/(na+2)
    #print(paHat)
    pbHat<-(sb+1)/(nb+2)
    ComSuc<-(s+2)/(d+4)
#with CIS
    pHat_estimate[n,1]=paHat
    pHat_estimate[n,2]=pbHat
    mu=pa-pb
    MuHat=paHat-pbHat
    SdHat=sqrt(paHat*(1-paHat)/(na)+pbHat*(1-pbHat)/(nb))
    E = qnorm(1-(alpha/2))*SdHat
    CI<-c(MuHat-E, MuHat + E)
    MuHat_RSHIR[n]=MuHat
    CI_RSIHR[n,]=CI
    if (CI_RSIHR[n,1]<=mu && mu<=CI_RSIHR[n,2]){countCI=countCI+1}
#With zdist
    Def<-paHat-pbHat
    varDef<-sqrt(paHat*(1-paHat)/na+pbHat*(1-pbHat)/nb)
    z<-Def/varDef
    zDistr[n]<-z
    c2<-0
```

```
    if (z<cutZ){c2<-1} #one-sided test
    if (c2==0){zCount<-zCount+1}
  }
  mean_length=mean(CI_RSIHR[,2]-CI_RSIHR[,1])
  p_cum_r=p_cum/r #mean(NA/n)
  p_sd_r=sd(p_sd) #SD(NA/n)
  probNA<-naCount/r
  probS<-sCount/r
  powerZ<-zCount/r
  CountCI_r=countCI/r
  return(list(zDistr=zDistr,
     xNA=xNA,probNA=probNA,
     probS=probS,
     CI_RSIHR=CI_RSIHR,
     powerZ=powerZ,
     pa=pa,pb=pb,cutZ=cutZ,gamma=gamma,
     p_cum_r=p_cum_r,CountCI_r=CountCI_r,p_sd_r=p_sd_r,
     pHat_estimate=pHat_estimate,mean_length=mean_length,
     MuHat_RSHIR=MuHat_RSHIR))}
#Bootstrap
#RSHIR
distrSDBC2_boot<-function(alpha,B,BB,gamma,d,data,cutZ){
 p_cum=0
  p_sd=c()
  naCount<-rep(0,d+1)
  xNA<-1:(d+1)
  sCount<-rep(0,d+3)
  xS<-0:(d+2)
  zCount<-0
  zDistr<-rep(0,B)
  CI=c()
```

```
SD=c()
CI_RSIHR=matrix(0,BB,2)
boot_count=0
for(b in 1:BB){ print(b)
  pa=data[b,1]
  pb=data[b,2]
  pa_boot=pb_boot=c()
  theta1=as.vector(NULL)
  theta2=as.vector(NULL)
  theta=as.vector(NULL)
  p_theorya=as.vector(NULL)
  p_theory=as.vector(NULL)
  Quantile_Boot=as.vector(NULL)
  for (n in 1:B){   #print(n)##Count the simulation run
  sa<-0
  sb<-0
  na<-1
  nb<-1
  u1<-runif(2,0,1)
  if (u1[1]<data[b,1]) {sa<-sa+1}
  if (u1[2]<data[b,2]) {sb<-sb+1}
  paHat<-(sa+1)/(na+2)
  pbHat<-(sb+1)/(nb+2)
  EstiProp<-TargProp2(paHat,pbHat}
  x1<-1/2
  y<-EstiProp
  g1<-y*(y/x1)^gamma/(y*(y/x1)^gamma+(1-y)*((1-y)/(1-x1))^gamma)
  p<-g1
  for (i in 1:d){
  u<-runif(3,0,1)
  e<-0
```

```
    if (u[1]<p){na<-na+1
    if (u[2]<data[b,1]){sa<-sa+1}
    e<-1}
    if (e==0){nb<-nb+1
    if (u[3]<data[b,2]){sb<-sb+1}}
    paHat<-(sa+1)/(na+2)
    pbHat<-(sb+1)/(nb+2)
    EstiProp<-TargProp2(paHat,pbHat)
    x1<-na/(i+2)
    y<-EstiProp
    g1<-y*(y/x1)^gamma/(y*(y/x1)^gamma+(1-y)*((1-y)/(1-x1))^gamma)
    p<-g1
    }
    s<-sa+sb
    naCount[na]<-naCount[na]+1
    sCount[s+1]<-sCount[s+1]+1
    paHat<-(sa+1)/(na+2)
    pbHat<-(sb+1)/(nb+2)
    ComSuc<-(s+2)/(d+4)
    pa_boot=c(pa_boot,paHat)
    pb_boot=c(pb_boot,pbHat)
    }
    theta1=c(theta1,pa_boot)
    theta2=c(theta2,pb_boot)
    theta=sort(theta1-theta2)
    Quantile_Boot<- quantile(theta,probs =c(alpha/2,1-(alpha/2)))
    CI_RSIHR[b,]=Quantile_Boot
    if(CI_RSIHR[b,1]<=0.2 && 0.2<=CI_RSIHR[b,2]){boot_count=boot_count+1}
  }
  p_boot_count=boot_count/(BB)
  mean_length=mean(CI_RSIHR[,2]-CI_RSIHR[,1])
```

```
  return(list(CI_RSIHR=CI_RSIHR,
         p_boot_count=p_boot_count,
         mean_length=mean_length
         ))}
#Estimate pa and pb
#YW
distrSDBC3_matrix<-function(alpha,pa,pb,gamma,d,r,cutZ){
  pHat_estimate=matrix(NA,r,2)
  p_cum=0
  p_sd=c()
  naCount<-rep(0,d+1)
  xNA<-1:(d+1)
  sCount<-rep(0,d+3)
  xS<-0:(d+2)
  zCount<-0
  scCount<-0
  zDistr<-rep(0,r)
  CI=c()
  CI_YW=matrix(0,r,2)
  MuHat_YW=c()
  countCI=0
#SD=c()
  for (n in 1:r){ print(n)
    sa<-0
    sb<-0
    na<-1
    nb<-1
    u1<-runif(2,0,1)
    if (u1[1]<pa) {sa<-sa+1}
    if (u1[2]<pb) {sb<-sb+1}
    paHat<-(sa+1)/(na+2)
```

```
    pbHat<-(sb+1)/(nb+2)
    EstiProp<-TargProp3(paHat,pbHat)
    x1<-1/2
    y<-EstiProp
    g1<-y*(y/x1)^gamma/(y*(y/x1)^gamma+(1-y)*((1-y)/(1-x1))^gamma)
    p<-g1
    for (i in 1:d){
    u<-runif(3,0,1)
    e<-0
    if (u[1]<p){na<-na+1
    if (u[2]<pa){sa<-sa+1}
    e<-1}
    if (e==0){nb<-nb+1
    if (u[3]<pb){sb<-sb+1}}
    paHat<-(sa+1)/(na+2)
    pbHat<-(sb+1)/(nb+2)
    EstiProp<-TargProp3(paHat,pbHat)
    x1<-na/(i+2)
    y<-EstiProp
    g1<-y*(y/x1)^gamma/(y*(y/x1)^gamma+(1-y)*((1-y)/(1-x1))^gamma)
    p<-g1
    }
    s<-sa+sb
    naCount[na]<-naCount[na]+1
    sCount[s+1]<-sCount[s+1]+1
#SD and NA/n
    p_cum=p_cum+(na/(na+nb))
    p_sd=append(p_sd,(na/(na+nb)))
    paHat<-(sa+1)/(na+2)
    pbHat<-(sb+1)/(nb+2)
    ComSuc<-(s+2)/(d+4)
```

```
#with CIS
    pHat_estimate[n,1]=paHat
    pHat_estimate[n,2]=pbHat
    ####
    mu=pa-pb
    MuHat=paHat-pbHat
    SdHat=sqrt(paHat*(1-paHat)/(na)+pbHat*(1-pbHat)/(nb))
    E = qnorm(1-(alpha/2))*SdHat
    CI<-c(MuHat-E, MuHat + E)
    MuHat_YW[n]=MuHat
    CI_YW[n,]=CI
    if (CI_YW[n,1]<=mu && mu<=CI_YW[n,2]){countCI=countCI+1}
#With zdist
    Def<-paHat-pbHat
    varDef<-sqrt(paHat*(1-paHat)/na+pbHat*(1-pbHat)/nb)
    z<-Def/varDef
    zDistr[n]<-z
    c2<-0
    if (z<cutZ){c2<-1} #one-sided test
    if (c2==0){zCount<-zCount+1}
  }
  mean_length=mean(CI_YW[,2]-CI_YW[,1])
  p_cum_r=p_cum/r #mean(NA/n)
  p_sd_r=sd(p_sd) #SD(NA/n)
  probNA<-naCount/r
  probS<-sCount/r
  powerZ<-zCount/r
  CountCI_r=countCI/r
  return(list(zDistr=zDistr,
     xNA=xNA,probNA=probNA,
     probS=probS,
```

```
     CI_YW=CI_YW,
     powerZ=powerZ,
     pa=pa,pb=pb,cutZ=cutZ,gamma=gamma,
     p_cum_r=p_cum_r,CountCI_r=CountCI_r,p_sd_r=p_sd_r
    ,pHat_estimate=pHat_estimate,mean_length=mean_length,
    MuHat_YW=MuHat_YW))}

#Bootstrap
#YW
distrSDBC3_boot<-function(alpha,B,BB,gamma,d,data,cutZ){
  p_cum=0
  p_sd=c()
  naCount<-rep(0,d+1)
  xNA<-1:(d+1)
  sCount<-rep(0,d+3)
  xS<-0:(d+2)
  zCount<-0
  zDistr<-rep(0,B)
  CI=c()
  SD=c()
  CI_YW=matrix(0,BB,2)
  boot_count=0
  #Cov_countb=0
  for(b in 1:BB){ print(b)
    pa=data[b,1]
    pb=data[b,2]
    pa_boot=pb_boot=c()
    theta1=as.vector(NULL)
    theta2=as.vector(NULL)
    theta=as.vector(NULL)
    p_theorya=as.vector(NULL)
```

```
p_theory=as.vector(NULL)
Quantile_Boot=as.vector(NULL)
for (n in 1:B){
sa<-0
sb<-0
na<-1
nb<-1
u1<-runif(2,0,1)
if (u1[1]<data[b,1]) {sa<-sa+1}
if (u1[2]<data[b,2]) {sb<-sb+1}
paHat<-(sa+1)/(na+2)
pbHat<-(sb+1)/(nb+2)
EstiProp<-TargProp3(paHat,pbHat)
x1<-1/2
y<-EstiProp
g1<-y*(y/x1)^gamma/(y*(y/x1)^gamma+(1-y)*((1-y)/(1-x1))^gamma)
p<-g1
for (i in 1:d){
u<-runif(3,0,1)
e<-0
if (u[1]<p){na<-na+1
if (u[2]<data[b,1]){sa<-sa+1}
e<-1}
if (e==0){nb<-nb+1
if (u[3]<data[b,2]){sb<-sb+1}}
paHat<-(sa+1)/(na+2)
pbHat<-(sb+1)/(nb+2)
EstiProp<-TargProp3(paHat,pbHat)
x1<-na/(i+2)
y<-EstiProp
g1<-y*(y/x1)^gamma/(y*(y/x1)^gamma+(1-y)*((1-y)/(1-x1))^gamma)
```

```
  p<-g1
}
s<-sa+sb
naCount[na]<-naCount[na]+1
sCount[s+1]<-sCount[s+1]+1
paHat<-(sa+1)/(na+2)
pbHat<-(sb+1)/(nb+2)
ComSuc<-(s+2)/(d+4)
pa_boot=c(pa_boot,paHat)
pb_boot=c(pb_boot,pbHat)
  }
  theta1=c(theta1,pa_boot)
  theta2=c(theta2,pb_boot)
  theta=sort(theta1-theta2)
  Quantile_Boot<- quantile(theta,probs =c(alpha/2,1-(alpha/2)))
  CI_YW[b,]=Quantile_Boot
  if(CI_YW[b,1]<=0.2 && 0.2<=CI_YW[b,2]){boot_count=boot_count+1}
}
p_boot_count=boot_count/(BB)
mean_length=mean(CI_YW[,2]-CI_YW[,1])
return(list(CI_YW=CI_YW,
      p_boot_count=p_boot_count,
      mean_length=mean_length))}
```

Bibliography

[1] Louis HY Chen. Poisson approximation for dependent trials. *The Annals of Probability*, pages 534–545, 1975.

[2] Louis HY Chen, Larry Goldstein, and Qi-Man Shao. *Normal approximation by Stein's method*. Springer Science & Business Media, 2010.

[3] Louis HY Chen and Qi-Man Shao. Stein's method for normal approximation. *An introduction to Stein's method*, 4:1–59, 2005.

[4] Jeffrey R Eisele. The doubly adaptive biased coin design for sequential clinical trials. *Journal of Statistical Planning and Inference*, 38(2):249–261, 1994.

[5] Robert E Gaunt, Alastair M Pickett, Gesine Reinert, et al. Chi-square approximation by stein's method with application to pearson's statistic. *The Annals of Applied Probability*, 27(2):720–756, 2017.

[6] Robert E Gaunt and Gesine Reinert. The rate of convergence of some asymptotically chi-square distributed statistics by stein's method. *arXiv preprint arXiv:1603.01889*, 2016.

[7] Feifang Hu and William F Rosenberger. Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons. *Journal of the American Statistical Association*, 98(463):671–678, 2003.

[8] Feifang Hu and William F Rosenberger. *The theory of response-adaptive randomization in clinical trials*, volume 525. John Wiley & Sons, 2006.

[9] Feifang Hu, Li-Xin Zhang, et al. Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials. *The Annals of Statistics*, 32(1):268–301, 2004.

[10] Christophe Ley, Gesine Reinert, Yvik Swan, et al. Approximate computation of expectations: a canonical stein operator. *arXiv preprint arXiv:1408.2998*, 2014.

[11] Christophe Ley, Gesine Reinert, Yvik Swan, et al. Stein's method for comparison of univariate distributions. *Probability Surveys*, 14:1–52, 2017.

[12] Vincent F Melfi, Connie Page, and Margarida Geraldes. An adaptive randomized design with application to estimation. *Canadian Journal of Statistics*, 29(1):107–116, 2001.

[13] William Rosenberger and John Lachin. *Randomization in Clinical Trials: Theory and Practice*. 01 2002.

[14] William F Rosenberger, Nancy Flournoy, and Stephen D Durham. Asymptotic normality of maximum likelihood estimators from multiparameter response-driven designs. *Journal of Statistical Planning and Inference*, 60(1):69–76, 1997.

[15] William F Rosenberger and Feifang Hu. Bootstrap methods for adaptive designs. *Statistics in medicine*, 18(14):1757–1767, 1999.

[16] William F Rosenberger, Nigel Stallard, Anastasia Ivanova, Cherice N Harper, and Michelle L Ricks. Optimal adaptive designs for binary response trials. *Biometrics*, 57(3):909–913, 2001.

[17] Charles Stein. Approximate computation of expectations. *Lecture Notes-Monograph Series*, 7:i–164, 1986.

[18] Charles Stein et al. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*. The Regents of the University of California, 1972.

[19] LJ Wei and S Durham. The randomized play-the-winner rule in medical trials. *Journal of the American Statistical Association*, 73(364):840–843, 1978.

[20] Yanqing Yi and Xuan Li. Response adaptive designs with asymptotic optimality. *Canadian Journal of Statistics*, 46(3):458–469, 2018.

[21] Yanqing Yi and Xikui Wang. Goodness-of-fit test for response adaptive clinical trials. *Statistics & probability letters*, 77(10):1014–1020, 2007.

[22] Yanqing Yi and Xikui Wang. Response adaptive designs with a variance-penalized criterion. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 51(5):763–773, 2009.

[23] Lanju Zhang and William F Rosenberger. Response-adaptive randomization for clinical trials with continuous outcomes. *Biometrics*, 62(2):562–569, 2006.