12-2015

Simulations of a New Response-Adaptive Biased Coin Design

Aleksandra Stein
University of Nebraska-Lincoln, aleksandra.stein@gmail.com

SIMULATIONS OF A NEW RESPONSE-ADAPTIVE BIASED COIN DESIGN

by

Aleksandra M. Stein

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Statistics

Under the Supervision of Professor Kent Eskridge

Lincoln, Nebraska

December, 2015

SIMULATIONS OF A NEW RESPONSE-ADAPTIVE BIASED COIN DESIGN

Aleksandra M. Stein, Ph. D.

University of Nebraska, 2015

Adviser: Kent M. Eskridge

Modern medical experiments accrue and treat patients—hence obtain treatment response data—throughout a trial. Designs which prospectively plan to modify patient allocation by leveraging accumulating data are response-adaptive randomization (RAR) designs. Many such designs attempt to balance the desire to bias assignment proportions towards a treatment which is performing better against the need to maintain randomization in the face of continued equipoise.

This dissertation consists of simulated investigations into frequentist and ethical properties of an new RAR biased coin design. Chapter 2 proposes a new adaptive design for phase III clinical trials, a modification of the 2001 Bandyopadhyay and Biswas biased coin design. Simulations show how the new design continues to ethically expose patients to the better treatment while simultaneously mitigating power loss inherent in the original design. Chapters 2 and 3 expand the applicability of the new design to scenarios where treatment variances or covariate-treatment impacts are unequal. In Chapter 4, simulations demonstrate that the new response-adaptive biased coin design can be more ethical than equal allocation, even when patient outcomes are not immediately available. Each chapter illustrates the utility and benefits of the new design through a real-world application of an HIV treatment adherence intervention. Asymptotic results are applied to a special case of the BBS design and small sample implications are compared with simulated outcomes in Chapter 5.

ACKNOWLEDGMENTS

Mahalo nui loa to my advisor and to my life partner for the years and years of encouragement. Thank you for helping me keep at it.

Mahalo also to my committee for the valuable feedback and the repeated readings. I appreciate your tireless support from across the ocean!

Table of Contents

Chapter 1

Introduction

1.1 Motivation

1.1.1 Curing Scurvy

In 1747, James Lind performed one of the earliest recorded clinical trials [69]. Aboard the *Salisbury*, Lind selected 12 sailors beset with a homogeneous display of the symptoms of scurvy and assigned them haphazardly to six proposed treatment regimens—with otherwise identical living conditions and diet. Dr. Lind believed that the methods were the only distinguishing factor among his patients, so differences in their recoveries would be due solely to their medications. As with modern medical trials, his goal was to determine the best of the treatments and henceforth promote it as the standard therapy for afflicted mariners.

Considering the state of medicine at that epoch, Lind's *Treatise on the scurvy* provides outstanding descriptive documentation of the accumulated evidence in favor of a treatment. Employing a liberal definition of randomization, James Lind's investigation falls under the umbrella of equal-allocation, randomized clinical trials.

According to modern medical standards, however, this experiment deserves further scrutiny. On the one hand, since Dr. Lind already suspects citrus fruits will yield better results, he is arguably not in a state of equipoise and thus is ethically bound to offer his patients the best available medication [3, 106]. From this perspective, the *Salisbury* trial violated the basic human rights of its crew. On the other hand, since Dr. Lind's suspicions are based solely on anecdotal evidence and previously unjustified beliefs, it is also reasonable to think that Dr. Lind and the medical community at large are in a state of equipoise [1, 2, 37, 84]. With this approach, Lind is right to offer multiple treatment options and does not breach his patients' rights in randomizing their care.

These conflicting viewpoints trouble clinicians and researchers even today. Is it acceptable for medical personnel to randomize patients equally across all treatments if there is a suspicion that one treatment is superior? Does the answer hold if that suspicion is completely unconfirmed by data? What happens when enough information exists to suggest the superiority of one treatment but not to confirm it? Is equal randomization still ethical? Is any randomization still acceptable? Adaptive clinical trial designs and particularly response-adaptive randomization schema explore ethical patient allocations. In particular, such techniques attempt to balance the desire to bias assignment proportions towards a treatment which is performing better while maintaining the necessary randomization in the face of continued equipoise.

1.1.2 Equipoise and Accumulating Evidence

Randomized clinical trials (RCTs) still enjoy "gold standard" status in clinical research [14, 54, 64, 84], despite design origins in agricultural settings. This is likely due in large part to the desire to achieve a sound scientific answer based on statistical

properties of power and balance—and perhaps due in small part to historical precedence and blind faith [48, 74, 81]. The ethics of such contemporary clinical trials are hotly debated, covering the entire spectrum of beliefs prioritizing individual patient care at one extreme and collective benefit at the other. Topics of dissension include minor facets such as the benefits of covariates or interim analyses, as well as substantial issues like equipoise and randomization. One well-packaged debate on equipoise, the personal care principle, and RCTs may be viewed in the article by Royall [84], ensuing comments—particularly those of Byar [37] and Simes [87]—and Royall's rejoinder [85] in 1991. For over 80 ethical essays in 10 parts, consult the compilation edited by Emanuel, Crouch, Arras, and Moreno [47] in 2004. A full array of *Ethics of Medical Research on Humans* is presented by Foster [54] including three distinct approaches to moral beliefs, three different levels of equipoise, and discussion about the personal care principle for doctors versus for researchers, including which role should trump the other. While many older and classical sources of ethical exposition exist in the literature, the latter two resources [47, 54] broach novel ethical topics which have only been considered recently, due to technological advances, including statistical concerns.

Dovetailing on the issues of equipoise is the problem of accumulating evidence. Flehinger and Louis summarize the quandary in [50]:

> Once the decision to initiate the trials is made, there is an obligation to assign effectively equal numbers of patients to the two competing methods of treatment until a criterion determined in advance by significance and power considerations is met. On the other hand, if there is a real difference between the two treatments, it often becomes apparent from the accumulating data long before the criterion end-point is reached. When

> this happens the physician must choose between treating very sick patients with a method which appears to be inferior and terminating the trials before a statistically valid conclusion has been reached.

With the aim of striking a balance between the conflicting goals of offering the best individual patient care and confidently discerning the best treatment option for the collective patient horizon, researchers began considering adaptive allocation designs [42, 50, 106]. Pioneers of adaptive randomization schemes include Thompson in 1933 with his article "On the likelihood that one unknown probability exceeds another in view of the evidence of the two samples" [96] and in 1952 Robbins' "Some aspects of the sequential design of experiments" [78]. Broader study of sequential designs ensues [2, 100] with prominent statisticians both encouraging [84, 88, 105] and discouraging [76] the use of such designs in the early stages of study. Throughout the late 1900's, the statistical literature continues to accumulate [1, 2, 3, 42, 50, 101, 103, 102, 104], but medical applications lag behind [48, 79, 82]. An early view of the benefits of the compromise gained through adaptive randomization is that of Cornfield, Halperin, and Greenhouse in 1969 [43]:

> The usual ethical justification for not administering an agent of possible efficacy to all patients is the absence of definite information about its effectiveness. However satisfactory this justification may be before the trial starts it rapidly loses cogency as evidence for or against the agent accumulates during the course of the trial. ... the allocation of proportionately more and more of the future patients to the apparently better treatment at least reduces this ethical problem.

Despite statistical enthusiasm, even the most intuitive designs are received with misapprehension by the medical community. Clinicians not only reject the use of such

designs, but also refuse to absorb the results of trials which employed them.

1.1.3 Response-Adaptive Randomization as a Problematic Solution

A large and discouraging black mark for early response-adaptive randomization designs is the poor reception of adaptive allocation designs involving extracorporeal membrane oxygenation (ECMO) in infants with respiratory failure. The ECMO trials (sequentially [23, 72, 98]) remain the most widely debated and most frequently cited application of adaptive allocation even in recent literature [22, 74, 82, 84, 93]. Because the doctors who performed the first trial believed that ECMO was more effective than the conventional therapy at the time, they opted to employ a randomized play-the-winner design derived from [104] and [106].

Bartlett *et al.* employ an urn model allocation scheme which assigns the first patient a 50% probability of being exposed to each treatment. In the model, one ball labeled "ECMO" and one ball labeled "standard care" are placed into the urn. A ball is chosen at random from the urn and the first patient is allocated to the treatment type of the selected ball—in this case, ECMO. The urn is updated by adding a ball labeled "ECMO" when a patient randomized to ECMO survived or when one receiving conventional therapy died and vice-versa for adding a ball labeled "standard care". The total sample size of the trial is undefined and dependent on the undetermined allocations as well as the patient results. Hence, the researchers implement a stopping rule, "which detected a high probability of selection of the better treatment." Namely, the trial terminates after ten balls of one type are added. During the study, one patient dies while receiving standard care and eleven survive on ECMO [23].

Retrospectively, the clinicians suggest that a similar initial allocation might have been better provided by multiple pairs of balls of each type in the urn, rather than a single starting pair [23]. Such a change could have placed more than one patient on conventional therapy for a more typical trial randomization; however, exposing more patients to the standard care may also have increased the number of unnecessary deaths [82]. Indeed, while the ECMO results encourage many doctors and medical centers [22], others express incredulity at an RCT with only a single patient on conventional treatment [82]. Ultimately, a second, adaptive experiment is performed at Harvard [72] followed by a third, nonadaptive investigation implemented throughout the UK [98]. Harvard's evaluation reaches the same conclusions as its predecessor. The UK assessment is forced to stop early due to the unethical nature of continuing an equal-allocation RCT in view of the overwhelmingly positive results of ECMO in the first part of the experiment.

The authors of the first study muse that the later studies are, "criticized for being unnecessary and unethical, just as ours was criticized as being unbelievable and unethical," [22]. The phenomenon of disapproval from both sides is to be expected, suggest Hu and Rosenberger [59]:

> We believe that because [response-adaptive randomization] represents a middle ground between the community benefit and the individual patient benefit, it is subject to attack from either side.

Despite their rocky introduction to the medical community, response-adaptive randomization (RAR) designs do provide a compromise between the collective ethics of equal-allocation RCTs and the personal care principle demanded by individual ethics. In the past decade, researchers and regulatory agencies alike have requested further review of adaptive designs and urged their implementation with few objections

[8]. Refer to Berry's arguments for Bayesian adaptive designs [24, 25, 26] (or that of Thall [95]); Hu and Rosenberger's prolific writings including [79, 80, 93] which focus on RAR; as well as a couple contributions from Baldi Antognini and Giovagnoli, also on RAR [12, 14]. Additional supplications surface in discussion of the necessity of data-dependent designs by Palmer [74] and in direction provided by the Food and Drug Administration (FDA) and the U.S. Department of Health and Human Services (DHHS) [52, 53, 99]. Moreover, at least five adaptive design books have appeared in the past decade [27, 39, 57, 59, 77], all advocating application of adaptive designs in current clinical trials.

While recent literature leads us to conclude that particular RAR designs and their properties are becoming well-understood [93], there remains vast room for improvement in adaptive allocation rules. For example, the majority of clinical trials today are not interested in a single, immediately observable binary response from a large number of strictly homogeneous patients assigned to either the treatment or control arm. Instead, RCTs may have multiple treatments each with multiple endpoints, continuous and delayed responses, and a patient population replete with covariates which change over time.

1.1.4 Layout

This dissertation consists of simulated investigations into frequentist and ethical properties of an new response-adaptive randomization biased coin design under three important clinical trial scenarios. The first section of Chapter 1 illustrates the conflict medical researchers face with clinical trials and motivates the necessity of response-adaptive randomization. The remainder of Chapter 1 provides a thorough literature review covering other approaches to patient allocation, a statistical framework for

response-adaptive randomization, and some history surrounding the Bandyopadhyay and Biswas biased coin design of 2001.

Chapter 2 proposes a new adaptive design for phase III clinical trials, a modification of the Bandyopadhyay and Biswas design with the constant tuning parameter replaced by an adaptive estimate. Simulations show how the new design continues to ethically expose patients to the better treatment when a treatment difference exists while simultaneously mitigating power loss inherent in the original design. The new design also expands the applicability of the new response-adaptive randomization to more real-world clinical trial scenarios, particularly ones where the treatment variances are unequal and/or known before the study. The utility and benefits of the modified design are illustrated through a real-world application of an HIV treatment adherence intervention.

Chapter 3 further broadens the scope of the new design's applicability. Simulations in this chapter confirm the design's ethicality even when covariate-treatment interactions are present. The utility and benefits of the modified design are applied to the HIV trial from Chapter 2 with the additional complexity of having covariate impacts to patient outcomes which vary by treatment.

In Chapter 4, simulations demonstrate that the new response-adaptive biased coin design can be more ethical than equal allocation, even when patient outcomes are not immediately available. Moreover, the design's power improves as patient responses are delayed and the design tends towards complete randomization. Delay in patient responses are incorporated into the HIV trial from Chapter 2, once again transforming a balanced assignment trial into an ethical, response-adaptive study.

Chapter 5 discusses the results and implications of the findings of this dissertation. Asymptotic results are applied to a special case of the BBS design and small sample implications are compared with simulated outcomes from earlier Chapters. Further

areas of study are suggested to continue the important progress in this area of ethical research.

1.2 Background

1.2.1 Balance

As noted previously, balanced randomized clinical trials are the favored design for today's medical experiments [14, 74, 81, 84]. Equal allocation is frequently performed in the style of a completely randomized design (CRD) or a permuted block design. Other approaches include biased coin designs and urn models which force balance even in small trials.

1.2.1.1 Complete Randomization

The CRD is adopted from the methods in the 1935 book of R. A. Fisher regarding the design of (mostly agricultural) experiments [49]. A CRD allocates patients to treatments such that each patient has an equal opportunity of being randomized to each treatment in the trial. While this randomization scheme is asymptotically balanced, randomizing each patient unconditionally may cause a significant imbalance in treatments for smaller trials. For this reason, restricted randomization schemes have been proposed which allocate subjects to treatments randomly, conditioned on previous patient assignments. Patient outcomes do not factor into restricted randomization. The two main types of restricted randomization are permuted block designs and biased coin designs [68].

Biased coin designs (BCDs) target balanced patient allocation by assigning subjects to treatments with equal probability if the past assignments are balanced, or

with unequal probability if past assignments are not balanced—biasing allocations towards equal treatment exposure. As implied by the two-sided nature of a coin, BCDs tend to be restricted to two treatments; however, expansions exist, particularly in the form of urn models [9, 10, 11, 92, 113]. While permuted block designs are often implemented in practice, BCDs have not received such a warm welcome. Because block designs are simple and well-understood, we briefly describe them first in Section 1.2.1.2. Since BCDs are, to many, the foundation of response-adaptive randomization rules, we next introduce them briefly from the vantage of Efron and Wei—both early proponents but with distinct perspectives—in Sections 1.2.1.3 and 1.2.1.4, respectively.

1.2.1.2 Permuted Block Designs

The most commonly applied alternative to a CRD is a permuted block design. These designs assign patients to treatments in predetermined permuted block schemes. For example, a block of size 4 aimed at assigning patients to one of two treatments could contain any of the following combinations of letters: AABB, ABAB, ABBA, BBAA, BABA, BAAB.

Suppose a block comes up ABBA. Such a block would assign the next 4 patients to enroll in the trial to the treatments A, B, B, and A, sequentially. One particularly nice advantage of permuted block designs is that balance is achieved at the end of each randomization block. If an experiment employs blocks of size six, balance is attained—at a minimum—after each sixth subject enrollment. Hence, smaller block sizes are desirable for higher or more frequent balance, especially if the patient population may drift over time. Note that blocks of size two are too small; such a design corresponds to a sequentially paired and highly deterministic design. Even so, paired designs effectively minimize treatment imbalance as well as drift in patient

characteristics over time [2], both desirable traits.

Like deterministic pairing, all blocking eliminates the random nature of the patient assignment for patients assigned at the end of a block. Any unblinding of patient assignments due to the increased predictability towards the ends of the blocks allows increased opportunity for selection bias in the study. From this perspective, larger blocks are desirable and smaller block sizes should be avoided to minimize this risk.

A trial may thus try to benefit from harnessing blocks of different sizes throughout its randomization, gaining the benefits of both small and large blocks within a single study. Unfortunately, even this solution also has a drawback, namely that balance occurs at unknown times, making interim monitoring inconvenient. Moreover, the reduction in opportunity for selection bias is minimal [81]. Choosing a permuted blocking design and employing randomization by blocks can be complex. An introduction to permuted block designs in the context of restricted randomization is given by Rosenberger and Lachin [81] and a more in depth review presented by Zelen [107].

1.2.1.3 Biased Coin Designs

A complete randomization scheme for an equal-allocation RCT with two treatments is equivalent to assigning a patient to one of two treatments by flipping a fair coin. This allocation is asymptotically equal, but may not be balanced for any given trial. The likelihood of imbalance worsens for smaller sample sizes. Bradley Efron addressed this issue of complete randomization designs in 1971 by offering a biased coin design (BCD) which has the property of, "Forcing a sequential experiment to be balanced," [46]. Instead of sequentially assigning treatments with probability each $\frac{1}{2}$, Efron proposes a biased coin which sequentially assigns patients to treatments with probability $\frac{1}{2}$ each if the past treatment assignment is balanced, and otherwise skews the treatment

allocation in favor of the less-assigned option. In particular, let $S_n = \sum_{i=1}^{n} T_i$ where

$$T_i = \begin{cases} 1 & \text{if the } i^{th} \text{ patient is assigned to Treatment A} \\ -1 & \text{if the } i^{th} \text{ patient is assigned to Treatment B} \end{cases}.$$

Efron's BCD with bias γ, denoted $BCD(\gamma)$, randomizes the n^{th} patient to Treatment A with probability

$$\phi_n = \begin{cases} \gamma & \text{if } S_{n-1} < 0 \\ \frac{1}{2} & \text{if } S_{n-1} = 0 \\ 1-\gamma & \text{if } S_{n-1} > 0 \end{cases}$$

for a fixed biasing probability, $\gamma \in \left[\frac{1}{2}, 1\right)$. A bias of $\gamma = \frac{2}{3}$ is suggested for its asymptotic probability of balance (50% chance for even samples and 75% chance for being unbalanced by only one patient for odd samples). Efron [46] argues that $BCD\left(\frac{2}{3}\right)$ can be as effective at forcing balance as a permuted block design with blocks of size 10, but is easier to implement. Moreover investigators have a lower probability of being able to predict future treatment assignments. In fact, $BCD\left(\frac{2}{3}\right)$ is as predictable as a design with blocks of size 16-18. Exact properties of $BCD(\gamma)$ for all values of γ are only recently ascertained by Markaryan and Rosenberger [70]. In 2010, the authors determine that for some values of γ, a BCD may display some undesirable traits. Nevertheless, BCDs are generally better at meeting sequential enrollment RCT objectives than permuted block designs. In addition to low risk of selection bias, Efron's BCD has the smallest asymptotic variability of all two-treatment, equal-allocation procedures [62].

Other biased coin designs with fixed skewing allocations include Chen's BCD with imbalance tolerance [41] and the big stick design (BSD) of Soares and Wu [90].

The first extension, the BCD with imbalance tolerance, denoted $BCDWIT(\gamma, c)$ randomizes the n^{th} patient to Treatment A with probability

$$\phi_n = \begin{cases} 0 & \text{if } c \leq S_{n-1} \\ 1-\gamma & \text{if } 0 < S_{n-1} < c \\ \frac{1}{2} & \text{if } S_{n-1} = 0 \\ \gamma & \text{if } -c < S_{n-1} < 0 \\ 1 & \text{if } S_{n-1} \leq -c \end{cases}$$

for a fixed biasing probability, $\gamma \in \left[\frac{1}{2}, 1\right)$, and positive constant c. The value of c is the cutoff after which, if $BCD(\gamma)$ fails to adequately address the growing imbalance, the next patient is deterministically assigned to the treatment with fewer exposures. The second extension is a special case of the first. That is, the big stick design $BSD(c)$ is actually $BCDWIT\left(\frac{1}{2}, c\right)$,

$$\phi_n = \begin{cases} 1 & \text{if } c \leq S_{n-1} \\ \frac{1}{2} & \text{if } -c < S_{n-1} < c \\ 0 & \text{if } S_{n-1} \leq -c \end{cases} .$$

In this case, subjects are assigned to treatment with equal probability until the difference threshold c is crossed. The next enrollee is then automatically allocated to the treatment with fewer patients.

1.2.1.4 Dynamic Biased Coin Designs

In 1977, Wei [101] points out that Efron's BCD only accounts for the existence of an imbalance in treatment assignments, but not for the magnitude of the imbalance. That is, γ is fixed whether one treatment has been assigned two extra patients or 200

extra patients. In designs motivated by Wei's approach, γ varies based on the value of S_n, rather than simply its sign. Biased coin designs which account for the magnitude of treatment differences by altering the assignment proportions conditionally are referred to as generalized biased coin designs (GBCDs) [89], adjustable biased coin designs [12], or adaptive biased coin designs [103], despite not being adaptive procedures in the current sense of the word. See Section 1.3.1 for more details on terminology. This class of biased coin designs are actually urn models and hence are approached from the context of a generalized Friedman's urn (GFU), also known as a generalized Polya urn.

The urn design described by Wei [102], $UD(\omega, \alpha, \beta)$ with ω, α, β nonnegative, may be applied to sequential randomization of a clinical trial with K treatments, as follows. An urn is filled with ω balls of each of the K colors. When a patient arrives, draw a ball from the urn with replacement, noting its color. For a ball of color k, with $k = 1, \ldots, K$, assign the patient to treatment k. Update the urn by adding α balls of color k and β balls of each of the other colors. If $\omega = 0$, for the first patient, select a color k with probability $\frac{1}{K}$, assign the patient to treatment k, and update the urn as above. The allocation of such a design has a higher degree of randomization early on for larger ω, while the design favors balance when ω is smaller or when α is small relative to β. When $\alpha = \beta$, $UD(\omega, \alpha, \beta)$ is a CRD.

In 1977 [101], Wei shows that the subclass of designs $UD(\omega, 0, \beta)$ with $K = 2$ can force a RCT to be balanced when the number of patients allocated is small, but will relax towards complete randomization as the sample grows. Wei writes:

> Therefore, the $[UD(\omega, 0, \beta)]$ forces the experiment to be more balanced when severe imbalance occurs and also forces small-sized experiments to be balanced, but when n is large enough, the $UD(\alpha, \beta)$ puts less weight on

balancing the experiment and tends toward the complete randomization scheme.

In 1978, Wei proves the trade-off between forcing balance for a small number of assignments and acting like complete randomization in larger samples still holds [103]. He also presents a multi-urn method of incorporating covariates into the allocation rule, by having distinct urns for individual levels of prognostic factors [102]. This suggestion is equivalent to randomizing independently within each stratum in order to balance similarly across population strata. Wei reiterates that a GBCD of this type shares the balancing properties of Efron's BCD, but, "behaves more and more like the complete randomization scheme as the size of the experiment increases." Additionally, for a predetermined minimum sample size, Wei's GBCD requires fewer excess patients than the BCD to achieve both the minimum sample size and complete balance in the trial.

Extensions of the GBCD begin in the 1980s and include developments in the last decade. Smith, for instance, leverages the martingale invariance principle to study generalizations of Wei's GBCD [89]. Atkinson leverages optimal design theory to suggest conditional allocation proportions under treatment imbalance [5]. The adjustable BCD of Baldi Antognini and Giovagnolli [12] generalizes Efron's [46] and Wei's [102] BCDs to obtain a GBCD of the form

$$\phi_n = \begin{cases} g(S_{n-1}) & \text{if } S_{n-1} > 0 \\ \frac{1}{2} & \text{if } S_{n-1} = 0 \\ 1 - g(S_{n-1}) & \text{if } S_{n-1} < 0 \end{cases}$$

based on the function $g(S_n)$ which takes on values in $[0, 1]$ and has the property $g(x) = 1 - g(-x)$ for all integers x. From a coin perspective, the randomization

scheme in [12] allows an infinite number of thresholds c and biased probabilities γ for a BCD with continuous levels of imbalance intolerance. From an urn perspective, the link function $g(\cdot)$ in [12] defines the values of α and β—specifying a continuum of fractional numbers of balls to place into the generalized urn after each patient assignment.

1.2.2 Ethics

Despite the favor curried by balanced RCTs and their rampant use in modern settings, ethical trial design proposals date back to the mid-1900's. For example, in sequentially enrolled clinical trial design proposals such as [96], Bayesian approaches are solicited to help bias patient allocation toward the better-performing treatment arm. This allows more patients to be exposed to the superior therapy throughout the course of a medical experiment. Other ethical proposals for clinical trial design surfaced shortly after 1950 [1, 2, 42, 43, 78]. These ideas include multi-phased studies, early stopping rules, flexible designs, and minimizing patient exposure to the inferior treatment—all of which are gaining popular consideration and real-world application only in recent years [51, 82, 94].

1.2.2.1 Bayesian Approaches

William Thompson's "On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples" pioneered the field of response-adaptive randomization [96]:

> there can be no objection to the use of data, however meagre, as a guide to action required before more can be collected ... If [RAR] were adopted, ... it seems apparent that a considerable saving of individuals otherwise

> sacrificed to the inferior treatment might be effected. This would be important in cases where either the rate of accumulation of data is slow or the individuals treated are valuable, or both.

In 1933, he proposed that, in view of the probability that one treatment is better than another, more patients should be allocated to that treatment. Thompson then examines a particular function of allocation which adheres to such a principle. For this function, he derives its properties using Bayesian analysis and provides tables which may be used to determine the probability with which one should allocate the subsequently enrolled individual to the first of the two treatments.

The simplicity of the Bayesian ideology as well as the ability to incorporate—and subsequently, automatically update—uninformative prior distributions in the face of a complete lack of reliable information contribute to the current popularity of Bayesian designs in medical studies. Advances in modern computing power, including programing of dynamic algorithms, allow researchers to consider Bayesian procedures far more complex than tracking sequential results through a static table. Overarching views of current Bayesian approaches to medical research are summarized by [91] in 2004 and by [27] in 2010.

1.2.2.2 Multistage Approaches and Optimal Allocations

In 1952, Herbert Robbins also writes about how to best allocate patients to treatment [78]. Robbins' article on sequential design expands on the scant literature of the time (namely *Sequential Analysis* of a single population by Wald [100]) by considering how best to sample from two normal populations whose means and variances are unknown. He notes that for a fixed sample size, the variance for the estimator $\overline{x}_1 - \overline{x}_2$ of $\mu_1 - \mu_2$, $\sigma^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_1^2}{n_1}$, is minimized when $\frac{n_1}{n_2} = \frac{\sigma_1}{\sigma_2}$. Robbins then suggests a two-stage

approach in which $2m$ samples are taken and the parameters μ_i and σ_i are estimated for $i = 1, 2$ in the first stage. During the second stage, the appropriate number of subsequent samples are drawn until the entire sample of n_1 from population 1 and n_2 from population 2 has been drawn such that $\frac{n_1}{n_2} = \frac{\widehat{\sigma}_1}{\widehat{\sigma}_2}$. The choice of m, he notes, must balance the fact that smaller m will provide weaker estimates of σ_1 and σ_2 but larger m may make it difficult to efficiently utilize said estimates. Robbins goes on to suggest that a two-stage design could be generalized to a multi-stage design by estimating the variances after each sequential observation is sampled.

The open problem Robbins summarizes is equivalent to that of choosing the optimal RAR scheme to minimize the number of treatment failures for two treatments with binary outcomes. In 1963, Theodore Colton considers a different slant on the optimization conundrum, focusing on minimizing the number of patients assigned to the inferior treatment [42]. Colton selects a loss function in terms of treatment assignment rather than treatment failures, comparing "minimax, maximin, and Bayesian approaches" to determine the appropriate two-stage sequential trial design. Six years later, Cornfield *et al.* extend the two-stage, optimal allocation approach by considering multiple phases, each with different allocation proportions [43]. While the authors admit that generalizing a multi-stage procedure by simply continuing a two-step process is not optimal, their approach is practical, ethical, and effective.

Multi-stage testing research takes many forms in current clinical research. Consider Jennison and Turnbull's text [66] for an introduction to group sequential designs, interim analyses, and early stopping rules. Chapter 14 in [44] tackles the same topics from an oncology perspective, including real-world examples. In Chapter 5 of [77], Jennison and Turnbull present an adaptive perspective on group sequential designs. Chapter 13 from [77] provides a thorough but succinct overview of seamless Phase II/III designs.

Optimal allocation research continues today as well. For the target allocation of a randomization scheme to be optimal, it must be the solution to an optimization problem with formal optimization criteria. Intuitive optimality criteria include maximizing power, minimizing allocation variance, maximizing the number of patients assigned to the superior treatment, minimizing patient failures, as well as timing and cost measures. Sylvie wrote the book on *Optimal Design* [86], but Hu and Rosenberger [58] provide an article specifically for RAR. Robbins' approach of minimizing the number of treatment failures is commonly referred to as *RSIHR* in RAR literature, named after the authors who derived the optimal solution to this allocation problem for binary treatment outcomes in [83]. For optimal allocations minimizing treatment failures in clinical trials with continuous outcomes, see [33] and [35]. Myriad other criteria are actively being pursued such as compound ethical optimality [13, 28], optimality across more than two treatments [36, 67], and determinant optimality [56].

1.2.2.3 Adaptive Biased Coin Designs

Sections 1.2.1.3 and 1.2.1.4 describe biased coin designs (BCDs) and urn models that are nonadaptive, targeting balanced treatment allocations. These randomizations bias exposure probabilities toward the under-assigned treatment based on the difference in treatment assignments S_{n-1}. If the skewing proportions are conditioned on patient responses, however, instead of on differences in the number of subjects exposed to each arm, these BCDs become RAR designs.

Perhaps the most famous RAR is Zelen's *play-the-winner* rule [106] wherein patients are enrolled sequentially to one of two treatment arms. Patient responses are binary (success or failure) and are observed before the next patient arrives. If a patient is successfully treated, the subsequent patient is exposed to the same arm. If

a patient fails, the subsequent patient is exposed to the other arm. Each subject in the study is thus assigned to the treatment which "won" (or at least did not fail) in the prior round of testing. In this deterministic coin design, treatment assignment is known with certainty based on previous participant outcomes, allowing for potentially large selection biases. Nevertheless, the design is simple and intuitive, requiring no computational or implementation support.

The randomized version of the play-the-winner rule, proposed by Wei and Durham [104], follows the $UD(\omega, \alpha, \beta)$ urn model described in Section 1.2.1.4. An initial number ω of balls of each treatment type are in an urn. Patients are allocated by selecting a ball from the urn with replacement and exposing the patient to that treatment type. Throughout the experiment, α balls of one type and β balls of the other types are added to the urn. The biggest difference is the design's adaptive nature—instead of adding balls to the urn based on which treatment a patient is assigned, balls are added to the urn based on a patient's response to the assigned treatment.

In its simplest form, the randomized play-the-winner has only two treatment arms. If a patient responds successfully after being exposed to treatment A, one ball of type A is added and no balls of type B are added. If a patient fails to respond after being exposed to treatment A, no balls of type A are added and one ball of type B is added. That is, $\alpha = 1$ ball is added for the "winning" therapy and $\beta = 0$ balls are added for the "losing" therapy. The adaptive $UD(1, 1, 0)$ is the randomization design harnessed in the initial ECMO trial [23] which terminated after ten balls of one type were added to the urn. In generalized versions, ω, α, and β assume any nonzero values for a trial with $K \geq 2$ treatments [79].

The companion model to the randomized play-the-winner is the drop-the-loser design by Ivanova [64]. An initial number ω of balls of each treatment type are in an

urn. Patients are allocated by selecting a ball from the urn and exposing the patient to the corresponding therapy. If the outcome is a success, the ball is replaced. If the outcome is a failure, the ball is not replaced; that ball is "dropped" from the urn. To ensure a degree of *regularization* in this randomization scheme—i.e., that the design does not eliminate a treatment arm—one can also include immigration balls. Immigration balls, when drawn, are immediately replaced. The only action taken with an immigration ball selection is to add a predetermined but equal number of balls of each treatment type to the urn. The larger ω and the immigration components of an urn are, the less impactful removing balls from the urn will be on the trial allocations. Even so, the drop-the-loser design boasts minimal allocation variance among similar RAR designs and increased power over equal allocation [58, 60, 64, 80, 92, 111]. Additional adaptive BCDs are discussed in Section 1.3.3 after the presentation of RAR statistical framework.

1.3 Adaptive Framework

1.3.1 Terminology

Many researchers point out that the language describing adaptive designs, including those for RARs, is inconsistent within the literature [5, 6, 17, 38]. In 2006 Dragalin muses [45],

> as often happens with novel approaches, there has been substantial confusion over what these designs are and when they are most applicable. They are known as adaptive, sequential, flexible, self-designing, multistage, dynamic, response-driven, smart, novel designs.

This dissertation follows the lexicon of Dragalin for general adaptive design terminology. Whenever possible, these chapters conform to conventions proposed by Rosenberger and fellow authors (e.g., [59, 81, 93]) for respond-adaptive nomenclature and statistical notation.

Begining 2004, regulatory agencies such as the Food and Drug Administration (FDA) and the Department of Health and Human Services (DHHS) encourage the use of adaptive designs to "streamline the clinical trials process." The agencies stipulate, however, that this does not include poorly planned protocols or ad hoc analyses [53]. Similarly, Dragalin [45] declares,

> *Adaptive design* is defined as a multistage study design that uses accumulating data to decide how to modify aspects of the study without undermining the validity and integrity of the trial.

An adaptive clinical trial, therefore, is one which employs prospectively approved modifications to its design. Approved alterations focus on one of the following four categories [45, 93].

- **Allocation rules** dictate the randomization processes or proportions for assigning patients to treatments.
- **Sampling rules** determine the number of patients to be included.
- **Stopping rules** designate when a trial should end.
- **Decision rules** are a catch-all category for rules not detailed above. These include rules on decisions to be made at interim and final analyses; adjustments to biomarkers, endpoints, model estimates, populations; alterations of analyses or hypotheses; etc.

As this dissertation proposes and examines a new allocation rule, Section 1.3.2 establishes the definitions and statistical conventions for RAR.

1.3.2 Allocation Rules

1.3.2.1 Mathematical Framework

Statistical experiments include information on investigative exposures, relevant covariates, and measurable outcomes of each experimental unit, as the particulars become available. Allocation rules are randomization algorithms based on various subsets of the above available information. This section explicitly defines these terms before describing how such factors are employed in adaptive randomization procedures.

Randomization Sequence. The matrix $\boldsymbol{T} = (\boldsymbol{T}_1, \ldots, \boldsymbol{T}_n)'$ is a randomization sequence, with $\boldsymbol{T}_i = \boldsymbol{e}_k$ for $k = 1, \ldots, K$, and $i = 1, \ldots, n$.

Example. For $K = 2$ treatments and n patients, one possible randomization sequence is

$$\boldsymbol{T} = \begin{bmatrix} 0 & 0 & 1 & 0 & \cdots & 1 \\ 1 & 1 & 0 & 1 & \cdots & 0 \end{bmatrix}'.$$

Covariate Information. The matrix $\boldsymbol{Z} = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)'$ represents the covariate information for all n patients with $\boldsymbol{Z}_i = (z_{i1}, \ldots, z_{iS})$ describing the S observed prognostic factors in patient i.

As subjects accrue in the trial, knowledge of their covariates becomes available and does not change thereafter. The information for each prognostic factor z_{is} may describe a patient's level of a discretely coded covariate or a continuous characteristic.

Example. When the sole prognostic factor ($S = 1$) is sex in a trial with n patients, the i^{th} subject's covariate information could be $\boldsymbol{Z}_i = \begin{bmatrix} 1 & 0 \end{bmatrix}$ if the patient is female and $\boldsymbol{Z}_i = \begin{bmatrix} 0 & 1 \end{bmatrix}$ if the patient is male.

Response Variable. The matrix $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)'$ with $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iK})$ is the hypothetical sequences of responses which would be observed if every treatment were assigned to every patient, independently. Realistically, only one Y_{ik} is possible in a typical non-crossover setting.

Clearly the response variables depend on the treatment assignment, and likely on patient covariates as well. Thus response variables may be considered as functions of the randomization sequence and prognostic factors.

Example. For $K = 2$ treatments and n patients, the response matrix would be

$$\boldsymbol{Y}' = \begin{bmatrix} y_{11} & y_{21} \cdots & y_{n1} \\ y_{12} & y_{22} \cdots & y_{n2} \end{bmatrix}.$$

In the case that 1 represents a success on a treatment, -1 a failure, and 0 for non-assignment, the matrix at the trial's end might be

$$\boldsymbol{Y}' = \begin{bmatrix} 0 & 0 & -1 & 0 & \cdots & -1 \\ 1 & -1 & 0 & 1 & \cdots & 0 \end{bmatrix}.$$

Having established the components on which a randomization procedure might be conditioned, a randomization procedure is statistically defined as follows.

Randomization Procedure. Given σ-fields (also known as Borel fields)

$$\mathcal{T}_n = \sigma(\boldsymbol{T}_1, \ldots, \boldsymbol{T}_n),\ \mathcal{Y}_n = \sigma(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n), \text{ and } \mathcal{Z}_n = \sigma(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n),$$

let $\mathcal{F}_n = \mathcal{T}_n \otimes \mathcal{Y}_n \otimes \mathcal{Z}_{n+1}$. Then for all $\mathcal{F}_{n-1}$-measurable functions ϕ_n,

$$\phi_n = \mathbb{E}[T_n | \mathcal{F}_{n-1}]$$

is the conditional probability of assigning each treatment to the n^{th} patient, given all previous and current information available. The information contained in $\mathcal{F}_{n-1}$ may include previous randomizations, previous results, and n^{th} patient's known covariate information. Alternatively, $\mathcal{F}_{n-1}$ may be the empty set. Randomization procedures are partitioned into five categories: complete, restricted, response-adaptive, covariate-adaptive, and covariate-adjusted response-adaptive [45, 59, 81]. The subsequent section describes these five classifications.

1.3.2.2 Randomization Designs

Complete Randomization. Classic randomization procedures such as those employed in equal-allocation CRDs are often likened to flipping a fair coin or rolling a fair die. That is, CRDs perform unconditional allocation of each subject, ignoring all other information:

$$\phi_i = \mathbb{E}[T_i | \mathcal{F}_{i-1}] = \mathbb{E}[T_i].$$

In this case, $\mathcal{F}_i$ is empty and ϕ is constant across all patients $\left(\text{e.g. } \frac{1}{2}\right)$.

Restricted Randomization. Because complete randomization may in fact assign patients in unequal proportions (e.g. 30-70 or 60-40), clinical trials tend to imple-

ment restricted randomization procedures rather than complete ones. In particular, many medical trials attempt to randomize patients in such a way as to protect balance, assigning each subject to a treatment so that the number of patients assigned to each treatment tends to be equally distributed over all treatments. Restricted randomization procedures take the form

$$\phi_i = \mathbb{E}[T_i|\mathcal{F}_{i-1}] = \mathbb{E}[T_i|\mathcal{T}_{i-1}],$$

with randomization of the i^{th} patient taking into account the treatment allocations of the $i-1$ previous patients—or a function of the prior assignments such as S_{n-1}. Previously discussed examples of restricted randomization procedures include permuted block designs (see, for example, [81]) and biased coin designs (e.g. Efron [46] and Wei [101]) from Section 1.2.1.

Response-Adaptive Randomization. Expanding the randomization conditioning to include available estimates of the treatments produces a method of response-adaptive randomization (RAR),

$$\phi_i = \mathbb{E}[T_i|\mathcal{F}_{i-1}] = \mathbb{E}[T_i|\mathcal{T}_{i-1}, \mathcal{Y}_{i-1}].$$

This will allocate the i^{th} patient to a treatment while accounting for the previous $i-1$ treatment allocations as well as the previous $i-1$ responses.

The goal of RAR is often to assign more patients to the best treatment, based on previous patient responses, without compromising the goals of the study. Response-adaptive randomization designs also incorporate the impact due to potential delays in the availability of information [11, 61, 63, 92, 111]. This will cause randomization procedures to be conditioned on information accrued up through the j^{th} patient with

$j \leq i$ determined by the amount of delay in responses.

Recently, adaptive allocation research embraced an additional factor: patient covariates [18, 17, 19, 34, 110, 112]. In cases where stratification over covariates is desirable, patient covariate information may be included in the conditional randomizations, up through the i^{th} patient. That is, the current subject's covariate information may be considered in his or her randomization. Typically, randomization has been assumed to be performed separately within each stratum [18, 68, 102]; however, this requires the number of prognostic factors to be small and moreover requires the factors to be discrete. Incorporating covariate-based adaptations into a restricted randomization scheme produces a covariate-adaptive randomization procedure. Similarly, adjusting for covariates in RAR produces a covariate-adjusted response-adaptive randomization.

Covariate-Adaptive Randomization. Covariate-adaptive randomizations allocate subjects conditionally with the goal of balancing covariates across treatment assignments. The procedure

$$\phi_i = \mathbb{E}[T_i|\mathcal{F}_{i-1}] = \mathbb{E}[T_i|\mathcal{T}_{i-1}, \mathcal{Z}_{i-1}]$$

will randomize the i^{th} patient knowing the previous $i-1$ treatment allocations and the prognostic factors of all patients past and present.

Covariate-Adjusted Response-Adaptive Randomization. A covariate-adjusted response-adaptive procedure combines the goals of response-adaptive randomization and covariate-adaptive randomization. That is, the procedure

$$\phi_i = \mathbb{E}[T_i|\mathcal{F}_{i-1}] = \mathbb{E}[T_i|\mathcal{T}_{i-1}, \mathcal{Y}_{i-1}, \mathcal{Z}_i]$$

frequently aims to allocate more patients to the best treatment in a manner specific to each level of prognostic factor combinations, based on previous patient assignments, responses, and covariates—including the current patient's covariate information.

1.3.3 History of the Response-Adaptive Biased Coin Design

In 1933 Thompson discusses the "likelihood that one unknown probability exceeds another in view of the evidence of the two samples" [96] and suggests that as a patient arrives for study enrollment, the current information from previously sampled subjects be used to dictate how the new patient is treated. One intuitive example would be to assign patient i to treatment A in accordance with the probability that treatment A is superior to treatment B—or the current estimate of that probability given information from the previous $i-1$ patients. Let $\widehat{p}_A(i-1)$ denote the estimate of the probability that treatment A is superior to treatment B based on the outcomes of subjects 1 through i. Thompson's proposal is then equivalent to flipping a biased coin to assign patient i to treatment A with probability $\phi_i = \widehat{p}_A(i-1)$.

Thompson goes on to note that this algorithm is merely a special case of the design which assigns patient i to treatment A proportional to a function of the probability that treatment A is superior to treatment B—or the current estimate. That is, for a function $f : [0,1] \to [0,1]$ which is monotonically increasing on its domain, one can flip a biased coin and assign patient i to treatment A with probability $\phi_i = f\left(\widehat{p}_A(i-1)\right)$. For particular optimal functions, consider optimality criteria for binomial responses [58, 83, 97]; an optimal allocation for continuous treatment responses [71]; and optimal targets generalizable to trials with greater than two treatment arms [16, 30, 33, 55, 108, 109, 114].

A non-optimal function which appears frequently in the literature beginning in

the 1990's (for starters, [4, 7, 6, 17, 19, 29, 34, 36, 108]) is the probit link function $\Phi(\cdot)$. Despite its fame as the cumulative distribution function of a standard normal variable, $\Phi(\cdot)$ does not meet any formal optimality criteria. Nevertheless, researches focusing on biased coin designs with link function $\Phi(\cdot)$ aim to allocate more patients to the treatment with a better response. Assuming large responses are desirable, the design randomizes patient i to treatment A with probability

$$\phi_i = \Phi\left(\frac{\widehat{\mu}_A(i-1) - \widehat{\mu}_B(i-1)}{T}\right),$$

where $\widehat{\mu}_k(i-1)$ is the current estimate of the mean of treatment k for $k = A, B$ based on the data collected from the first $i-1$ patients. The value of the scaling parameter T must be positive and is most often a fixed constant [17, 34]. The value of T has been arbitrarily set as a small, positive integer in various simulation comparisons [6, 65, 108]. The value of T mitigates the effect of the estimated treatment differences: larger values of T promote equal allocation, smaller values promote ethical allocation.

In recent decades, researchers have suggested various design enhancements. For instance, the allocation rule can be broadened from a two-arm study to multi-treatment trials by comparing each treatment to the mean of all treatments. Atkinson [4] suggests randomizing the i^{th} patient to Treatment k with probability

$$\phi_i = \Phi\left(\frac{\widehat{\mu}_k(i-1) - \overline{\mu}(i-1)}{T}\right),$$

where $\overline{\mu}(i-1)$ is the estimated mean of all treatments based on the first $i-1$ patients.

Another augmentation includes leveraging a different cumulative distribution function (CDF) to alter a design's adaptive nature [17]. Any symmetric distribution's CDF can be leveraged as a link function. Moreover, a design may leverage multiple

link functions throughout the course of a single trial. For example, the link function may change as more data is collected, say, from a Cauchy CDF in the initial data collection stage to a normal CDF when the estimates are deemed appropriately trustworthy [17]. Centrally weighted distributions will have CDFs that, when employed as link functions, cause adaptive designs to be more data-driven—that is, more responsive to accumulated findings. This trait is appealing when the data and the treatment estimates are reliable, but risky when little data or prior knowledge of the treatments are available. Harnessing heavy-tailed distributions' CDFs as link functions minimizes the initial risks due to high variability, providing a design with a higher degree of randomization, but potentially less ethical patient allocations. Exposing an initial m patients to each treatment prior to commencing adaptive randomization is another way to alter the design for a similar effect, basing initial estimates on more adequate quantities of accumulated data.

Modifying the tuning parameter T is a further area of deliberation. On the one hand, keeping the scaling parameter constant throughout the trial is common in the literature [6, 7, 19, 31, 34, 40, 73, 75, 108]. On the other hand, the original paper notes that the value of T need not be constant and may evolve with the data [17]. Biswas, Huang, and Huang propose—but do not pursue—the idea that T might be replaced with the current estimate of the treatment standard deviation, assuming treatment variances are equal [34]. Other authors employ $T = \sqrt{\widehat{\sigma}_A^2 + \widehat{\sigma}_B^2}$ or $T = \sqrt{(\widehat{\sigma}_A^2 + \widehat{\sigma}_B^2)/2}$ in comparisons of competing and related designs [15, 20, 21, 32, 65]. Still other designs have consider weighting functions of various forms. For example, Bandyopadhyay and De suggest that any adaptive design which incorporates the function c_i will conform to the typical asymptotic behaviors provided $c_i \to 0$ as $i \to \infty$ and $\sum_{i=1}^{\infty} c_i = \infty$ [21]. Biswas and Bhattacharya also embraced this form for weighting information in designs with dual constraints [28].

Chapters 2 – 4 extend the design proposed by Bandyopadhyay and Biswas in 2001 by replacing the constant tuning parameter T with the current estimate of the pooled treatment standard deviation. This idea was initially proposed in [17], wherein the authors assume that treatment variances are known and equal. Despite this proposal, using the pooled treatment standard deviation estimate as an adaptive value of T has not appeared or been studied in any subsequent works. This modification expands the applicability of the response-adaptive biased coin by eliminating the need for known and equal treatment variances, improving the ethicality of the allocation, and even increasing the design power in some cases. Cited reference searches performed as recently as September 2015 corroborate that no other literature examines the impact of replacing T with the current estimates of the pooled standard deviation in the Bandyopadhyay and Biswas design.

This research is valuable because, as of yet, no adaptive design can target a highly ethical randomization ratio while preserving trial properties including power, type I error rates, sample size, and allocation variance. The Bandyopadhyay and Biswas procedure is highly ethical, but in its current form can sustain large losses in power and inflated type I error risk for many values of treatment parameters. Finding appropriate replacements for T—for example, the pooled treatment standard deviation—can mitigate the design's loss of power while simultaneously preserving the ability to expose more patients to the better treatment. Adopting an ethical adaptive design for use in clinical trials would be a great step forward for the medical community, minimizing individual patient risk while performing research necessary to maximize collective patient benefit.

1.4 References

[1] Anscombe, F.J. "Sequential medical trials". In: *Journal of the American Statistical Association* 58.302 (1963), pp. 365–384.

[2] Armitage, P. *Sequential medical trials*. 1st ed. Oxford: Blackwell Scientific Publications, 1960.

[3] Armitage, P. "Sequential medical trials: Some Comments on F.J. Anscombe's Paper". In: *Journal of the American Statistical Association* 58.302 (1963), pp. 384–387.

[4] Atkinson, A.C. "Adaptive Biased-Coin Designs for Clinical Trials with Several Treatments". In: *Discussiones Mathematicae - Probability and Statistics* 24.1 (2004), pp. 85–108.

[5] Atkinson, A.C. "Optimum biased coin designs for sequential clinical trials with prognostic factors". In: *Biometrika* 69.1 (1982), pp. 61–67.

[6] Atkinson, A.C. and Biswas, A. "Adaptive biased-coin designs for skewing the allocation proportion in clinical trials with normal responses". In: *Statistics in Medicine* 24.16 (2005), pp. 2477–2492.

[7] Atkinson, A.C. and Biswas, A. "Bayesian Adaptive Biased-Coin Designs for Clinical Trials with Normal Responses". In: *Biometrics* 61 (2005), pp. 118–125.

[8] Azriel, D., Mandel, M., and Rinott, Y. *On optimal allocation in binary response trials; is adaptive design really necessary?* Discussion Paper Series. 2011.

[9] Bai, Z.D. and Hu, F. "Asymptotic theorems for urn models with nonhomogeneous generating matrix". In: *Stochastic Processes and Their Applications* 80 (1999), pp. 87–101.

[10] Bai, Z.D. and Hu, F. "Asymptotics in randomized urn models". In: *Annals of Applied Probability* 15.1b (2005), pp. 914–940.

[11] Bai, Z.D., Hu, F., and Rosenberger, W.F. "Asymptotic properties of adaptive designs for clinical trials with delayed response". In: *The Annals of Statistics* 30.1 (2002), pp. 122–139.

[12] Baldi Antognini, A. and Giovagnoli, A. "A new 'biased coin design' for the sequential allocation of two treatments". In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 53.4 (2004), pp. 651–664.

[13] Baldi Antognini, A. and Giovagnoli, A. "Compound optimal allocation for individual and collective ethics in binary clinical trials". In: *Biometrika* 97.4 (2010), pp. 935–946.

[14] Baldi Antognini, A. and Giovagnoli, A. "Ethical Adaptive Designs for Binary Clinical Trials". In: *Societa Italiana di Statistica* (2007).

[15] Bandyopadhyay, Uttam and Bhattacharya, Rahul. "On an ethical cum optimal adaptive allocation design". In: *Statistics* 41.6 (2007), pp. 471–483.

[16] Bandyopadhyay, Uttam and Bhattacharya, Rahul. "Response adaptive procedures with dual optimality". In: *Statistica Neerlandica* 63.3 (2009), pp. 353–367.

[17] Bandyopadhyay, Uttam and Biswas, A. "Adaptive designs for normal response with prognostic factors". In: *Biometrika* 88 (2001), pp. 409–419.

[18] Bandyopadhyay, Uttam and Biswas, A. "Allocation by randomized play-the-winner rule in the presence of prognostic factors". In: *Sankhya* 61.3 (1999), pp. 397–412.

[19] Bandyopadhyay, Uttam, Biswas, A., and Bhattacharya, Rahul. "A covariate adjusted two-stage allocation design for binary responses in randomized clinical trials". In: *Statistics in Medicine* 26 (2007), pp. 4386–4399.

[20] Bandyopadhyay, Uttam, Biswas, A., and Bhattacharya, Rahul. "A new response-adaptive design for continuous treatment responses for phase III clinical trials". In: *Journal of Statistical Planning and Inference* 141.7 (2011), pp. 2256–2265.

[21] Bandyopadhyay, Uttam and De, Saurav. "On a two-treatment adaptive allocation rule for continuous response". In: *Statistical Methodology* 4 (2007), pp. 434–447.

[22] Bartlett, R.H. and Cornell, R.G. "Ethics and statistics in randomized clinical trials: Comment". In: *Statistical Science* 6.1 (1991), pp. 63–65.

[23] Bartlett, R.H. et al. "Extracorporeal circulation in neonatal respiratory failure: a prospective randomized study". In: *Pediatrics* 76.4 (1985), pp. 479–487.

[24] Berry, D.A. "Bayesian clinical trials". In: *Nature Reviews Drug Discovery* 5 (2006), pp. 27–36.

[25] Berry, D.A. "Bayesian Statistics and the Efficiency and Ethics of Clinical Trials". In: *Statistical Science* 19.1 (2004), pp. 175–187.

[26] Berry, D.A. and Eick, S.G. "Adaptive assignment versus balanced randomization in clinical trials: a decision analysis". In: *Statistics in Medicine* 14 (1995), pp. 231–246.

[27] Berry, S.M. et al. *Bayesian Adaptive Methods for Clinical Trials*. 1st ed. Chapman & Hall/CRC Biostatistics Series, 2010.

[28] Biswas, A. and Bhattacharya, Rahul. "An optimal response-adaptive design with dual constraints". In: *Statistics and Probability Letters* 80.3 (2010), pp. 177–185.

[29] Biswas, A. and Bhattacharya, Rahul. "Optimal response-adaptive allocation designs in phase III clinical trials: Incorporating ethics in optimality". In: *Statistics and Probability Letters* 81.8 (2011), pp. 1155–1160.

[30] Biswas, A. and Bhattacharya, Rahul. "Optimal response-adaptive designs for normal responses". In: *Biometrical Journal* 51.1 (2009), pp. 193–202.

[31] Biswas, A. and Bhattacharya, Rahul. "Reduction of Variability of Response-Adaptive Designs for Continuous Treatment Responses in Phase 3 Clinical Trials". In: *Drug Information Journal* 44 (2010), pp. 9–19.

[32] Biswas, A. and Bhattacharya, Rahul. "Response-adaptive designs for continuous treatment responses in phase III clinical trials: A review". Epub ahead of print. DOI: 10.1177/0962280212441424. 2012.

[33] Biswas, A., Bhattacharya, Rahul, and Zhang, L. "Optimal Response-Adaptive Designs for Continuous Responses in Phase III Trials". In: *Biometrical Journal* 49.6 (2007), pp. 928–940.

[34] Biswas, A., Huang, H.-H., and Huang, W.-T. "Covariate-adjusted adaptive designs for continuous responses in a phase III clinical trial: recommendation for practice". In: *Journal of Biopharmaceutical Statistics* 16.2 (2006), pp. 227–239.

[35] Biswas, A. and Mandal, Saumen. "Multi-treatment optimal response-adaptive designs for phase III clinical trials". In: *mODa 7 – Advances in Model-Oriented Design and Analysis*. Ed. by Di Bucchianico, A., Lauter, H., and Wynn, H.P. Physica-Verlag Heidelberg, 2004, pp. 51–59.

[36] Biswas, A., Mandal, Saumen, and Bhattacharya, Rahul. "Multi-treatment optimal response-adaptive designs for phase III clinical trials". In: *Journal of the Korean Statistical Society* 40.1 (2011), pp. 33–44.

[37] Byar, D.P. "Ethics and statistics in randomized clinical trials: Comment". In: *Statistical Science* 6.1 (1991), pp. 65–68.

[38] Chambaz, Antoine and van der Laan, Mark J. "Targeting the Optimal Design in Randomized Clinical Trials with Binary Outcomes and No Covariate: Theoretical Study". In: *The International Journal of Biostatistics* 7.1 (2011), Article 10.

[39] Chang, M. *Adaptive Design Theory and Implementation Using SAS and R*. 1st ed. Chapman & Hall/CRC Biostatistics Series, 2007.

[40] Chang, Y.I. and Park, Eunisk. "Sequential estimation for covariate-adjusted response-adaptive designs". In: *Journal of the Korean Statistical Society* 42 (2013), pp. 105–116.

[41] Chen, Y.P. "Biased coin design with imbalance tolerance". In: *Communications in Statistics. Stochastic Models* 15.5 (1999), pp. 953–975.

[42] Colton, T. "A model for selecting one of two medical treatments". In: *Journal of the American Statistical Association* 58.302 (1963), pp. 388–400.

[43] Cornfield, Jerome, Halperin, Max, and Greenhouse, Samuel W. "An Adaptive Procedure for Sequential Clinical Trials". In: *Journal of the American Statistical Association* 64.327 (1969), pp. 759–770.

[44] Crowley, J. and Pauler, D., eds. *Handbook of Statistics in Clinical Oncology*. 2nd ed. Chapman & Hall/CRC Biostatistics Series, 2006.

[45] Dragalin, V. "Adaptive designs: terminology and classification". In: *Drug Information Journal* 40.4 (2006), pp. 425–435.

[46] Efron, B. "Forcing a sequential experiment to be balanced". In: *Biometrika* 58.3 (1971), pp. 403–417.

[47] Ezekiel J. Emanuel et al., eds. *Ethical and Regulatory Aspects of Clinical Research: Readings and Commentary*. Baltimore, MD: The Johns Hopkins University Press, 2004.

[48] Fisher, Lloyd D. "Advances in clinical trials in the twentieth century". In: *Annual Review of Public Health* 20.1 (1999), pp. 109–125.

[49] Fisher, R. A. *The Design of Experiments*. 1st ed. Edinburgh: Oliver and Boyd, 1935.

[50] Flehinger, B.J. and Louis, T.A. "Sequential Treatment Allocation in Clinical Trials". In: *Biometrika* 58.3 (1971), pp. 419–426.

[51] Flight, L., Julious, S., and Goodacre, S. "Assessing the Current and Potential Use of Adaptive Study Designs in Emergency Medicine Clinical Trials". In: *Clinical Trial Design II*. Ed. by Ko, C.W. 259. American Statistical Association. 2015.

[52] Food and Drug Administration. *Critical Path Initiative: Transforming the way FDA-regulated products are developed, evaluated, and manufactured*. Report on Key Achievements in 2009. 2010.

[53] Food and Drug Administration. "Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics". Draft Document. 2010.

[54] Foster, C. *The Ethics of Medical Research on Humans*. 1st ed. University Press, Cambridge, UK: Cambridge University Press, 2001.

[55] Gwise, T., Hu, J., and Hu, F. "Optimal biased coins for two-arm clinical trials". In: *Statistics and its Interface* 1 (2008), pp. 125–135.

[56] Gwise, T., Zhou, J., and Hu, F. "An optimal response adaptive biased coin design with k heteroscedastic treatments". In: *Journal of Statistical Planning and Inference* 141.1 (2011), pp. 235–242.

[57] Hinkelmann, K. *Design and Analysis of Experiments, Vol. 3: Special Designs and Applications*. 1st ed. Vol. 3. Wiley-Blackwell, 2011.

[58] Hu, F. and Rosenberger, W.F. "Optimality, Variability, Power: Evaluating Response-Adaptive Randomization Procedures for Treatment Comparisons". In: *Journal of the American Statistical Association* 98.463 (2003), pp. 671–678.

[59] Hu, F. and Rosenberger, W.F. *The Theory of Response-Adaptive Randomization in Clinical Trials*. 1st ed. Hoboken, NJ: John Wiley & Sons, Inc., 2006.

[60] Hu, F., Rosenberger, W.F., and Zhang, L.-X. "Asymptotically best response-adaptive randomization procedures". In: *Journal of Statistical Planning and Inference* 136 (2006), pp. 1911–1922.

[61] Hu, F. and Zhang, L.-X. "Asymptotic normality of urn models for clinical trials with delayed response". In: *Bernoulli* 10.3 (2004), pp. 447–463.

[62] Hu, F., Zhang, L.-X., and He, X. "Efficient randomized-adaptive designs". In: *The Annals of Statistics* 37 (2009), pp. 2543–2560.

[63] Hu, F. et al. "Doubly-Adaptive Biased Coin Designs with Delayed Responses". In: *Canadian Journal of Statistics* 36.4 (2008), pp. 541–559.

[64] Ivanova, A. V. "A play-the-winner type urn model with reduced variability". In: *Metrika* 58.1 (2003), pp. 1–14.

[65] Ivanova, A. V., Biswas, A., and Lurie, A. "Response-adaptive designs for continuous outcomes". In: *Journal of Statistical Planning and Inference* 136.6 (2006), pp. 1845–1852.

[66] Jennison, C. and Turnbull, B.W. *Group Sequential Methods with Applications to Clinical Trials*. 1st ed. Interdisciplinary Statistics. Chapman & Hall/CRC, 1999.

[67] Jeon, Y. and Hu, F. "Optimal adaptive designs for binary response trials with three treatments". In: *Statistics in Biopharmaceutical Research* 2.3 (2010), pp. 310–318.

[68] Lachin, J. M. "Statistical properties of randomization in clinical trials". In: *Controlled Clinical Trials* 9 (1988), pp. 289–311.

[69] Lind, James. "A treatise on the scurvy. In three parts. Containing an inquiry into the nature, causes and cure, of that disease. Together with a critical and chronological view of what has been published on the subject." In: Edinburgh: Sands, Murray and Cochran, 1753. Chap. Part II, Chapter IV.

[70] Markarayan, Tigran and Rosenberger, W.F. "Exact Properties of Efron's Biased Coin Randomization Procedure". In: *The Annals of Statistics* 38.3 (2010), pp. 1546–1567.

[71] Melfi, V. F., Page, C., and Geraldes, M. "An Adaptive Randomized Design with Application to Estimation". In: *Canadian Journal of Statistics* 29.1 (2001), pp. 107–116.

[72] O'Rourke, P.P. et al. "Extracorporeal membrane oxygenation and conventional medical therapy in neonates with persistent pulmonary hypertension of the newborn: a prospective randomized study". In: *Pediatrics* 84.6 (1989), pp. 957–963.

[73] Paganoni, A.M. and Secchi, P. "A numerical study for comparing two response-adaptive designs for continuous treatment effects". In: *Statistical Methods and Applications* 17.3 (2007), pp. 321–346.

[74] Palmer, C.R. "Ethics, data-dependent designs, and the strategy of clinical trials: time to start learning-as-we-go?" In: *Statistical Methods in Medical Research* 11 (2002), pp. 381–402.

[75] Park, E. and Chang, Y.I. "Multiple-stage sampling procedure for covariate-adjusted response-adaptive designs". Epub ahead of print. 2013.

[76] Pocock, S.J. "Allocation of Patients to Treatment in Clinical Trials". In: *Biometrics* 35.1 (1979), pp. 183–197.

[77] Pong, A. and Chow, S.C., eds. *Handbook of Adaptive Designs in Pharmaceutical and Clinical Development*. Boca Raton: CRC Press, 2011.

[78] Robbins, H. "Some aspects of the sequential design of experiments". In: *Bulletin of the American Mathematical Society* 58.5 (1952), pp. 527–535.

[79] Rosenberger, W.F. "New Directions in Adaptive Designs". In: *Statistical Science* 11 (1996), pp. 137–149.

[80] Rosenberger, W.F. and Hu, F. "Maximizing power and minimizing treatment failures in clinical trials". In: *Clinical Trials* 1.2 (2004), pp. 141–147.

[81] Rosenberger, W.F. and Lachin, J. M. *Randomization in Clinical Trials Theory and Practice*. 1st ed. Wiley-Interscience, 2002.

[82] Rosenberger, W.F. and Lachin, J. M. "The use of response-adaptive designs in clinical trials". In: *Controlled Clinical Trials* 14.6 (1993), pp. 471–484.

[83] Rosenberger, W.F. et al. "Optimal Adaptive Designs for Binary Response Trials". In: *Biometrics* 57 (2001), pp. 909–913.

[84] Royall, R.M. "Ethics and statistics in randomized clinical trials". In: *Statistical Science* 6.1 (1991), pp. 52–62.

[85] Royall, R.M. "Ethics and statistics in randomized clinical trials: Rejoinder". In: *Statistical Science* 6.1 (1991), pp. 83–88.

[86] Silvey, S.D. *Optimal Designs*. London: Chapman and Hall, 1980.

[87] Simes, R.J. "Ethics and statistics in randomized clinical trials: Comment". In: *Statistical Science* 6.1 (1991), pp. 78–80.

[88] Simon, Richard. "Adaptive Treatment Assignment Methods and Clinical Trials". In: *Biometrics* 33.4 (1977), pp. 743–749.

[89] Smith, R.L. "Properties of biased coin designs in sequential clinical trials". In: *The Annals of Statistics* 12.3 (1984), pp. 1018–1034.

[90] Soares, J.F. and Wu, C.F.J. "Some restricted randomization rules in sequential designs". In: *Communications in Statistics, Series A* 12 (1983), pp. 2017–2034.

[91] Spiegelhalter, D.J., Abrams, K.R., and Myles, J.P. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. New York, NY: Wiley & Sons, 2004.

[92] Sun, R., Cheung, S.H., and Zhang, L.-X. "A generalized drop-the-loser urn for multi-treatment clinical trials". In: *Journal of Statistical Planning and Inference* 137.6 (2007), pp. 2011–2023.

[93] Sverdlov, Oleksandr, Rosenberger, W.F., and Hu, F. "Adaptive Randomization for Clinical Trials". In: (2011).

[94] Tamura, R.N. et al. "A Case Study of an Adaptive Clinical Trial in the Treatment of Out-Patients with Depressive Disorder". In: *Journal of the American Statistical Association* 89.427 (1994), pp. 768–776.

[95] Thall, P.F. and Wathen, J.K. "Practical Bayesian adaptive randomisation in clinical trials". In: *European Journal of Cancer* 43 (2007), pp. 859–866.

[96] Thompson, W. R. "On the likelihood that one unknown probability exceeds another in view of the evidence of the two samples". In: *Biometrika* 25.3/4 (1933), pp. 285–294.

[97] Tymofyeyev, Y., Rosenberger, W.F., and Hu, F. "Implementing optimal allocation in sequential binary response experiments". In: *Journal of the American Statistical Association* 102.477 (2007), pp. 224–234.

[98] UK Collaborative ECMO Trial Group. "UK collaborative randomised trial of neonatal extracorporeal membrane oxygenation". In: *The Lancet* 348.9020 (1996), pp. 75–82.

[99] U.S. Department of Health and Human Services and Food and Drug Administration. *Innovation/stagnation: Challenge and opportunity on the critical path to new medical products*. 2004.

[100] Wald, Abraham. *Sequential Analysis*. 1st ed. New York, NY: John Wiley & Sons, Inc., 1947.

[101] Wei, L.J. "A class of designs for sequential clinical trials". In: *Journal of the American Statistical Association* 72.358 (1977), pp. 382–386.

[102] Wei, L.J. "An application of an urn model to the design of sequential controlled clinical trials". In: *Journal of the American Statistical Association* 73.363 (1978), pp. 559–563.

[103] Wei, L.J. "The Adaptive Biased Coin Design for Sequential Experiments". In: *The Annals of Statistics* 6.1 (1978), pp. 92–100.

[104] Wei, L.J. and Durham, S. "The randomized play-the-winner rule in medical trials". In: *Journal of the American Statistical Association* 73.364 (1978), pp. 840–843.

[105] Weinstein, M. C. "Allocation of subjects in medical experiments". In: *New England Journal of Medicine* 291.24 (1974), pp. 1278–1285.

[106] Zelen, M. "Play-the-winner rule and the controlled clinical trial". In: *Journal of the American Statistical Association* 64.325 (1969), pp. 131–146.

[107] Zelen, M. "The randomization and stratification of patients to clinical trials". In: *Journal of Chronic Diseases* 27.7-8 (1974), pp. 365–375.

[108] Zhang, L. and Rosenberger, W.F. "Response-Adaptive Randomization for Clinical Trials with Continuous Outcomes". In: *Biometrics* 62.2 (2006), pp. 562–569.

[109] Zhang, L. and Rosenberger, W.F. "Response-adaptive randomization for survival trials: the parametric approach". In: *Journal of Applied Statistics* 56.2 (2007), pp. 153–165.

[110] Zhang, L.-X. and Hu, F. "A new family of covariate-adjusted response-adaptive designs and their properties". In: *Applied Mathematics - A Journal of Chinese Universities* 24.1 (2009), pp. 1–13.

[111] Zhang, L.-X. et al. "A generalized drop-the-loser urn for clinical trials with delayed responses". In: *Statistica Sinica* 17 (2007), pp. 387–409.

[112] Zhang, L.-X. et al. "Asymptotic properties of covariate-adjusted response-adaptive designs". In: *The Annals of Statistics* 35.3 (2007), pp. 1166–1182.

[113] Zhang, L.-X. et al. "Immigrated Urn Models-Theoretical Properties and Applications". In: *The Annals of Statistics* 39.1 (2011), pp. 643–671.

[114] Zhu, H. and Hu, F. "Implementing optimal allocation for sequential continuous responses with multiple treatments". In: *Journal of Statistical Planning and Inference* 139.7 (2009), pp. 2420–2430.

Chapter 2

Response-Adaptive Biased Coin Design with Unknown, Unequal Treatment Variances

2.1 Summary

Modern medical experiments accrue patients—and hence response data—throughout the duration of a trial. Designs which prospectively plan to modify patient allocation by leveraging accumulating data are response-adaptive. This paper examines a response-adaptive design that randomizes patients relative to current treatment estimates, scaled by an arbitrary positive constant T. The design is only suitable in a limited number of situations: covariate effects must be the same for both treatments, patient responses must be immediately observable, and treatment variances must be known and equal. This article proposes an intuitive replacement for T that allocates more patients to the superior treatment while mitigating the loss of power and increased bias inherent in many response-adaptive designs. This article also ex-

pands the applicability of the new design by relaxing the assumption of known, equal variances. The utility and benefits of the new design are illustrated by a real-world application of an HIV treatment adherence intervention.

Keywords: response-adaptive randomization, biased coin, power, Type I error

2.2 Introduction

2.2.1 Response-Adaptive Randomization

Unlike traditional agricultural experiments where data for all experimental units are collected nearly simultaneously at the end of a trial, modern medical experiments accrue and treat patients—hence obtain treatment response data—throughout the duration of a trial. Researchers may leverage accumulating data to modify various aspects of the experiment including the experiment's population, randomization rates, available treatments, duration, and/or analysis based on pre-specified criteria. Designs that evolve in a prospectively planned manner based on accumulated data are known as adaptive designs.

Adaptive designs gained theoretical support towards the end of the century among academic statisticians [35]. Towards the end of the 2000s, the US Food and Drug Administration called for further research on and implementation of adaptive designs to improve patient care and accelerate the treatment approval process [16, 17]. The commercial appeal of harnessing adaptive designs such as response-adaptive randomization (RAR) enticed the pharmaceutical industry to encourage research and leverage approved adaptive designs. One established advantage of RAR designs is their potential to assign significantly more patients to the better treatment. Other RAR designs have the ability to maximize a design's power (alternatively maintain power

relative to a nonadaptive design but decrease the requisite sample size) [7, 21, 34]. Unfortunately, no known adaptive design provides both benefits simultaneously [19, 22, 32].

Section 2.5 illustrates an ethical advantage to RARs by transforming a nonadaptive HIV intervention into a response-adaptive clinical trial. This real-world reconstruction highlights the number of additional patients who could benefit from the more successful treatment under a RAR design while still correctly identifying the superior treatment.

2.2.2 Biased Coin Designs

One popular adaptive design family is that of the biased coin design (BCD). BCDs for medical experiments originally surface under the context of encouraging balance during patient randomization by skewing allocation probabilities towards the therapy with fewer patient assignments [2, 14, 39, 44]. When a trial has two treatments (A and B), equal allocation corresponds to an unbiased, nonadaptive coin where the probability of assignment to either treatment is $\phi = \frac{1}{2}$. Other adaptive BCDs randomize patients to treatment A with probability $\phi = \frac{1}{2}$ only if the prior allocations are equally distributed between A and B, with probability $\phi = g(x) \in \left(\frac{1}{2}, 1\right]$ if there are more prior patients assigned to treatment B, and with probability $\phi = 1 - g(x)$ otherwise. The function $g(x)$ need only be rotationally symmetric about the point $(0, \frac{1}{2})$—i.e., $g(-x) = 1 - g(x)$ for all $x \in \mathbb{R}$; it may be a constant as in [14], a step function like [12] and [40], or any continuous link function per [44].

In adaptive, non-RAR BCDs, x tracks previous treatment assignments to determine the current allocation imbalance. Such BCDs harness this information to skew the incoming patient's allocation towards the lesser allocated treatment. RAR

BCDs are structured similarly but instead of the randomization depending on past assignments, x may be a function of treatment performance and/or patient covariates. Adaptive BCDs target covariate stratification [3, 36], maximize design power and efficiency [18, 42, 48], estimate optimal allocation proportions [15, 24, 29], and assign more patients to superior or successful treatment [4, 34].

2.2.3 Response-Adaptive Biased Coin Designs

Bandyopadhyay and Biswas present a RAR BCD for comparing two treatments with normal responses having known, equal variances and a common prognostic factor [8]. The design, referred to as BB, assigns a patient to treatment A with probability $\Phi\left(\frac{\mu_A-\mu_B}{T}\right)$ and to treatment B with probability $1-\Phi\left(\frac{\mu_A-\mu_B}{T}\right)$, where μ_k is the mean response of treatment k for $k = A, B$; T is an arbitrary positive constant; and Φ is the cumulative distribution function of the standard normal distribution. Selecting at value for T is discussed in detail in Section 2.3.2. In reality, if the treatment means or the difference between treatments were known, no experimentation would be necessary. Therefore, [8] propose using the currently accrued patient responses to estimate the treatment differences, assigning incoming patients to treatment based on the best estimates available at enrollment.

The BB RAR correctly allocates a larger proportions of patients to the better treatment in simulations. This ethical behavior comes at a cost. The design sustains a potentially large loss of power to detect a true difference in treatments and a slight increase to the risk of finding a nonexistent treatment difference compared to equal allocation. Moreover, both the assignment probability and the Type I and Type II error rates depend on the arbitrarily selected parameter T. On top of this, the original BB design is only suitable in a limited number of situations: covariate effects must

be the same for both treatments, patient responses must be immediately observable, and—most restrictively—treatment variances must be known and equal.

This article suggests a new design based on BB which embodies similar ethical advantages but has preferable statistical properties. A large component of the new design, called BBS, comprises replacing the arbitrary constant T with an intuitive adaptive statistic S, the pooled treatment standard deviation. With this modification, the BBS design exposes more patients to the superior treatment while mitigating the loss of power and increased bias inherent in the BB design. This article also expands the applicability of the BBS design by relaxing the need for known or equal treatment variances.

2.3 Clinical Trial Model and Design

2.3.1 Overarching Model

Consider a clinical trial with two treatments that sequentially enrolls and treats patients. Patients have P prognostic factors $\boldsymbol{Z} = (Z_1, \ldots, Z_P)$ that are independently and identically distributed. For example, Z_1 might be sex and Z_2 might be weight. The values of patient j's covariates $\boldsymbol{z}_j = (z_{j1}, \ldots, z_{jP})$ are observed at enrollment. Patient responses follow a normal distribution conditional on the treatment exposure, the prognostic factors, and the treatment-covariate interaction:

$$y_{jk} = \mu_k + \boldsymbol{\beta}_k \boldsymbol{z}_j' + \epsilon_{jk}$$

for patients $j = 1, \ldots, N$ and treatments $k = A, B$. In the model, μ_k is the mean effect of treatment k; $\boldsymbol{\beta}_k$ is the set of $(P \times 1)$ slopes scaling the covariate impacts on treatment k's response; $\boldsymbol{z}_j$ is the vector of prognostic factors for patient j; and

ϵ_{jk} is the random error for patient j on treatment k, independently and identically distributed $N(0, \sigma_k^2)$.

2.3.2 BB Design Details

The BB design only applies to situations when the treatment variances are known and equal ($\sigma_A^2 = \sigma_B^2$). Additionally, the covariate slopes—although not known—are also equal ($\beta_A = \beta_B$). Hence the BB model simplifies to

$$y_{jk} = \mu_k + \boldsymbol{\beta} \boldsymbol{z}_j' + \epsilon_j$$

for patients $j = 1, \ldots, N$ and treatments $k = A, B$. In the BB model, μ_k is the mean effect of treatment k; $\boldsymbol{\beta}$ is the vector of covariate slopes; $\boldsymbol{z}_j$ is the vector of prognostic factors for patient j; and ϵ_j is the random error for patient j, independently and identically distributed $N(0, \sigma^2)$.

The first $2m$ patients to arrive are arbitrarily assigned so that m patients receive treatment A and m receive treatment B. When patient j enrolls in the trial ($j = 2m + 1, \ldots, N$), the immediately observed responses of the previous $j - 1$ patients are leveraged to estimate the treatment difference $\widehat{\mu}_{A,j-1} - \widehat{\mu}_{B,j-1}$. Patient j is then randomly assigned to treatment A with probability $\Phi\left(\frac{\widehat{\mu}_{A,j-1} - \widehat{\mu}_{B,j-1}}{T}\right)$ and to treatment B with probability $1 - \Phi\left(\frac{\widehat{\mu}_{A,j-1} - \widehat{\mu}_{B,j-1}}{T}\right)$, with T an arbitrary positive constant. Patient j's response is immediately observed and the treatment difference is re-estimated so that the next patient $(j + 1)$ is allocated according to the latest estimates. This process is continued until the experiment concludes.

While arbitrary, the choice of T is—not surprisingly—a highly influential factor in the BB design. In the literature, values of T implemented in simulations vary with all authors commenting on the importance of the choice of scaling constant. Commonly

employed values include $T = 1$ [4, 5, 8, 10, 11, 31], $T = 2$ [8, 10, 11, 47], and $T = 3$ [8, 10, 11], but $T = 5$ and $T = 10$ are also leveraged [9, 11, 31]. The BB coin's ability to assign more patients to the better treatment is enhanced by selecting smaller values of T and mitigated by larger T. On the other hand, the same is true of the variance of the treatment estimators; smaller values of T increase the variability of the estimates. Thus the power to detect a treatment difference is higher in simulations with larger T for all but one of [8]'s trials of 40 patients. In fact, [8] conclude, "at the initial stages with inadequate data a larger value of T is preferred. One can start with a larger value of T and switch over to progressively smaller values at suitable stages." For this reason, an adaptive modification of T is proposed that (1) is a more intuitive, less arbitrary choice and (2) decreases as patients accrue throughout the trial.

2.3.3 BBS Design Details

For a study scenario as described in Section 2.3.1, assume that the treatment variances are unknown but thought to be equal or similar. Instead of the arbitrary constant T as the denominator of the RAR BCD, define

$$S = \sqrt{\frac{(N_A - 1)s_A^2 + (N_B - 1)s_B^2}{N_A + N_B - 2}},$$

the pooled treatment standard deviation estimator, to be the new scaling constant. As before, assign the first set of patients to treatment so that at least m patients receive and respond to each treatment. When patient j enrolls in the trial ($j \geq 2m + 1, \ldots, N$), the immediately observed responses of the previous $j - 1$ patients are leveraged to estimate the treatment difference $\widehat{\mu}_{A,j-1} - \widehat{\mu}_{B,j-1}$ and the pooled

standard deviation

$$\widehat{S} = \sqrt{\frac{(N_{A,j-1}-1)s^2_{A,j-1} + (N_{B,j-1}-1)s^2_{B,j-1}}{N_{A,j-1}+N_{B,j-1}-2}}.$$

Patient j is then randomly assigned to treatment A or B with probabilities

$$\Phi\left(\frac{\widehat{\mu}_{A,j-1}-\widehat{\mu}_{B,j-1}}{\widehat{S}}\right) \quad \text{or} \quad 1-\Phi\left(\frac{\widehat{\mu}_{A,j-1}-\widehat{\mu}_{B,j-1}}{\widehat{S}}\right),$$

respectively. Patient j's response is immediately observed and the treatment difference is re-estimated so that the next patient $(j+1)$ is allocated according to the latest estimates. This process continues until all patients are treated. Denote this modification of the BB design as BBS.

Leveraging the BBS design provides several advantages. First, S is an unbiased estimator under the assumption of equal variances with N_k the number of patients who have responded to treatment k and s^2_k the sample variance estimate for the mean of treatment k. Second, S is an intuitive choice for scaling treatment estimates, reminiscent of standardized estimators and forming a statistically pleasing ratio with the numerator estimates. Third, as the number of patients enrolled increases, the precision of the estimators increase as well. S, therefore, is largest when the sample size is small and decreases as patient responses accrue. Hence when the treatment differences are least precise, they are moderated by suitably sized variance estimators. As the numerator decreases, the dampening effect is mitigated by increased precision in the denominator, allowing the BBS coin to be adaptive and assign patients to the better treatment with higher probability.

2.4 Simulations and Results

2.4.1 Simulation Specifics

Sections 2.4.2 and 2.4.3 describe the results of simulated clinical trials under equal allocation, BB, and the newest competitor BBS. Treatment mean pairs (μ_A, μ_B) are set to no effect $(0, 0)$ and three levels of positive effect in treatment A $(0.1, 0)$, $(0.5, 0)$, and $(1, 0)$. For select scenarios, a positive effect of $\mu_B = 1$ in treatment B and no effect in treatment A is also simulated. Individual patient errors are simulated under equal and unequal conditions; $\sigma_B = 1$ while σ_A ranges from 0.5 to 1.5 in increments of 0.2. For select scenarios, σ_A is held constant at 1 while σ_B varies from 0.5 to 1.5 in increments of 0.2. For the BB design, constant values of $T = 1$, 2, and 3 are selected and are referred to as BB_1, BB_2, and BB_3, respectively. Further adopting simulation conventions from [8], one normally-distributed prognostic factor is simulated. While estimated separately for each treatment, the true covariate effects, variances, and slopes are identical for both treatments with $Z \sim N(1, 1)$ and $\beta = 2$.

In the BB design, rejection of the null hypothesis $H_0 : \mu_A \leq \mu_B$ is calculated with a one-sided, two-sample t-test with significance level $\alpha = 0.05$. Under the assumption of equal variances, the t-statistic standard error employs the sample pooled variance estimates. Relaxing the assumption of equal variances expands the applicability of the proposed BBS design to potentially unequal variance situations. Hence, in the equal allocation and BBS designs, the t-statistic standard error is calculated using treatment-specific sample variances. In these cases, the underlying hypothesis distribution leverages the Welch-Satterthwaite approximated pooled degrees of freedom [38, 45].

Sample sizes mimic small to large Phase III clinical trials (N = 50; 100; 500;

1,000; 5,000). The initial number of response estimates necessary to begin adaptive randomization is a minimum of $m = 10$ for each treatment. That is, at least 10 patients must be exposed to treatment A and another 10 patients must be exposed to treatment B before adaptive allocation estimates can be leveraged. These first patients are assigned to study arms via equal allocation and respond to treatment immediately. All patient outcomes are assumed to be instantaneously observed; no delay is incorporated into these trials. For each scenario described, 1,000 replications are simulated in SAS IML [37].

2.4.2 Patient Allocation

Tables 2.1 – 2.5 contain proportions and standard deviations of patients assigned to treatment A by design and treatment effect values. Table 2.1 details the average patient allocations for BBS; BB under $T = 1, 2, 3$; and equal allocation when treatment variances are equal ($\sigma_A = \sigma_B = 1.0$). In Table 2.1, the left most column indicates which design simulations results are being described. The second column specifies which value of μ_A applies to a particular row with all rows having the same value of $\mu_B = 0.0$. The remaining five columns are the mean (SD) proportion of patients assigned to treatment A over 1,000 simulations for each of the five trial sizes $N = 50$; 100; 500; 1,000; and 5,000.

The first section of data in Table 2.1 summarizes allocations from simulations where patients are randomized to treatment under the BBS design. The first row represents results for treatment means $\mu_A = \mu_B = 0.0$. Across all five values of clinical study sizes, the treatment A allocation is 50%—that is, patient assignments are balanced between treatments. Equal exposure is appropriate in this case as there is no difference between treatment A and treatment B mean effects. The allocation

variance is highest when trial sizes are small (0.12 for $N = 50$ and 0.13 for $N = 100$) and the variance decreases for moderate and large studies (0.07 for $N = 500$ and 0.02 for $N = 5{,}000$). The second through fourth rows represent results when the treatment means differ. Specifically, row two reports results when $\mu_A = 0.1$, row three when $\mu_A = 0.5$, and row four when $\mu_A = 1.0$. In each of these rows, the proportion of patients allocated to treatment A increases for a given study size. For example, when $N = 500$, allocation to treatment A grows from 50% in row one to 54% in row two, then jumps to 69% in row three and 83% in row four.

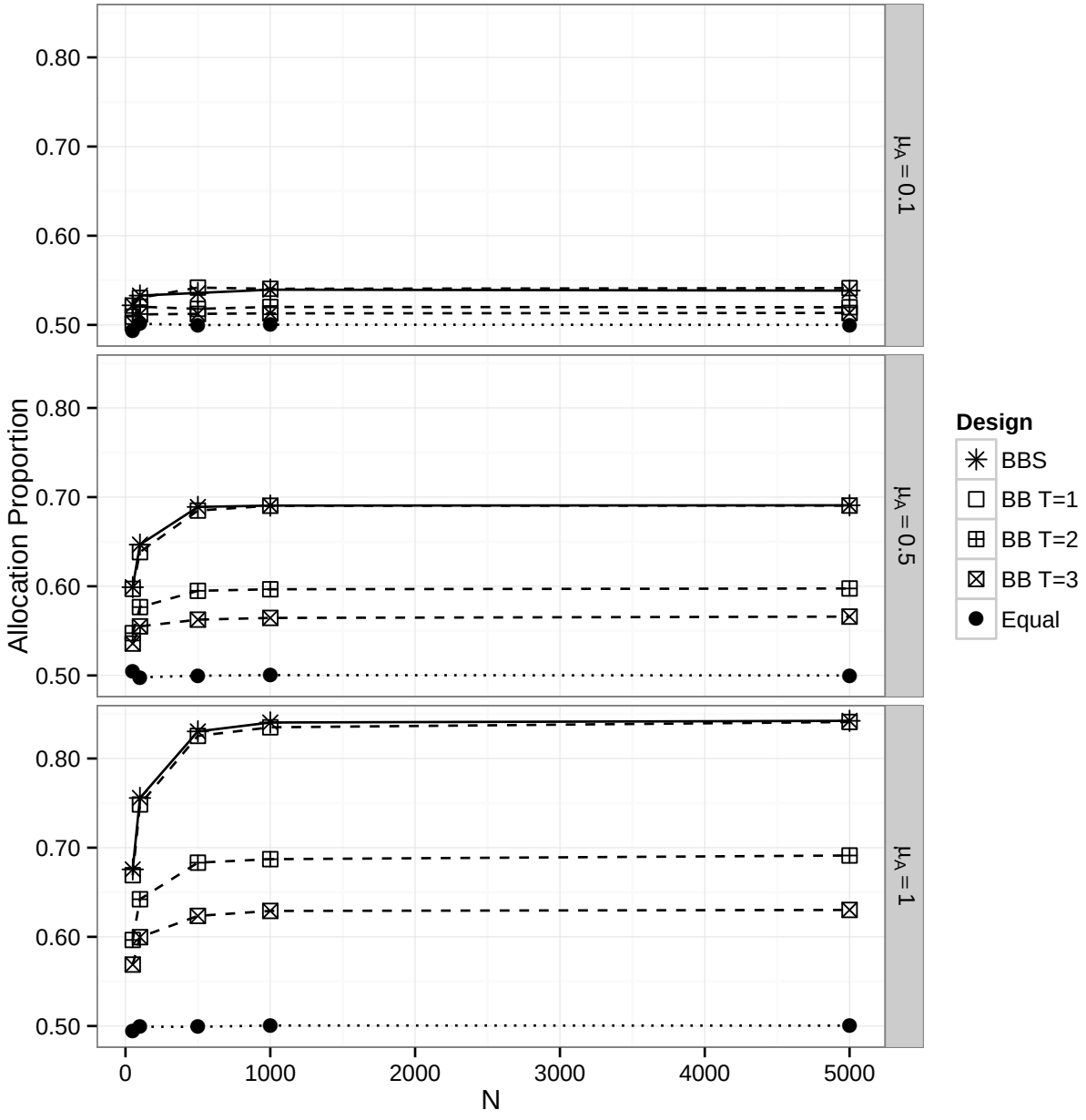
The subsequent sections of Table 2.1 are structured similarly for each simulated design, beginning with BB_1. The values in the second section closely echo those of the BBS design. As the scaling parameter T increases from 1 to 3, however, the ability of the BB design to allocate more patients to the treatment with larger mean decreases. For example, when $N = 500$, both BBS and BB_1 assign 83% of patients to treatment A when $\mu_A = 1.0$ and $\mu_B = 0.0$. When $T = 2$, BB_2's treatment A allocation drops to 68% for the same parameters. Furthermore, the BB_3 design only randomizes 62% of patients to treatment A when $N = 500$ and $\mu_A = 1.0$. Of course, equal allocation (Equal) assigns approximately equal proportions of patients to treatment A as to treatment B regardless of enrollment size or study parameters.

In addition to the declining patient randomization proportions across designs, the variance of the allocations also decreases across the designs in similar succession. The assignment proportions allocations decrease as trial size increases, as μ_A increases relative to μ_B, and as the adaptive design denominator increases—forcing the designs to be more like equal allocation. While the variance in the BBS design is larger than that of BB_2, BB_3, and equal allocation for the same sample size, the gains in the number of patients receiving better treatment can be substantial. For example, BBS and BB_1 report a 10% increase in patients exposed to the superior treatment in clinical

trials with $\mu_A = 0.5$ and $N = 50$ compared to a 4% increase in BB_3. That amounts to 5 additional patients on treatment A for the BBS design versus only 2 additional individuals for the BB design with $T > 1$. This extra ethicality is accompanied by a randomization standard deviation increase of merely 0.03. At $N = 500$, the patient benefit includes nearly 10 additional individuals exposed to the better treatment on BBS with a treatment difference of 0.5 compared to 3-5 additional patients on BB_2 or BB_3. Again, the difference in standard deviation is still only 0.03.

Figure 2.1 illustrates the trends in patient assignment from Table 2.1 across the five simulated trial sizes for all three treatment differences when $\sigma_A = \sigma_B = 1.0$. As seen in Table 2.1 and in Figure 2.1, patient allocation is impacted by the difference in treatment effects, the total enrollment of the study, and the particular randomization design when treatment variances are held constant. For a given combination of treatment mean effects, the allocation patterns across designs and trial sizes are consistent in their form, but vary in their magnitude. That is, all four RAR designs rapidly improve in terms of the ethicality of their assignments as studies grow from small to moderate sizes. This growth plateaus when moving from moderate to large trials. Moreover, the order and shape of the allocation curves across enrollment sizes is consistent for each treatment difference, but the curves are scaled relative to the magnitude of the difference. The BBS and BB_1 curves are the most ethical allocations. BB_2 lags behind the frontrunners, peaking in ethically approximately where the first two designs' minimum ethical allocations begin. BB_3 follows BB_2 with a much smaller gap than between BB_2 and BB_1. Equal allocation steadily assigns 50% of patients to treatment A regardless of clinical trial size or difference in treatment means. Overall, the BBS and BB_1 designs continue to expose more patients to the treatment with larger effect size compared to equal allocation even when the difference in treatment means is small relative to the treatment variance.

Figure 2.1: Proportion of patients assigned to treatment A by design and clinical trial size as μ_A varies for fixed treatment parameter values, $\mu_B = 0$, and $\sigma_A = \sigma_B = 1$. Data points are grouped by design; BBS is represented by $*$ with a solid line, BB by variations on $\square$ with a dashed line, and equal allocation by $\bullet$ with a dotted line.



Tables 2.2 – 2.5 present patient assignment proportions for BBS, BB, and equal allocation designs under different treatment effect scenarios ($\mu_A = 0.0$, 0.1, 0.5, and 1.0, respectively, with $\mu_B = 0.0$ for all four tables) when treatment variances are

unequal. In Tables 2.2 – 2.5, the first column on the left designates the design for the section. The second column identifies which value of σ_A applies to each row. All rows across all four tables represent $\sigma_B = 1.0$. The remaining five columns are the mean (SD) proportion of patients assigned to treatment A over 1,000 simulations for each of the five trial sizes $N = 50; 100; 500;$ 1,000; and 5,000.

The first section in Table 2.2 indicates that when treatment means are equal, the BBS design randomizes patients to treatment in a nearly balanced allocation, subject to slight fluctuations tied to treatment variance. For example, the first row demonstrates that when $\sigma_A = 0.5$, the BBS design assigns 51–52% of patients to treatment A in small to moderately sized clinical trials and 50% of patients to treatment A for large studies ($N =$ 5,000). The standard deviation of these allocations decreases as N increases. For $\sigma_A = 0.5$ and $\sigma_B = 1.0$, the BBS allocation standard deviation starts at 0.13 when $N = 50$, reaches 0.09 when $N = 500$, and finishes at 0.03 when $N =$ 5,000. In the reverse scenario, row six in Table 2.2 demonstrates that when $\sigma_A = 1.5$ and $\sigma_B = 1.0$, the BBS design assigns 49% of patients to treatment A in all but one scenario and 50% of patients are allocated to each treatment when $N =$ 5,000. As with row one, the randomization standard deviation decreases as N increases: 0.12 when $N = 50$, 0.08 when $N = 500$, and 0.03 when $N =$ 5,000. When the treatment standard deviations are approximately equal ($\sigma_A = 0.9$ or 1.1 in rows three and four), the BBS assignment is balanced and the data are similar to the first row in Table 2.1.

The following four sections of Table 2.2 share the same organization for each design, beginning with BB_1 and ending with equal allocation. BB_1 simulated behavior follows similar allocation patterns as BBS but with different variance patterns. In terms of the randomization proportions, when $\sigma_A = 0.5$, slightly more patients are exposed to treatment A; when $\sigma_A = 1.5$, slightly fewer patients are exposed to treatment A; when σ_A is closer to 1.0 or when $N =$ 5,000, treatment A and treatment B

assignment proportions are balanced. BB_1 randomization variance is distinct from BBS. As N increases, the allocation standard deviations decrease; however, within a fixed value of N, BB_1 allocation standard deviation increases as σ_A increases. For example, when $N = 500$, the standard deviation increases from 0.06 when $\sigma_A = 0.5$ to 0.10 when $\sigma_A = 1.5$. The BB_2 design shares this allocation variance pattern, but to a smaller degree. BB_3 randomization variance appears predominantly consistent within a trial size across treatment A variance levels. Moreover, the allocation proportions from BB_2 and BB_3 are nearly all 50%—no relationship with σ_A is evident. As anticipated, equal allocation exposes equal proportions of patients to treatments A and B. Randomization variance for the nonadaptive design decreases as clinical trial size increases, but is not altered by treatment standard deviations.

Tables 2.3 – 2.5 adhere to the same format as Table 2.2, except in these tables a treatment difference exists. In the simulations represented, BBS now allocates more patients to treatment A. The relationships in randomization proportions and allocation variance across values of σ_A and N persist. Across the four tables, the larger the difference in treatments, the higher the proportion of patients assigned to treatment A. The BB designs also expose more patients to the superior treatment. BB_1 allocations are closest to BBS, without the same sensitivity to changes in σ_A. BB_2 and BB_3 are each less ethical in their assignments than BB_1 but still manage to randomize more patients to the treatment with larger mean than equal allocation.

To confirm the symmetry of the RAR allocations across values of σ_A and σ_B, all four adaptive designs are also simulated with $\sigma_A = 1.0$ and $\sigma_B = 0.5, \ldots, 1.5$ as well as with $\mu_A = 0.0$ and $\mu_B = 1.0$. Although the simulation results are not presented in a table, Figure 2.2 illustrates the patient allocation proportions for BBS and BB_1 when $N = 100$. Other clinical trial enrollment sizes demonstrate patterns which follow those exhibited by $N = 100$. Figure 2.2 contains four graphs depicting

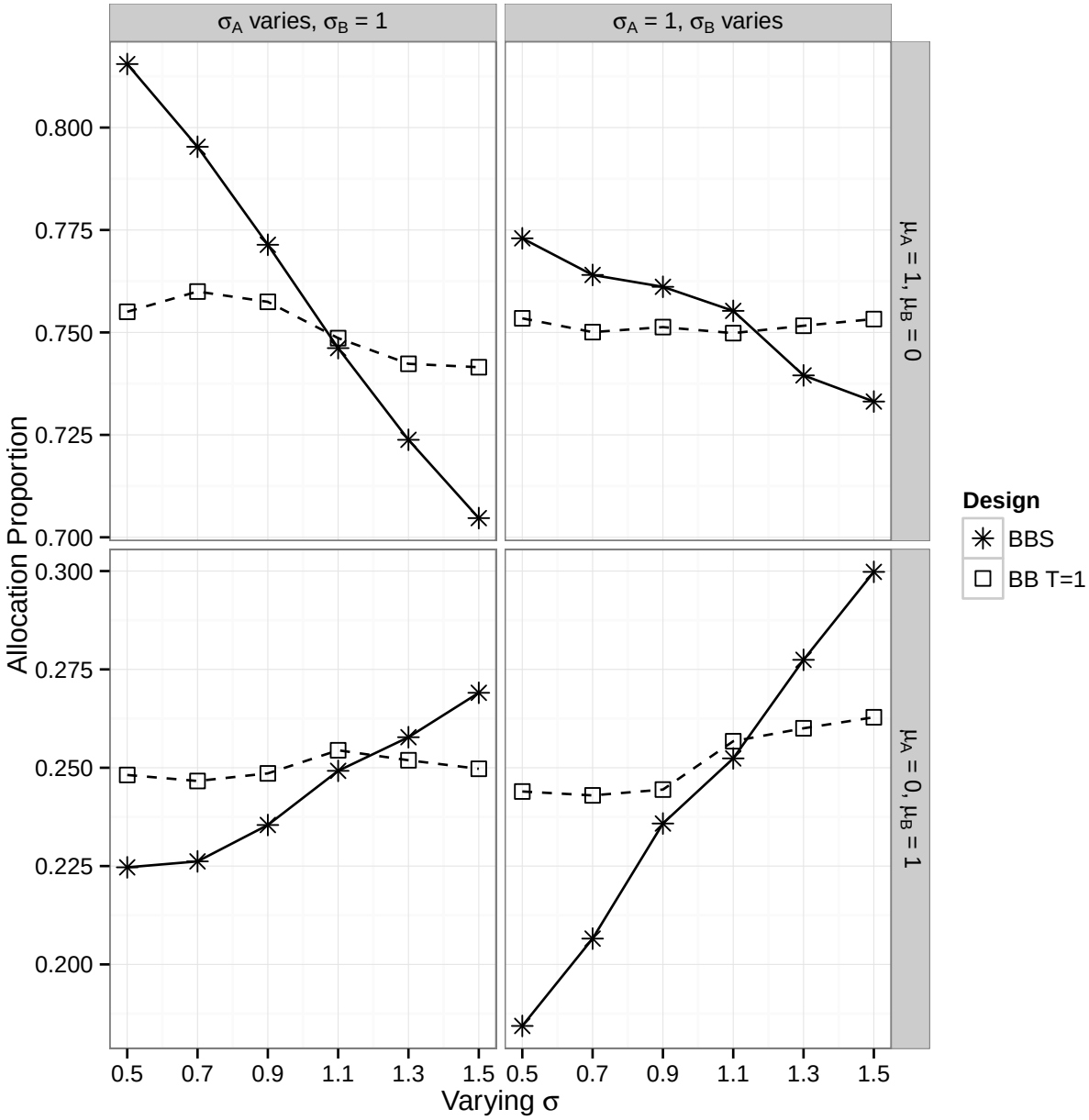
treatment allocation rates at the four combinations of $\mu_A = 0.0$ and $\mu_B = 1.0$ or $\mu_A = 1.0$ and $\mu_B = 0.0$ crossed with σ_A varying and $\sigma_B = 1.0$ or $\sigma_A = 1.0$ and σ_B varying. For example, the top right quadrant of Figure 2.2 contains the proportion of patients allocated to treatment A by the BBS and BB_1 designs when $\mu_A = 1.0$, $\mu_B = 0.0$, $\sigma_A = 1.0$, and σ_B varies from 0.5 to 1.5 along the horizontal axis. In the top left quadrant of Figure 2.2, $\mu_A = 1.0$, $\mu_B = 0.0$, $\sigma_B = 1.0$, and σ_A varies from 0.5 to 1.5 along the horizontal axis.

All four graphs illustrate two major trends in the BBS and BB designs. First of all, for a given set of treatment variances, these ethical RAR designs successfully expose more patients to the treatment with larger mean effect. Secondly, more patients can be randomized to the better treatment when the total design variance is minimized relative to a fixed difference in treatment means.

When considering the graphs in Figure 2.2 in horizontal pairs, an additional pattern emerges for the BBS design. For a fixed set of treatment mean effects and total treatment variance, BBS exposes more patients to the superior treatment when the treatment with larger mean effect has smaller variance. For example, when $\mu_A = 1.0$ and $\mu_B = 0.0$, 82% of patients can be exposed to treatment A on the BBS design when $\sigma_A = 0.5$ and $\sigma_B = 1.0$ (left most point in the top left quadrant) versus only 77% when $\sigma_A = 1.0$ and $\sigma_B = 0.5$ (left most point in the top right quadrant). Similarly, when $\mu_A = 1.0$ and $\mu_B = 0.0$, 73% of patients can be exposed to treatment A on the BBS design when $\sigma_A = 1.0$ and $\sigma_B = 1.5$ (right most point in the top right quadrant) versus 70% when $\sigma_A = 1.5$ and $\sigma_B = 1.0$ (right most point in the top left quadrant).

Examining diagonal pairs of graphs from Figure 2.2, a final pattern of BBS allocation becomes obvious. The BBS designs is symmetric in its ethicality, after accounting for total variance. For instance, when the treatment with larger mean effect

Figure 2.2: Proportion of patients assigned to treatment A by the BBS and BB_1 designs for $N = 100$ patients as μ_A, μ_B, σ_A, and σ_B vary. Data points are grouped by design; BBS is represented by $*$ with a solid line and BB_1 by $\square$ with a dashed line.



has smaller variance, an additional 32% of patients can be assigned to that treatment (left most points of the top left and bottom right quadrants), regardless of whether that larger treatment mean is μ_A or μ_B. By contrast, when the treatment with larger mean effect has larger variance, an additional 27% of patients can be assigned to that

treatment (left most points of the top right and bottom left quadrants), regardless of whether that larger treatment mean is μ_A or μ_B.

2.4.3 Power and Type I Error

Tables 2.6 – 2.10 contain proportions and standard deviations of trials that reject the null hypothesis that the mean effect of treatment A is no better than the mean effect of treatment B. Table 2.6 details the average rejection rates for the BBS; BB under $T = 1, 2, 3$; and equal allocation rejection rates when treatment variances are equal ($\sigma_A = \sigma_B = 1.0$). The left most column of Table 2.6 specifies which design's results are being described. The second column specifies the clinical trial size N associated with a particular row. The next three columns contain the mean (SD) proportion of trials that correctly conclude that $\mu_A > \mu_B$ when $\mu_A = 0.1, 0.5$, and 1.0, respectively, with $\mu_B = 0.0$. The final column reports the mean (SD) proportion of trials that incorrectly conclude $\mu_A > \mu_B$ when, in fact, $\mu_A = \mu_B = 0.0$.

The first section of data in Table 2.6 summarizes rejection rates from simulations where patients are assigned to treatment under the BBS design. The first row represents results for small studies with $N = 50$. As expected, the probability of the BBS accurately rejecting the null hypothesis increases as the difference in treatment means increases. Specifically, the BBS rejection power across 1,000 simulations is 11% when $(\mu_A, \mu_B) = (0.1, 0.0)$ versus 40% when $(\mu_A, \mu_B) = (0.5, 0.0)$ and 76% when $(\mu_A, \mu_B) = (1.0, 0.0)$ with standard deviations 0.31, 0.49, and 0.43, respectively. The probability of falsely rejecting the null hypothesis in a clinical trial of size $N = 50$ is 7% for the BBS with a standard deviation of 0.26. On the other end of the spectrum, the fifth row in Table 2.6 represents BBS results for large studies with $N = 5{,}000$. In particular, power when $(\mu_A, \mu_B) = (0.1, 0.0)$ is only 79% with standard deviation

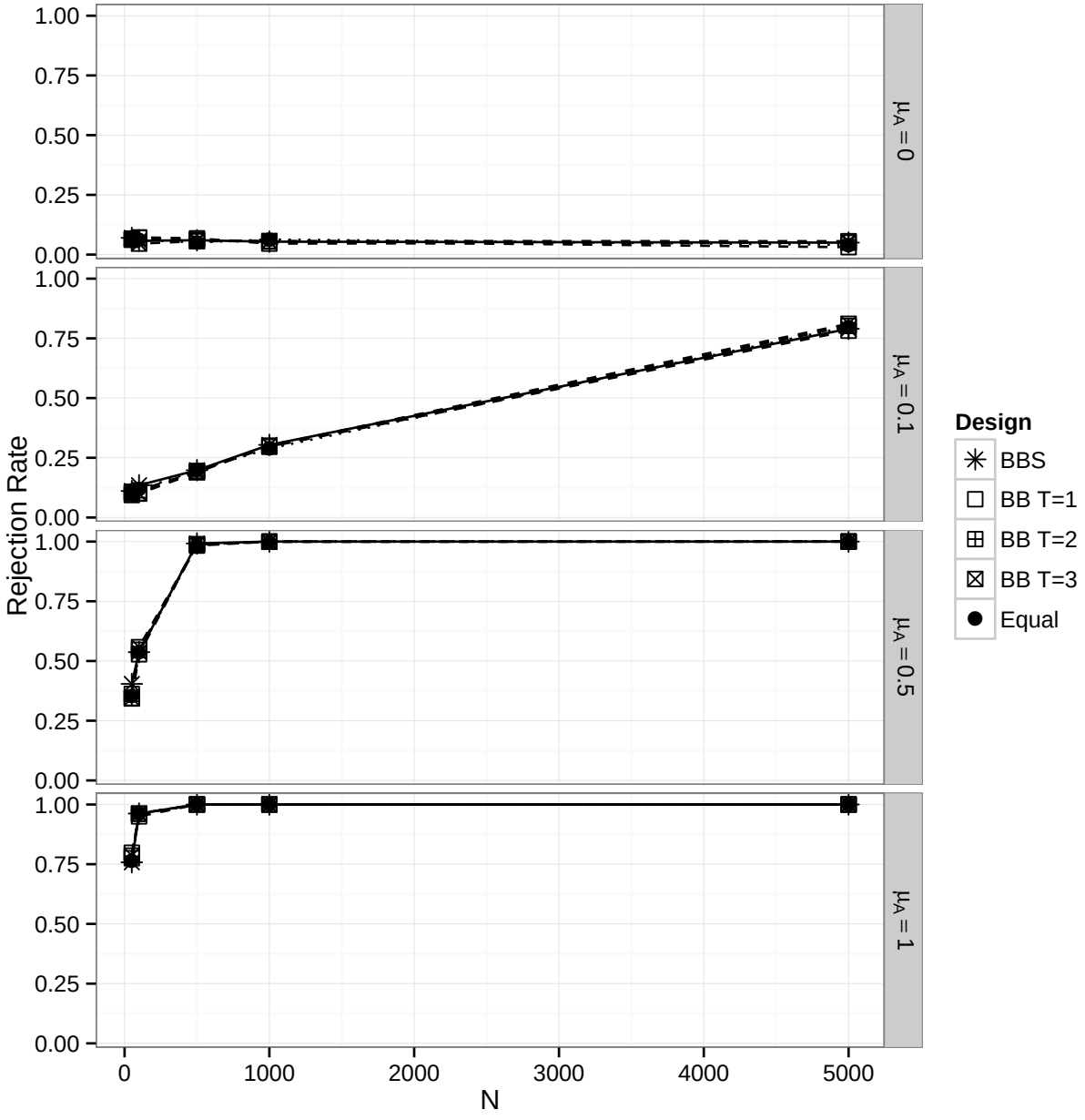
0.41, but power when $(\mu_A, \mu_B) = (0.5, 0.0)$ or $(1.0, 0.0)$ is 100% with negligible variance. Moreover, the type I error rate is 5% for the BBS design when $N = 5{,}000$ with a standard deviation of 0.22. Between enrollments of 50 patients and 5,000 patients, rejection rates increase within a fixed treatment difference and standard deviations generally decrease. In contrast, rejection rates and their standard deviations decrease for the BBS as trial size increases when there is no difference in treatment mean effects.

The remaining four sections of Table 2.6 are structured similarly for each simulated design, displaying similar power and type I error patterns, as well. All designs present power ranging from approximately 10% when $N = 50$ to roughly 80% when $N = 5{,}000$ for a 0.1 difference in treatment means and from over 75% when $N = 50$ to 100% when $N \geq 500$ for a 1.0 difference in treatment mean effects. All five designs also demonstrate slightly inflated type I errors of 6–7% when $N = 50$ falling to 5% or below when $N = 5{,}000$ and standard deviations ranging from approximately 0.25 to just over 0.20. Overall, all four RAR designs exhibit similar rejection rates to those of the nonadaptive equal allocation design, after adjusting for study size and treatment effects when $\sigma_A = \sigma_B = 1.0$.

Figure 2.3 illustrates the trends in trial rejection from Table 2.6 across the five simulated trial sizes for all four treatment effect combinations when $\sigma_A = \sigma_B = 1.0$. As seen in Table 2.6 and in Figure 2.3, when treatment variances are held constant, rejection rates are impacted by the difference in treatment means and the total clinical study enrollment numbers. The particular randomization design employed has minimal effect on the proportion of simulated trials rejected when the study sizes are moderate or large.

When no difference in treatments exists ($\mu_A = \mu_B = 0.0$), rejection rates range from to 5–7% in small trials ($N = 50$ or 100) to as low as 3 or 4% when $N = 5{,}000$. The

Figure 2.3: Proportion of trials which reject the null hypothesis by design when μ_A varies, $\mu_B = 0.0$, $\sigma_A = \sigma_B = 1$, and N varies. Data points are grouped by design; BBS is represented by $*$ with a solid line, BB by variations on $\square$ with a dashed line, and equal allocation by $\bullet$ with a dotted line.



total range of false rejection rates are small. Moreover, there are no clear patterns between the five designs' type I error rate rankings across the different values of sample size N. BBS is in the middle of the pack for most of the clinical trial sizes;

BB_1 runs high when trial sizes are small but achieves the lowest false rejection rate when $N = 5{,}000$; and equal allocation is at both extremes for moderate clinical trial sizes.

When the effect size is small relative to the treatment standard deviation ($\mu_A = 0.1$, $\mu_B = 0.0$, and $\sigma_A = \sigma_B = 1.0$), the power of all designs is low for small clinical trials and increases slowly as trial size increases. The total power in this cases peaks at approximately 80%. When the treatment effect size is modest relative to the treatment error ($\mu_A = 0.5$, $\mu_B = 0.0$, and $\sigma_A = \sigma_B = 1.0$), minimum rejection rates more than double for small clinical trials, rising swiftly to a plateau of near perfect rejection rates once a moderate number of patients are treated. When the treatment effect size equals treatment error ($\mu_A = 1.0$, $\mu_B = 0.0$, and $\sigma_A = \sigma_B = 1.0$), minimum rejection rates are high even for small clinical trials. Power quickly surpasses 95% even when $N = 100$, reaching 100% for $N \geq 500$. Additional simulation results not presented in tabular or graphical form confirm that there is no significant loss in design efficiency in the equal allocation design or the BBS design due to incorporating the potential for unequal variances when variances are, in fact, equivalent. That is, leveraging a pooled standard error estimate neither increases power nor decreases Type I error rates in equal allocation or BBS.

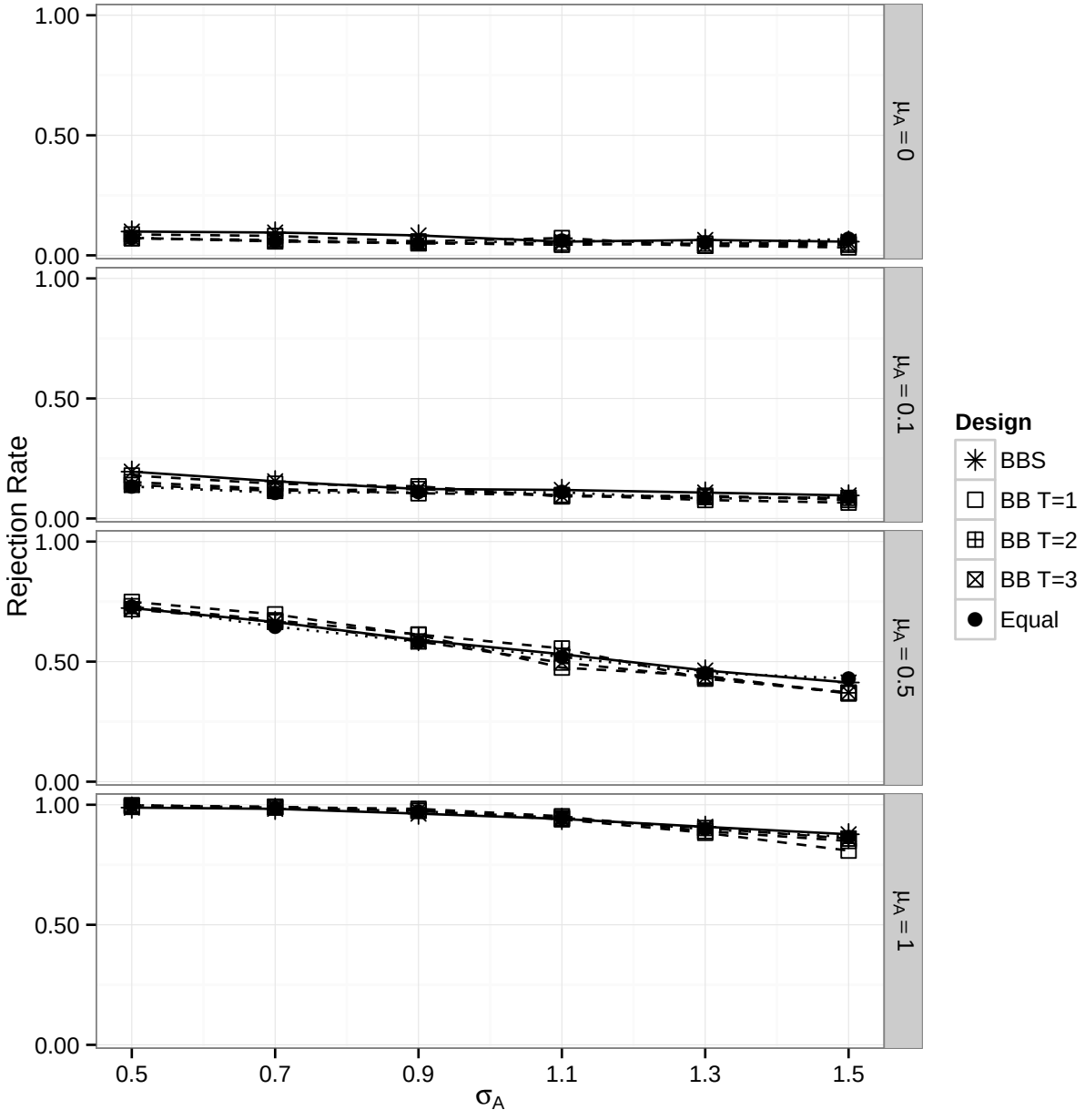
Tables 2.7 – 2.10 contain comprehensive results of rejection rates and standard deviations of these rates when $\mu_A = 0.0$, 0.1, 0.5, and 1.0, respectively; $\mu_B = 0.0$ for all four tables; and treatment variances are unequal under five different clinical trial sizes. In Tables 2.7 through 2.10, the first column on the left designates the design for the section. The second column identifies which value of σ_A applies to each row. All rows across all four tables represent $\sigma_B = 1.0$. The remaining five columns are the mean (SD) proportion of 1,000 simulated clinical trials for each of the five study sizes $N = 50; 100; 500;$ 1,000; and 5,000.

The first section in Table 2.7 represents the rate at which the BBS design incorrectly rejects the null hypothesis. For example, the first row indicates that when $\sigma_A = 0.5$, the BBS design rejects 10% of small trial null hypotheses ($N = 50$ or 100) even when treatment means are equal. Fewer false rejections occur for the BBS when clinical trials are moderate (9% when $N = 500$) or larger (6% when $N = 1{,}000$ and 7% when $N = 5{,}000$). The last row in the BBS section indicates that when $\sigma_A = 1.5$, BBS rejection rates are closer to the expected type I error rates for hypothesis tests using $\alpha = 0.05$: 6% of small trials incorrectly reject the hypothesis of equal treatment means and only 5% of moderate or large trials falsely conclude that $\mu_A > \mu_B$. The variance of the rejection rates decreases as clinical trial sizes increase. Additionally, the standard deviation of rejections decreases as σ_A increases from 0.5 to 1.5. In the case of the BBS design, standard deviations range from 0.29 to 0.23 when $N = 50$, from 0.29 to 0.22 when $N = 500$, and from 0.25 to 0.21 when $N = 5{,}000$.

The four subsequent sections of Table 2.7 follow the same organization for each of the simulated designs. BB_1 clinical trials reject the null hypothesis at similar rates to BBS and with similar variance patterns. BB_2 type I errors rates only reach a maximum of 7% when clinical trials are small and σ_A is small relative to σ_B. BB_3 rejection behaviors are similar as well, peaking at 9% when $N = 50$ and $\sigma_A = 0.5$ and decreasing as N increases and as σ_A increases. Equal allocation false rejection rates also vary from 5% in some simulations, particularly when the study size is small (6% for all trials of size $N = 50$) and treatment variance A differs moderately from treatment B variance. The rejection variance patterns are similar across all five designs.

Tables 2.8 – 2.10 adhere to the same format as Table 2.7, except in these tables a treatment difference exists. In Tables 2.8 through 2.10, larger differences in treatment means translate into greater power to correctly reject the null hypothesis. Moreover,

Figure 2.4: Proportion of trials which reject the null hypothesis by design when μ_A varies, $\mu_B = 0.0$, σ_A varies, $\sigma_B = 1.0$, and $N = 100$. Data points are grouped by design; BBS is represented by $*$ with a solid line, BB by variations on $\square$ with a dashed line, and equal allocation by $\bullet$ with a dotted line.



all designs have increased power rates when σ_A is small relative to σ_B regardless of treatment effect size. The RAR designs' power levels are comparable to those of equal allocation, especially when $N \geq 500$. When clinical trial sizes are moderate or large,

the power to correctly reject the null hypothesis exceeds 90%, growing further as σ_A decreases, as N increases, and as μ_A increases.

Figure 2.4 highlights the trends in the five design's rates of rejecting the null hypothesis from Tables 2.7 – 2.10 when $N = 100$ and $\sigma_A \neq \sigma_B$. Rejection patterns are similar in other clinical trial enrollment sizes, except when power levels rise to 100% for all designs. As seen in Tables 2.7 – 2.10 and in Figure 2.4, when $\mu_A = \mu_B = 0.0$ and σ_A is small relative to σ_B, all designs' false rejection rates are slightly inflated, particularly those of BBS and BB_1. Type I error rates approach 5% in all designs once $\sigma_A \geq 0.9$ and $\sigma_B = 1.0$. When the difference in treatment mean effects is small ($\mu_A = 0.1$ and $\mu_B = 0.0$), BBS—and to a lesser extent the BB designs—correctly rejects the null hypothesis more frequently than equal allocation when σ_A is less than σ_B. Although the BBS design still maintains the highest rejection rates for a treatment difference of 0.1 when $\sigma_A > \sigma_B$, all designs rejection rates are similar. Power is comparable across the five designs and changing values of σ_A for $\mu_A = 0.5$ and 1.0, as well. Figure 2.4 illustrates how the probability of rejection, correct or incorrect, decreases as σ_A increases from 0.5 to 1.5 across all designs.

2.5 Application

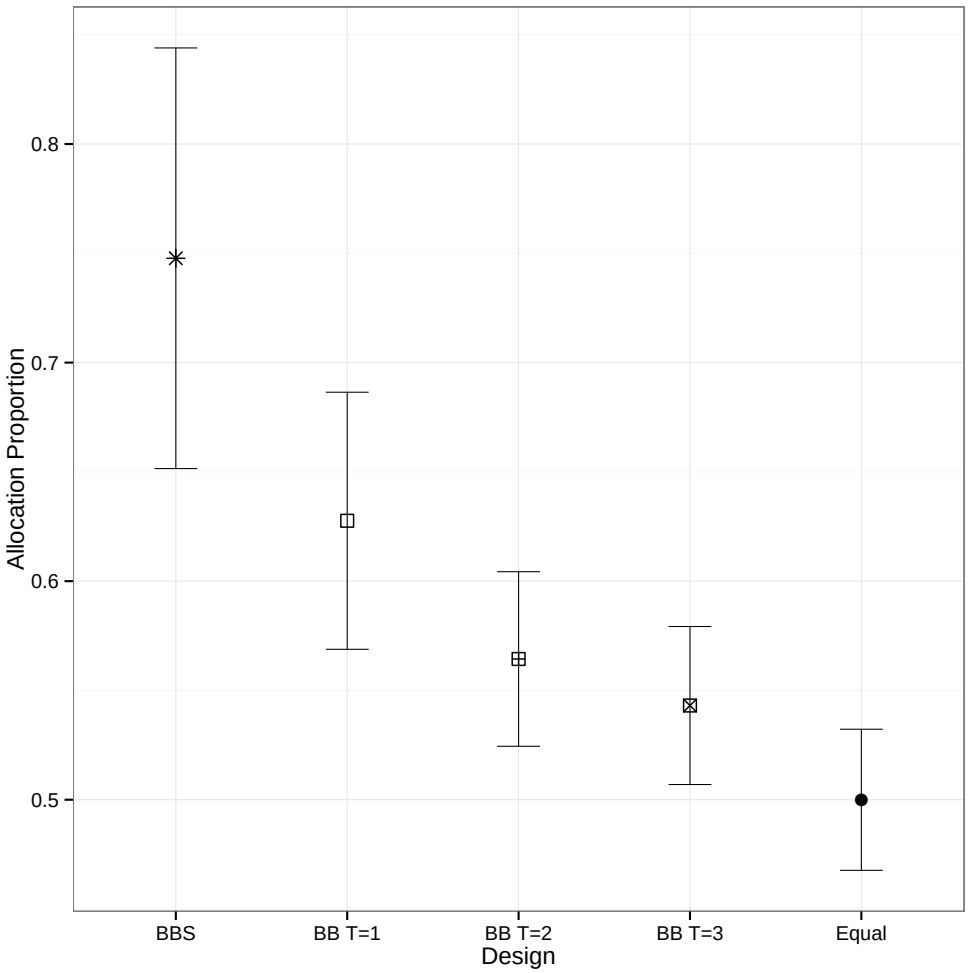
Since the 1980s, an estimated 80 million people have been infected with the Human Immunodeficiency Virus (HIV) and there have been 40 million HIV-related deaths [20, 46]. Antiretroviral therapies help to control the spread of this epidemic, extend the expected lifespan of infected persons, and increase the quality of life for those living with the virus—provided adequate adherence to the prescribed therapies is attained [13, 27]. Inversely, non-adherence may negatively impact not only the individual patient but the population in general, leading to poor patient outcomes as well as

development and transmission of treatment-resistant viral strains.

While many interventions consider adherence above an arbitrary threshold as their binary outcome of success, there is a lack of consensus on what constitutes adequate adherence and whether or not a measure of adherence is valid [1, 26, 30, 43]. Instead, it may be more prudent and relevant to consider the actual HIV RNA viral load suppression as a continuous measure of success in a clinical trial. Moreover, as adherence affects treatment outcomes, baseline adherence can be leveraged as a useful prognostic factor. Adjusting for baseline adherence gives a more accurate measure of intervention effectiveness [28, 33].

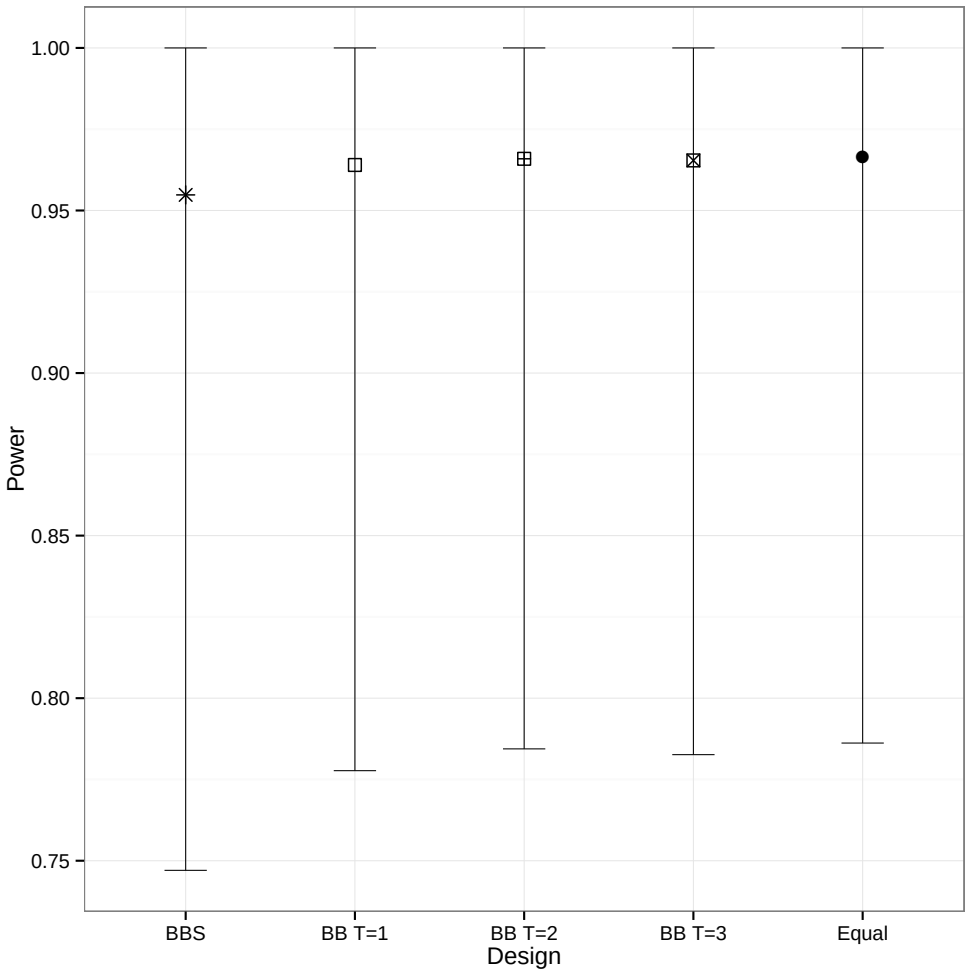
Modifying [33] based on reasonable estimates from [28, 33] as baseline adherence covariate and treatment outcome HIV RNA suppression parameters, 244 HIV patients on antiretroviral therapy are prospectively randomized to standard medication alone (treatment M) or medication with additional educational and counseling (treatment C). Patients who received medication alone see a treatment effect (standard deviation) of $\mu_M = 0.22$ ($\sigma_M = 0.54$) fewer HIV RNA copies/mL versus $\mu_C = 0.58$ ($\sigma_C = 0.47$) fewer HIV RNA copies/mL when medication is coupled with educational and counseling therapy. Additionally, baseline adherence is normally distributed for patients regardless of treatment assignment with mean $\mu_Z = 0.60$ and standard deviation $\sigma_Z = 0.49$. The treatment-covariate interaction also does not differ across treatments with a slope of $\beta = 1.11$. Combining the treatment effects and the covariate treatment interactions, medication alone only decreases HIV RNA by 0.89 copies/mL while the combined therapy decreases HIV RNA by 1.25 copies/mL, a difference which is clinically and statistically significant in [33]. The one-sided null hypothesis (combination treatment is no more effective than medication treatment alone) is tested at significance $\alpha = 0.05$. Delay in response time is ignored for this simulation, but discussed further in Section 2.6.3.

Figure 2.5: Proportion (and standard deviation) of 244 HIV antiretroviral therapy adherence intervention patients assigned to the more effective combined therapy of medication, education, and counseling ($\mu_C = 0.58$ fewer RNA copies/mL versus a reduction of only $\mu_M = 0.22$ RNA copies/mL on medication alone). Data points are grouped by design; BBS is represented by $*$, BB by variations on $\square$, and equal allocation by $\bullet$.



All five randomization schemes (BBS, BB with $T = 1, 2, 3$, and equal allocation) undergo 10,000 trial simulations for the above parameters. Table 2.11 summarizes the patient allocations and power of each design. In particular, the first row of Table 2.11 indicates that, under BBS, 75% of simulated patients are exposed to treatment C with a standard deviation of 0.10. Hence, an expected 182 patients would have

Figure 2.6: Proportion (and standard deviation) of 10,000 simulated HIV antiretroviral therapy clinical trials which correctly reject the null hypothesis and conclude that an education and counseling intervention in combination with medication is more effective at decreasing HIV RNA copies/mL than medication treatment alone (reductions of $\mu_C = 0.58$ versus $\mu_M = 0.22$ RNA copies/mL). Data points are grouped by design; BBS is represented by $*$, BB by variations on $\square$, and equal allocation by $\bullet$.



received the medication, education, and counseling therapy out of the 244 enrollees. That means 60 additional patients would have been randomized to the more effective intervention through the BBS design than by equal allocation. Furthermore, BBS correctly rejects the null hypothesis in 95% of 10,000 simulations with a standard deviation of 0.21. In comparison, the last row of Table 2.11 reveals how equal allo-

cation exposes 50% of simulated patients to treatment C with a standard deviation of 0.03. Half of all simulated patients receive the superior treatment and half remain on the inferior treatment resulting in zero additional ethical allocations. The equal allocation achieves 97% simulated power with a standard deviation of 0.18. Under the BBS design, more patients would have been exposed to the superior treatment than through any other design and minimal loss of power predicted.

Figure 2.5 illustrates the simulated assignment proportions of each design as described in Table 2.11. The BBS design would achieve the most ethical randomization, assigning 75% of patients to the superior treatment, followed in ethicality by the BB designs with $T = 1$, 2, and 3, respectively. Equal allocation would not deviate from a 50% exposure rate for each treatment. While exposing the most patients to treatment C, BBS would also have the highest variance in allocation, followed again by the BB_1, BB_2, and BB_3 designs. Equal allocation standard deviation would be smallest at 0.03.

Figure 2.6 displays the proportion of simulated trials which correctly reject the null hypothesis for each design as detailed in Table 2.11. The BBS design concludes that treatment C is more effective in 95% of simulated trials versus a 96% power for BB_1 and a 97% rejection rate for the other three designs. The simulated standard deviations of the five designs are also similar: 0.21 for BBS, 0.19 for BB_1, and 0.18 for BB_2, BB_3, and equal allocation. In Figure 2.6, the rejection rates appear comparable across all designs.

2.6 Discussion

Response-adaptive randomization biased coin designs leverage accumulating data to ethically expose patients to treatment during a clinical trial. The BB design from the

2001 Bandyopadhyay and Biswas paper [8] assigns patients to treatment A with probability $\Phi\left(\frac{\widehat{\mu}_A - \widehat{\mu}_B}{T}\right)$ where $\widehat{\mu}_k$ is the current estimate of treatment $k = A, B$ and $\Phi(\cdot)$ is the CDF of the normal distribution. The BB design has limitations including an arbitrary choice of scaling parameter T, restricted application to real-world situations, and a potential trade-off between ethical assignment and power to detect a treatment difference. This paper proposes the BBS design as an improved alternative to the BB randomization—replacing the constant scaling parameter T from the BB design with the current estimate of the pooled treatment standard deviation S. Simulations help examine the BBS capacity to ethically assign patients to treatment, the design's applicability to additional clinical scenarios such as unknown and unequal treatment variances, and the BBS rejection rates. BBS design outcomes are compared to those of the BB design with values of T fixed at 1, 2, and 3 (referred to as BB_1, BB_2, and BB_3, respectively), as well as to equal allocation trial behaviors.

2.6.1 Ethical Patient Allocation

In all scenarios simulated with a treatment difference, the BBS design allocates more patients to the better treatment, regardless of treatment effect size and clinical trial enrollment levels. In fact, the treatment with larger mean effect has smaller treatment variance, the BBS performs as well as or better than BB_1. When the treatment with larger mean effect also has larger treatment variance, the BBS design still allocates more patients to the better treatment than equal allocation, BB_2, or BB_3; however, BBS does not perform ethically as BB_1.

A clinical experiment need not be large to benefit from the BBS design. The proportion of patients assigned to the better treatment appears to peak in all simulated adaptive designs between trials of size $N = 500$ and $N = 1{,}000$. The proportion in-

creases a mere 1% as exposure doubles from 500 to 1,000 patients and, in fact, there is no further gain as trial size quintuples from 1,000 to 5,000.

One element to note is that variance in allocation rates grows in adaptive designs compared to fixed allocation, particularly when the treatment variance is large relative to the treatment difference. In all designs, allocation variance is mitigated as the trial population grows. The trade-off of the three underlying factors—treatment difference, treatment variance, and sample size—must be considered when selecting an appropriate design and choosing design parameters.

2.6.2 Rejection Rates

Under the original assumptions of the BB design, treatment variances must be both known and equal. These constraints limit the feasibility of the original design. Relaxing these requirements for the BBS design expands the relevant scenarios under which BBS might be applicable. The BBS design employs a t-statistic that estimates both treatment variances separately but utilizes the Welch-Satterthwaite degrees of freedom approximation for unequal variances [38, 45].

The BBS design appears to reject the null hypothesis ($H_0 : \mu_A \leq \mu_B$) at similar or higher rates as equal allocation. In particular, the BBS design demonstrates comparable power to equal allocation with a possible increased rejection rate when the difference in treatment effects is small relative to treatment variance. Additionally, the BBS design also appears to have a slightly inflated risk of Type I error compared to the other designs.

2.6.3 Delay

The HIV adherence intervention trial in [33] enrolled 244 patients over the course of three months, provided zero to three educational and counseling sessions during the four months following a patient's enrollment, and obtained RNA suppression results at 6 months post-randomization. If patients are only assigned to treatment after the previous patient's results are obtained, this intervention would require 122 years to adaptively allocate and sequentially treat 244 patients. Fortunately, mechanisms to incorporate delay exist which do not require postponing patient enrollment or treatment. While Section 2.5 ignores the delay in patient responses to simulate the adaptive designs described in this paper, a practical solution must be found to enable application of the BBS design to trials without significantly extending the study's lifespan.

The original BB design assumes that patient responses are instantaneous [8]. In fact, it is only required that after a patient is enrolled and exposed to treatment, that patient's response is observed and incorporated into future randomization criteria prior to assigning the next patient to treatment. The immediate response requirement can thus be circumvented in two possible manners. First, a clinical trial protocol can be altered to incorporate delays due to halting future patient enrollment and/or exposure until the most recently treated patient responds. This would result in extending a nine month trial such as [33] by an extra century, as described above. Second, the randomization scheme can be revised so that estimates leveraged in assigning patients to treatment depend only on currently responded patients at the time of a new patient enrollment.

The latter is a common modification, being more practical and ethical than an unduly delayed clinical trial conclusion [6, 23, 25, 41, 49]. While the aforementioned

articles prove that delay has minimal impact on the asymptotic characteristics of select adaptive designs—namely patient allocation proportion and power—convergence to these properties may slow. Thus small to moderate trials may not behave similarly under delayed response conditions. Moreover both large and small-sample behavior may differ by adaptive randomization and delay mechanisms. Further study of the effect of delay on the BBS design for various sample sizes is warranted.

2.6.4 Recommendations

The goal of most Phase III clinical trials is to determine which of two or more treatments is superior so that more patients gain access to a better standard of care. Current practice is that, during the clinical trial process, patients are exposed to treatments in equal proportion, regardless of evidence gathered suggesting the superiority of one treatment. As with [33], in many of these clinical scenarios the treatment variances are unknown and/or unequal, the outcome is not binary but instead continuous, patient covariates must be considered when evaluating treatment effect, and—most importantly—each patient enrolled can significantly benefit from receiving the more effective intervention.

The BBS design extends clinicians and researchers the opportunity to leverage data collected throughout such a clinical trial to expose many more patients to the better of two treatments *during* the trial rather than simply after its conclusion. This ethical advantage is accompanied by an inflated risk of incorrectly concluding one treatment is superior when, in fact, both treatments are equally effective. While the risk of a Type I error in the BBS design is less than double that of equal allocation for each sample size and is anticipated in fewer than 10% of all trials where no treatment difference is present, an erroneous conclusion of treatment superiority could

still be detrimental to future patients. For example, if the treatment falsely deemed superior is considerably more expensive, causes more adverse events, or has decreased propensity for patient adherence and would be the new standard of care for a large number of patients, minimizing a Type I error should be appropriately weighted when selecting a design. On the other hand, such an error could be minimally impactful if, for instance, the reign of any particular treatment is anticipated to be short, as when a research pipeline contains continually improving and quickly evolving therapies, frequently renewing the standard of care. The benefit of increased patient exposure during a trial must be weighted against the increased risk of incorrectly selecting a treatment as superior when it is not.

Continued investigation is suggested to better understand the potential influence of covariates on BBS design outcomes, especially covariates whose impact on patient outcomes differs by treatment. To expand the applicability of the BBS design to clinical trials with a delay in patient responses—for example the HIV adherence intervention described in Section 2.5—it is imperative to ascertain how the design fares under delayed response conditions. Finally, other modifications of the BB design should be considered, for instance replacing T with the unpooled standard deviation estimator or the estimate for the standard deviation of the difference in treatment effects.

2.7 Appendix

Table 2.1: Proportion (SD) of patients assigned to Treatment A when μ_A varies, $\mu_B = 0.0$, and $\sigma_A = \sigma_B = 1.0$. Each section denotes a different randomization design—BBS; BB with $T = 1$, 2, or 3; and equal allocation (Equal). Values for treatment parameter μ_A are described in the second column on the left. Allocation proportion and standard deviation across 1,000 simulations of trial size N are given in the remaining five columns.

		$\sigma_A = \sigma_B = 1.0$				
Design	$\mu_B = 0.0$	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
BBS	$\mu_A = 0.0$	0.50 (0.12)	0.50 (0.13)	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)
	$\mu_A = 0.1$	0.52 (0.12)	0.53 (0.13)	0.54 (0.07)	0.54 (0.05)	0.54 (0.02)
	$\mu_A = 0.5$	0.60 (0.11)	0.65 (0.11)	0.69 (0.06)	0.69 (0.05)	0.69 (0.02)
	$\mu_A = 1.0$	0.68 (0.09)	0.76 (0.09)	0.83 (0.06)	0.84 (0.04)	0.84 (0.02)
BB_1	$\mu_A = 0.0$	0.50 (0.12)	0.50 (0.12)	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)
	$\mu_A = 0.1$	0.52 (0.12)	0.53 (0.12)	0.54 (0.08)	0.54 (0.05)	0.54 (0.02)
	$\mu_A = 0.5$	0.60 (0.11)	0.64 (0.11)	0.68 (0.07)	0.69 (0.05)	0.69 (0.02)
	$\mu_A = 1.0$	0.67 (0.09)	0.75 (0.08)	0.83 (0.05)	0.83 (0.04)	0.84 (0.02)
BB_2	$\mu_A = 0.0$	0.50 (0.08)	0.50 (0.07)	0.50 (0.04)	0.50 (0.03)	0.50 (0.01)
	$\mu_A = 0.1$	0.51 (0.09)	0.52 (0.08)	0.52 (0.04)	0.52 (0.03)	0.52 (0.01)
	$\mu_A = 0.5$	0.55 (0.08)	0.58 (0.07)	0.59 (0.04)	0.60 (0.03)	0.60 (0.01)
	$\mu_A = 1.0$	0.60 (0.08)	0.64 (0.07)	0.68 (0.04)	0.69 (0.03)	0.69 (0.01)
BB_3	$\mu_A = 0.0$	0.50 (0.08)	0.50 (0.06)	0.50 (0.03)	0.50 (0.02)	0.50 (0.01)
	$\mu_A = 0.1$	0.50 (0.08)	0.51 (0.06)	0.51 (0.03)	0.51 (0.02)	0.51 (0.01)
	$\mu_A = 0.5$	0.54 (0.08)	0.56 (0.06)	0.56 (0.03)	0.56 (0.02)	0.57 (0.01)
	$\mu_A = 1.0$	0.57 (0.07)	0.60 (0.06)	0.62 (0.03)	0.63 (0.02)	0.63 (0.01)
Equal	$\mu_A = 0.0$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
	$\mu_A = 0.1$	0.49 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
	$\mu_A = 0.5$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
	$\mu_A = 1.0$	0.49 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)

Table 2.2: Proportion (SD) of patients assigned to Treatment A when $\mu_A = \mu_B = 0.0$, σ_A varies, and $\sigma_B = 1.0$. Each section denotes a different randomization design—BBS; BB with $T = 1$, 2, or 3; and equal allocation (Equal). Values for treatment parameter σ_A are described in the second column on the left. Allocation proportion and standard deviation across 1,000 simulations of trial size N are given in the remaining five columns.

Design	$\sigma_B = 1.0$	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
BBS	$\sigma_A = 0.5$	0.51 (0.13)	0.52 (0.13)	0.52 (0.09)	0.52 (0.07)	0.50 (0.03)
	$\sigma_A = 0.7$	0.51 (0.12)	0.51 (0.13)	0.51 (0.08)	0.51 (0.06)	0.50 (0.02)
	$\sigma_A = 0.9$	0.50 (0.12)	0.50 (0.13)	0.50 (0.08)	0.50 (0.05)	0.50 (0.03)
	$\sigma_A = 1.1$	0.50 (0.13)	0.50 (0.13)	0.50 (0.08)	0.50 (0.05)	0.50 (0.02)
	$\sigma_A = 1.3$	0.49 (0.12)	0.49 (0.13)	0.50 (0.08)	0.50 (0.05)	0.50 (0.02)
	$\sigma_A = 1.5$	0.49 (0.12)	0.49 (0.13)	0.49 (0.08)	0.49 (0.06)	0.50 (0.03)
BB_1	$\sigma_A = 0.5$	0.51 (0.11)	0.51 (0.10)	0.51 (0.06)	0.51 (0.04)	0.50 (0.02)
	$\sigma_A = 0.7$	0.50 (0.11)	0.50 (0.11)	0.50 (0.06)	0.50 (0.05)	0.50 (0.02)
	$\sigma_A = 0.9$	0.50 (0.11)	0.50 (0.12)	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)
	$\sigma_A = 1.1$	0.50 (0.12)	0.50 (0.12)	0.50 (0.08)	0.50 (0.06)	0.50 (0.03)
	$\sigma_A = 1.3$	0.50 (0.12)	0.49 (0.13)	0.50 (0.09)	0.50 (0.06)	0.50 (0.03)
	$\sigma_A = 1.5$	0.49 (0.13)	0.49 (0.15)	0.49 (0.10)	0.49 (0.07)	0.50 (0.03)
BB_2	$\sigma_A = 0.5$	0.50 (0.08)	0.50 (0.07)	0.50 (0.03)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 0.7$	0.50 (0.08)	0.50 (0.07)	0.50 (0.04)	0.50 (0.03)	0.50 (0.01)
	$\sigma_A = 0.9$	0.50 (0.09)	0.50 (0.07)	0.50 (0.04)	0.50 (0.03)	0.50 (0.01)
	$\sigma_A = 1.1$	0.49 (0.09)	0.50 (0.08)	0.50 (0.04)	0.50 (0.03)	0.50 (0.01)
	$\sigma_A = 1.3$	0.50 (0.09)	0.50 (0.08)	0.50 (0.04)	0.50 (0.03)	0.50 (0.01)
	$\sigma_A = 1.5$	0.50 (0.10)	0.50 (0.09)	0.50 (0.05)	0.50 (0.03)	0.50 (0.02)
BB_3	$\sigma_A = 0.5$	0.50 (0.08)	0.50 (0.06)	0.50 (0.03)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 0.7$	0.50 (0.07)	0.50 (0.06)	0.50 (0.03)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 0.9$	0.50 (0.08)	0.50 (0.06)	0.50 (0.03)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 1.1$	0.50 (0.08)	0.50 (0.06)	0.50 (0.03)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 1.3$	0.50 (0.08)	0.50 (0.07)	0.50 (0.03)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 1.5$	0.50 (0.08)	0.50 (0.07)	0.50 (0.03)	0.50 (0.03)	0.50 (0.01)
Equal	$\sigma_A = 0.5$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 0.7$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 0.9$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 1.1$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 1.3$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 1.5$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)

Table 2.3: Proportion (SD) of patients assigned to Treatment A when $\mu_A = 0.1$, $\mu_B = 0.0$, σ_A varies, and $\sigma_B = 1.0$. Each section denotes a different randomization design—BBS; BB with $T = 1$, 2, or 3; and equal allocation (Equal). Values for treatment parameter σ_A are described in the second column on the left. Allocation proportion and standard deviation across 1,000 simulations of trial size N are given in the remaining five columns.

Design	$\sigma_B = 1.0$	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
BBS	$\sigma_A = 0.5$	0.54 (0.12)	0.56 (0.13)	0.57 (0.10)	0.57 (0.08)	0.56 (0.04)
	$\sigma_A = 0.7$	0.53 (0.12)	0.54 (0.13)	0.56 (0.08)	0.55 (0.06)	0.55 (0.03)
	$\sigma_A = 0.9$	0.53 (0.12)	0.54 (0.13)	0.54 (0.07)	0.54 (0.05)	0.54 (0.02)
	$\sigma_A = 1.1$	0.52 (0.12)	0.53 (0.13)	0.54 (0.07)	0.53 (0.05)	0.54 (0.02)
	$\sigma_A = 1.3$	0.51 (0.12)	0.52 (0.13)	0.53 (0.07)	0.53 (0.06)	0.53 (0.02)
	$\sigma_A = 1.5$	0.51 (0.12)	0.51 (0.13)	0.52 (0.08)	0.52 (0.05)	0.53 (0.03)
BB_1	$\sigma_A = 0.5$	0.52 (0.11)	0.54 (0.10)	0.54 (0.06)	0.55 (0.04)	0.54 (0.02)
	$\sigma_A = 0.7$	0.52 (0.11)	0.53 (0.11)	0.54 (0.06)	0.54 (0.05)	0.54 (0.02)
	$\sigma_A = 0.9$	0.53 (0.12)	0.53 (0.12)	0.54 (0.07)	0.54 (0.05)	0.54 (0.02)
	$\sigma_A = 1.1$	0.52 (0.12)	0.53 (0.13)	0.54 (0.07)	0.54 (0.06)	0.54 (0.03)
	$\sigma_A = 1.3$	0.51 (0.12)	0.52 (0.14)	0.53 (0.09)	0.53 (0.06)	0.54 (0.03)
	$\sigma_A = 1.5$	0.51 (0.14)	0.52 (0.15)	0.52 (0.10)	0.53 (0.07)	0.54 (0.03)
BB_2	$\sigma_A = 0.5$	0.51 (0.08)	0.52 (0.07)	0.52 (0.04)	0.52 (0.02)	0.52 (0.01)
	$\sigma_A = 0.7$	0.51 (0.09)	0.52 (0.07)	0.52 (0.04)	0.52 (0.03)	0.52 (0.01)
	$\sigma_A = 0.9$	0.51 (0.09)	0.51 (0.07)	0.52 (0.04)	0.52 (0.03)	0.52 (0.01)
	$\sigma_A = 1.1$	0.51 (0.09)	0.52 (0.07)	0.52 (0.04)	0.52 (0.03)	0.52 (0.01)
	$\sigma_A = 1.3$	0.51 (0.09)	0.51 (0.08)	0.52 (0.04)	0.52 (0.03)	0.52 (0.01)
	$\sigma_A = 1.5$	0.51 (0.09)	0.51 (0.09)	0.52 (0.05)	0.52 (0.03)	0.52 (0.02)
BB_3	$\sigma_A = 0.5$	0.51 (0.08)	0.51 (0.06)	0.51 (0.03)	0.51 (0.02)	0.51 (0.01)
	$\sigma_A = 0.7$	0.51 (0.08)	0.51 (0.06)	0.51 (0.03)	0.51 (0.02)	0.51 (0.01)
	$\sigma_A = 0.9$	0.51 (0.08)	0.51 (0.06)	0.51 (0.03)	0.51 (0.02)	0.51 (0.01)
	$\sigma_A = 1.1$	0.51 (0.08)	0.51 (0.06)	0.51 (0.03)	0.51 (0.02)	0.51 (0.01)
	$\sigma_A = 1.3$	0.51 (0.08)	0.51 (0.07)	0.51 (0.03)	0.51 (0.02)	0.51 (0.01)
	$\sigma_A = 1.5$	0.51 (0.08)	0.51 (0.07)	0.51 (0.03)	0.51 (0.02)	0.51 (0.01)
Equal	$\sigma_A = 0.5$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 0.7$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 0.9$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 1.1$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 1.3$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 1.5$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)

Table 2.4: Proportion (SD) of patients assigned to Treatment A when $\mu_A = 0.5$, $\mu_B = 0.0$, σ_A varies, and $\sigma_B = 1.0$. Each section denotes a different randomization design—BBS; BB with $T = 1$, 2, or 3; and equal allocation (Equal). Values for treatment parameter σ_A are described in the second column on the left. Allocation proportion and standard deviation across 1,000 simulations of trial size N are given in the remaining five columns.

Design	$\sigma_B = 1.0$	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
BBS	$\sigma_A = 0.5$	0.63 (0.11)	0.70 (0.12)	0.79 (0.10)	0.79 (0.09)	0.79 (0.06)
	$\sigma_A = 0.7$	0.62 (0.11)	0.68 (0.12)	0.74 (0.09)	0.74 (0.07)	0.74 (0.04)
	$\sigma_A = 0.9$	0.60 (0.11)	0.65 (0.11)	0.70 (0.07)	0.70 (0.05)	0.71 (0.02)
	$\sigma_A = 1.1$	0.59 (0.11)	0.64 (0.12)	0.68 (0.07)	0.68 (0.05)	0.68 (0.02)
	$\sigma_A = 1.3$	0.58 (0.11)	0.62 (0.11)	0.65 (0.07)	0.66 (0.05)	0.66 (0.02)
	$\sigma_A = 1.5$	0.57 (0.12)	0.60 (0.12)	0.64 (0.06)	0.64 (0.05)	0.64 (0.02)
BB_1	$\sigma_A = 0.5$	0.60 (0.10)	0.65 (0.09)	0.69 (0.06)	0.69 (0.04)	0.69 (0.02)
	$\sigma_A = 0.7$	0.60 (0.10)	0.65 (0.10)	0.69 (0.06)	0.69 (0.05)	0.69 (0.02)
	$\sigma_A = 0.9$	0.59 (0.11)	0.65 (0.10)	0.69 (0.07)	0.69 (0.05)	0.69 (0.02)
	$\sigma_A = 1.1$	0.60 (0.11)	0.63 (0.11)	0.69 (0.07)	0.69 (0.05)	0.69 (0.02)
	$\sigma_A = 1.3$	0.59 (0.12)	0.64 (0.13)	0.68 (0.07)	0.69 (0.06)	0.69 (0.03)
	$\sigma_A = 1.5$	0.58 (0.13)	0.63 (0.13)	0.68 (0.09)	0.68 (0.06)	0.69 (0.03)
BB_2	$\sigma_A = 0.5$	0.55 (0.08)	0.58 (0.07)	0.59 (0.04)	0.6 (0.02)	0.60 (0.01)
	$\sigma_A = 0.7$	0.55 (0.09)	0.58 (0.07)	0.59 (0.04)	0.6 (0.03)	0.60 (0.01)
	$\sigma_A = 0.9$	0.55 (0.09)	0.58 (0.07)	0.59 (0.04)	0.6 (0.03)	0.60 (0.01)
	$\sigma_A = 1.1$	0.55 (0.09)	0.58 (0.08)	0.59 (0.04)	0.6 (0.03)	0.60 (0.01)
	$\sigma_A = 1.3$	0.55 (0.09)	0.58 (0.08)	0.59 (0.04)	0.6 (0.03)	0.60 (0.01)
	$\sigma_A = 1.5$	0.54 (0.10)	0.57 (0.08)	0.59 (0.05)	0.6 (0.03)	0.60 (0.01)
BB_3	$\sigma_A = 0.5$	0.54 (0.07)	0.55 (0.06)	0.56 (0.03)	0.56 (0.02)	0.57 (0.01)
	$\sigma_A = 0.7$	0.54 (0.08)	0.55 (0.06)	0.56 (0.03)	0.56 (0.02)	0.57 (0.01)
	$\sigma_A = 0.9$	0.53 (0.08)	0.55 (0.06)	0.56 (0.03)	0.57 (0.02)	0.57 (0.01)
	$\sigma_A = 1.1$	0.53 (0.08)	0.55 (0.07)	0.56 (0.03)	0.56 (0.02)	0.57 (0.01)
	$\sigma_A = 1.3$	0.53 (0.08)	0.55 (0.07)	0.56 (0.03)	0.56 (0.02)	0.57 (0.01)
	$\sigma_A = 1.5$	0.53 (0.08)	0.54 (0.07)	0.56 (0.03)	0.56 (0.03)	0.57 (0.01)
Equal	$\sigma_A = 0.5$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 0.7$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 0.9$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 1.1$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 1.3$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 1.5$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)

Table 2.5: Proportion (SD) of patients assigned to Treatment A when $\mu_A = 1.0$, $\mu_B = 0.0$, σ_A varies, and $\sigma_B = 1.0$. Each section denotes a different randomization design—BBS; BB with $T = 1$, 2, or 3; and equal allocation (Equal). Values for treatment parameter σ_A are described in the second column on the left. Allocation proportion and standard deviation across 1,000 simulations of trial size N are given in the remaining five columns.

Design	$\sigma_B = 1.0$	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
BBS	$\sigma_A = 0.5$	0.71 (0.09)	0.82 (0.07)	0.93 (0.05)	0.95 (0.04)	0.97 (0.02)
	$\sigma_A = 0.7$	0.69 (0.09)	0.79 (0.08)	0.89 (0.05)	0.91 (0.04)	0.92 (0.03)
	$\sigma_A = 0.9$	0.68 (0.09)	0.77 (0.08)	0.85 (0.06)	0.86 (0.05)	0.86 (0.02)
	$\sigma_A = 1.1$	0.66 (0.10)	0.75 (0.09)	0.81 (0.05)	0.82 (0.04)	0.82 (0.02)
	$\sigma_A = 1.3$	0.65 (0.10)	0.72 (0.09)	0.78 (0.05)	0.79 (0.04)	0.79 (0.02)
	$\sigma_A = 1.5$	0.63 (0.11)	0.70 (0.09)	0.75 (0.05)	0.76 (0.04)	0.76 (0.02)
BB_1	$\sigma_A = 0.5$	0.67 (0.09)	0.76 (0.08)	0.83 (0.05)	0.84 (0.04)	0.84 (0.02)
	$\sigma_A = 0.7$	0.67 (0.09)	0.75 (0.08)	0.83 (0.05)	0.84 (0.04)	0.84 (0.02)
	$\sigma_A = 0.9$	0.67 (0.09)	0.75 (0.08)	0.83 (0.05)	0.84 (0.04)	0.84 (0.02)
	$\sigma_A = 1.1$	0.66 (0.09)	0.75 (0.09)	0.83 (0.06)	0.84 (0.04)	0.84 (0.02)
	$\sigma_A = 1.3$	0.66 (0.10)	0.74 (0.10)	0.83 (0.06)	0.84 (0.05)	0.84 (0.02)
	$\sigma_A = 1.5$	0.65 (0.11)	0.74 (0.10)	0.82 (0.06)	0.83 (0.05)	0.84 (0.02)
BB_2	$\sigma_A = 0.5$	0.60 (0.08)	0.65 (0.06)	0.68 (0.03)	0.69 (0.03)	0.69 (0.01)
	$\sigma_A = 0.7$	0.60 (0.08)	0.65 (0.07)	0.68 (0.04)	0.69 (0.03)	0.69 (0.01)
	$\sigma_A = 0.9$	0.60 (0.08)	0.65 (0.07)	0.68 (0.04)	0.69 (0.03)	0.69 (0.01)
	$\sigma_A = 1.1$	0.60 (0.08)	0.64 (0.07)	0.68 (0.04)	0.69 (0.03)	0.69 (0.01)
	$\sigma_A = 1.3$	0.60 (0.09)	0.65 (0.07)	0.68 (0.04)	0.69 (0.03)	0.69 (0.01)
	$\sigma_A = 1.5$	0.59 (0.09)	0.64 (0.08)	0.68 (0.04)	0.69 (0.03)	0.69 (0.01)
BB_3	$\sigma_A = 0.5$	0.57 (0.07)	0.60 (0.06)	0.63 (0.03)	0.63 (0.02)	0.63 (0.01)
	$\sigma_A = 0.7$	0.57 (0.07)	0.60 (0.06)	0.63 (0.03)	0.63 (0.02)	0.63 (0.01)
	$\sigma_A = 0.9$	0.57 (0.08)	0.60 (0.06)	0.63 (0.03)	0.63 (0.02)	0.63 (0.01)
	$\sigma_A = 1.1$	0.57 (0.08)	0.60 (0.06)	0.62 (0.03)	0.63 (0.02)	0.63 (0.01)
	$\sigma_A = 1.3$	0.57 (0.08)	0.60 (0.06)	0.62 (0.03)	0.63 (0.02)	0.63 (0.01)
	$\sigma_A = 1.5$	0.57 (0.08)	0.60 (0.07)	0.62 (0.03)	0.63 (0.02)	0.63 (0.01)
Equal	$\sigma_A = 0.5$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 0.7$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 0.9$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 1.1$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 1.3$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
	$\sigma_A = 1.5$	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)

Table 2.6: Proportion (SD) of trials where the null hypothesis is correctly rejected for $(\mu_A, \mu_B) = (0.1, 0.0)$, $(0.5, 0.0)$, and $(1.0, 0.0)$ or incorrectly rejected for $(\mu_A, \mu_B) = (0.0, 0.0)$ under equal treatment variances ($\sigma_A = \sigma_B = 1.0$). Each section denotes a different randomization design—BBS; BB with $T = 1$, 2, or 3; and equal allocation (Equal). Values for clinical trial enrollment size N are described in the second column on the left. Rejection rates and standard deviations across 1,000 simulations are given in the remaining four columns.

		Power (μ_A, μ_B)			Type I Error
Design	N	$(0.1, 0.0)$	$(0.5, 0.0)$	$(1.0, 0.0)$	$(0.0, 0.0)$
BBS	50	0.11 (0.31)	0.40 (0.49)	0.76 (0.43)	0.07 (0.26)
	100	0.14 (0.34)	0.54 (0.50)	0.96 (0.19)	0.06 (0.23)
	500	0.20 (0.40)	0.99 (0.09)	1.00 (0.00)	0.06 (0.24)
	1,000	0.30 (0.46)	1.00 (0.00)	1.00 (0.00)	0.05 (0.23)
	5,000	0.79 (0.41)	1.00 (0.00)	1.00 (0.00)	0.05 (0.22)
BB_1	50	0.09 (0.29)	0.36 (0.48)	0.80 (0.40)	0.07 (0.25)
	100	0.10 (0.30)	0.55 (0.50)	0.95 (0.22)	0.07 (0.26)
	500	0.20 (0.40)	0.98 (0.13)	1.00 (0.00)	0.06 (0.24)
	1,000	0.30 (0.46)	1.00 (0.00)	1.00 (0.00)	0.06 (0.23)
	5,000	0.80 (0.40)	1.00 (0.00)	1.00 (0.00)	0.03 (0.17)
BB_2	50	0.11 (0.31)	0.34 (0.47)	0.78 (0.42)	0.06 (0.24)
	100	0.11 (0.32)	0.53 (0.50)	0.96 (0.19)	0.07 (0.26)
	500	0.19 (0.39)	0.99 (0.10)	1.00 (0.00)	0.07 (0.25)
	1,000	0.30 (0.46)	1.00 (0.00)	1.00 (0.00)	0.05 (0.21)
	5,000	0.78 (0.41)	1.00 (0.00)	1.00 (0.00)	0.05 (0.22)
BB_3	50	0.10 (0.30)	0.35 (0.48)	0.79 (0.41)	0.06 (0.24)
	100	0.10 (0.30)	0.56 (0.50)	0.96 (0.19)	0.05 (0.21)
	500	0.19 (0.39)	0.99 (0.11)	1.00 (0.00)	0.06 (0.23)
	1,000	0.30 (0.46)	1.00 (0.03)	1.00 (0.00)	0.06 (0.23)
	5,000	0.81 (0.39)	1.00 (0.00)	1.00 (0.00)	0.06 (0.23)
Equal	50	0.10 (0.31)	0.35 (0.48)	0.77 (0.42)	0.07 (0.25)
	100	0.12 (0.32)	0.54 (0.50)	0.96 (0.19)	0.06 (0.24)
	500	0.20 (0.40)	0.98 (0.13)	1.00 (0.00)	0.05 (0.22)
	1,000	0.29 (0.45)	1.00 (0.00)	1.00 (0.00)	0.06 (0.24)
	5,000	0.80 (0.40)	1.00 (0.00)	1.00 (0.00)	0.04 (0.20)

Table 2.7: Proportion (SD) of trials where the null hypothesis is incorrectly rejected when $\mu_A = \mu_B = 0.0$, σ_A varies, and $\sigma_B = 1.0$. Each section denotes a different randomization design—BBS; BB with $T = 1$, 2, or 3; and equal allocation (Equal). Values for treatment parameter σ_A are described in the second column on the left. Rejection rates and standard deviations across 1,000 simulations of trial size N are given in the remaining five columns.

Design	$\sigma_B = 1.0$	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
BBS	$\sigma_A = 0.5$	0.10 (0.29)	0.10 (0.30)	0.09 (0.29)	0.06 (0.24)	0.07 (0.25)
	$\sigma_A = 0.7$	0.09 (0.29)	0.10 (0.29)	0.06 (0.23)	0.07 (0.25)	0.06 (0.23)
	$\sigma_A = 0.9$	0.07 (0.25)	0.08 (0.28)	0.06 (0.24)	0.05 (0.21)	0.06 (0.23)
	$\sigma_A = 1.1$	0.06 (0.24)	0.06 (0.23)	0.05 (0.22)	0.04 (0.20)	0.04 (0.20)
	$\sigma_A = 1.3$	0.07 (0.26)	0.06 (0.24)	0.04 (0.20)	0.06 (0.24)	0.05 (0.22)
	$\sigma_A = 1.5$	0.06 (0.23)	0.06 (0.23)	0.05 (0.22)	0.05 (0.22)	0.05 (0.21)
BB_1	$\sigma_A = 0.5$	0.10 (0.30)	0.09 (0.28)	0.08 (0.26)	0.07 (0.26)	0.05 (0.22)
	$\sigma_A = 0.7$	0.09 (0.29)	0.08 (0.27)	0.06 (0.24)	0.06 (0.23)	0.05 (0.22)
	$\sigma_A = 0.9$	0.08 (0.28)	0.06 (0.23)	0.05 (0.21)	0.04 (0.20)	0.06 (0.24)
	$\sigma_A = 1.1$	0.06 (0.23)	0.07 (0.26)	0.05 (0.21)	0.05 (0.22)	0.05 (0.22)
	$\sigma_A = 1.3$	0.05 (0.22)	0.04 (0.20)	0.05 (0.22)	0.05 (0.22)	0.05 (0.22)
	$\sigma_A = 1.5$	0.05 (0.22)	0.03 (0.18)	0.04 (0.19)	0.04 (0.20)	0.06 (0.23)
BB_2	$\sigma_A = 0.5$	0.07 (0.25)	0.07 (0.26)	0.05 (0.22)	0.06 (0.23)	0.06 (0.24)
	$\sigma_A = 0.7$	0.07 (0.26)	0.06 (0.24)	0.06 (0.24)	0.06 (0.24)	0.06 (0.23)
	$\sigma_A = 0.9$	0.06 (0.23)	0.05 (0.22)	0.05 (0.21)	0.05 (0.21)	0.06 (0.23)
	$\sigma_A = 1.1$	0.05 (0.21)	0.04 (0.21)	0.04 (0.20)	0.03 (0.18)	0.05 (0.21)
	$\sigma_A = 1.3$	0.05 (0.22)	0.05 (0.22)	0.04 (0.19)	0.05 (0.21)	0.06 (0.24)
	$\sigma_A = 1.5$	0.05 (0.21)	0.05 (0.22)	0.04 (0.19)	0.04 (0.20)	0.05 (0.21)
BB_3	$\sigma_A = 0.5$	0.09 (0.29)	0.07 (0.26)	0.05 (0.22)	0.06 (0.24)	0.05 (0.22)
	$\sigma_A = 0.7$	0.06 (0.24)	0.06 (0.23)	0.06 (0.23)	0.04 (0.20)	0.06 (0.24)
	$\sigma_A = 0.9$	0.08 (0.28)	0.05 (0.22)	0.05 (0.22)	0.05 (0.22)	0.05 (0.22)
	$\sigma_A = 1.1$	0.07 (0.25)	0.05 (0.22)	0.05 (0.21)	0.05 (0.22)	0.05 (0.23)
	$\sigma_A = 1.3$	0.05 (0.21)	0.04 (0.20)	0.04 (0.21)	0.05 (0.21)	0.06 (0.23)
	$\sigma_A = 1.5$	0.03 (0.17)	0.04 (0.21)	0.04 (0.19)	0.05 (0.21)	0.04 (0.21)
Equal	$\sigma_A = 0.5$	0.06 (0.24)	0.07 (0.26)	0.07 (0.25)	0.05 (0.21)	0.06 (0.24)
	$\sigma_A = 0.7$	0.06 (0.24)	0.06 (0.24)	0.05 (0.21)	0.06 (0.23)	0.05 (0.22)
	$\sigma_A = 0.9$	0.06 (0.24)	0.05 (0.22)	0.06 (0.23)	0.05 (0.21)	0.05 (0.22)
	$\sigma_A = 1.1$	0.06 (0.23)	0.06 (0.24)	0.07 (0.25)	0.06 (0.23)	0.04 (0.19)
	$\sigma_A = 1.3$	0.06 (0.24)	0.05 (0.22)	0.05 (0.22)	0.04 (0.21)	0.05 (0.21)
	$\sigma_A = 1.5$	0.06 (0.24)	0.07 (0.25)	0.05 (0.21)	0.04 (0.19)	0.05 (0.22)

Table 2.8: Proportion (SD) of trials where the null hypothesis is correctly rejected when $\mu_A = 0.1$, $\mu_B = 0.0$, σ_A varies, and $\sigma_B = 1.0$. Each section denotes a different randomization design—BBS; BB with $T = 1$, 2, or 3; and equal allocation (Equal). Values for treatment parameter σ_A are described in the second column on the left. Rejection rates and standard deviations across 1,000 simulations of trial size N are given in the remaining five columns.

Design	$\sigma_B = 1.0$	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
BBS	$\sigma_A = 0.5$	0.16 (0.36)	0.20 (0.40)	0.30 (0.46)	0.44 (0.50)	0.94 (0.24)
	$\sigma_A = 0.7$	0.12 (0.33)	0.16 (0.36)	0.29 (0.45)	0.36 (0.48)	0.90 (0.30)
	$\sigma_A = 0.9$	0.12 (0.33)	0.12 (0.33)	0.22 (0.41)	0.30 (0.46)	0.85 (0.36)
	$\sigma_A = 1.1$	0.10 (0.30)	0.12 (0.32)	0.20 (0.40)	0.29 (0.45)	0.75 (0.43)
	$\sigma_A = 1.3$	0.10 (0.30)	0.11 (0.31)	0.17 (0.37)	0.28 (0.45)	0.70 (0.46)
	$\sigma_A = 1.5$	0.08 (0.27)	0.10 (0.29)	0.18 (0.38)	0.20 (0.40)	0.63 (0.48)
BB_1	$\sigma_A = 0.5$	0.14 (0.35)	0.18 (0.38)	0.31 (0.46)	0.47 (0.50)	0.94 (0.24)
	$\sigma_A = 0.7$	0.13 (0.33)	0.15 (0.35)	0.25 (0.43)	0.39 (0.49)	0.91 (0.29)
	$\sigma_A = 0.9$	0.12 (0.33)	0.13 (0.34)	0.22 (0.41)	0.31 (0.46)	0.82 (0.38)
	$\sigma_A = 1.1$	0.09 (0.29)	0.09 (0.29)	0.18 (0.38)	0.30 (0.46)	0.77 (0.42)
	$\sigma_A = 1.3$	0.07 (0.25)	0.10 (0.29)	0.15 (0.35)	0.22 (0.42)	0.68 (0.47)
	$\sigma_A = 1.5$	0.06 (0.24)	0.08 (0.27)	0.12 (0.32)	0.20 (0.40)	0.61 (0.49)
BB_2	$\sigma_A = 0.5$	0.15 (0.35)	0.15 (0.36)	0.30 (0.46)	0.43 (0.49)	0.94 (0.23)
	$\sigma_A = 0.7$	0.12 (0.32)	0.12 (0.33)	0.27 (0.44)	0.37 (0.48)	0.91 (0.29)
	$\sigma_A = 0.9$	0.10 (0.30)	0.11 (0.31)	0.23 (0.42)	0.30 (0.46)	0.83 (0.37)
	$\sigma_A = 1.1$	0.08 (0.27)	0.10 (0.30)	0.18 (0.38)	0.26 (0.44)	0.79 (0.41)
	$\sigma_A = 1.3$	0.09 (0.28)	0.08 (0.27)	0.18 (0.39)	0.24 (0.43)	0.69 (0.46)
	$\sigma_A = 1.5$	0.08 (0.28)	0.07 (0.25)	0.14 (0.35)	0.21 (0.41)	0.62 (0.49)
BB_3	$\sigma_A = 0.5$	0.12 (0.33)	0.14 (0.35)	0.29 (0.46)	0.42 (0.49)	0.93 (0.25)
	$\sigma_A = 0.7$	0.11 (0.31)	0.12 (0.32)	0.26 (0.44)	0.37 (0.48)	0.89 (0.31)
	$\sigma_A = 0.9$	0.10 (0.29)	0.12 (0.33)	0.23 (0.42)	0.35 (0.48)	0.83 (0.38)
	$\sigma_A = 1.1$	0.11 (0.31)	0.09 (0.29)	0.20 (0.40)	0.27 (0.45)	0.79 (0.41)
	$\sigma_A = 1.3$	0.09 (0.28)	0.09 (0.28)	0.16 (0.37)	0.25 (0.43)	0.70 (0.46)
	$\sigma_A = 1.5$	0.07 (0.26)	0.09 (0.28)	0.14 (0.34)	0.20 (0.40)	0.63 (0.48)
Equal	$\sigma_A = 0.5$	0.11 (0.31)	0.13 (0.34)	0.29 (0.45)	0.43 (0.49)	0.95 (0.22)
	$\sigma_A = 0.7$	0.10 (0.30)	0.11 (0.31)	0.23 (0.42)	0.39 (0.49)	0.91 (0.29)
	$\sigma_A = 0.9$	0.09 (0.28)	0.11 (0.31)	0.21 (0.41)	0.33 (0.47)	0.82 (0.38)
	$\sigma_A = 1.1$	0.10 (0.29)	0.11 (0.31)	0.19 (0.39)	0.27 (0.44)	0.77 (0.42)
	$\sigma_A = 1.3$	0.10 (0.29)	0.08 (0.28)	0.18 (0.39)	0.26 (0.44)	0.69 (0.46)
	$\sigma_A = 1.5$	0.09 (0.29)	0.09 (0.28)	0.16 (0.37)	0.22 (0.41)	0.61 (0.49)

Table 2.9: Proportion (SD) of trials where the null hypothesis is correctly rejected when $\mu_A = 0.5$, $\mu_B = 0.0$, σ_A varies, and $\sigma_B = 1.0$. Each section denotes a different randomization design—BBS; BB with $T = 1$, 2, or 3; and equal allocation (Equal). Values for treatment parameter σ_A are described in the second column on the left. Rejection rates and standard deviations across 1,000 simulations of trial size N are given in the remaining five columns.

	$\sigma_B = 1.0$	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
BBS	$\sigma_A = 0.5$	0.52 (0.50)	0.72 (0.45)	1.00 (0.03)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 0.7$	0.45 (0.50)	0.67 (0.47)	1.00 (0.06)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 0.9$	0.39 (0.49)	0.59 (0.49)	1.00 (0.07)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.1$	0.31 (0.46)	0.53 (0.50)	0.98 (0.15)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.3$	0.33 (0.47)	0.46 (0.50)	0.96 (0.20)	1.00 (0.03)	1.00 (0.00)
	$\sigma_A = 1.5$	0.28 (0.45)	0.41 (0.49)	0.96 (0.20)	1.00 (0.06)	1.00 (0.00)
BB_1	$\sigma_A = 0.5$	0.57 (0.50)	0.75 (0.43)	1.00 (0.04)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 0.7$	0.45 (0.50)	0.70 (0.46)	1.00 (0.05)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 0.9$	0.40 (0.49)	0.61 (0.49)	0.99 (0.10)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.1$	0.34 (0.47)	0.48 (0.50)	0.98 (0.14)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.3$	0.26 (0.44)	0.44 (0.50)	0.96 (0.20)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.5$	0.22 (0.41)	0.37 (0.48)	0.91 (0.29)	1.00 (0.00)	1.00 (0.00)
BB_2	$\sigma_A = 0.5$	0.51 (0.50)	0.73 (0.44)	1.00 (0.03)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 0.7$	0.43 (0.50)	0.67 (0.47)	0.99 (0.08)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 0.9$	0.38 (0.49)	0.61 (0.49)	0.99 (0.09)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.1$	0.32 (0.47)	0.56 (0.50)	0.97 (0.16)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.3$	0.28 (0.45)	0.43 (0.50)	0.95 (0.22)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.5$	0.23 (0.42)	0.37 (0.48)	0.94 (0.25)	1.00 (0.06)	1.00 (0.00)
BB_3	$\sigma_A = 0.5$	0.51 (0.50)	0.72 (0.45)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 0.7$	0.46 (0.50)	0.67 (0.47)	1.00 (0.03)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 0.9$	0.34 (0.47)	0.58 (0.49)	0.99 (0.08)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.1$	0.32 (0.47)	0.50 (0.50)	0.98 (0.14)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.3$	0.27 (0.45)	0.44 (0.50)	0.97 (0.17)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.5$	0.24 (0.43)	0.37 (0.48)	0.91 (0.28)	0.99 (0.08)	1.00 (0.00)
Equal	$\sigma_A = 0.5$	0.50 (0.50)	0.73 (0.44)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 0.7$	0.44 (0.50)	0.65 (0.48)	1.00 (0.03)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 0.9$	0.37 (0.48)	0.58 (0.49)	0.99 (0.08)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.1$	0.33 (0.47)	0.52 (0.50)	0.99 (0.11)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.3$	0.29 (0.45)	0.45 (0.50)	0.97 (0.18)	1.00 (0.04)	1.00 (0.00)
	$\sigma_A = 1.5$	0.24 (0.43)	0.43 (0.50)	0.93 (0.26)	0.99 (0.08)	1.00 (0.00)

Table 2.10: Proportion (SD) of trials where the null hypothesis is correctly rejected when $\mu_A = 1.0$, $\mu_B = 0.0$, σ_A varies, and $\sigma_B = 1.0$. Each section denotes a different randomization design—BBS; BB with $T = 1$, 2, or 3; and equal allocation (Equal). Values for treatment parameter σ_A are described in the second column on the left. Rejection rates and standard deviations across 1,000 simulations of trial size N are given in the remaining five columns.

	$\sigma_B = 1.0$	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
BBS	$\sigma_A = 0.5$	0.90 (0.30)	0.99 (0.11)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 0.7$	0.84 (0.37)	0.98 (0.13)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 0.9$	0.83 (0.38)	0.96 (0.19)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.1$	0.77 (0.42)	0.94 (0.24)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.3$	0.67 (0.47)	0.91 (0.29)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.5$	0.63 (0.48)	0.88 (0.33)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
BB_1	$\sigma_A = 0.5$	0.91 (0.28)	0.99 (0.09)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 0.7$	0.89 (0.31)	0.99 (0.11)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 0.9$	0.82 (0.38)	0.98 (0.16)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.1$	0.74 (0.44)	0.94 (0.24)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.3$	0.66 (0.47)	0.88 (0.32)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.5$	0.55 (0.50)	0.81 (0.39)	1.00 (0.03)	1.00 (0.00)	1.00 (0.00)
BB_2	$\sigma_A = 0.5$	0.92 (0.28)	1.00 (0.04)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 0.7$	0.89 (0.31)	0.99 (0.10)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 0.9$	0.81 (0.39)	0.98 (0.13)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.1$	0.73 (0.44)	0.95 (0.21)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.3$	0.68 (0.47)	0.89 (0.31)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.5$	0.62 (0.49)	0.85 (0.36)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
BB_3	$\sigma_A = 0.5$	0.92 (0.27)	1.00 (0.07)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 0.7$	0.89 (0.32)	0.99 (0.09)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 0.9$	0.80 (0.40)	0.98 (0.16)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.1$	0.73 (0.44)	0.95 (0.23)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.3$	0.64 (0.48)	0.90 (0.30)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.5$	0.60 (0.49)	0.86 (0.35)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
Equal	$\sigma_A = 0.5$	0.92 (0.27)	1.00 (0.04)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 0.7$	0.88 (0.32)	0.99 (0.11)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 0.9$	0.80 (0.40)	0.97 (0.16)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.1$	0.75 (0.44)	0.95 (0.23)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.3$	0.69 (0.46)	0.90 (0.30)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	$\sigma_A = 1.5$	0.64 (0.48)	0.87 (0.34)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)

Table 2.11: HIV antiretroviral therapy adherence intervention patient allocation and power by design. P_C (SD) represents the proportion (standard deviation) of 244 enrolled patients randomized to the more effective therapy of combined medication, education, and counseling ($\mu_C = 0.58$ versus $\mu_M = 0.22$ fewer RNA copies/mL on medication alone). N_C reports the expected number of patients assigned to treatment C while Diff is the difference between N_C and 122—the number of patients expected to receive each treatment under equal allocation. Power (SD) details the proportion (standard deviation) of 10,000 clinical trials which correctly reject the null hypothesis and conclude that combined medication, education, and counseling is more effective than medication intervention alone.

Design	P_C (SD)	N_C	Diff	Power (SD)
BBS	0.75 (0.10)	182	60	0.95 (0.21)
BB_1	0.63 (0.06)	153	31	0.96 (0.19)
BB_2	0.56 (0.04)	138	16	0.97 (0.18)
BB_3	0.54 (0.04)	133	11	0.97 (0.18)
Equal	0.50 (0.03)	122	0	0.97 (0.18)

2.8 References

[1] British HIV Association. *BHIVA - Adherence 2001*. URL: `http://www.bhiva.org/Adherence2001.aspx` (visited on 02/16/2015).

[2] Atkinson, A.C. "Optimum biased coin designs for sequential clinical trials with prognostic factors". In: *Biometrika* 69.1 (1982), pp. 61–67.

[3] Atkinson, A.C. "Optimum biased coin designs for sequential treatment allocation with covariate information". In: *Statistics in Medicine* 18 (1999), pp. 1741–1752.

[4] Atkinson, A.C. and Biswas, A. "Adaptive biased-coin designs for skewing the allocation proportion in clinical trials with normal responses". In: *Statistics in Medicine* 24.16 (2005), pp. 2477–2492.

[5] Atkinson, A.C. and Biswas, A. "Bayesian Adaptive Biased-Coin Designs for Clinical Trials with Normal Responses". In: *Biometrics* 61 (2005), pp. 118–125.

[6] Bai, Z.D., Hu, F., and Rosenberger, W.F. "Asymptotic properties of adaptive designs for clinical trials with delayed response". In: *The Annals of Statistics* 30.1 (2002), pp. 122–139.

[7] Baldi Antognini, A. "A theoretical analysis of the power of biased coin designs". In: *Journal of Statistical Planning and Inference* 138.6 (2008), pp. 1792–1798.

[8] Bandyopadhyay, Uttam and Biswas, A. "Adaptive designs for normal response with prognostic factors". In: *Biometrika* 88 (2001), pp. 409–419.

[9] Bandyopadhyay, Uttam, Biswas, A., and Bhattacharya, Rahul. "A covariate adjusted two-stage allocation design for binary responses in randomized clinical trials". In: *Statistics in Medicine* 26 (2007), pp. 4386–4399.

[10] Biswas, A. and Bhattacharya, Rahul. "Reduction of Variability of Response-Adaptive Designs for Continuous Treatment Responses in Phase 3 Clinical Trials". In: *Drug Information Journal* 44 (2010), pp. 9–19.

[11] Biswas, A., Huang, H.-H., and Huang, W.-T. "Covariate-adjusted adaptive designs for continuous responses in a phase III clinical trial: recommendation for practice". In: *Journal of Biopharmaceutical Statistics* 16.2 (2006), pp. 227–239.

[12] Chen, Y.P. "Biased coin design with imbalance tolerance". In: *Communications in Statistics. Stochastic Models* 15.5 (1999), pp. 953–975.

[13] M. A. Chesney, M. Morin, and L. Sherr. "Adherence to HIV combination therapy". In: *Soc Sci Med* 50.11 (2000), pp. 1599–1605.

[14] Efron, B. "Forcing a sequential experiment to be balanced". In: *Biometrika* 58.3 (1971), pp. 403–417.

[15] Eisele, J. R. "The doubly adaptive biased coin design for sequential clinical trials". In: *Journal of Statistical Planning and Inference* 38.2 (1994), pp. 249–261.

[16] Food and Drug Administration. *Critical Path Initiative: Transforming the way FDA-regulated products are developed, evaluated, and manufactured*. Report on Key Achievements in 2009. 2010.

[17] Food and Drug Administration. "Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics". Draft Document. 2010.

[18] Gwise, T., Hu, J., and Hu, F. "Optimal biased coins for two-arm clinical trials". In: *Statistics and its Interface* 1 (2008), pp. 125–135.

[19] Hinkelmann, K. *Design and Analysis of Experiments, Vol. 3: Special Designs and Applications*. 1st ed. Vol. 3. Wiley-Blackwell, 2011.

[20] Joint United Nations Programme on HIV and AIDS (UNAIDS). *Fact sheet*. URL: `http://www.unaids.org/en/resources/campaigns/globalreport2013/factsheet` (visited on 02/16/2015).

[21] Hu, F. and Rosenberger, W.F. "Optimality, Variability, Power: Evaluating Response-Adaptive Randomization Procedures for Treatment Comparisons". In: *Journal of the American Statistical Association* 98.463 (2003), pp. 671–678.

[22] Hu, F. and Rosenberger, W.F. *The Theory of Response-Adaptive Randomization in Clinical Trials*. 1st ed. Hoboken, NJ: John Wiley & Sons, Inc., 2006.

[23] Hu, F. and Zhang, L.-X. "Asymptotic normality of urn models for clinical trials with delayed response". In: *Bernoulli* 10.3 (2004), pp. 447–463.

[24] Hu, F. and Zhang, L.-X. "Asymptotic Properties of Doubly Adaptive Biased Coin Designs for Multitreatment Clinical Trials". In: *The Annals of Statistics* 32.1 (2004), pp. 268–301.

[25] Hu, F. et al. "Doubly-Adaptive Biased Coin Designs with Delayed Responses". In: *Canadian Journal of Statistics* 36.4 (2008), pp. 541–559.

[26] F. S. Julian, P. Martin, and S. R. Erickson. "Validation of the Special Projects of National Significance adherence tool in HIV/AIDS patients". In: *Ann Pharmacother* 44.6 (2010), pp. 1003–1009.

[27] H. Knobel et al. "Validation of a simplified medication adherence questionnaire in a large cohort of HIV-infected patients: the GEEMA Study". In: *AIDS* 16.4 (2002), pp. 605–613.

[28] K. Lamiraud et al. "Adherence to and effectiveness of highly active antiretroviral treatment for HIV infection: assessing the bidirectional relationship". In: *Med Care* 50.5 (2012), pp. 410–418.

[29] Melfi, V. F., Page, C., and Geraldes, M. "An Adaptive Randomized Design with Application to Estimation". In: *Canadian Journal of Statistics* 29.1 (2001), pp. 107–116.

[30] A. Nelsen et al. "Development and validation of a measure for intention to adhere to HIV treatment". In: *AIDS Patient Care STDS* 26.6 (2012), pp. 329–334.

[31] Paganoni, A.M. and Secchi, P. "A numerical study for comparing two response-adaptive designs for continuous treatment effects". In: *Statistical Methods and Applications* 17.3 (2007), pp. 321–346.

[32] Pong, A. and Chow, S.C., eds. *Handbook of Adaptive Designs in Pharmaceutical and Clinical Development*. Boca Raton: CRC Press, 2011.

[33] C. Pradier et al. "Efficacy of an educational and counseling intervention on adherence to highly active antiretroviral therapy: French prospective controlled study". In: *HIV Clin Trials* 4.2 (2003), pp. 121–131.

[34] Rosenberger, W.F. and Hu, F. "Maximizing power and minimizing treatment failures in clinical trials". In: *Clinical Trials* 1.2 (2004), pp. 141–147.

[35] Rosenberger, W.F. and Lachin, J. M. "The use of response-adaptive designs in clinical trials". In: *Controlled Clinical Trials* 14.6 (1993), pp. 471–484.

[36] Rosenberger, W.F. and Sverdlov, Oleksandr. "Handling Covariates in the Design of Clinical Trials". In: *Statistical Science* 23.3 (2008), pp. 404–419.

[37] SAS Institute Inc. *SAS/IML Software, Version 9.2*. Cary, NC, 2011. URL: `http://www.sas.com/`.

[38] F. E. Satterthwaite. "An Approximate Distribution of Estimates of Variance Components". In: *Biometrics Bulletin* 2.6 (Dec. 1946), pp. 110–114. ISSN: 00994987. DOI: `10.2307/3002019`. URL: `http://dx.doi.org/10.2307/3002019`.

[39] Smith, R.L. "Properties of biased coin designs in sequential clinical trials". In: *The Annals of Statistics* 12.3 (1984), pp. 1018–1034.

[40] Soares, J.F. and Wu, C.F.J. "Some restricted randomization rules in sequential designs". In: *Communications in Statistics, Series A* 12 (1983), pp. 2017–2034.

[41] Sun, R., Cheung, S.H., and Zhang, L.-X. "A generalized drop-the-loser urn for multi-treatment clinical trials". In: *Journal of Statistical Planning and Inference* 137.6 (2007), pp. 2011–2023.

[42] Tymofyeyev, Y., Rosenberger, W.F., and Hu, F. "Implementing optimal allocation in sequential binary response experiments". In: *Journal of the American Statistical Association* 102.477 (2007), pp. 224–234.

[43] Viswanathan, S. et al. "Level of Adherence and HIV RNA Suppression in the Current Era of Highly Active Antiretroviral Therapy (HAART)". In: *AIDS Behav* (2014).

[44] Wei, L.J. "The Adaptive Biased Coin Design for Sequential Experiments". In: *The Annals of Statistics* 6.1 (1978), pp. 92–100.

[45] B. L. Welch. "The Generalization of 'Student's' Problem when Several Different Population Variances are Involved". In: *Biometrika* 34.1/2 (1947), pp. 28–35. ISSN: 00063444. DOI: 10.2307/2332510. URL: http://dx.doi.org/10.2307/2332510.

[46] World Health Organization (WHO). *WHO Global Health Observatory: HIV/AIDS*. URL: http://www.who.int/gho/hiv/en/ (visited on 02/16/2015).

[47] Zhang, L. and Rosenberger, W.F. "Response-Adaptive Randomization for Clinical Trials with Continuous Outcomes". In: *Biometrics* 62.2 (2006), pp. 562–569.

[48] Zhang, L.-X. and Hu, F. "A new family of covariate-adjusted response-adaptive designs and their properties". In: *Applied Mathematics - A Journal of Chinese Universities* 24.1 (2009), pp. 1–13.

[49] Zhang, L.-X. et al. “A generalized drop-the-loser urn for clinical trials with delayed responses”. In: *Statistica Sinica* 17 (2007), pp. 387–409.

Chapter 3

Response-Adaptive Biased Coin Design with Unknown, Unequal Covariate Slopes

3.1 Summary

Adaptive designs—experiments which leverage accumulating data to modify various aspects of the investigative plan in a prospectively defined manner—are advocated by academic statisticians, clinicians and researchers, as well as the US Food and Drug Administration to improve patient care and accelerate the treatment approval process. This paper focuses on an adaptive design which progressively allocates more patients to the treatment with larger effect size while preserving randomization and even blinding processes. Unlike its predecessor, this new design is applicable when treatment variances are unknown or unequal. Simulations confirm that the new design continues to be preferable even when covariate-treatment interactions are present. The utility and benefits of the modified design are illustrated with a real-world application of an

HIV treatment adherence intervention.

Keywords: adaptive randomization, ethical allocation, covariate-treatment interaction

3.2 Introduction

3.2.1 Adaptive Designs

Adaptive designs—experiments which leverage accumulating data to modify various aspects of the investigative plan including population, randomization, trial duration, and/or analysis in a prospectively defined manner—have been advocated by academic statisticians and select clinicians since the 1970s [8, 27, 30, 32, 36, 43]. More recently, the US Food and Drug Administration called for further research on and implementation of such designs to improve patient care and accelerate the treatment approval process [15, 16].

One advantage of certain adaptive designs is the potential to treat patients more effectively while preserving randomization and even blinding processes. For example, this paper focuses on a modified design which progressively allocates more patients to the treatment with larger effect size [4]. Other adaptive designs have the ability to maximize a design's power (alternatively maintain power relative to a nonadaptive design but decrease the requisite sample size) [3, 19, 31]. Unfortunately, no known adaptive design provides both benefits simultaneously [17, 20, 28].

3.2.2 BBS Design

In Chapter 2, the author introduces an improvement to the adaptive biased coin design proposed by [4]. The original design, referred to as BB, assigns a patient to treatment A with probability $\Phi\left(\frac{\widehat{\mu}_A - \widehat{\mu}_B}{T}\right)$, where Φ is the cumulative distribution function of the standard normal distribution; $\widehat{\mu}_k$ is the current estimate at time of randomization of the mean effect of treatment $k = A, B$; and T is an arbitrary, positive constant.

The enhanced design—called BBS—is juxtaposed against equal allocation (assignment to treatment A with constant probability $\frac{1}{2}$) and BB with $T = 1$, $T = 2$, and $T = 3$. The five designs are compared via simulation, calculating patient allocation to the better treatment and evaluating rejection rates of the null hypothesis (power and Type I error). Overall, BBS performs favorably when compared to equal allocation. This remains true when compared to BB at all levels, including the most aggressive of the BB design parameters, $T = 1$. The BBS design also performs favorably when the BB assumption of equal treatment variances is violated. In particular, BBS assigns patients to the better treatment more aggressively than BB with $T = 1$ when the pooled treatment variance is smaller than 1, more conservatively when the pooled treatment variance is larger than 1, and in equal proportion to BB with $T = 1$ when both treatment variances are 1. For these reasons, only BB with $T = 1$—the best competitor from the original design—is kept for comparison in the simulations described below. The BB design with $T = 1$ will be denoted as BB_1 for this paper. The only potential pitfall of the BBS design is an increased risk of falsely concluding a treatment difference exists.

3.2.3 Covariate-Treatment Interaction

Recent research highlights many examples of clinical trial with a covariate-treatment interaction. See [12] for a thorough overview of the topic; [5] for examples in research; [40] for a framework for identification of treatment effect heterogeneity within a clinical trial; and [49] or [9] for individualized medicine/personalized treatment approaches to the topic. The BB design requires that no covariate-treatment interaction exist which limits the adaptive design's applicability. The BBS design is less restrictive and, in particular, is applicable even when covariate-treatment interaction is present. In this article, the impact on the designs due to covariate influences which differ across treatments is examined.

This article further demonstrates that BBS is a suitable candidate for use in clinical trials where exposing patients to the superior treatment is a priority. Moreover, this article confirms that BBS continues to be preferable to BB due to the ability to safely relax the BB assumptions of known, equal treatment variances and identical treatment-covariate effects across treatments. Finally, like Chapter 2, this article illustrates the utility and benefits of the modified BBS design with a real-world application of an HIV treatment adherence intervention.

3.3 Design and Simulation Parameters

3.3.1 Theory

Sections 2.3.1 to 2.3.3 in Chapter 2 detail the clinical trial model and design employed throughout this article. To summarize, a clinical trial seeks to determine which of two treatments has a larger treatment effect, after removing the covariate impact on patient response. Prognostic factors are independently and identically distributed,

known for each patient upon enrollment. Patient responses, conditional on each patient's covariate and unknown covariate slope, are normally distributed.

Three randomization designs are considered in this paper. The simplest of the three is the randomized clinical trial "gold standard" design, equal allocation [26]. With a fixed probability of assignment, equal allocation assigns patients to treatment A with probability $\frac{1}{2}$ and to treatment B with probability $\frac{1}{2}$, regardless of what patient outcome data may be available. Equal allocation designs help to minimize selection bias, maximize power, decrease trial size, and promote balance asymptotically [27, 31, 33, 36]. Nevertheless, as the design name implies, equal allocation randomizes patients to treatment in asymptotically equal proportions, even if one treatment performs substantially better than the other treatment for the duration of a trial.

The second design is an adaptive biased coin that allocates more patients to the treatment with larger effect size by updating its randomization proportions over time. The BB design assigns a fixed number of patients to each treatment, estimates the difference in treatment effects, and then randomizes each subsequent patient to treatment A with probability $\Phi\left(\frac{\widehat{\mu}_A-\widehat{\mu}_B}{T}\right)$ and to treatment B with probability $1-\Phi\left(\frac{\widehat{\mu}_A-\widehat{\mu}_B}{T}\right)$, where the estimates are continually updated after each patient is treated and immediately responds. In the original BB proposal, T may be any positive constant. Keeping with the assumptions that treatment variances are known and both equal to 1.0, this article only considers BB_1, the best contender against the BBS design in Chapter 2.

Like the BB design, BBS assigns a minimum number of patients to each treatment, estimates the difference in treatment effects, and randomizes subsequent patients to treatment A with probability $\Phi\left(\frac{\widehat{\mu}_A-\widehat{\mu}_B}{S}\right)$ and to treatment B with probability $1-\Phi\left(\frac{\widehat{\mu}_A-\widehat{\mu}_B}{S}\right)$, where the estimates are regularly updated after each patient is treated and immediately responds. In the BBS design, however, S is not an arbitrary constant

but another adaptive estimator, the pooled standard deviation estimate

$$S = \sqrt{\frac{(N_A - 1)s_A^2 + (N_B - 1)s_B^2}{N_A + N_B - 2}}.$$

Chapter 2 compares the distinct designs based on which randomization scheme allocates more patients to the better treatment without sacrificing power or inflating Type I error rates. This article considers what happens to the three metrics (allocation, power, and Type I error) when another BB design restriction is relaxed and the covariate slopes are heterogenous across treatments. Specifically, the effect a prognostic factor has on patient outcome is different between treatment A and treatment B.

3.3.2 Simulation

Section 3.4 describes the results of simulated clinical trials under equal allocation, BB_1, and BBS. Treatment mean pairs (μ_A, μ_B) are set to no effect $(0, 0)$ and positive effect in treatment A $(1, 0)$. Individual patient errors are simulated under equal and unequal conditions; $\sigma_B = 1$ while σ_A ranges from 0.5 to 1.5 by half-steps. The symmetry of varying treatment A versus treatment B parameters are discussed in Chapter 2.

One normally-distributed prognostic factor is simulated. While estimated separately for each treatment, the covariate effects and variances are identical for both treatments with $Z \sim N(1, 1)$. The covariate slopes, however, vary by treatment. While the covariate-treatment interaction effect for treatment A holds constant at $\beta_A = 2$, the slope for treatment B varies from $\beta_B = 1$ to $\beta_B = 3$ in unit intervals—that is, $\beta_B = 1, 2, 3$. Similarly, while β_B is held constant, β_A ranges from 1 – 3.

In the BB design, rejection of the null hypothesis $H_0 : \mu_A \leq \mu_B$ is calculated with

a one-sided two-sample t-test with significance level $\alpha = 0.05$. Under the assumption of equal variances, the t-statistic standard error employs the sample pooled variance estimates. Relaxing the equal variance assumption for the equal allocation and BBS designs, the t-statistic standard error is calculated using treatment-specific sample variances. In these cases, the underlying hypothesis distribution leverages the Welch-Satterthwaite approximated pooled degrees of freedom [35, 44].

Sample sizes mimic small to large Phase III clinical trials ($N = 50$; 100; 500; 1,000; 5,000). The initial number of patients enrolled prior to switching to an adaptive method is fixed at $m = 10$ for each treatment. Patient responses are assumed to be instantaneous; no delay is incorporated into these trials. For each scenario described, 1,000 replications are simulated in SAS IML [34].

3.4 Results

3.4.1 Patient Allocation

Tables 3.1 – 3.6 contain proportions and standard deviations of patients assigned to treatment A by design and treatment effect values. Tables 3.1 – 3.3 detail average patient allocation for equal allocation, the BB_1 design, and BBS—respectively—when neither treatment A nor treatment B have an effect ($\mu_A = \mu_B = 0$). Tables 3.4 – 3.6 detail average patient allocation for the three designs when treatment A has a positive impact and treatment B has none ($\mu_A = 1, \mu_B = 0$). The first two columns of each table indicate the treatment parameters for the row; (σ_A, σ_B) describes the standard deviations for each treatment while (β_A, β_B) lists the treatment-covariate slope by treatment. The remaining five columns are the mean (SD) percent of patients assigned to treatment A over 1,000 simulations for each of the five trial sizes $N =$

50; 100; 500; 1,000; and 5,000.

The top half of each table describes simulations where the relationship between the covariate and treatment A varies but the relationship between covariate and treatment B is held constant at $\beta_B = 2$. In the bottom half of each table, it is β_A which is held constant at $\beta_A = 2$ while the relationship between the covariate and treatment B varies. Each half is separated by a dashed line for reading convenience.

Tables 3.1 and 3.4 summarize allocations from simulations where patients are assigned using equal allocation. In the first row of Table 3.1, there are no differences in treatment effects, but treatment A variance is 0.5 while treatment B variance is 1.0 and the covariate-treatment slope for treatment A is $\beta_A = 1$ while the covariate-treatment slope for treatment B is $\beta = 2$. As anticipated for equal allocation, the probability of assignment to treatment A is 50% with negligible variation across 1,000 simulations at all clinical trial sizes. Table 3.4 reveals similar results: even when treatment A differs from treatment B in effect size ($\mu_A = 1, \mu_B = 0$), allocation to treatment A is consistently half of patients with standard deviation no greater than 0.01.

The first row of Table 3.6 contains the mean proportion of patients assigned to treatment A from simulations where patients are randomized using BBS, treatment A has positive effect $\mu_A = 1$ while treatment B has no effect, and treatment variances and covariate-treatment slopes are the same as described above. Unlike equal allocation, BBS assigns more patients to the better treatment: 71% of patients are randomized to treatment A when $N = 50$, 82% when $N = 100$, and more than 90% of patients receive the superior treatment when trial enrollment is 500 or larger.

Figure 3.1 illustrates the trends in patient assignment from Tables 3.1 – 3.3 across the five simulated trial sizes for all nine combinations of covariate-treatment slope and treatment A standard deviation when $\mu_A = \mu_B = 0$. The covariate-treatment

interaction of treatment A varies with the non-varying covariate slope of treatment B set to 2.0. That is, Figure 3.1 contains data from the top half of Tables 3.1 – 3.3.

Figure 3.1: Proportion of patients assigned to treatment A by design when $\mu_A = \mu_B = 0$, σ_A varies, $\sigma_B = 1$, and N varies. Note that total patient enrollment N is not to scale. Treatment covariate slopes vary for β_A and are fixed for $\beta_B = 2$. Trends appear similar when $\beta_A = 2$ and β_B varies. Data points are grouped by design; BBS is represented by $*$, BB_1 by $\square$, and equal allocation by $\bullet$.

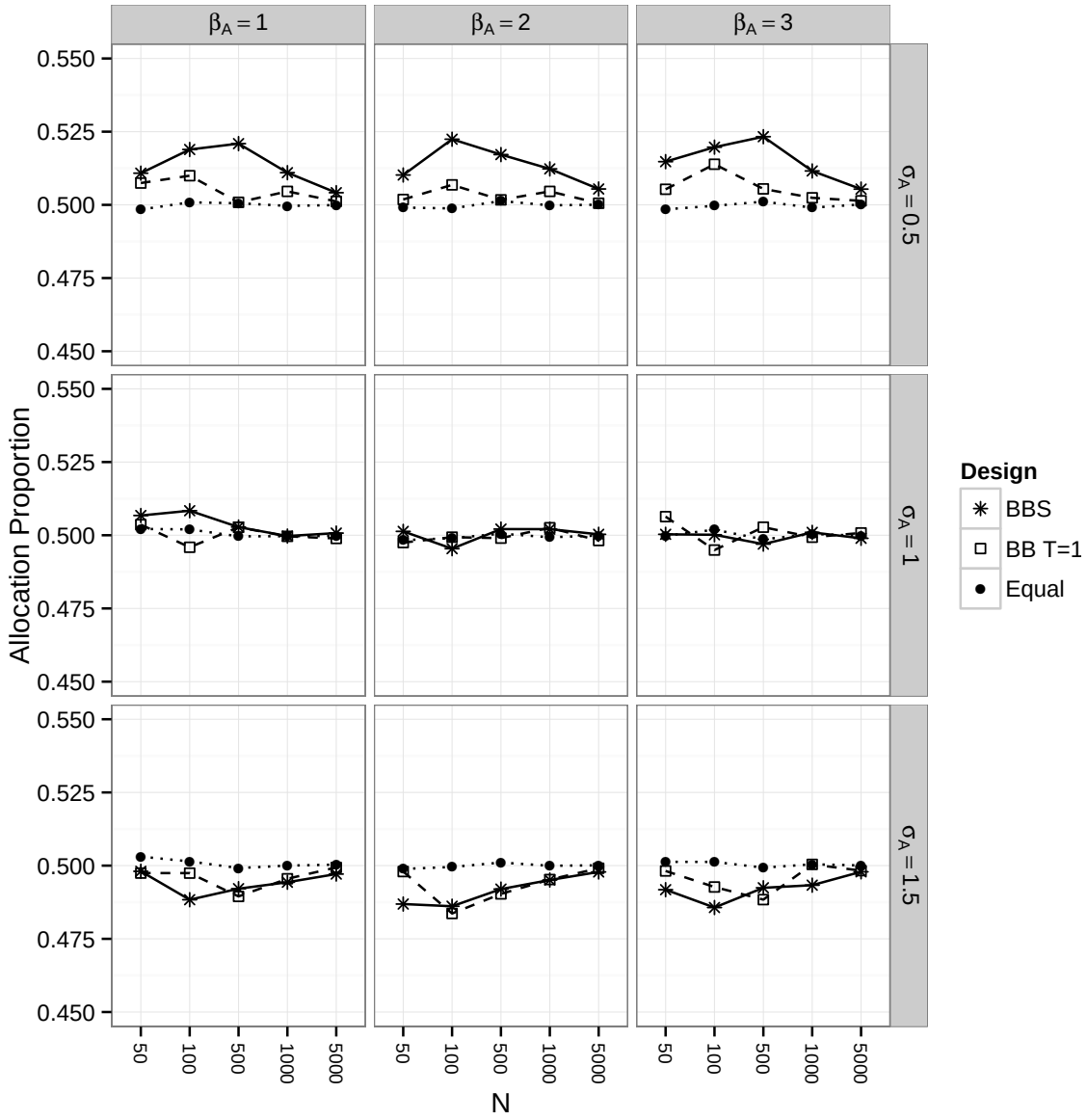


Figure 3.2 illustrates the trends in patient assignment from Tables 3.4 – 3.6 across

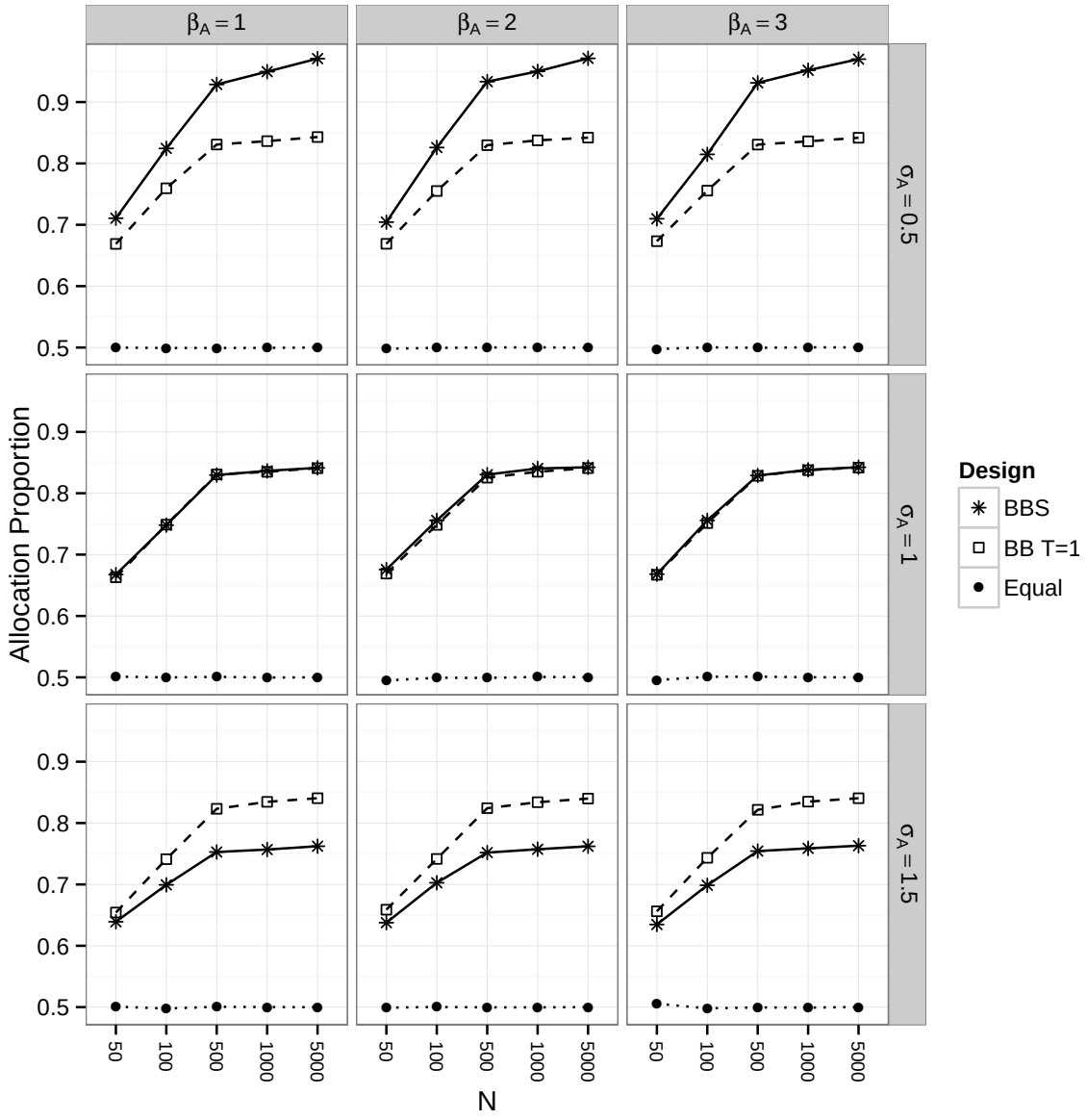
the five simulated trial sizes for all nine combinations of covariate-treatment slope and treatment A standard deviation when $\mu_A = 1$, $\mu_B = 0$. The covariate-treatment interaction of treatment A varies with the non-varying covariate slope of treatment B set to 2.0. That is, Figure 3.2 contains data from the top half of Tables 3.4 – 3.6.

As seen in Tables 3.4 – 3.6 and in Figures 3.1 and 3.2, patient allocation is impacted by treatment effect and variance as well as trial enrollment. Treatment-covariate slope, however, has negligible impact on patient allocation, regardless of design. Overall, the BBS and BB_1 designs continue to expose more patients to the treatment with larger effect size compared to equal allocation.

Simulations indicate that the largest influence on patient allocation in the BBS design (given a difference in treatment effects and sample size) is treatment variance. For example, in Table 3.6 and in Figure 3.2, as treatment A standard deviation increases from 0.5 to 1.5 (with standard deviation of treatment B fixed at 1), the maximum proportion of patients assigned to the superior treatment decreases from nearly 100% of patients to just over 75%, achieved when $N = 5,000$. The minimum proportion, achieved when $N = 50$, decreases from 71% to 63%. Little variation in patient allocation, if any, appears to be associated with changes in covariate slope. This is evidenced throughout Tables 3.3 and 3.6 and reflected in Figures 3.1 and 3.2: when treatment effects (μ_A, μ_B), treatment variances (σ_A, σ_B), and trial size (N) are held constant, patient allocation proportions and standard deviations are similar across all six combinations of treatment-covariate impact (β_A, β_B).

Trends in the BB_1 design also appear to be due to total trial size, difference in treatment effects, and treatment variances. The impact of treatment variance on patient allocation is dampened in BB_1 compared to BBS. For example, in Table 3.5 the proportion of patients assigned to the superior treatment by BB_1 only fluctuates by 2% when $N = 50$ and by 3% when $N = 100$ compared to the differential noted

Figure 3.2: Proportion of patients assigned to treatment A by design when $\mu_A = 1$, $\mu_B = 0$, σ_A varies, $\sigma_B = 1$, and N varies. Note that total patient enrollment N is not to scale. Treatment covariate slopes vary for β_A and are fixed for $\beta_B = 2$. Trends appear similar when $\beta_A = 2$ and β_B varies. Data points are grouped by design; BBS is represented by $*$, BB_1 by $\square$, and equal allocation by $\bullet$.



in the BBS design in Table 3.6 of 8% when $N = 50$ and 13% when $N = 100$. As with BBS, variation in patient allocation does not appear to be associated with changes in covariate slope. Tables 3.2 and 3.5 contain similar allocation proportions

and standard deviations across treatment-covariate impacts after accounting for other factors (treatment effect and variance, patient enrollment).

While equal allocation treatment randomization proportions are noisier in trials with small N than with large N, equal allocation exposes asymptotically equal proportions of patients to treatment A as to treatment B[13, 47]. In Tables 3.1 and 3.4, assignment of patients to treatment A under equal allocation deviates only minimally from $\frac{1}{2}$, regardless of changes to treatment effect, treatment variance, treatment-covariate interaction, or sample size. In each of Figures 3.1 and 3.2, equal allocation randomization proportions deviate from 0.50 negligibly throughout all 45 simulation parameter combinations.

When no treatment differences exist, patient allocations by the BBS design are closer to equal; however, assignments may still be biased due to differences in treatment variances. For example, when $\sigma_A = 0.5$, more patients were assigned to treatment A by BBS than to treatment B across all trial sizes. On the other hand, when $\sigma_A = 1.5$, fewer patients were randomized by BBS to treatment A regardless of enrollment size. The largest variation is seen when $N = 100$ or 500. When $N = 50$, at least 40% of patients are initially assigned using equal allocation before adaptive randomize begins. This may be responsible for mitigating some of the skewing due to treatment variances. At the other extrema, when $N = 5,000$, patient exposure differs by less than 1% between treatments.

When no treatment differences exist, the BB_1 design behaves similarly to the BBS design. As seen in the top row of graphs in Figure 3.1, under equivalent treatment means, BBS deviates from 50–50 allocation more than BB_1 when $\sigma_A = 0.5$. The two designs are comparable for $\sigma_A = 1.0$ and for $\sigma_A = 1.5$.

3.4.2 Power and Type I Error

Tables 3.7 – 3.12 contain proportions and standard deviations of trials in which the null hypothesis that treatment A was no better than treatment B was rejected. Each table consists of data for a different randomization design/treatment effect combination. Tables 3.7 – 3.9 detail simulated type I error rates for equal allocation, BB_1, and BBS—respectively. That is, these tables relate the average rejection rates when neither treatment A nor treatment B have an effect ($\mu_A = \mu_B = 0$). Tables 3.10 – 3.12 detail simulated power for the three designs; the ability of a trial to correctly reject the null hypothesis when treatment A has a positive impact and treatment B has none ($\mu_A = 1, \mu_B = 0$). The first two columns of each table indicate the treatment parameters for the row; (σ_A, σ_B) describes the standard deviations for each treatment while (β_A, β_B) lists the treatment-covariate slope by treatment. The remaining five columns are the mean (SD) proportion of trials which rejected the null hypothesis over 1,000 simulations for each of the five trial sizes $N = 50; 100; 500;$ 1,000; and 5,000.

The top half of each table describes simulations where the relationship between the covariate and treatment A varies but the relationship between covariate and treatment B is held constant at $\beta_B = 2$. In the bottom half of each table, it is β_A which is held constant at $\beta_A = 2$ while the relationship between the covariate and treatment B varies. Each half is separated by a dashed line for reading convenience.

The first row of Table 3.7 summarizes rejection rates from simulations where patients are assigned using equal allocation, there are no differences in treatment effects, but treatment A variance is 0.5 while treatment B variance is 1.0 and the covariate-treatment slope for treatment A is $\beta_A = 1$ while the covariate-treatment slope for treatment B is $\beta = 2$. The probability of falsely rejecting the null hypothesis under these conditions is dependent on clinical trial size: 0.06 when $N = 50$, 100,

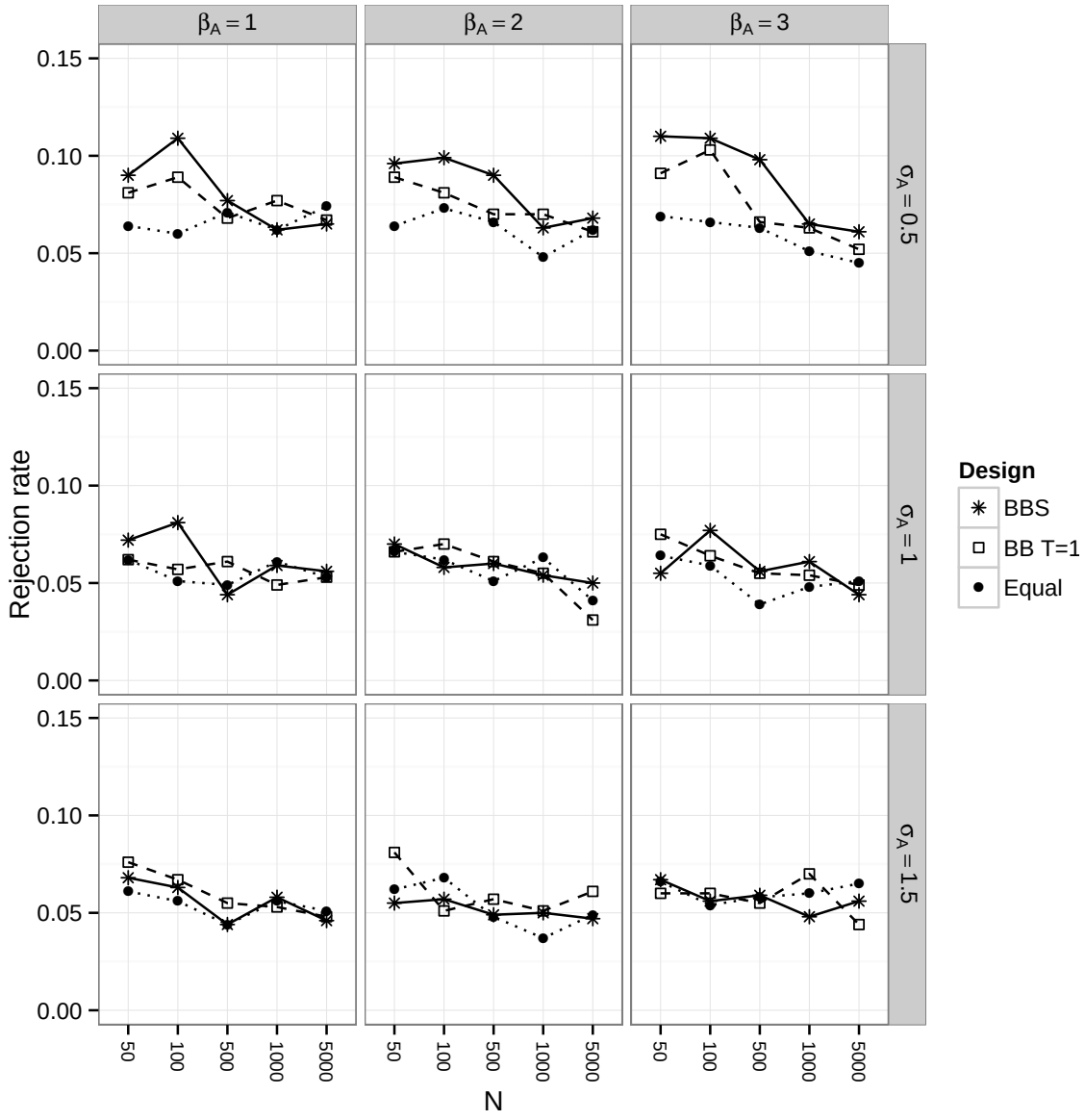
or 1,000 and 0.07 when $N = 500$ or 5,000. By contrast, in equal allocation based Table 3.10, treatment A differs from treatment B in effect size ($\mu_A = 1, \mu_B = 0$) while treatment variances and covariate-treatment slopes remain consistent with those in Table 3.7. The first row of Table 3.10 indicates that the power to correctly reject the null hypothesis under equal allocation is 0.93 when $N = 50$ and 1 for larger values of N.

Figure 3.3 illustrates the trends in type I error from Tables 3.7 – 3.9 across the five simulated trial sizes for all nine combinations of covariate-treatment slope and treatment A standard deviation when $\mu_A = \mu_B = 0$. The covariate-treatment interaction of treatment A varies with the non-varying covariate slope of treatment B set to 2.0. That is, Figure 3.3 contains data from the top half of Tables 3.7 – 3.9.

Figure 3.4 illustrates the trends in power from Tables 3.10 – 3.12 across the five simulated trial sizes for all nine combinations of covariate-treatment slope and treatment A standard deviation when $\mu_A = 1$, $\mu_B = 0$. The covariate-treatment interaction of treatment A varies with the non-varying covariate slope of treatment B set to 2.0. That is, Figure 3.4 contains data from the top half of Tables 3.10 – 3.12.

As seen in Tables 3.10 – 3.12 and in Figures 3.3 and 3.4, the probability with which the null hypothesis of a trial will be rejected is impacted by treatment effect and variance as well as trial enrollment. Treatment-covariate slope, however, has negligible impact on a trial's rejection of the null hypothesis, regardless of design. Overall, the BBS and the BB_1 designs continue to reject the null hypothesis that treatment A is no more effective than treatment B at comparable rates to equal allocation. The biggest distinction noted in rejection rate simulations involves an inflated Type I error in BBS and in BB_1 when the sample size N is small, especially when the standard deviation of treatment A is also small. The adaptive designs incorrectly reject the null hypothesis at similar rates to equal allocation when N is

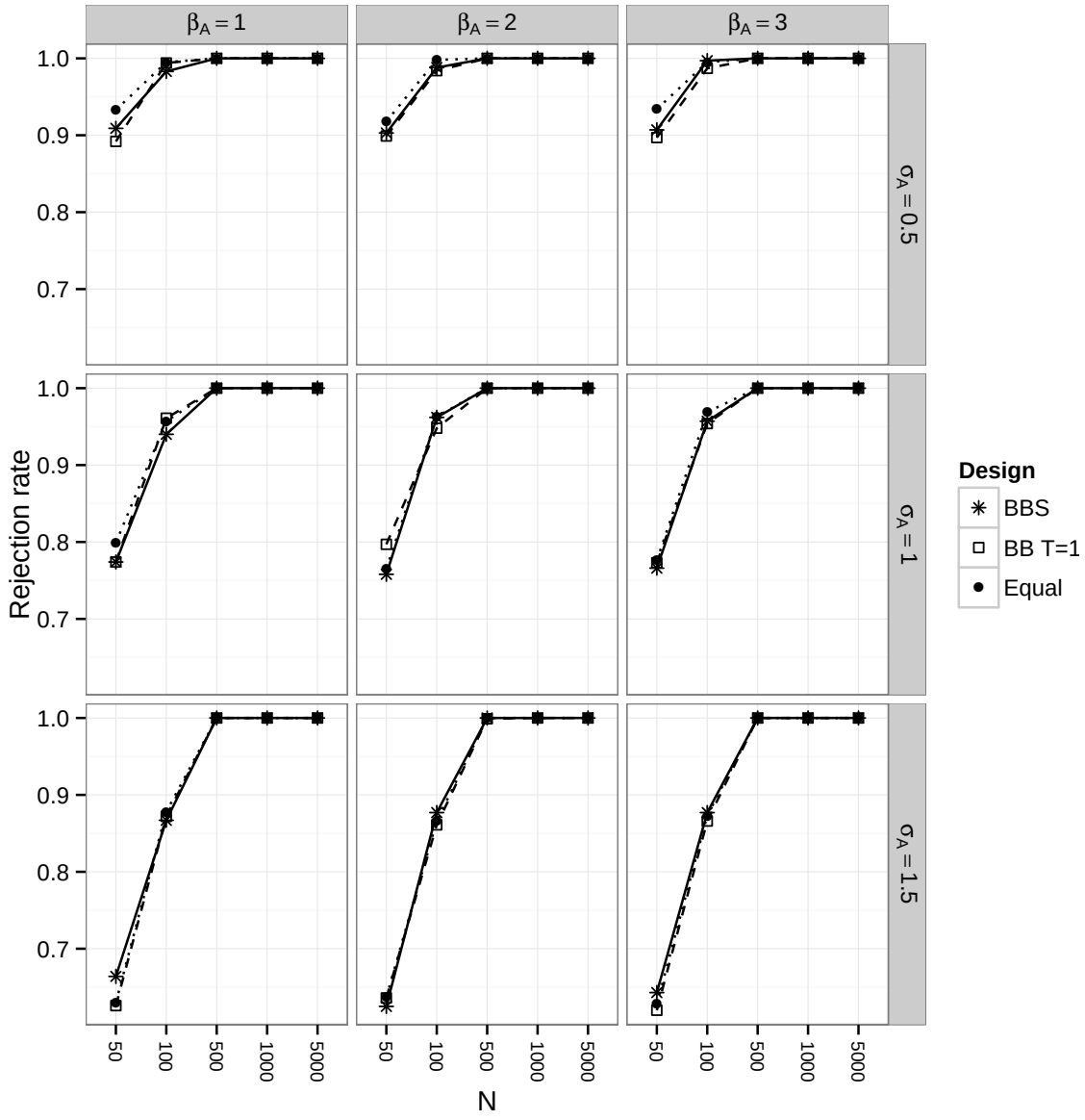
Figure 3.3: Proportion of trials which incorrectly reject the null hypothesis by design when $\mu_A = \mu_B = 0$, σ_A varies, $\sigma_B = 1$, and N varies. Note that total patient enrollment N is not to scale. Treatment covariate slopes vary for β_A and are fixed for $\beta_B = 2$. Trends appear similar when $\beta_A = 2$ and β_B varies. Data points are grouped by design; BBS is represented by $*$, BB_1 by $\square$, and equal allocation by $\bullet$.



large and when the standard deviation of treatment A is large.

Simulations indicate that the largest influences on power in the BBS design are treatment variance and clinical trial size, for a fixed difference in treatment effects.

Figure 3.4: Proportion of trials which correctly reject the null hypothesis by design when $\mu_A = 1$, $\mu_B = 0$, σ_A varies, $\sigma_B = 1$, and N varies. Note that total patient enrollment N is not to scale. Treatment covariate slopes vary for β_A and are fixed for $\beta_B = 2$. Tends appear similar when $\beta_A = 2$ and β_B varies. Data points are grouped by design; BBS is represented by $*$, BB_1 by $\square$, and equal allocation by $\bullet$.



For example, in Table 3.12 and in Figure 3.4, as treatment A standard deviation increases from 0.5 to 1.5 (with standard deviation of treatment B fixed at 1 and both covariate-treatment slopes at 2), the BBS design's power to reject the null hypothesis

decreases from 0.90 to 0.63 when $N = 50$, from 0.99 to 0.88 when $N = 100$, and is constant at 1.0 for larger N. These numbers hold when treatment A standard deviation is fixed and treatment B standard deviation varies.

Likewise, the largest influences on Type I error rates for the BBS design appear to be treatment variance and clinical trial size. As treatment A standard deviation decreases from 1.5 to 0.5 with the standard deviation of treatment B fixed at 1, the BBS rate of false rejection the null hypothesis increases, especially for small values of N. For example, in Table 3.9 and in Figure 3.3, when $N = 50$, the risk of a BBS trial producing a Type I error ranges from 0.06 to 0.11; when $N = 500$ the risk ranges from 0.04 to 0.10; and when $N = 5,000$ the risk increases only from 0.04 to 0.08 across all values of treatment variance and covariate-treatment slope.

Furthermore, little variation in rejection rates appear to be associated with changes in covariate slope. This is evidenced throughout Tables 3.9 and 3.12 and reflected in Figures 3.3 and 3.4: when covariate-treatment interactions are simulated, the power to reject the null hypothesis varies by no more than three percent while treatment effects (μ_A, μ_B), treatment variances (σ_A, σ_B), and trial size (N) are held constant.

The influences on power and type I error appear similar for the BB_1 design. For example, in Table 3.11 and in Figure 3.4, as treatment A standard deviation increases from 0.5 to 1.5, the power of BB_1 trials to reject the null hypothesis decreases from approximately 0.91 to 0.53 when $N = 50$, from 1.0 to 0.81 when $N = 100$, and is constant at 1.0 for larger N—comparable to the trends displayed by BBS power. Like the BBS, the BB_1 design also sees an increased risk of false rejection, particularly for small N and small treatment A variance. For example, in Table 3.8 and in Figure 3.3, when $N = 50$, the risk of a BB_1 trial producing a type I error ranges from 0.04 to 0.11; when $N = 500$ the risk ranges from 0.03 to 0.08; and when $N = 5,000$ the risk ranges from 0.03 to 0.07—again comparable to BBS type I error. As with

BBS, variation in BB_1 rejection rates do not appear to be associated with changes in covariate slope. Tables 3.8 and 3.11 contain similar type I error and power rates across treatment-covariate impacts after accounting for other factors (treatment effect and variance, patient enrollment).

Both adaptive designs' power differ only minimally from that of equal allocation. For example, in Table 3.10 and in Figure 3.4, as treatment A standard deviation increases from 0.5 to 1.5, equal allocation power decreases from approximately 0.93 to 0.63 when $N = 50$, from 1.0 to 0.84 when $N = 100$, and is constant at 1.0 for larger N—similar to the trends displayed by both adaptive designs. The biggest difference between equal allocation rejection rates and those of the adaptive designs is the aforementioned inflation of adaptive type I error rates. While the BBS and BB_1 false rejection rates can vary from 5% of trials to one out of every ten trials, the equal allocation false rejection rates remain steady. For example, in Table 3.7 and in Figure 3.3, when $N = 50$, the risk of an equal allocation trial producing a type I error ranges only from 0.05 to 0.08; for all larger N, the risk ranges from 0.04 to 0.07; and when $N = 5,000$ the risk ranges from 0.03 to 0.07—again comparable to BBS type I error. Again, little variation in rejection rates appear to be associated with changes in covariate slope. This is evidenced throughout Tables 3.7 and 3.10 and reflected in Figures 3.3 and 3.4: when covariate-treatment interactions are simulated, the power to reject the null hypothesis in an equal allocation trial varies by no more than one percent while treatment effects (μ_A, μ_B), treatment variances (σ_A, σ_B), and trial size (N) are held constant.

3.5 Application

As of 2013 estimates, the HIV/AIDS epidemic has seen 80 million individuals worldwide infected with HIV, half of whom have died from HIV-related causes [46, 18]. Due to breakthroughs in medical research, transmission of the virus can be curbed, life expectancies can be extended, and quality of life can be improved—provided patients achieve adequate adherence to prescribed therapies [7, 23]. Antiretroviral courses are among the most successful treatments currently offered; however, the complexity of the medication and dosage combinations, the incessant and inflexible timing of treatments, and the side effects of medication make compliance difficult [7, 41].

In a 2003 adherence intervention, 244 HIV patients on antiretroviral therapy are prospectively randomized to standard medication alone or medication with additional educational and counseling in [29]. The authors examine the intervention's effect on adherence as a binary measure and—more directly related to therapeutic effectiveness—HIV RNA suppression (in RNA copies/mL) as a continuous measure. Both outcomes are conditioned on baseline adherence as a prognostic factor.

A simulation is performed which modifies [29] based on reasonable estimates from [24, 29] of baseline adherence covariate and treatment outcome HIV RNA suppression parameters. Baseline adherence is independently and normally distributed for each patient regardless of treatment assignment with mean 0.60 and standard deviation 0.49. Three treatment-covariate interaction combinations are tested: medication alone (M) and combined medication/education/counseling (C) slope pairs $(\beta_C, \beta_M) = (1.2, 1.0); (1.1, 1.1); (1.0, 1.2)$. Treatment effect and variance parameters are chosen so that patient responses (treatment effects with the additional impact due to covariates) align with [29]'s clinically and statistically significant outcomes. That is, medication alone only decreases HIV RNA by 0.89 copies/mL with a standard

deviation of 0.90 while the combined therapy decreases HIV RNA by 1.25 copies/mL (standard deviation 0.86). The one-sided null hypothesis (combined therapy is no more effective than medication alone) is tested at significance $\alpha = 0.05$. Delay in response time is ignored in these simulations but discussed in Section 3.6.3. Ten-thousand trials are simulated for each scenario.

Equal allocation, BB_1, and BBS are leveraged to assign 244 simulated patients to treatment. Table 3.13 records simulated patient assignment (counts and proportions) and power of each design. Data are grouped by design in sets of three rows. The first two columns (β_C, β_M) and $\mu_C - \mu_M$ distinguish for which covariate slope pair and treatment effect difference simulation outcomes are being summarized. The first set of rows testify that equal allocation assigns 122 or 50% of simulated patients to each treatment irrespective of changes to the covariate slopes. There would have been no additional patients (zero patients and a zero percent increase) assigned to the superior treatment over equal allocation using this method of complete randomization. The simulated power to detect a difference in treatments ranges from 0.75 when the superior treatment also has a larger covariate-treatment interaction (hence treatment effect alone is smallest, 0.24 copies/mL) to 1.00 when the superior treatment has a smaller covariate-treatment interaction (hence difference in treatment effects is largest at 0.48 copies/mL). The second set of rows demonstrate that BB_1 has nearly identical simulated power as equal allocation after accounting for covariate slopes. BB_1 would have randomized an additional 17, 26, and 33% of patients to the more effective intervention therapy than equal allocation, based on the difference in treatment effects. The third set of rows indicate that, although the BBS design rejects the null hypothesis in only 1% fewer simulated trials in each scenario than equal allocation, BBS would have exposed 169, 182, and 192 patients to the treatment with larger effect size and superior outcome. That is, 38% to 57% more patients would have

been exposed to the superior therapy via BBS randomization than through equal allocation. The differences in patient allocation across the three designs are significant ($p < 0.001$) for each covariate slope combination; the differences in power are not ($p = 0.923, 0.361, 0.896$ for $(\beta_C, \beta_M) = (1.2, 1.0); (1.1, 1.1); (1.0, 1.2)$, respectively).

Figures 3.5 and 3.6 illustrate the statistics described in Table 3.13. Figure 3.5 emphasizes the increase in proportion of patients that would have been allocated to the superior treatment using the BBS design compared to both equal allocation and BB_1, even as the treatment effect size shrinks due to varying covariate treatment effects. Figure 3.6 demonstrates how similar the probabilities of rejecting the null hypothesis would have been for each design, especially with respect to the standard deviation of the probabilities, after accounting for the difference in treatment effect sizes.

3.6 Discussion

3.6.1 Ethical Patient Allocation

As covariate effects on patient outcomes vary within a reasonable range, their effect on patient allocation is small. When a difference in treatment effects exists, both adaptive designs expose more patients to the better treatment than equal allocation. Chapter 2 illustrates the impact of the magnitude of the difference between treatment effects on patient allocation by adaptive designs; the larger the treatment difference relative to the design denominator (T or S), the more patients are likely to be assigned to the treatment with larger effect size. Chapter 2 also describes how the proportion of patients exposed to a treatment with larger effect size increases quickly as total enrollment changes from small to mid-sized clinical trials, but plateaus for medium

Figure 3.5: Allocation proportions for HIV antiretroviral therapy adherence intervention by randomization design through 10,000 simulated trials. The points (bars) represent the proportion (standard deviation) of 244 patients assigned to the more effective combined therapy of medication, education, and counseling (C) versus the less effective treatment of medication alone (M). The proportion (SD) is calculated for each pair of covariate slope values. Data points are grouped by design; BBS is represented by $*$, BB_1 by $\square$, and equal allocation by $\bullet$. For reference, the difference in treatment effects ($\mu_C - \mu_M$) across the three covariate slope combinations from left to right are 0.48, 0.36, and 0.24 HIV RNA copies/mL.

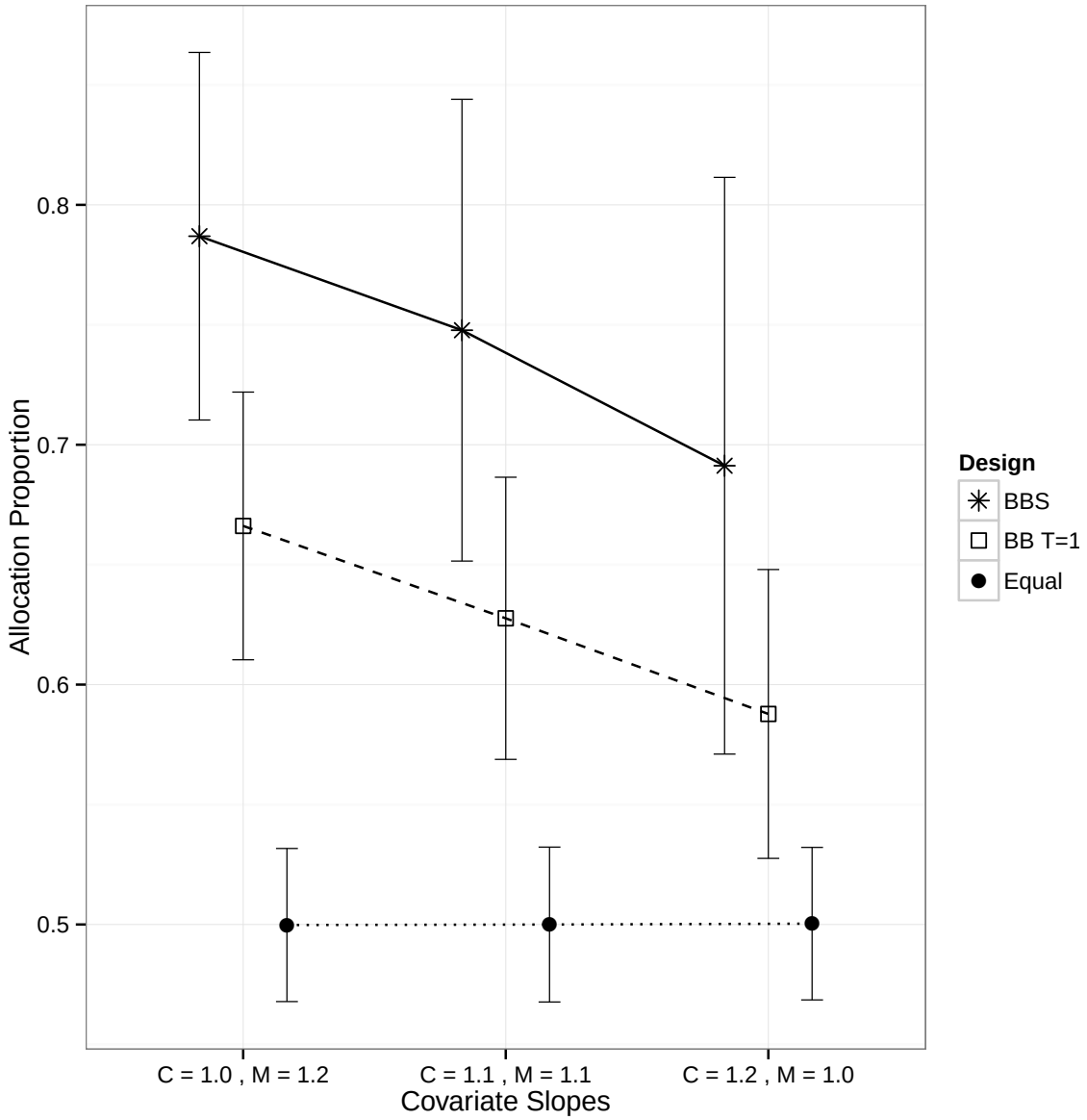
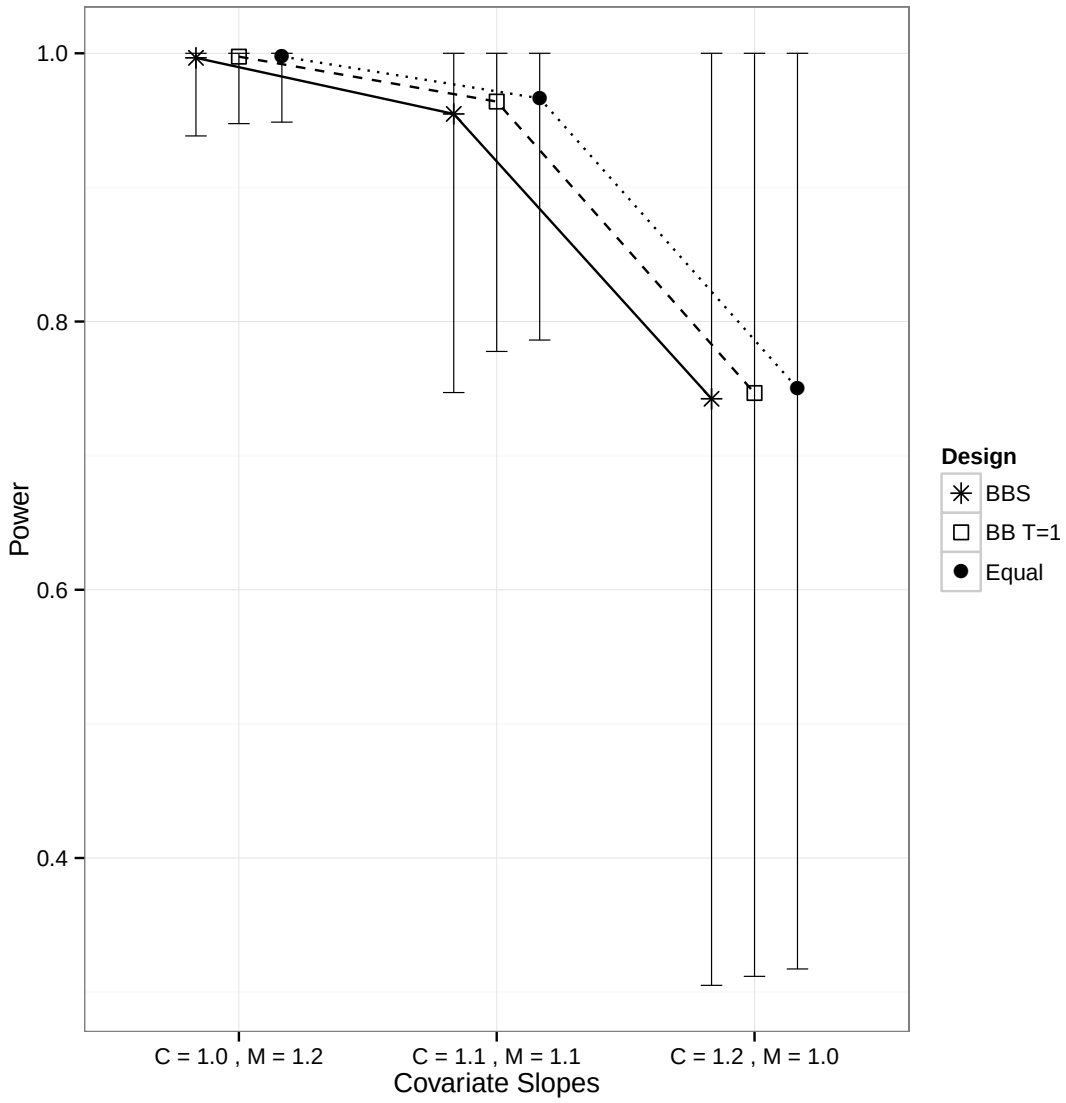


Figure 3.6: Rejection rates for HIV antiretroviral therapy adherence intervention by randomization design through 10,000 simulated trials. The point (bars) represent the proportion (standard deviation) of simulated antiretroviral therapy clinical trials which correctly reject the null hypothesis and conclude that an education and counseling intervention in conduction with medication (C) is more effective than medication treatment alone (M). The proportion (SD) is calculated for each pair of covariate slope values. Data points are grouped by design; BBS is represented by $*$, BB_1 by $\square$, and equal allocation by $\bullet$. For reference, the difference in treatment effects ($\mu_C - \mu_M$) across the three covariate slope combinations from left to right are 0.48, 0.36, and 0.24 HIV RNA copies/mL.



to large trials.

Section 3.4.1 reinforces that another factor in the proportion of patients allocated to treatment by the BBS design is treatment variance. While the BB_1 design randomizes patients irrespective of values of σ_A and σ_B, BBS randomization responds to changes in total variance. This behavior is easily anticipated by considering the denominator of the allocation proportion in both adaptive designs. In the BB design, T is constant. Hence, for a fixed difference in treatment effects, the randomization probability of BB_1 is constant. In the BBS design, S is the current estimate of pooled standard deviation. When either treatment variance grows, the pooled standard deviation grows. Hence, for a fixed difference in treatment effects, the randomization probability of the BBS shrinks. Depending on whether the pooled treatment standard deviation is smaller or larger than 1, the BBS design assigns more or fewer patients to the better treatment, respectively, than the BB_1 design.

Like many biased coin designs, both adaptive designs considered here attempt to strike a balance between complete randomization and another desirable condition with conflicting properties such as uniformity [13, 42, 6], optimal power or allocation variance [14, 25, 19], and patient performance via treatment failure minimization or maximizing assignment to the superior treatment [31, 1]. In particular, the BBS and BB designs are concerned with the latter: maintaining complete randomization in the face of uncertainty but assigning increasing proportions of patients to the better treatment based on the credence of superiority developed by accumulating trial data. The way both designs attempt to achieve this balance is by capitalizing on a ratio common throughout the history of statistics, the ratio of the difference in treatment effect estimated means—a proxy for superior treatment—to the variance in the observations—a proxy for certainty around the former measure of difference. The biggest difference between the two designs are the underlying assumptions about

how much or little is known about the treatment effects on the patient population. In the BB design, researchers are assumed to know the true treatment variances (and furthermore the variances are assumed to be identical for both treatments). Therefore, if the BB design can be applied, a constant value for T and moreover a fixed allocation proportion is reasonable and justified. If, however, those assumptions are not met and researchers are in a state of uncertainty regarding the true variance due to treatment effects—or if those treatment variances are not equal, then holding assignment probabilities constant is illogical and the BBS design is preferable.

The simulation results presented in Sections 3.4.1 and 3.5 confirm that total trial enrollment, the difference between treatment effects, and the treatment variances (via the pooled treatment standard deviation) determine the expected patient allocation proportions for the BBS design. Covariate slopes, including slopes that differ by treatment, have minimal impact on assignment probabilities. For a given positive difference in treatment effects, the BBS design provides increased patient exposure to a superior treatment with minimal loss of power compared to equal allocation. When there is no difference in treatment effects, the BBS design assigns asymptotically equal proportions of patients to each treatment with only a marginal increase in risk of type I error for specific parameter combinations. These behaviors are particularly desirable in cases where the BB_1 design is not an appropriate randomization alternative, for example when treatment variances (or covariate slopes) are unknown or unequal.

3.6.2 Rejection Rates

As with patient assignment, varying covariate effects within a reasonable range have little direct impact on rejection rates. Influential factors include total sample size, difference in treatment effects, and treatment variance. Equal allocation, BB_1, and BBS

all have similar power conditional on the three aforementioned considerations. Correct rejection rates increase as N increases and as total treatment variance decreases relative to a fixed difference in treatment effects.

In contrast, when treatment effects are equal, adaptive design behavior may differ somewhat from equal allocation. With equal allocation, Type I error rates are relatively stable at 0.05 regardless of treatment variance and with only minimal additional noise accompanying smaller trials and smaller treatment variance. For adaptive designs, false rejection rates also decrease as N increases and as total variance increases. In smaller trials, however, adaptive designs can incorporate nearly double the risk of a Type I error compared to equal allocation, particularly when treatment variance is small. On the other hand, this is only an increased risk of three or four percent to an error which is typically quite benign. Researchers and clinicians must be mindful of the increased risk of incorrectly concluding a difference in treatment effects exists and the tradeoff between a potentially better chance of detecting a true difference.

The simulation results presented in Sections 3.4.2 and 3.5 support that total trial enrollment, difference between treatment effects, and the treatment variances determine the power or risk of type I error of a BBS experiment. Covariate slopes, including slopes that differ by treatment, have minimal impact on rejection rates. For a given positive difference in treatment effects, the BBS design endures minimal power loss over that of equal allocation, while assigning more patients to the better treatment. With no difference in treatment effects, BBS assumes only a marginal increase in risk of type I error for specific parameter combinations. These behaviors are desirable when optimizing patient care while maintaining randomization in a controlled clinical trial.

3.6.3 Delay

As noted in Chapter 2, enrolling 244 patients such that each patient finishes a 6 month intervention and a treatment response is observed prior to randomizing the next patient would take over a century. While this may sound drastic, a clinical trial does not have to choose between giving up all adaptivity and acquiring egregious delays. Many response-adaptive randomization schemes can be revised so that estimates leveraged in assigning patients to treatment depend only on currently responded patients at the time of a new patient enrollment. In fact, this is a commonly recommended modification, being simultaneously practical and increasing ethicality compared to an unduly delayed clinical trial conclusion [2, 21, 22, 37, 48]. While the aforementioned articles prove that delay has minimal impact on the asymptotic characteristics of select adaptive designs—namely patient allocation proportion and power—convergence to these properties may slow. In particular, small to moderate trials may not behave similarly under delayed response conditions. As trial behavior may differ across combinations of sample size, adaptive randomization, and delay mechanisms, further study of the effect of delay on the BBS design is encouraged.

The simulations in Section 3.5 sidestep the issue of delay by assuming patient response is instantaneous so that patient responses can be immediately incorporated into the allocation algorithm. There are many applications, however, where a delay in patient responses would not pose an obstacle. For example, within healthcare, an adaptive design can easily be implemented when testing new cures for aggressive diseases such as ebola [10, 45] or a new vaccination to stem acute or airborne viral outbreaks such as influenza or SARS [38, 11, 39]. In these cases, there may be a delay in patient responses but the delay is at most a matter of days. Waiting an extra day to update the randomization ratio is worthwhile when considering the ben-

efit of adapting the randomization ratio to optimally treat severe cases or maximize outbreak containment while continuing to accumulate more data on longterm best intervention options. Incorporating BBS into the toolbox of resource-efficient, ethical randomization schemes in scenarios where the impact of delay is negligible or minimal relative to the risk of a complete randomization trial is also encouraged.

3.6.4 Recommendations

As discussed in Chapter 2, the primary goal of many Phase III clinical trials is to establish one treatment's superiority and hence determine the best standard of care for future patients. Following current standards, patients are exposed to treatments in equal proportions until superiority is concluded or until the trial is terminated. This holds true regardless of accumulating evidence suggesting the superiority of one treatment. As with [29], in many of these clinical scenarios each patient enrolled can significantly benefit from receiving the more effective intervention. Allocation proportions in these cases should not remain either constant or uniform, but instead should reflect the accumulating evidence tempered by continued equipoise.

The BBS design is one way to implement a more ethical randomization procedures with continuous patient outcomes. Results from Chapter 2 indicate that this design is recommended over the BB design at all values of T and over complete randomization in the face of unknown and unequal treatment variances, provided covariates have identical impacts on both treatments. Chapter 2 concludes with a recommendation to further investigate the potential influence of unknown covariates and covariates whose impact differs by treatment on the BBS design outcomes. That recommendation is the focus of Chapter 3 and the results discussed above.

Even when all patients have similar covariates, covariate-treatment interactions

are common [5, 9, 12, 40]. In fact, the existence of a sizable covariate-treatment interaction could mean that the sign differs between the difference in treatment effects and the difference in patient outcomes including covariate impact. For example, treatment A may have a larger treatment mean than treatment B ($\mu_A > \mu_B$), but if treatment A has a small covariate impact relative to that of treatment B then the outcome of treatment B including covariate impact would be larger than the outcome of treatment A including covariate impact ($\mu_B + \boldsymbol{\beta}_B \boldsymbol{z} > \mu_A + \boldsymbol{\beta}_A \boldsymbol{z}$). That is, looking at the treatment means alone, one might consider treatment A to be preferable while patients on treatment B would actually have overall better outcomes than patients on treatment A. Of course, if a relatively large covariate-treatment interaction is present, designs which focus on the difference in treatment effects should not be employed, regardless of whether the design is adaptive or static. That is, neither the BB nor the BBS would be suitable randomization candidates in scenarios where large covariate-treatment interactions exist as both designs highlight the difference in treatment effects rather than the difference in actual patient performance on a given treatment. Nevertheless, the simulation results described throughout Chapter 3 lend confidence to leveraging BBS when small to moderate covariate-treatment interactions exist. That is, when the benefit of a treatment effect is representative of or similar to the benefit in patient outcome on that treatment, the BBS design allows researchers to assign patients to treatment more ethically than complete randomization, without requiring equal, known treatment variances and uniform covariate effects across treatments.

The ethical advantage of this adaptive design is accompanied by an inflated risk of incorrectly concluding one treatment is superior. This risk and its impacts are discussed further in Chapter 2 and should be considered thoroughly by researchers prior to choosing any clinical trial design. Additionally, one other area of consideration

put forth in Chapter 2 bears mentioning: expanding the applicability of the BBS design to include scenarios with a delay in patient responses. As described in Section 3.6.3, many clinical scenarios have minimally delayed patient outcomes, so not all clinical trials require immediate responses. When such a testing situation arises, the BBS design should be strong candidate for ethical randomization. On the other hand, when the time between treatment and response observation is long—for example the HIV adherence intervention described in Sections 2.5 and 3.5—it is imperative to ascertain how the design fares under delayed response conditions prior to being able to recommend it as an integral part of a researcher's or clinician's design toolbox.

3.7 Appendix

Table 3.1: Proportion (SD) of patients assigned to Treatment A under equal allocation randomization with $\mu_A = \mu_B = 0$. Model variables σ_A, σ_B, β_A, β_B are given in the two leftmost columns. Allocation proportion and standard deviation across 1,000 simulations of trial size N given in remaining five columns.

(σ_A, σ_B)	(β_A, β_B)	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
$(0.5, 1.0)$	$(1, 2)$	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(0.5, 1.0)$	$(2, 2)$	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(0.5, 1.0)$	$(3, 2)$	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.0, 1.0)$	$(1, 2)$	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.0, 1.0)$	$(2, 2)$	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.0, 1.0)$	$(3, 2)$	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.5, 1.0)$	$(1, 2)$	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.5, 1.0)$	$(2, 2)$	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.5, 1.0)$	$(3, 2)$	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(0.5, 1.0)$	$(2, 1)$	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(0.5, 1.0)$	$(2, 2)$	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(0.5, 1.0)$	$(2, 3)$	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.0, 1.0)$	$(2, 1)$	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.0, 1.0)$	$(2, 2)$	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.0, 1.0)$	$(2, 3)$	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.5, 1.0)$	$(2, 1)$	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.5, 1.0)$	$(2, 2)$	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)

Continued on next page

Table 3.1 Allocation proportion under equal allocation with $\mu_A = \mu_B = 0$ – *Cont.*

(σ_A, σ_B)	(β_A, β_B)	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
$(1.5, 1.0)$	$(2, 3)$	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)

Table 3.2: Proportion (SD) of patients assigned to Treatment A under BB_1 randomization with $\mu_A = \mu_B = 0$. Model variables σ_A, σ_B, β_A, β_B are given in the two leftmost columns. Allocation proportion and standard deviation across 1,000 simulations of trial size N given in remaining five columns.

(σ_A, σ_B)	(β_A, β_B)	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
$(0.5, 1.0)$	$(1, 2)$	0.51 (0.01)	0.51 (0.01)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(0.5, 1.0)$	$(2, 2)$	0.50 (0.01)	0.51 (0.01)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(0.5, 1.0)$	$(3, 2)$	0.51 (0.01)	0.51 (0.01)	0.51 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.0, 1.0)$	$(1, 2)$	0.50 (0.01)	0.50 (0.02)	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)
$(1.0, 1.0)$	$(2, 2)$	0.50 (0.01)	0.50 (0.01)	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)
$(1.0, 1.0)$	$(3, 2)$	0.51 (0.01)	0.49 (0.02)	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)
$(1.5, 1.0)$	$(1, 2)$	0.50 (0.02)	0.50 (0.02)	0.49 (0.01)	0.50 (0.00)	0.50 (0.00)
$(1.5, 1.0)$	$(2, 2)$	0.50 (0.02)	0.48 (0.02)	0.49 (0.01)	0.50 (0.00)	0.50 (0.00)
$(1.5, 1.0)$	$(3, 2)$	0.50 (0.02)	0.49 (0.02)	0.49 (0.01)	0.50 (0.00)	0.50 (0.00)
$(0.5, 1.0)$	$(2, 1)$	0.51 (0.01)	0.51 (0.01)	0.51 (0.00)	0.50 (0.00)	0.50 (0.00)
$(0.5, 1.0)$	$(2, 2)$	0.50 (0.01)	0.51 (0.01)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(0.5, 1.0)$	$(2, 3)$	0.50 (0.01)	0.51 (0.01)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.0, 1.0)$	$(2, 1)$	0.51 (0.01)	0.51 (0.01)	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)
$(1.0, 1.0)$	$(2, 2)$	0.50 (0.01)	0.50 (0.01)	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)
$(1.0, 1.0)$	$(2, 3)$	0.50 (0.01)	0.50 (0.01)	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)

Continued on next page

Table 3.2 Allocation proportion under BB_1 with $\mu_A = \mu_B = 0$ – *Continued*

(σ_A, σ_B)	(β_A, β_B)	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
$(1.5, 1.0)$	$(2, 1)$	0.49 (0.02)	0.48 (0.02)	0.48 (0.01)	0.49 (0.00)	0.50 (0.00)
$(1.5, 1.0)$	$(2, 2)$	0.50 (0.02)	0.48 (0.02)	0.49 (0.01)	0.50 (0.00)	0.50 (0.00)
$(1.5, 1.0)$	$(2, 3)$	0.49 (0.02)	0.49 (0.02)	0.49 (0.01)	0.49 (0.00)	0.50 (0.00)

Table 3.3: Proportion (SD) of patients assigned to Treatment A under BBS randomization with $\mu_A = \mu_B = 0$. Model variables σ_A, σ_B, β_A, β_B are given in the two leftmost columns. Allocation proportion and standard deviation across 1,000 simulations of trial size N given in remaining five columns.

(σ_A, σ_B)	(β_A, β_B)	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
$(0.5, 1.0)$	$(1, 2)$	0.51 (0.01)	0.52 (0.02)	0.52 (0.01)	0.51 (0.00)	0.50 (0.00)
$(0.5, 1.0)$	$(2, 2)$	0.51 (0.01)	0.52 (0.02)	0.52 (0.01)	0.51 (0.00)	0.51 (0.00)
$(0.5, 1.0)$	$(3, 2)$	0.51 (0.02)	0.52 (0.02)	0.52 (0.01)	0.51 (0.00)	0.51 (0.00)
$(1.0, 1.0)$	$(1, 2)$	0.51 (0.01)	0.51 (0.01)	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)
$(1.0, 1.0)$	$(2, 2)$	0.50 (0.01)	0.50 (0.02)	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)
$(1.0, 1.0)$	$(3, 2)$	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)
$(1.5, 1.0)$	$(1, 2)$	0.50 (0.02)	0.49 (0.02)	0.49 (0.01)	0.49 (0.00)	0.50 (0.00)
$(1.5, 1.0)$	$(2, 2)$	0.49 (0.02)	0.49 (0.02)	0.49 (0.01)	0.50 (0.00)	0.50 (0.00)
$(1.5, 1.0)$	$(3, 2)$	0.49 (0.02)	0.49 (0.02)	0.49 (0.01)	0.49 (0.00)	0.50 (0.00)
$(0.5, 1.0)$	$(2, 1)$	0.51 (0.02)	0.52 (0.02)	0.52 (0.01)	0.51 (0.00)	0.51 (0.00)
$(0.5, 1.0)$	$(2, 2)$	0.51 (0.01)	0.52 (0.02)	0.52 (0.01)	0.51 (0.00)	0.51 (0.00)
$(0.5, 1.0)$	$(2, 3)$	0.51 (0.02)	0.52 (0.02)	0.52 (0.01)	0.51 (0.00)	0.50 (0.00)
$(1.0, 1.0)$	$(2, 1)$	0.49 (0.02)	0.49 (0.01)	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)

Continued on next page

Table 3.3 Allocation proportion under BBS with $\mu_A = \mu_B = 0$ – *Continued*

(σ_A, σ_B)	(β_A, β_B)	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
$(1.0, 1.0)$	$(2, 2)$	0.50 (0.01)	0.50 (0.02)	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)
$(1.0, 1.0)$	$(2, 3)$	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)
$(1.5, 1.0)$	$(2, 1)$	0.50 (0.01)	0.48 (0.02)	0.49 (0.01)	0.49 (0.00)	0.50 (0.00)
$(1.5, 1.0)$	$(2, 2)$	0.49 (0.02)	0.49 (0.02)	0.49 (0.01)	0.50 (0.00)	0.50 (0.00)
$(1.5, 1.0)$	$(2, 3)$	0.50 (0.02)	0.48 (0.02)	0.49 (0.01)	0.50 (0.00)	0.50 (0.00)

Table 3.4: Proportion (SD) of patients assigned to Treatment A under equal allocation randomization with $\mu_A = 1$ and $\mu_B = 0$. Model variables σ_A, σ_B, β_A, β_B are given in the two leftmost columns. Allocation proportion and standard deviation across 1,000 simulations of trial size N given in remaining five columns.

(σ_A, σ_B)	(β_A, β_B)	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
$(0.5, 1.0)$	$(1, 2)$	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(0.5, 1.0)$	$(2, 2)$	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(0.5, 1.0)$	$(3, 2)$	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.0, 1.0)$	$(1, 2)$	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.0, 1.0)$	$(2, 2)$	0.49 (0.01)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.0, 1.0)$	$(3, 2)$	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.5, 1.0)$	$(1, 2)$	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.5, 1.0)$	$(2, 2)$	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.5, 1.0)$	$(3, 2)$	0.51 (0.01)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(0.5, 1.0)$	$(2, 1)$	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(0.5, 1.0)$	$(2, 2)$	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)

Continued on next page

Table 3.4 Allocation proportion under equal allocation with $\mu_A = 1$, $\mu_B = 0$ – *Cont.*

(σ_A, σ_B)	(β_A, β_B)	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
$(0.5, 1.0)$	$(2, 3)$	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.0, 1.0)$	$(2, 1)$	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.0, 1.0)$	$(2, 2)$	0.49 (0.01)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.0, 1.0)$	$(2, 3)$	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.5, 1.0)$	$(2, 1)$	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.5, 1.0)$	$(2, 2)$	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)
$(1.5, 1.0)$	$(2, 3)$	0.50 (0.01)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)	0.50 (0.00)

Table 3.5: Proportion (SD) of patients assigned to Treatment A under BB_1 randomization with $\mu_A = 1$ and $\mu_B = 0$. Model variables σ_A, σ_B, β_A, β_B are given in the two leftmost columns. Allocation proportion and standard deviation across 1,000 simulations of trial size N given in remaining five columns.

(σ_A, σ_B)	(β_A, β_B)	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
$(0.5, 1.0)$	$(1, 2)$	0.67 (0.01)	0.76 (0.01)	0.83 (0)	0.84 (0)	0.84 (0)
$(0.5, 1.0)$	$(2, 2)$	0.67 (0.01)	0.76 (0.01)	0.83 (0)	0.84 (0)	0.84 (0)
$(0.5, 1.0)$	$(3, 2)$	0.67 (0.01)	0.76 (0.01)	0.83 (0)	0.84 (0)	0.84 (0)
$(1.0, 1.0)$	$(1, 2)$	0.66 (0.01)	0.75 (0.01)	0.83 (0)	0.83 (0)	0.84 (0)
$(1.0, 1.0)$	$(2, 2)$	0.67 (0.01)	0.75 (0.01)	0.83 (0)	0.83 (0)	0.84 (0)
$(1.0, 1.0)$	$(3, 2)$	0.67 (0.01)	0.75 (0.01)	0.83 (0)	0.84 (0)	0.84 (0)
$(1.5, 1.0)$	$(1, 2)$	0.65 (0.01)	0.74 (0.01)	0.82 (0)	0.83 (0)	0.84 (0)
$(1.5, 1.0)$	$(2, 2)$	0.66 (0.01)	0.74 (0.01)	0.82 (0)	0.83 (0)	0.84 (0)
$(1.5, 1.0)$	$(3, 2)$	0.66 (0.01)	0.74 (0.01)	0.82 (0)	0.83 (0)	0.84 (0)

Continued on next page

Table 3.5 Allocation proportion under BB_1 with $\mu_A = 1$, $\mu_B = 0$ – *Continued*

(σ_A, σ_B)	(β_A, β_B)	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
$(0.5, 1.0)$	$(2, 1)$	0.67 (0.01)	0.76 (0.01)	0.83 (0)	0.84 (0)	0.84 (0)
$(0.5, 1.0)$	$(2, 2)$	0.67 (0.01)	0.76 (0.01)	0.83 (0)	0.84 (0)	0.84 (0)
$(0.5, 1.0)$	$(2, 3)$	0.67 (0.01)	0.76 (0.01)	0.83 (0)	0.84 (0)	0.84 (0)
$(1.0, 1.0)$	$(2, 1)$	0.67 (0.01)	0.75 (0.01)	0.83 (0)	0.84 (0)	0.84 (0)
$(1.0, 1.0)$	$(2, 2)$	0.67 (0.01)	0.75 (0.01)	0.83 (0)	0.83 (0)	0.84 (0)
$(1.0, 1.0)$	$(2, 3)$	0.67 (0.01)	0.75 (0.01)	0.83 (0)	0.84 (0)	0.84 (0)
$(1.5, 1.0)$	$(2, 1)$	0.66 (0.01)	0.75 (0.01)	0.82 (0)	0.83 (0)	0.84 (0)
$(1.5, 1.0)$	$(2, 2)$	0.66 (0.01)	0.74 (0.01)	0.82 (0)	0.83 (0)	0.84 (0)
$(1.5, 1.0)$	$(2, 3)$	0.65 (0.01)	0.74 (0.01)	0.82 (0)	0.83 (0)	0.84 (0)

Table 3.6: Proportion (SD) of patients assigned to Treatment A under BBS randomization with $\mu_A = 1$ and $\mu_B = 0$. Model variables σ_A, σ_B, β_A, β_B are given in the two leftmost columns. Allocation proportion and standard deviation across 1,000 simulations of trial size N given in remaining five columns.

(σ_A, σ_B)	(β_A, β_B)	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
$(0.5, 1.0)$	$(1, 2)$	0.71 (0.01)	0.82 (0.01)	0.93 (0.00)	0.95 (0.00)	0.97 (0.00)
$(0.5, 1.0)$	$(2, 2)$	0.70 (0.01)	0.83 (0.00)	0.93 (0.00)	0.95 (0.00)	0.97 (0.00)
$(0.5, 1.0)$	$(3, 2)$	0.71 (0.01)	0.81 (0.01)	0.93 (0.00)	0.95 (0.00)	0.97 (0.00)
$(1.0, 1.0)$	$(1, 2)$	0.67 (0.01)	0.75 (0.01)	0.83 (0.00)	0.84 (0.00)	0.84 (0.00)
$(1.0, 1.0)$	$(2, 2)$	0.68 (0.01)	0.76 (0.01)	0.83 (0.00)	0.84 (0.00)	0.84 (0.00)
$(1.0, 1.0)$	$(3, 2)$	0.67 (0.01)	0.76 (0.01)	0.83 (0.00)	0.84 (0.00)	0.84 (0.00)
$(1.5, 1.0)$	$(1, 2)$	0.64 (0.01)	0.70 (0.01)	0.75 (0.00)	0.76 (0.00)	0.76 (0.00)
$(1.5, 1.0)$	$(2, 2)$	0.64 (0.01)	0.70 (0.01)	0.75 (0.00)	0.76 (0.00)	0.76 (0.00)

Continued on next page

Table 3.6 Allocation proportion under BBS with $\mu_A = 1$, $\mu_B = 0$ – *Continued*

(σ_A, σ_B)	(β_A, β_B)	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
$(1.5, 1.0)$	$(3, 2)$	0.63 (0.01)	0.70 (0.01)	0.75 (0.00)	0.76 (0.00)	0.76 (0.00)
$(0.5, 1.0)$	$(2, 1)$	0.71 (0.01)	0.82 (0.01)	0.93 (0.00)	0.95 (0.00)	0.97 (0.00)
$(0.5, 1.0)$	$(2, 2)$	0.70 (0.01)	0.83 (0.00)	0.93 (0.00)	0.95 (0.00)	0.97 (0.00)
$(0.5, 1.0)$	$(2, 3)$	0.71 (0.01)	0.82 (0.01)	0.93 (0.00)	0.95 (0.00)	0.97 (0.00)
$(1.0, 1.0)$	$(2, 1)$	0.67 (0.01)	0.75 (0.01)	0.83 (0.00)	0.84 (0.00)	0.84 (0.00)
$(1.0, 1.0)$	$(2, 2)$	0.68 (0.01)	0.76 (0.01)	0.83 (0.00)	0.84 (0.00)	0.84 (0.00)
$(1.0, 1.0)$	$(2, 3)$	0.67 (0.01)	0.76 (0.01)	0.83 (0.00)	0.84 (0.00)	0.84 (0.00)
$(1.5, 1.0)$	$(2, 1)$	0.63 (0.01)	0.70 (0.01)	0.75 (0.00)	0.76 (0.00)	0.76 (0.00)
$(1.5, 1.0)$	$(2, 2)$	0.64 (0.01)	0.70 (0.01)	0.75 (0.00)	0.76 (0.00)	0.76 (0.00)
$(1.5, 1.0)$	$(2, 3)$	0.64 (0.01)	0.70 (0.01)	0.75 (0.00)	0.76 (0.00)	0.76 (0.00)

Table 3.7: Proportion (SD) of trials where the null hypothesis is incorrectly rejected under equal allocation randomization with $\mu_A = \mu_B = 0$. Model variables σ_A, σ_B, β_A, β_B are given in the two leftmost columns. Rejection rate and standard deviation across 1,000 simulations of trial size N given in remaining five columns.

(σ_A, σ_B)	(β_A, β_B)	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
$(0.5, 1.0)$	$(1, 2)$	0.06 (0.06)	0.06 (0.06)	0.07 (0.07)	0.06 (0.06)	0.07 (0.07)
$(0.5, 1.0)$	$(2, 2)$	0.06 (0.06)	0.07 (0.07)	0.07 (0.06)	0.05 (0.05)	0.06 (0.06)
$(0.5, 1.0)$	$(3, 2)$	0.07 (0.06)	0.07 (0.06)	0.06 (0.06)	0.05 (0.05)	0.05 (0.04)
$(1.0, 1.0)$	$(1, 2)$	0.06 (0.06)	0.05 (0.05)	0.05 (0.05)	0.06 (0.06)	0.05 (0.05)
$(1.0, 1.0)$	$(2, 2)$	0.07 (0.06)	0.06 (0.06)	0.05 (0.05)	0.06 (0.06)	0.04 (0.04)
$(1.0, 1.0)$	$(3, 2)$	0.06 (0.06)	0.06 (0.06)	0.04 (0.04)	0.05 (0.05)	0.05 (0.05)

Continued on next page

Table 3.7 Rejection rate under equal allocation with $\mu_A = \mu_B = 0$ – *Continued*

(σ_A, σ_B)	(β_A, β_B)	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
$(1.5, 1.0)$	$(1, 2)$	0.06 (0.06)	0.06 (0.05)	0.04 (0.04)	0.06 (0.05)	0.05 (0.05)
$(1.5, 1.0)$	$(2, 2)$	0.06 (0.06)	0.07 (0.06)	0.05 (0.05)	0.04 (0.04)	0.05 (0.05)
$(1.5, 1.0)$	$(3, 2)$	0.07 (0.06)	0.05 (0.05)	0.06 (0.05)	0.06 (0.06)	0.07 (0.06)
$(0.5, 1.0)$	$(2, 1)$	0.07 (0.07)	0.07 (0.06)	0.06 (0.06)	0.06 (0.06)	0.06 (0.06)
$(0.5, 1.0)$	$(2, 2)$	0.06 (0.06)	0.07 (0.07)	0.07 (0.06)	0.05 (0.05)	0.06 (0.06)
$(0.5, 1.0)$	$(2, 3)$	0.08 (0.07)	0.07 (0.07)	0.06 (0.05)	0.06 (0.06)	0.06 (0.05)
$(1.0, 1.0)$	$(2, 1)$	0.05 (0.04)	0.06 (0.06)	0.05 (0.05)	0.06 (0.05)	0.05 (0.05)
$(1.0, 1.0)$	$(2, 2)$	0.07 (0.06)	0.06 (0.06)	0.05 (0.05)	0.06 (0.06)	0.04 (0.04)
$(1.0, 1.0)$	$(2, 3)$	0.06 (0.05)	0.05 (0.04)	0.06 (0.06)	0.05 (0.05)	0.05 (0.04)
$(1.5, 1.0)$	$(2, 1)$	0.05 (0.05)	0.07 (0.06)	0.05 (0.05)	0.06 (0.06)	0.06 (0.06)
$(1.5, 1.0)$	$(2, 2)$	0.06 (0.06)	0.07 (0.06)	0.05 (0.05)	0.04 (0.04)	0.05 (0.05)
$(1.5, 1.0)$	$(2, 3)$	0.06 (0.06)	0.06 (0.05)	0.06 (0.05)	0.05 (0.05)	0.04 (0.04)

Table 3.8: Proportion (SD) of trials where the null hypothesis is incorrectly rejected under BB_1 randomization with $\mu_A = \mu_B = 0$. Model variables σ_A, σ_B, β_A, β_B are given in the two leftmost columns. Rejection rate and standard deviation across 1,000 simulations of trial size N given in remaining five columns.

(σ_A, σ_B)	(β_A, β_B)	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
$(0.5, 1.0)$	$(1, 2)$	0.10 (0.09)	0.11 (0.10)	0.07 (0.06)	0.08 (0.07)	0.06 (0.06)
$(0.5, 1.0)$	$(2, 2)$	0.10 (0.09)	0.09 (0.08)	0.08 (0.07)	0.07 (0.07)	0.05 (0.05)
$(0.5, 1.0)$	$(3, 2)$	0.10 (0.09)	0.12 (0.10)	0.07 (0.06)	0.06 (0.06)	0.05 (0.04)
$(1.0, 1.0)$	$(1, 2)$	0.07 (0.06)	0.06 (0.05)	0.06 (0.06)	0.05 (0.05)	0.05 (0.05)

Continued on next page

Table 3.8 Rejection rate under BB_1 with $\mu_A = \mu_B = 0$ – *Continued*

(σ_A, σ_B)	(β_A, β_B)	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
$(1.0, 1.0)$	$(2, 2)$	0.07 (0.06)	0.07 (0.07)	0.06 (0.06)	0.06 (0.05)	0.03 (0.03)
$(1.0, 1.0)$	$(3, 2)$	0.07 (0.07)	0.06 (0.06)	0.06 (0.05)	0.05 (0.05)	0.05 (0.05)
$(1.5, 1.0)$	$(1, 2)$	0.05 (0.05)	0.05 (0.04)	0.04 (0.04)	0.05 (0.05)	0.05 (0.04)
$(1.5, 1.0)$	$(2, 2)$	0.05 (0.05)	0.03 (0.03)	0.04 (0.04)	0.04 (0.04)	0.06 (0.05)
$(1.5, 1.0)$	$(3, 2)$	0.04 (0.04)	0.04 (0.04)	0.05 (0.05)	0.06 (0.05)	0.04 (0.04)
$(0.5, 1.0)$	$(2, 1)$	0.11 (0.10)	0.09 (0.08)	0.08 (0.07)	0.07 (0.06)	0.07 (0.06)
$(0.5, 1.0)$	$(2, 2)$	0.10 (0.09)	0.09 (0.08)	0.08 (0.07)	0.07 (0.07)	0.05 (0.05)
$(0.5, 1.0)$	$(2, 3)$	0.10 (0.09)	0.09 (0.08)	0.07 (0.07)	0.07 (0.07)	0.05 (0.05)
$(1.0, 1.0)$	$(2, 1)$	0.08 (0.07)	0.08 (0.07)	0.05 (0.04)	0.06 (0.06)	0.04 (0.04)
$(1.0, 1.0)$	$(2, 2)$	0.07 (0.06)	0.07 (0.07)	0.06 (0.06)	0.06 (0.05)	0.03 (0.03)
$(1.0, 1.0)$	$(2, 3)$	0.07 (0.06)	0.08 (0.08)	0.04 (0.04)	0.05 (0.05)	0.06 (0.05)
$(1.5, 1.0)$	$(2, 1)$	0.06 (0.06)	0.05 (0.04)	0.03 (0.03)	0.04 (0.04)	0.04 (0.04)
$(1.5, 1.0)$	$(2, 2)$	0.05 (0.05)	0.03 (0.03)	0.04 (0.04)	0.04 (0.04)	0.06 (0.05)
$(1.5, 1.0)$	$(2, 3)$	0.05 (0.05)	0.04 (0.04)	0.05 (0.04)	0.05 (0.04)	0.04 (0.04)

Table 3.9: Proportion (SD) of trials where the null hypothesis is incorrectly rejected under BBS randomization with $\mu_A = \mu_B = 0$. Model variables σ_A, σ_B, β_A, β_B are given in the two leftmost columns. Rejection rate and standard deviation across 1,000 simulations of trial size N given in remaining five columns.

(σ_A, σ_B)	(β_A, β_B)	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
$(0.5, 1.0)$	$(1, 2)$	0.09 (0.08)	0.11 (0.10)	0.08 (0.07)	0.06 (0.06)	0.07 (0.06)
$(0.5, 1.0)$	$(2, 2)$	0.10 (0.09)	0.10 (0.09)	0.09 (0.08)	0.06 (0.06)	0.07 (0.06)

Continued on next page

Table 3.9 Rejection rate under BBS with $\mu_A = \mu_B = 0$ – *Continued*

(σ_A, σ_B)	(β_A, β_B)	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
$(0.5, 1.0)$	$(3, 2)$	0.11 (0.10)	0.11 (0.10)	0.10 (0.09)	0.07 (0.06)	0.06 (0.06)
$(1.0, 1.0)$	$(1, 2)$	0.07 (0.07)	0.08 (0.07)	0.04 (0.04)	0.06 (0.06)	0.06 (0.05)
$(1.0, 1.0)$	$(2, 2)$	0.07 (0.07)	0.06 (0.05)	0.06 (0.06)	0.05 (0.05)	0.05 (0.05)
$(1.0, 1.0)$	$(3, 2)$	0.06 (0.05)	0.08 (0.07)	0.06 (0.05)	0.06 (0.06)	0.04 (0.04)
$(1.5, 1.0)$	$(1, 2)$	0.07 (0.06)	0.06 (0.06)	0.04 (0.04)	0.06 (0.05)	0.05 (0.04)
$(1.5, 1.0)$	$(2, 2)$	0.06 (0.05)	0.06 (0.05)	0.05 (0.05)	0.05 (0.05)	0.05 (0.04)
$(1.5, 1.0)$	$(3, 2)$	0.07 (0.06)	0.06 (0.05)	0.06 (0.06)	0.05 (0.05)	0.06 (0.05)
$(0.5, 1.0)$	$(2, 1)$	0.11 (0.10)	0.11 (0.10)	0.07 (0.07)	0.09 (0.08)	0.08 (0.07)
$(0.5, 1.0)$	$(2, 2)$	0.10 (0.09)	0.10 (0.09)	0.09 (0.08)	0.06 (0.06)	0.07 (0.06)
$(0.5, 1.0)$	$(2, 3)$	0.10 (0.09)	0.11 (0.09)	0.09 (0.08)	0.08 (0.07)	0.05 (0.05)
$(1.0, 1.0)$	$(2, 1)$	0.07 (0.06)	0.04 (0.04)	0.05 (0.04)	0.05 (0.05)	0.06 (0.06)
$(1.0, 1.0)$	$(2, 2)$	0.07 (0.07)	0.06 (0.05)	0.06 (0.06)	0.05 (0.05)	0.05 (0.05)
$(1.0, 1.0)$	$(2, 3)$	0.07 (0.07)	0.06 (0.06)	0.05 (0.05)	0.05 (0.05)	0.05 (0.05)
$(1.5, 1.0)$	$(2, 1)$	0.06 (0.06)	0.06 (0.06)	0.05 (0.05)	0.05 (0.05)	0.05 (0.05)
$(1.5, 1.0)$	$(2, 2)$	0.06 (0.05)	0.06 (0.05)	0.05 (0.05)	0.05 (0.05)	0.05 (0.04)
$(1.5, 1.0)$	$(2, 3)$	0.06 (0.06)	0.05 (0.05)	0.04 (0.03)	0.06 (0.05)	0.05 (0.05)

Table 3.10: Proportion (SD) of trials where the null hypothesis is correctly rejected under equal allocation randomization with $\mu_A = 1$, $\mu_B = 0$. Model variables σ_A, σ_B, β_A, β_B are given in the two leftmost columns. Rejection rate and standard deviation across 1,000 simulations of trial size N given in remaining five columns.

(σ_A, σ_B)	(β_A, β_B)	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
$(0.5, 1.0)$	$(1, 2)$	0.93 (0.06)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(0.5, 1.0)$	$(2, 2)$	0.92 (0.08)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(0.5, 1.0)$	$(3, 2)$	0.93 (0.06)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.0, 1.0)$	$(1, 2)$	0.80 (0.16)	0.96 (0.04)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.0, 1.0)$	$(2, 2)$	0.77 (0.18)	0.96 (0.04)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.0, 1.0)$	$(3, 2)$	0.78 (0.17)	0.97 (0.03)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.5, 1.0)$	$(1, 2)$	0.63 (0.23)	0.88 (0.11)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.5, 1.0)$	$(2, 2)$	0.64 (0.23)	0.87 (0.12)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.5, 1.0)$	$(3, 2)$	0.63 (0.23)	0.87 (0.11)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(0.5, 1.0)$	$(2, 1)$	0.92 (0.07)	0.99 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(0.5, 1.0)$	$(2, 2)$	0.92 (0.08)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(0.5, 1.0)$	$(2, 3)$	0.93 (0.07)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.0, 1.0)$	$(2, 1)$	0.76 (0.18)	0.97 (0.03)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.0, 1.0)$	$(2, 2)$	0.77 (0.18)	0.96 (0.04)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.0, 1.0)$	$(2, 3)$	0.78 (0.17)	0.97 (0.03)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.5, 1.0)$	$(2, 1)$	0.63 (0.23)	0.89 (0.09)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.5, 1.0)$	$(2, 2)$	0.64 (0.23)	0.87 (0.12)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.5, 1.0)$	$(2, 3)$	0.64 (0.23)	0.84 (0.13)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)

Table 3.11: Proportion (SD) of trials where the null hypothesis is correctly rejected under BB_1 randomization with $\mu_A = 1$, $\mu_B = 0$. Model variables σ_A, σ_B, β_A, β_B are given in the two leftmost columns. Rejection rate and standard deviation across 1,000 simulations of trial size N given in remaining five columns.

(σ_A, σ_B)	(β_A, β_B)	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
$(0.5, 1.0)$	$(1, 2)$	0.91 (0.08)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(0.5, 1.0)$	$(2, 2)$	0.91 (0.08)	0.99 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(0.5, 1.0)$	$(3, 2)$	0.91 (0.08)	0.99 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.0, 1.0)$	$(1, 2)$	0.78 (0.17)	0.96 (0.04)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.0, 1.0)$	$(2, 2)$	0.80 (0.16)	0.95 (0.05)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.0, 1.0)$	$(3, 2)$	0.78 (0.17)	0.96 (0.04)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.5, 1.0)$	$(1, 2)$	0.56 (0.25)	0.82 (0.15)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.5, 1.0)$	$(2, 2)$	0.55 (0.25)	0.81 (0.15)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.5, 1.0)$	$(3, 2)$	0.53 (0.25)	0.81 (0.15)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(0.5, 1.0)$	$(2, 1)$	0.91 (0.08)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(0.5, 1.0)$	$(2, 2)$	0.91 (0.08)	0.99 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(0.5, 1.0)$	$(2, 3)$	0.91 (0.08)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.0, 1.0)$	$(2, 1)$	0.78 (0.17)	0.95 (0.05)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.0, 1.0)$	$(2, 2)$	0.80 (0.16)	0.95 (0.05)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.0, 1.0)$	$(2, 3)$	0.79 (0.17)	0.97 (0.03)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.5, 1.0)$	$(2, 1)$	0.56 (0.25)	0.84 (0.14)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.5, 1.0)$	$(2, 2)$	0.55 (0.25)	0.81 (0.15)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.5, 1.0)$	$(2, 3)$	0.54 (0.25)	0.83 (0.14)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)

Table 3.12: Proportion (SD) of trials where the null hypothesis is correctly rejected under BBS randomization with $\mu_A = 1$, $\mu_B = 0$. Model variables σ_A, σ_B, β_A, β_B are given in the two leftmost columns. Rejection rate and standard deviation across 1,000 simulations of trial size N given in remaining five columns.

(σ_A, σ_B)	(β_A, β_B)	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
$(0.5, 1.0)$	$(1, 2)$	0.91 (0.08)	0.98 (0.02)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(0.5, 1.0)$	$(2, 2)$	0.90 (0.09)	0.99 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(0.5, 1.0)$	$(3, 2)$	0.91 (0.08)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.0, 1.0)$	$(1, 2)$	0.77 (0.18)	0.94 (0.06)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.0, 1.0)$	$(2, 2)$	0.76 (0.18)	0.96 (0.04)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.0, 1.0)$	$(3, 2)$	0.77 (0.18)	0.96 (0.04)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.5, 1.0)$	$(1, 2)$	0.66 (0.22)	0.87 (0.12)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.5, 1.0)$	$(2, 2)$	0.63 (0.23)	0.88 (0.11)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.5, 1.0)$	$(3, 2)$	0.64 (0.23)	0.88 (0.11)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(0.5, 1.0)$	$(2, 1)$	0.88 (0.10)	0.99 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(0.5, 1.0)$	$(2, 2)$	0.90 (0.09)	0.99 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(0.5, 1.0)$	$(2, 3)$	0.89 (0.10)	0.99 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.0, 1.0)$	$(2, 1)$	0.75 (0.19)	0.95 (0.05)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.0, 1.0)$	$(2, 2)$	0.76 (0.18)	0.96 (0.04)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.0, 1.0)$	$(2, 3)$	0.77 (0.18)	0.96 (0.04)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.5, 1.0)$	$(2, 1)$	0.62 (0.24)	0.87 (0.11)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.5, 1.0)$	$(2, 2)$	0.63 (0.23)	0.88 (0.11)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$(1.5, 1.0)$	$(2, 3)$	0.66 (0.22)	0.88 (0.10)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)

Table 3.13: HIV adherence intervention summary characteristics by design across covariate slopes (β_C, β_M) that may differ by treatment. $\mu_C - \mu_M$ is the difference in treatment effects (measured in HIV RNA copies/mL) for a particular combination of covariate slopes, provided for reference. # is the number of additional patients assigned to the superior treatment over equal allocation. % is the percent of additional patients assigned to the superior treatment over equal allocation. P_C is the proportion of patients assigned to treatment C, combined medication, education, and counseling therapy. Not shown is $P_M = 1 - P_C$, the proportion of patients assigned to treatment M, medication alone.

Equal Allocation						
(β_C, β_M)	$\mu_C - \mu_M$	N_C	#	%	P_C (SD)	Power (SD)
$(1.2, 1.0)$	0.24	122	0	0%	0.50 (0.03)	0.75 (0.43)
$(1.1, 1.1)$	0.36	122	0	0%	0.50 (0.03)	0.97 (0.18)
$(1.0, 1.2)$	0.48	122	0	0%	0.50 (0.03)	1.00 (0.05)
BB Design						
(β_C, β_M)	$\mu_C - \mu_M$	N_C	#	%	P_C (SD)	Power (SD)
$(1.2, 1.0)$	0.24	143	21	17%	0.59 (0.06)	0.75 (0.43)
$(1.1, 1.1)$	0.36	153	31	26%	0.63 (0.06)	0.96 (0.19)
$(1.0, 1.2)$	0.48	163	41	33%	0.67 (0.06)	1.00 (0.05)
BBS Design						
(β_C, β_M)	$\mu_C - \mu_M$	N_C	#	%	P_C (SD)	Power (SD)
$(1.2, 1.0)$	0.24	169	47	38%	0.69 (0.12)	0.74 (0.44)
$(1.1, 1.1)$	0.36	182	60	50%	0.75 (0.10)	0.95 (0.21)
$(1.0, 1.2)$	0.48	192	70	57%	0.79 (0.08)	1.00 (0.06)

3.8 References

[1] Atkinson, A.C. and Biswas, A. "Adaptive biased-coin designs for skewing the allocation proportion in clinical trials with normal responses". In: *Statistics in Medicine* 24.16 (2005), pp. 2477–2492.

[2] Bai, Z.D., Hu, F., and Rosenberger, W.F. "Asymptotic properties of adaptive designs for clinical trials with delayed response". In: *The Annals of Statistics* 30.1 (2002), pp. 122–139.

[3] Baldi Antognini, A. "A theoretical analysis of the power of biased coin designs". In: *Journal of Statistical Planning and Inference* 138.6 (2008), pp. 1792–1798.

[4] Bandyopadhyay, Uttam and Biswas, A. "Adaptive designs for normal response with prognostic factors". In: *Biometrika* 88 (2001), pp. 409–419.

[5] Byar, D.P. "Assessing Apparent Treatment-Covariate Interactions in Randomized Clinical Trials". In: *Statistics in Medicine* 4.3 (1985), pp. 255–263.

[6] Chen, Y.P. "Biased coin design with imbalance tolerance". In: *Communications in Statistics. Stochastic Models* 15.5 (1999), pp. 953–975.

[7] M. A. Chesney, M. Morin, and L. Sherr. "Adherence to HIV combination therapy". In: *Soc Sci Med* 50.11 (2000), pp. 1599–1605.

[8] Cornfield, Jerome, Halperin, Max, and Greenhouse, Samuel W. "An Adaptive Procedure for Sequential Clinical Trials". In: *Journal of the American Statistical Association* 64.327 (1969), pp. 759–770.

[9] Crowley, J. and Pauler, D., eds. *Handbook of Statistics in Clinical Oncology*. 2nd ed. Chapman & Hall/CRC Biostatistics Series, 2006.

[10] Centers for Disease Control. *Ebola (Ebola Virus Disease)*. 2014. URL: `http://www.cdc.gov/vhf/ebola/treatment/index.html`.

[11] Centers for Disease Control and Prevention. *The 2009 H1N1 Pandemic: Summary Highlights, April 2009-April 2010*. 2010. URL: `http://www.cdc.gov/h1n1flu/cdcresponse.htm`.

[12] Dennis O. Dixon. "Treatment-Covariate Interaction". In: *Encyclopedia of Biostatistics*. Ed. by Armitage, P. and Colton, T. 2nd ed. Vol. 8. Wiley & Sons, 2005, pp. 5499–5502.

[13] Efron, B. "Forcing a sequential experiment to be balanced". In: *Biometrika* 58.3 (1971), pp. 403–417.

[14] Eisele, J. R. "The doubly adaptive biased coin design for sequential clinical trials". In: *Journal of Statistical Planning and Inference* 38.2 (1994), pp. 249–261.

[15] Food and Drug Administration. *Critical Path Initiative: Transforming the way FDA-regulated products are developed, evaluated, and manufactured*. Report on Key Achievements in 2009. 2010.

[16] Food and Drug Administration. "Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics". Draft Document. 2010.

[17] Hinkelmann, K. *Design and Analysis of Experiments, Vol. 3: Special Designs and Applications*. 1st ed. Vol. 3. Wiley-Blackwell, 2011.

[18] Joint United Nations Programme on HIV and AIDS (UNAIDS). *Fact sheet*. URL: `http://www.unaids.org/en/resources/campaigns/globalreport2013/factsheet` (visited on 02/16/2015).

[19] Hu, F. and Rosenberger, W.F. "Optimality, Variability, Power: Evaluating Response-Adaptive Randomization Procedures for Treatment Comparisons". In: *Journal of the American Statistical Association* 98.463 (2003), pp. 671–678.

[20] Hu, F. and Rosenberger, W.F. *The Theory of Response-Adaptive Randomization in Clinical Trials*. 1st ed. Hoboken, NJ: John Wiley & Sons, Inc., 2006.

[21] Hu, F. and Zhang, L.-X. "Asymptotic normality of urn models for clinical trials with delayed response". In: *Bernoulli* 10.3 (2004), pp. 447–463.

[22] Hu, F. et al. "Doubly-Adaptive Biased Coin Designs with Delayed Responses". In: *Canadian Journal of Statistics* 36.4 (2008), pp. 541–559.

[23] H. Knobel et al. "Validation of a simplified medication adherence questionnaire in a large cohort of HIV-infected patients: the GEEMA Study". In: *AIDS* 16.4 (2002), pp. 605–613.

[24] K. Lamiraud et al. "Adherence to and effectiveness of highly active antiretroviral treatment for HIV infection: assessing the bidirectional relationship". In: *Med Care* 50.5 (2012), pp. 410–418.

[25] Melfi, V. F., Page, C., and Geraldes, M. "An Adaptive Randomized Design with Application to Estimation". In: *Canadian Journal of Statistics* 29.1 (2001), pp. 107–116.

[26] Palmer, C.R. "Ethics, data-dependent designs, and the strategy of clinical trials: time to start learning-as-we-go?" In: *Statistical Methods in Medical Research* 11 (2002), pp. 381–402.

[27] Pocock, S.J. "Allocation of Patients to Treatment in Clinical Trials". In: *Biometrics* 35.1 (1979), pp. 183–197.

[28] Pong, A. and Chow, S.C., eds. *Handbook of Adaptive Designs in Pharmaceutical and Clinical Development*. Boca Raton: CRC Press, 2011.

[29] C. Pradier et al. "Efficacy of an educational and counseling intervention on adherence to highly active antiretroviral therapy: French prospective controlled study". In: *HIV Clin Trials* 4.2 (2003), pp. 121–131.

[30] Rosenberger, W.F. "New Directions in Adaptive Designs". In: *Statistical Science* 11 (1996), pp. 137–149.

[31] Rosenberger, W.F. and Hu, F. "Maximizing power and minimizing treatment failures in clinical trials". In: *Clinical Trials* 1.2 (2004), pp. 141–147.

[32] Rosenberger, W.F. and Lachin, J. M. "The use of response-adaptive designs in clinical trials". In: *Controlled Clinical Trials* 14.6 (1993), pp. 471–484.

[33] Royall, R.M. "Ethics and statistics in randomized clinical trials". In: *Statistical Science* 6.1 (1991), pp. 52–62.

[34] SAS Institute Inc. *SAS/IML Software, Version 9.2*. Cary, NC, 2011. URL: http://www.sas.com/.

[35] F. E. Satterthwaite. "An Approximate Distribution of Estimates of Variance Components". In: *Biometrics Bulletin* 2.6 (Dec. 1946), pp. 110–114. ISSN: 00994987. DOI: 10.2307/3002019. URL: http://dx.doi.org/10.2307/3002019.

[36] Simon, Richard. "Adaptive Treatment Assignment Methods and Clinical Trials". In: *Biometrics* 33.4 (1977), pp. 743–749.

[37] Sun, R., Cheung, S.H., and Zhang, L.-X. "A generalized drop-the-loser urn for multi-treatment clinical trials". In: *Journal of Statistical Planning and Inference* 137.6 (2007), pp. 2011–2023.

[38] The College of Physicians of Philadelphia. *Vaccines for Pandemic Threats*. 2014. URL: http://www.historyofvaccines.org/content/articles/vaccines-pandemic-threats.

[39] U.S. Department of Health and Human Services. *Pandemic Flu History*. 2009. URL: http://www.flu.gov/pandemic/history/.

[40] Varadhan, R. et al. "A framework for the analysis of heterogeneity of treatment effect in patient-centered outcomes research". In: *Journal of Clinical Epidemiology* 66.8 (2013), pp. 818–825.

[41] Viswanathan, S. et al. "Level of Adherence and HIV RNA Suppression in the Current Era of Highly Active Antiretroviral Therapy (HAART)". In: *AIDS Behav* (2014).

[42] Wei, L.J. "The Adaptive Biased Coin Design for Sequential Experiments". In: *The Annals of Statistics* 6.1 (1978), pp. 92–100.

[43] Weinstein, M. C. "Allocation of subjects in medical experiments". In: *New England Journal of Medicine* 291.24 (1974), pp. 1278–1285.

[44] B. L. Welch. "The Generalization of 'Student's' Problem when Several Different Population Variances are Involved". In: *Biometrika* 34.1/2 (1947), pp. 28–35. ISSN: 00063444. DOI: 10.2307/2332510. URL: http://dx.doi.org/10.2307/2332510.

[45] World Health Organization (WHO). *Ebola virus disease: Fact Sheet Number 103*. 2015. URL: http://www.who.int/mediacentre/factsheets/fs103/en/ (visited on 4–2015).

[46] World Health Organization (WHO). *WHO Global Health Observatory: HIV/AIDS*. URL: http://www.who.int/gho/hiv/en/ (visited on 02/16/2015).

[47] Zelen, M. "The randomization and stratification of patients to clinical trials". In: *Journal of Chronic Diseases* 27.7-8 (1974), pp. 365–375.

[48] Zhang, L.-X. et al. "A generalized drop-the-loser urn for clinical trials with delayed responses". In: *Statistica Sinica* 17 (2007), pp. 387–409.

[49] Zhao, Y. and Zeng, D. “Recent development on statistical methods for personalized medicine discovery”. In: *Frontiers of Medicine* 7.1 (2013), pp. 102–110.

Chapter 4

Response-Adaptive Biased Coin Design with Delay

4.1 Summary

Adaptive designs are growing in popularity amongst statisticians, clinicians, and regulatory agencies. These designs leverage accumulating data and pre-specified criteria to alter ongoing clinical investigations. This paper explores the relationship between the delay in recording patient outcomes and an adaptive design's continued ability to randomize patients more ethically than complete randomization. Simulations illustrate how, even under the constraint of delayed response, this adaptive design maintains an ethical advantage over equal allocation by transforming an actual HIV intervention into a response-adaptive clinical trial.

Keywords: adaptive randomization, ethical allocation, delayed responses

4.2 Introduction

4.2.1 Adaptive Designs

Adaptive designs, a new breed of clinical trial design, are growing in popularity amongst statisticians, clinicians, and regulatory agencies [19, 20, 25, 42, 44, 53, 56]. These designs allow clinical experiments to leverage accumulating data to alter various aspects of the investigative plan including patient population, randomization, trial duration, and/or analysis based on pre-specified criteria. Modifications are geared at ethical concerns—optimal targeting of or decreased variability in patient assignment, earlier elimination of treatment arms that are less effective, increasing a treatment's speed to market, improved treatment of patient sub-populations by considering appropriate covariates within the randomization scheme. Many adaptive modifications are financially beneficial to the pharmaceutical industry in addition to clinically ethical. See [27] and [40] for a thorough introduction to adaptive design evolutions, advantages, and areas of concern.

Chapters 2 and 3 focus on a response-adaptive biased coin design which assigns more patients to the more effective therapy. In particular, this randomization strategy is shown to be suitable for a continuous outcome with unknown and potentially unequal variances, unknown covariates, and possible covariate-treatment interaction. One aspect of adaptive designs which merits further consideration, however, is the impact of delayed responses on the beneficial factors of a design. This paper explicitly explores the relationship between the delay in recording patient outcomes and the design's continued ability to randomize patients more ethically than complete randomization, the nonadaptive gold standard. Moreover, Section 4.5 illustrates that, even under the constraint of delayed response, this design maintains an ethical ad-

vantage over a similar nonadaptive design by transforming a real HIV intervention into a response-adaptive clinical trial.

4.2.2 BBS Design

Chapter 2 introduces an improvement to the adaptive biased coin design proposed by [4]. The enhanced design, referred to as BBS, assigns a patient to treatment A with probability $\Phi\left(\frac{\widehat{\mu}_A - \widehat{\mu}_B}{S}\right)$, where Φ is the cumulative distribution function of the standard normal distribution; $\widehat{\mu}_k$ is the current estimate at time of randomization of the mean effect of treatment $k = A, B$; and S is the pooled standard deviation of the two treatments, also estimated at the time of randomization. Simulations confirm that, even when the treatment variances are unequal, the new design performs well in comparison to complete randomization and to multiple variations of the original design. The BBS design expands the scope of applicability from the previous design with similar or improved allocation ratios, with minimal loss in power, and with only a slight increased risk of Type I error.

Chapter 3 continues to push the bounds of applicability of BBS via simulation by additionally considering scenarios where the covariate values differ between the treatments. Covariate-treatment interactions are shown to have negligible impact on the design, its ethicality, and rejection rates. Application to a real-world HIV intervention illustrates the value of BBS in assigning more patients to the treatment with larger effect size, after adjusting for covariate impact.

Both chapters circumvent the potential impact of delayed responses on all designs. Instead, simulations assume that all patients immediately respond to treatment, that all responses are immediately observed, and that all past patient response are incorporated into future randomizations. This chapter addresses the impact of delay on the

adaptive benefits of the design. In particular, simulations explore the repercussions of delayed response on the ability of the design to continue assigning more patients to the better treatment.

4.2.3 Delay in Treatment Responses

Many fields of medical research involve treating patients whose outcomes will not be immediately observable. For example, managing depression [32, 55], skin graft healing [9, 23, 29], human immunodeficiency virus viral load suppression [2, 37, 52] are all domains with delayed patient responses which nevertheless undergo large-scale, randomized clinical trials. Unlike classical randomized experiments, response-adaptive randomization designs depend on the accumulation of patient outcomes to inform and influence future patient assignments. At the extreme, delayed patient responses may result in randomization rules which are adaptive by design but do not actually change from one patient to the next. This must be considered when selecting an appropriate allocation algorithm.

Delays between a treatment exposure and a patient response may be fixed or variable. Fixed delays are easily conceptualized. For example, ten hours after a pain relief medication is introduced, a clinician records how a patient feels on a scale of 1 to 10 [17, 47]. The outcome of interest is the ten-hour improvement and it occurs ten hours after treatment exposure. In another example, a clinician records the distance a patient can walk 90 days after a surgical procedure takes place [1, 6, 34]. The outcome of interest is the 90-day measure and it occurs 90 days after treatment administration.

Variable response times are also a common occurrence in phase III clinical trials. For instance, a response of primary interest in a study might be the final value of a biomarker where final is either a particular threshold if achieved prior to 90 days after

treatment exposure or the value of the biomarker at the 90 day mark [12, 14, 15, 24, 39, 60]. Then the time to response is the time to threshold or 90 days, whichever occurs first. Alternatively, an outcome of interest may be time-to-event such as hospital length of stay [30] or time until a particular activity is resumed [5, 16]. Of course, survival trials also fit into this model with response being censored or uncensored survival time—for example symptom-free survival, progression-free survival, or simply survival [1, 6, 33, 35, 50]. Then the time to response is the response itself: the survival time of interest. As the more complicated of the two types of delay, variable response times are the focus in this paper but results may likewise apply to fixed delays.

Delayed observation of patient responses have long hindered alternatives to classical randomization in clinical research [11]. Nevertheless, adaptive designs can be leveraged in conjunction with delayed patient outcomes in theory [3, 8, 63] as well as in practice [54]. On the other hand, asymptotic results do not always apply to finite sample sizes. Simulations are recommended to confirm the impact of the delayed responses on the trial properties [28, 62, 64]. In this article, the impact on the BBS design of delay in observing patient responses is examined for small to large Phase III clinical trials.

This article establishes that BBS is a suitable candidate for use in clinical trials where exposing patients to the superior treatment is a priority. Moreover, this article confirms that BBS continues to be more ethical than equal allocation if a treatment difference exists, even when patient outcomes are not immediately available. Finally, like Chapters 2 and 3, this article illustrates the utility and benefits of the modified BBS design with a real-world application of an HIV treatment adherence intervention.

4.3 Design and Simulation Parameters

4.3.1 Overarching Model

The BBS design applies to two-treatment studies where the outcome of interest is continuously defined and predetermined prognostic factors are measured at enrollment. Statistically, assume patient responses y_{jk} follow a normal distribution conditional on the treatment exposure, the covariates, and the treatment-covariate interaction:

$$y_{jk} = \mu_k + \boldsymbol{\beta}_k \boldsymbol{z}'_j + \epsilon_{jk}$$

for patients $j = 1, \ldots, N$ and treatments $k = A, B$. The leftmost term on the right hand side is μ_k, the mean effect of treatment k. The rightmost term is ϵ_{jk}, the random error for patient j on treatment k, and is independently and identically distributed $N(0, \sigma_k^2)$. The middle term is the product of $\boldsymbol{\beta}_k = (\beta_{k1}, \ldots, \beta_{kP})$, the slopes of the P covariate effects on treatment k's response, and $\boldsymbol{z}_j = (z_{j1}, \ldots, z_{jP})$, the actual covariates for patient j measured at enrollment. Each prognostic factor Z_p is independently and identically distributed for all $p = 1, \ldots, P$ but the distribution may vary for each value of p. For example, one covariate may be the binary factor sex while another covariate may be weight, a continuous measure.

4.3.2 Study Design

The overarching goal of many clinical trials is to determine which of two treatments has a superior treatment effect. In this paper, patient responses—conditional on an individual patient's covariates and unknown covariate slopes—are normally distributed with larger responses desirable. Each prognostic factor is independently

and identically distributed and all factors are known for each patient upon enrollment. Treatment variances are unknown but thought to be equal or similar. Patient recruitment, randomization, and treatment is assumed to occur simultaneously.

The design considered in this article is an adaptive biased coin [4] that allocates more patients to the treatment with larger effect size by updating the exposure probabilities over time. When patient responses are immediately observable, the BBS design begins by assigning a fixed, minimum number m of patients to each treatment in order to estimate the difference in treatment effects $\widehat{\mu}_A - \widehat{\mu}_B$ and the pooled standard deviation

$$S = \sqrt{\frac{(N_A - 1)s_A^2 + (N_B - 1)s_B^2}{N_A + N_B - 2}}.$$

Here $\widehat{\mu}_k$ is the current estimate of the mean effect of treatment k, excluding the covariate contribution to patient outcomes; N_k is the number of patients who have been exposed to and immediately responded to treatment k; and s_k^2 is the current estimated variance for treatment k for $k = A$ or B. The minimum number of patient responses required for estimation is arbitrarily chosen to be $m = 10$. Once these estimates are obtained using an adequate number m of patients on each treatment, subsequent patients are randomized to treatment A with probability $\Phi\left(\frac{\widehat{\mu}_A - \widehat{\mu}_B}{S}\right)$ and to treatment B with probability $1 - \Phi\left(\frac{\widehat{\mu}_A - \widehat{\mu}_B}{S}\right)$. These allocation probabilities are updated each time a patient response becomes available.

When a delay exists between treatment exposures and patient responses, the BBS design must be adjusted slightly. In a typical phase III clinical trial, patients are recruited sequentially with varying times between each new patient enrollment. Patient j arrives at time t_j, patient $j + 1$ arrives at time t_{j+1}, and the time between t_j and t_{j+1} is exponentially distributed with mean equal to the expected inter-arrival time. If the time between treatment exposures and patient responses varies across patients,

then treatment outcomes will typically be observed in an order different from that of patient enrollment and with varying inter-observation times. Patient j arrives, is randomized, and obtains treatment at time t_j; patient j responds to treatment at time $t_j + \tau_j$ where τ_j derives from an exponentially distributed random variable with mean equal to the expected delay in outcome observation. Table 4.1 represents one possible timeline for a clinical trial which enrolls 50 patients with exponentially distributed inter-arrival times having mean 1 and exponentially distributed patient response delays with a mean delay time of 40 time units. Table 4.1 is discussed further in Section 4.3.4.

Like its counterpart with immediately observed outcomes, the BBS design with delayed responses begins by assigning patients to treatment A with probability $\frac{1}{2}$ and to treatment B with probability $\frac{1}{2}$. This continues until at least m patients have responded to treatment A and at least m patients have responded to treatment B. Once an adequate number m of patients respond on each treatment, the allocation probabilities can be estimated based on observed outcomes and known patient covariates. That is, the parameters for the set of prognostic factors $\boldsymbol{Z}$ are estimated based on all previously recruited patient values. These statistics then support the estimation of the difference in treatment mean effects and pooled treatment standard deviation. When each subsequent patient is recruited, $\Phi\left(\frac{\widehat{\mu}_A - \widehat{\mu}_B}{S}\right)$ and $1 - \Phi\left(\frac{\widehat{\mu}_A - \widehat{\mu}_B}{S}\right)$ are re-estimated based on all previously enrolled patient covariates and all recorded patient responses. The randomization proportions continue to be updated until the study enrollment terminates. The proportion of fixed allocation assignments (using $\frac{1}{2}$) relative to the total patient enrollment should be approximately anticipated prior to the trial commencement based on the expected recruitment time and study size.

In this article, the BBS is juxtaposed with the "gold standard" clinical trial design, equal allocation [36]. Under equal allocation, patients are assigned to treatment

A or treatment B each with fixed probability $\frac{1}{2}$, resulting in asymptotically equal proportions of patients on each treatment [38]. While there are many beneficial qualities of a balanced randomization scheme [38, 43, 45, 51], the comparison of interest is BBS design ability to adaptively allocate patients to the superior treatment when patient responses are not immediately observable—that is, even as the adaptive mechanisms of the design are inhibited.

Chapter 2 demonstrates that the BBS design allocates more patients to the better treatment than equal allocation without sacrificing power or inflating Type I error rates. It also demonstrates how the BBS design is preferable to its predecessor due to its ability to account for the degree of certainty or variance present in the data at a given point during trial enrollment. Chapter 3 confirms that the three aforementioned BBS design properties are not significantly altered when covariate impact on patient outcomes differed by treatment within a reasonable range. This article considers the consequences to the three metrics (allocation, power, and Type I error) as the response-adaptive randomization is subject to variable delays in patient response.

4.3.3 Simulation

Section 4.4 describes the results of simulated clinical trials under the BBS design and equal allocation for various treatment mean and standard deviation parameters. Throughout the simulations, treatment mean pairs (μ_A, μ_B) are considered at no effect $(0, 0)$ and positive effect in treatment A $(1, 0)$. A positive effect in treatment B $(0, 1)$ is discussed in Chapter 2. Individual patient errors are simulated under equal and unequal conditions; σ_B is fixed at 1 in all simulations while σ_A ranges from 0.5 to 1.5 by half-step increments. One normally-distributed prognostic factor $Z \sim N(1, 1)$ is simulated for all patients. The covariate impact on treatment outcomes is a constant

slope of $\beta = 2$ for both treatments.

Additionally, the BBS design simulations are subject to multiple levels of delay in patient responses. Both patient enrollment and patient responses are unrelated arrival processes. Patient inter-arrival times are simulated using independent and identical exponential distributions with mean one ($T_{j+1} - T_j \sim \exp(1)$). When a nonzero response delay is simulated, the time before a patient responds will be independent and identically distributed for all patients following an exponential wait time with mean equal to the specified delay. Delay means are zero (patient responds to treatment immediately), 40, 400, and 4,000. As an illustration, when responses have a mean delay of 40, an average of 40 additional patients will enroll between when a patient is treated and when that patient's response is observed—i.e., $\mathcal{T}_j \sim \exp(40)$. Details of the exponential delays are discussed further in Section 4.3.4.

Rejection of the null hypothesis $H_0 : \mu_A \leq \mu_B$ is calculated with a one-sided two-sample t-test with significance level $\alpha = 0.05$. Under the assumption of possibly unequal variances, the t-statistic standard error is calculated using treatment-specific sample variances, leveraging the Welch-Satterthwaite approximated pooled degrees of freedom [49, 59]. Sample sizes mimic small to large Phase III clinical trials ($N = 50$; 100; 500; 1,000; 5,000). Initially, patients are enrolled under nonadaptive, equal allocation until at least $m = 10$ patients have responded on each treatment arm. At that point, the BBS design begins randomizing adaptively based on currently available treatment estimates whereas equal allocation continues to assign patients with a fixed ratio. For each scenario described, 1,000 replications are simulated in SAS IML [48].

4.3.4 Exponential Delay

The time between a patient's exposure to a study treatment and the observation of that patient's response is simulated as an exponentially distributed delay. An exponentially distributed random variable has probability distribution function

$$f(x, \lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{when } x \geq 0, \\ 0 & \text{when } x < 0, \end{cases}$$

with mean $\frac{1}{\lambda}$, median $\frac{\ln(2)}{\lambda}$, and variance $\frac{1}{\lambda^2}$. Recalling that $\ln(2) \approx 0.7$, the median of an exponential distribution is clearly less than its mean. That is, more than half of all patients are expected to respond to treatment in a period of time shorter than the mean delay.

To illustrate, when the expected delay in treatment response is 40 units of time, the variance in patient delay is 1,600. The median delay from the exponential distribution with mean 40 is 27.7 units of time—a difference of less than 13 units of time between the mean and the median. By comparison, the exponential distribution with mean 400 produces a 400 time unit expected delay in observing patient outcomes and a 160,000 time unit variance. This distribution has a 277.3 time unit median delay—a difference of more than 120 units of time between mean and median.

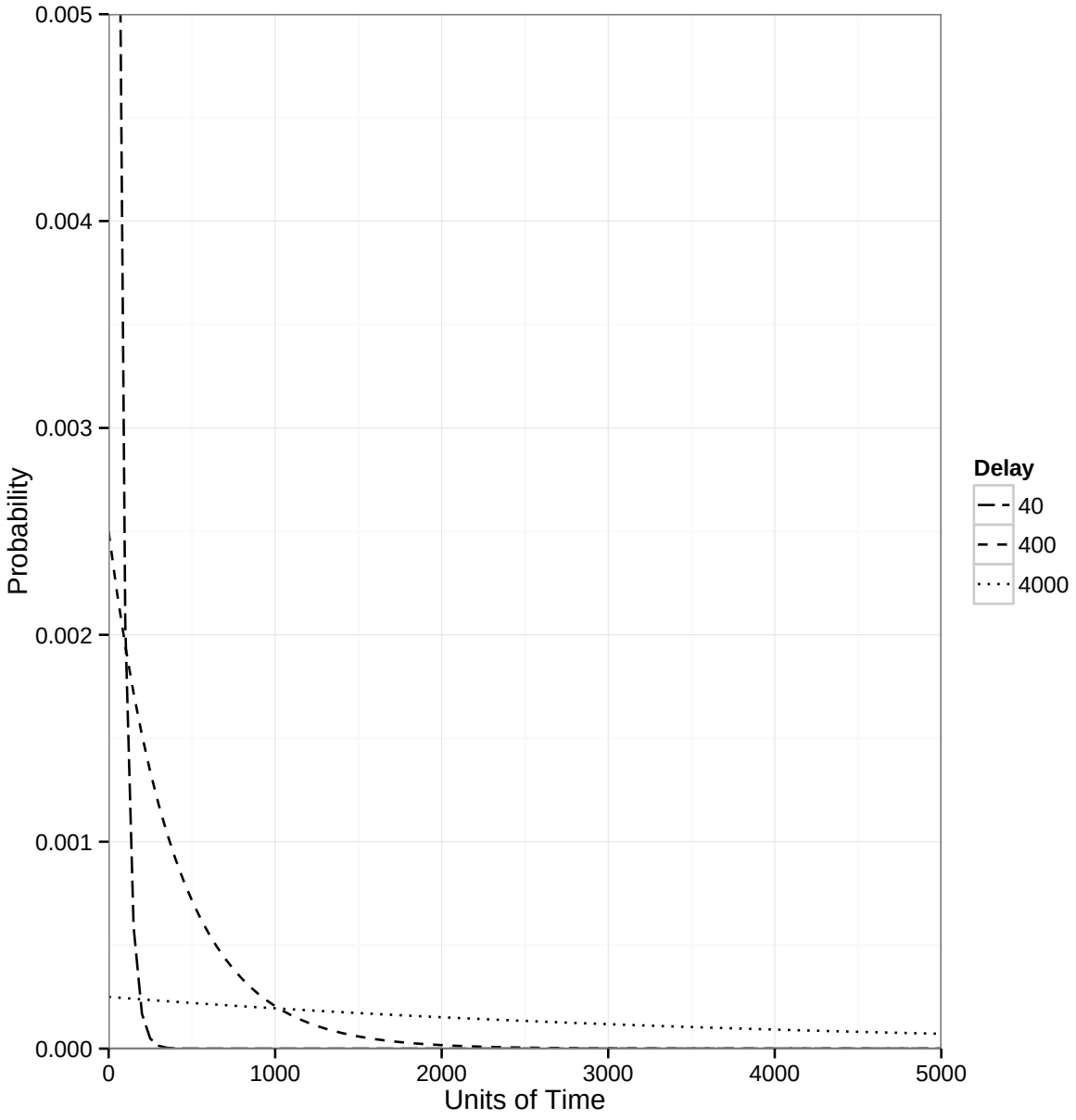
Figures 4.1 – 4.4 highlight the asymmetry of the three distributions simulated, i.e., exponential with means of 40; 400; and 4,000 units of time. Figure 4.1 juxtaposes all three exponential distributions on the same axes. This image accentuates the changes in the curvatures of the distributions as the scaling constant increases by a factor of ten. Figures 4.2 through 4.4 focus on the expected value and quartiles for each of the three distributions. In these images, the axes are proportional to the scaling constant

of the distribution; all three graphs now appear similarly shaped. In Figures 4.2 – 4.4, as units of time increase, the probability of a delay length being equivalent to a particular unit of time decreases monotonically. Moreover, larger expected values for this distribution correspond to increased variably and skew.

Table 4.1 represents a potential timeline for patient arrivals and responses during a clinical trial with sample size $N = 50$ when responses are delayed with an expected mean of 40 units of time. Starting on the left, the first column in the table is the patient arrival sequence. The second column contains the patient inter-arrival times. The first patient does not have an inter-arrival time. All other patients' inter-arrival times are derived from independently and identically distributed exponential random variables. The expected inter-arrival is one unit of time. The third column denotes the arrival, randomization, and treatment exposure time of each patient relative to patient 1. Patient enrollment, assignment, and treatment are assumed to happen concurrently. Patient 1 is considered to arrive, be randomized, and treated at time 0.0. The fourth column indicates the delay between the time a patient is exposed to treatment and the time the patient's outcome is observed. All response delays are values taken from independently and identically exponentially distributed random variables with mean equal to 40 units of time. The fifth column communicates the time at which a patient responds to treatment relative to the first patient's arrival. The final column illustrates the order in which the 50 patients respond to treatment according to these inter-arrival and response delay values.

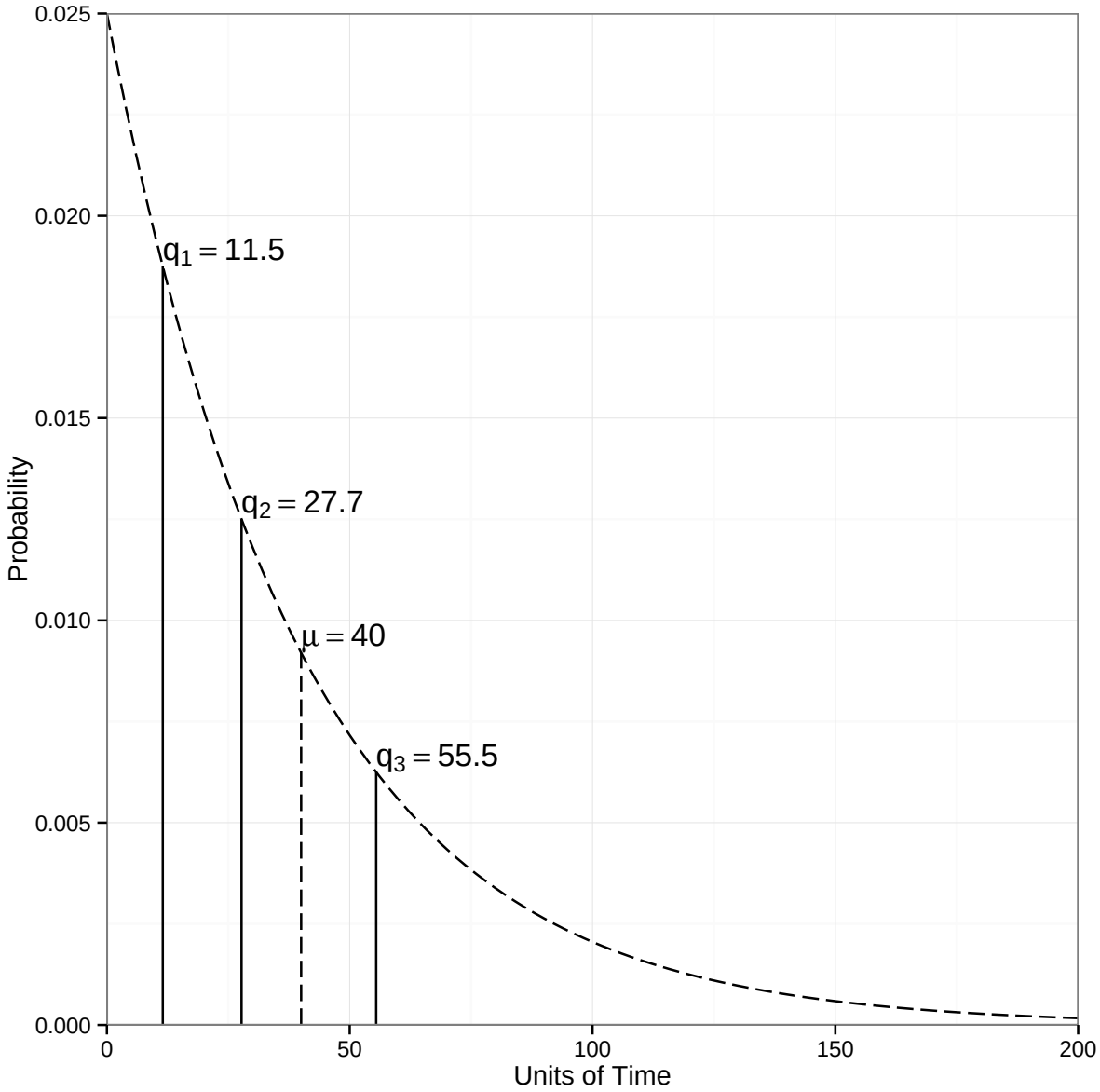
The first row of Table 4.1 does not contain an inter-arrival time for Patient 1 whose arrival and randomization time is considered to be 0.0, the start of trial enrollment. Patient 1 responds to treatment after 30.9 units of time and is the 11^{th} patient outcome observed. Patient 2 arrives 0.1 units after patient 1 at time 0.1. Patient 2's response delay is 33.9 time units. Hence, patient 2 responds to treatment at time 34.0 and is

Figure 4.1: The time until a patient responds to treatment follows an exponential distribution with mean delay 40; 400; or 4,000 units of time. For each expected delay, the probability distribution is graphed with increasingly dashed lines as the expected delay in treatment observation increases.



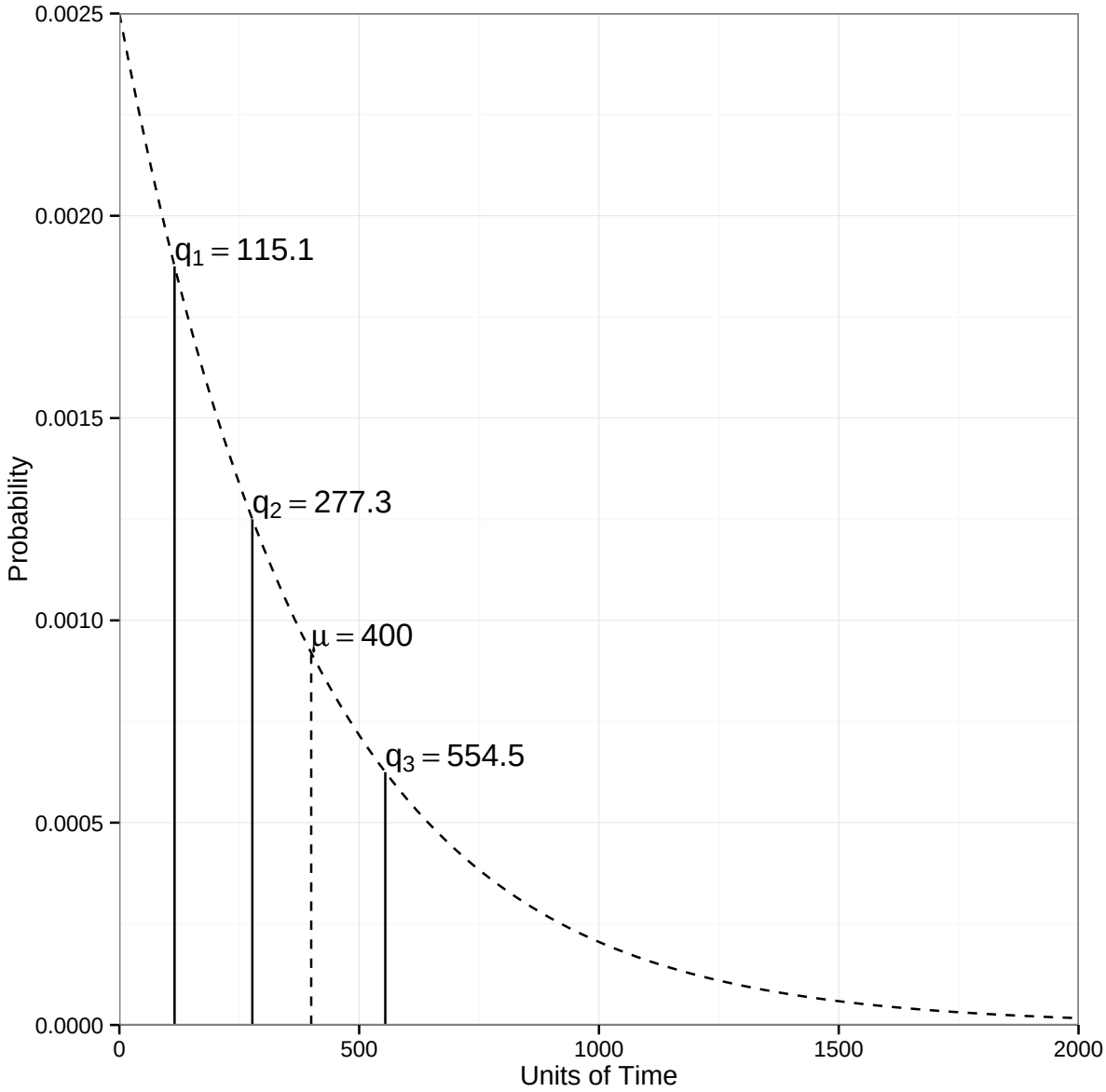
the 14^{th} patient to respond. Under this scenario, the 20^{th} observed outcome belongs to patient 3 and occurs at 44.2 units of time. If 10 of those 20 responses belong to patients on treatment A and 10 belong to patients on treatment B, then the minimum

Figure 4.2: For an expected delay of 40 units of time, the exponential probability distribution is identified with the first, second, and third quartiles highlighted. Half of patients will respond to treatment in fewer than 27.7 units of time.



$m = 10$ patients required to begin adaptive randomization would be achieved when patient 3 responds. Leveraging BBS, all patients enrolled after time 44.2 would be allocated with adaptively estimated assignment probabilities. In Table 4.1, patients 47 – 50 are enrolled after time 44.2 and thus could be adaptively randomized in a

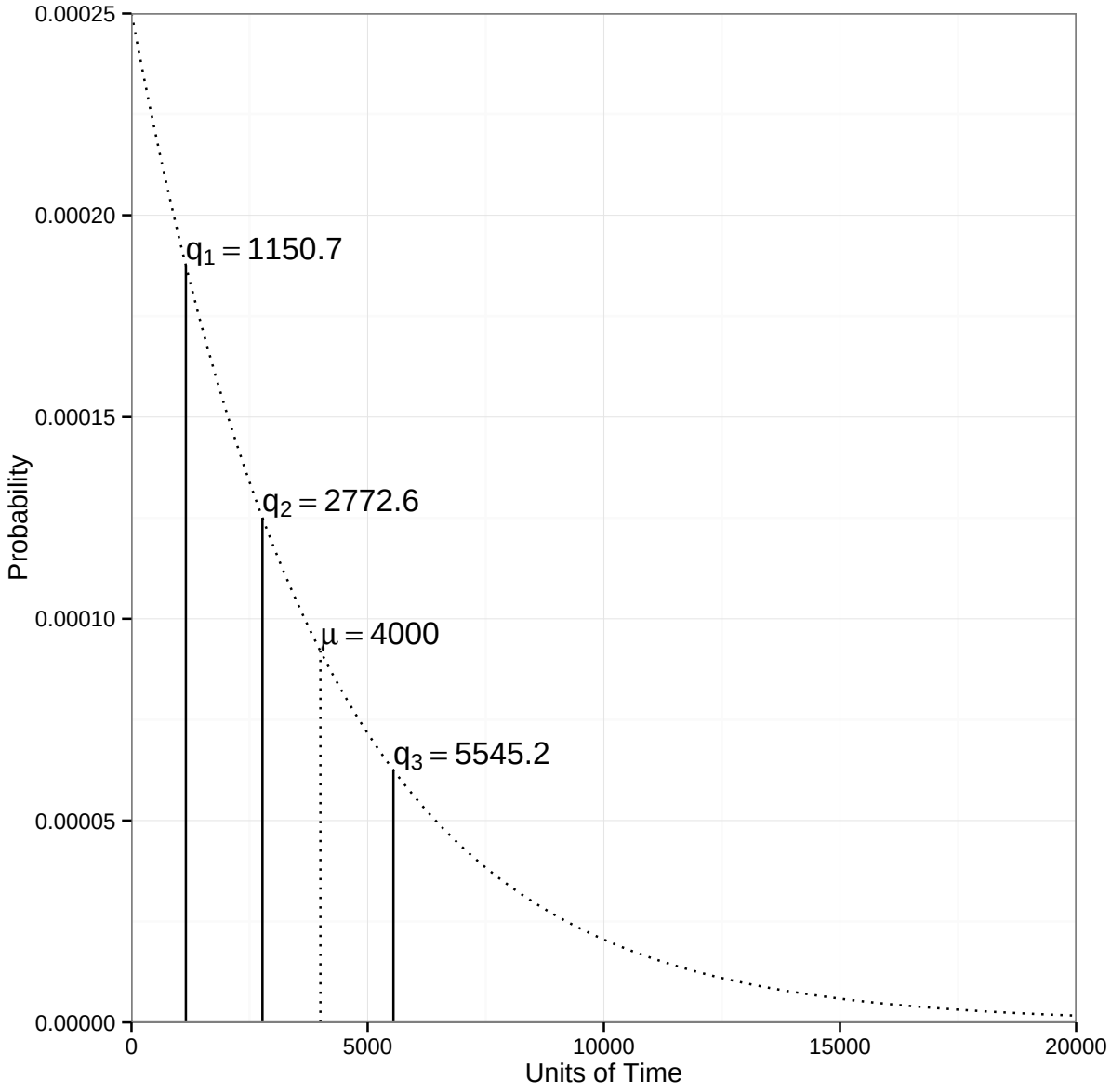
Figure 4.3: For an expected delay of 400 units of time, the exponential probability distribution is identified with the first, second, and third quartiles highlighted. Half of patients will respond to treatment in fewer than 277.3 units of time.



clinical trial employing the BBS design following this timeline.

For another example, consider a moderate sized clinical trial with $N = 500$ under the constraint that the mean patient delay in treatment responses is 4,000. Even though all patients have an expected delay in response of 4,000 units, 2.5% of patients

Figure 4.4: For an expected delay of 4,000 units of time, the exponential probability distribution is identified with the first, second, and third quartiles highlighted. Half of patients will respond to treatment in fewer than 2,772.6 units of time.



should be anticipated to respond before 100 units of time have passed; 4.9% of patients would be expected to respond before 200 units of time have passed; 7% of patients are predicted to respond before 300 units of time have passed; and 9.5% of patients are expected to respond before 400 units of time have passed. Working backwards, from

the first 100 patients to be assigned, at least 9.5 outcomes are expected to be observed before trial enrollment is complete; from the next 100 patients treated, a minimum of 7 more outcomes are predicted to present before enrollment is over; from patients 301 to 400, another 4.9 responses should be expected before enrollment ends. Without even considering the last 100 patients recruited, more than 20 responses are anticipated prior to all patients having been randomized. That is, despite individual patients having an expected delay equal to eight times the predicted trial enrollment period, the BBS design should still be able to adaptively randomize patients to treatment in a moderately sized study.

The simulated proportion of patients adaptively randomized by BBS are explored further in Section 4.4.1 for each clinical trial size and delay combination. As this last example illustrates, RAR can still positively impact patient assignment when the expected delay is large relative to the anticipated study enrollment period. The relationship between the exponential delay, the clinical trial size, and the minimum number m of responses required prior to beginning adaptive allocation will influence the design's ability to assign more patients to the better arm for a given set of treatment parameters.

4.4 Results

4.4.1 Patient Allocation

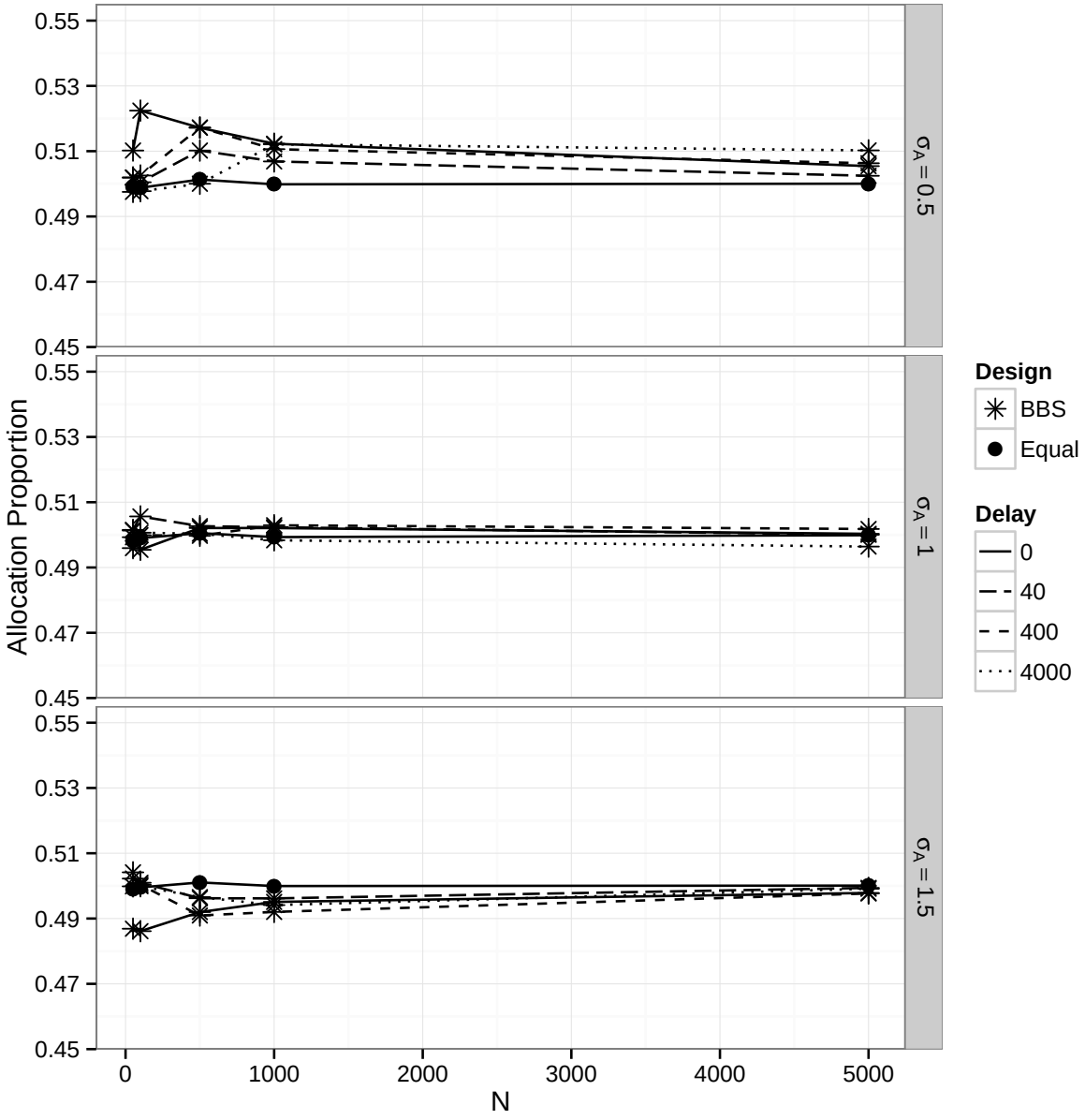
Tables 4.2 and 4.3 reflect the proportions and standard deviations of patients assigned to treatment A by design and delay for zero and non-zero treatment effects under the BBS design and equal allocation (Equal). Each table contains the average patient allocation as the BBS design ranges from immediately observed patient responses

(Delay = 0) up to a delay of 4,000 and under equal allocation with no delay in response across various treatment parameters and clinical trial sizes. The first two columns of each table indicate the design (BBS or Equal) and delay (0–4,000) parameters for the row. The remaining five columns are the mean (SD) proportion of patients assigned to treatment A over 1,000 simulations for each of the five trial sizes $N = 50; 100; 500;$ 1,000; and 5,000. Each set of five rows (BBS with delay 0, 40, 400, and 4,000; Equal with no delay) is grouped by treatment parameters μ_A, μ_B, σ_A, and σ_B.

For example, Table 4.2 details the average patient allocation when neither treatment A nor treatment B have an impact ($\mu_A = \mu_B = 0$). The first row of data represents BBS patient allocation when responses are immediately observable (Delay = 0), there are no differences in treatment effects, but $\sigma_A = 0.5$ while $\sigma_B = 1.0$. With a sample size of $N = 50$, BBS allocates 51% of patients to treatment A with a standard deviation of 0.12. With a sample size of $N = 100$, BBS allocates 52% of patients to treatment A with a standard deviation of 0.13. For sample sizes of $N = 500$, 1,000, or 5,000, BBS allocation with no delay remains close to 50% with no treatment mean difference. On the other hand, Table 4.3 details the average patient allocation when treatment A has a positive impact and treatment B has no impact ($\mu_A = 1$, $\mu_B = 0$). The first row of Table 4.3 describes how the BBS allocates at least 70% of patients to treatment A for all clinical trial enrollment sizes simulated when responses are immediately observed, $\sigma_A = 0.5$, and $\sigma_B = 1.0$.

Figure 4.5 illustrates the trends in patient assignment from Table 4.2 across the five simulated trial sizes for all three values of treatment A standard deviation when $\mu_A = \mu_B = 0$. Allocation means are represented by $*$ for the BBS design and by $\bullet$ for equal allocation. In both designs, immediate responses (Delay = 0) are represented by solid lines and for BBS increasing delay in outcomes are represented by increasingly dashed lines.

Figure 4.5: Proportion of patients assigned to treatment A by design when $\mu_A = \mu_B = 0$, σ_A varies, $\sigma_B = 1$, and N varies. Data points are grouped by design; BBS is represented by $*$ and equal allocation by $\bullet$. The time until a patient responds to treatment is an exponentially distributed delay. As delay in treatment observation increases, the line connecting delayed data points are increasingly dashed. No delay is represented by a solid line (—) and a mean delay of 4,000 is represented by a dotted line ($\cdots$).

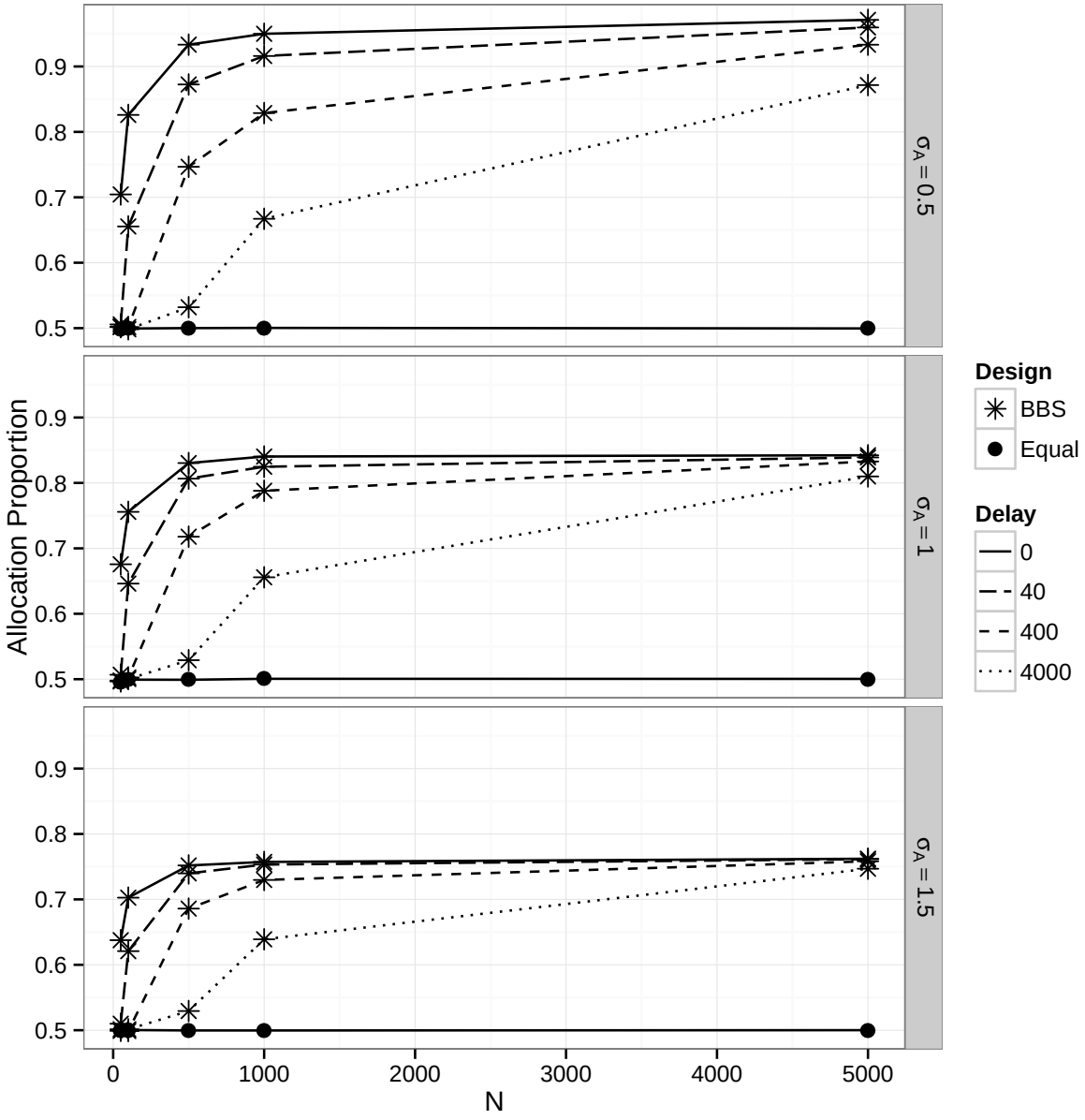


As seen in Table 4.2 and Figure 4.5, patient allocation with no treatment effect is roughly balanced between treatments regardless of design, delay, trial enrollment, or treatment variance. Both the BBS and equal allocation designs expose approximately half of patients to each treatment when no treatment effect is present. There is a slight tendency to expose more patients to treatment A when the total variance is small and to expose more patients to treatment B when the total variance is large.

Figure 4.6 illustrates the trends in patient assignment from Table 4.3 across the five simulated trial sizes for all three values of treatment A standard deviation when $\mu_A = 1$, $\mu_B = 0$. Allocation means are distinguished by design and delay parameter; $*$ for BBS and $\bullet$ for equal allocation connected by increasingly dashed lines as response delay increases.

As seen in Table 4.3 and Figure 4.6, patient allocation can improve dramatically on the BBS design versus balanced randomization when a treatment effect is present and there is no delay in observing patient outcomes. That is, when delay is nonexistent or small (0 or 40), more than 80% of patients can be assigned to the treatment with larger mean effect size when $\sigma_A = \sigma_B = 1.0$ (70% when $\sigma_A = 1.5$ and 90% when $\sigma_A = 0.5$). As the delay in patient response times grows, the ability of the BBS design to ethically assign patients to treatment shrinks. In fact, when responses are delayed so that no patients have responded to treatment prior to the conclusion of trial enrollment, the BBS design does not attain the ability to be adaptive. Instead, the BBS design randomizes patients to both treatments in equal proportions; BBS defaults to equal allocation when sample sizes are small relative to the delay in patent responses. Nevertheless, even when responses are delayed, as long as an adequate number of outcomes are observed before all subjects have been enrolled, the BBS design is still preferable to equal allocation in terms of ethical patient randomization. For example, when the delay factor is 400 for a trial containing only 500 patients, 72%

Figure 4.6: Proportion of patients assigned to treatment A by design when $\mu_A = 1$, $\mu_B = 0$, σ_A varies, $\sigma_B = 1$, and N varies. Data points are grouped by design; BBS is represented by $*$ and equal allocation by $\bullet$. The time until a patient responds to treatment is an exponentially distributed delay. As delay in treatment observation increases, the line connecting delayed data points are increasingly dashed. No delay is represented by a solid line (—) and a mean delay of 4,000 is represented by a dotted line ($\cdots$).



of patients are still randomized to the better therapy when treatment variances are equal compared to only 50% under a balanced design. Furthermore, even for a delay of 4,000 in a trial with a total enrollment of 500, an additional 3% of patients are assigned to the superior treatment under BBS than would be under equal allocation for all three combinations of treatment variance.

4.4.2 Power and Type I Error

Tables 4.4 and 4.5 reflect the proportions and standard deviations of simulated trials which reject the null hypothesis that the mean effect of treatment A is no larger than that of treatment B under the BBS design and equal allocation (Equal). Each table contains the rejection rates as the BBS design ranges from immediately observed patient responses (Delay = 0) up to a delay of 4,000 as well as under equal allocation with no delay in response. These rates are given for six distinct treatment parameters—three combinations per table—and five different enrollment levels. The first two columns of each table indicate the design (BBS or Equal) and delay (0–4,000) parameters for the row. The remaining five columns are the mean (SD) proportion of simulated trials which reject the null hypothesis over 1,000 repetitions of each clinical trial size $N = 50; 100; 500;$ 1,000; and 5,000. Each set of five rows (BBS with delay 0, 40, 400, and 4,000; Equal with no delay) is grouped by treatment parameters μ_A, μ_B, σ_A, and σ_B.

For example, Table 4.4 details the simulated rejection rates when neither treatment A nor treatment B have any impact ($\mu_A = \mu_B = 0$). That is, Table 4.4 relates type I error rates for BBS and equal allocation. The first row of data represents BBS type I error rates when responses are immediately observable (Delay = 0), there are no differences in treatment effects, but $\sigma_A = 0.5$ while $\sigma_B = 1.0$. With a sample size

of $N = 50$ or with $N = 100$, BBS incorrectly rejects 10% of trials with a standard deviation of 0.29 or 0.30, respectively. With a sample size of $N = 1{,}000$, however, BBS only rejects 6% of trials incorrectly with a standard deviation of 0.24.

On the other hand, Table 4.5 details the simulated rejection rates when treatment A has a positive impact and treatment B has no impact ($\mu_A = 1$, $\mu_B = 0$). That is, Table 4.5 reports the power for BBS and equal allocation designs. The first row of Table 4.5 describes how the BBS with no delay correctly rejects 90% of trials with a standard deviation of 0.30 for $N = 50$, 99% of trials with a standard deviation of 0.11 for $N = 100$, and 100% of trials with a standard deviation <0.005 for enrollment levels of 500 or larger when $\sigma_A = 0.5$ and $\sigma_B = 1.0$.

Figure 4.7 illustrates the trends in type I error from Table4.4 across the five simulated trial sizes for all three values of treatment A standard deviation when $\mu_A = \mu_B = 0$. Rejection rates are represented by $*$ for the BBS design and by $\bullet$ for equal allocation. In both designs, immediate responses (Delay = 0) are represented by solid lines and for BBS increasing delay in outcomes are represented by increasingly dashed lines.

As seen in Table 4.4 and Figure 4.7, rejection rates with no effect in either treatment hover near 6% for both designs, all levels of delay, all trial enrollment sizes, and all three values treatment A variance. Type I error rates tend to be inflated when the total variance is small ($\sigma_A = 0.5$, $\sigma_B = 1.0$) and when the clinical trial size is small to moderate ($N \leq 500$). This is particularly true under the BBS design with no delay: ten percent of trials incorrectly reject the null hypothesis following BBS randomization with immediately observed patient outcomes for enrollment levels of 50 or 100 patients.

Figure 4.8 illustrates the trends in power from Table 4.5 across the five simulated trial sizes for all three values of treatment A standard deviation when $\mu_A = 1$, $\mu_B = 0$.

Figure 4.7: Proportion of trials which incorrectly reject the null hypothesis by design when $\mu_A = \mu_B = 0$, σ_A varies, $\sigma_B = 1$, and N varies. Data points are grouped by design; BBS is represented by $*$ and equal allocation by $\bullet$. The time until a patient responds to treatment is an exponentially distributed delay. As delay in treatment observation increases, the line connecting delayed data points are increasingly dashed. No delay is represented by a solid line (—) and a mean delay of 4,000 is represented by a dotted line ($\cdots$).

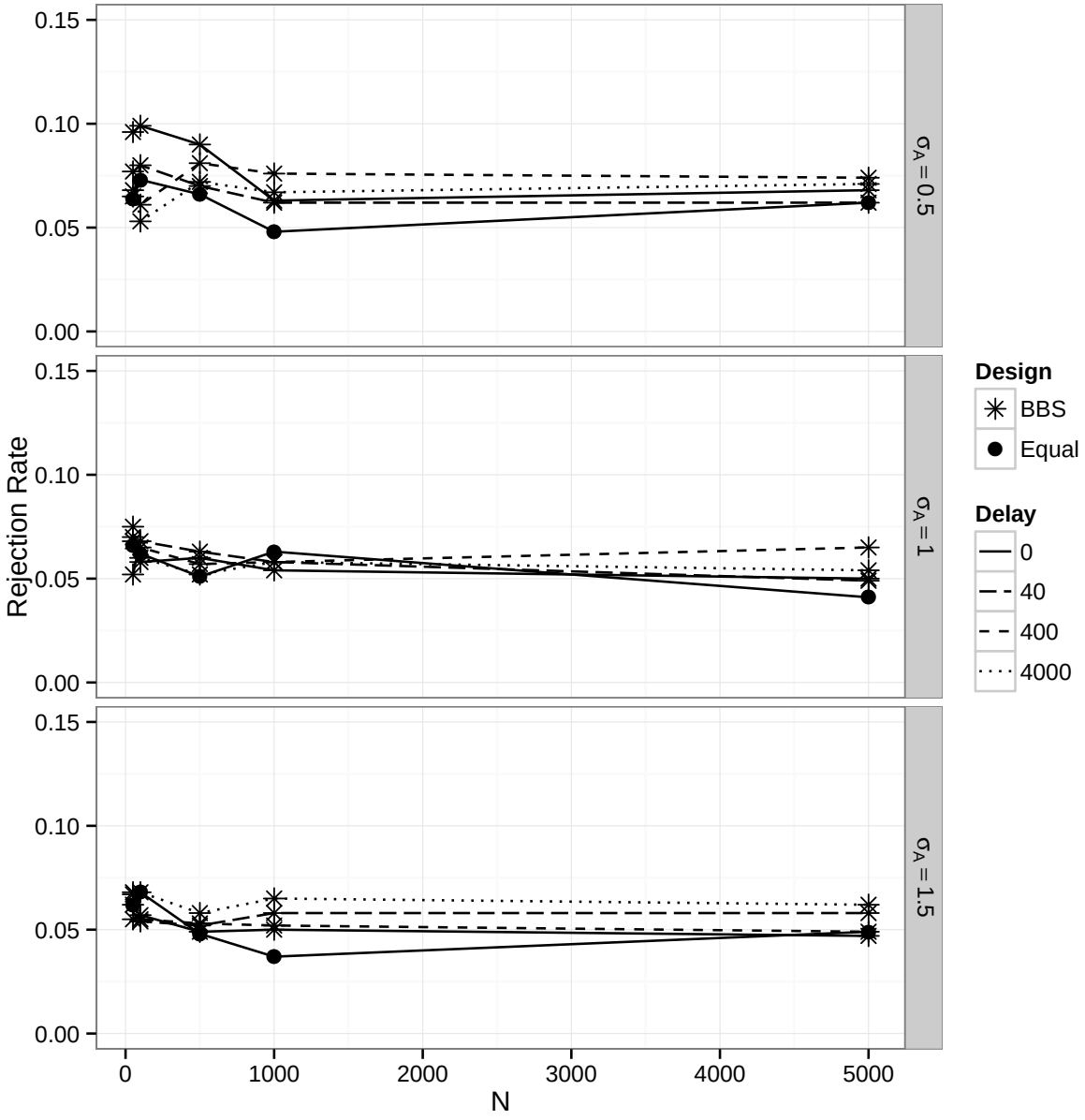
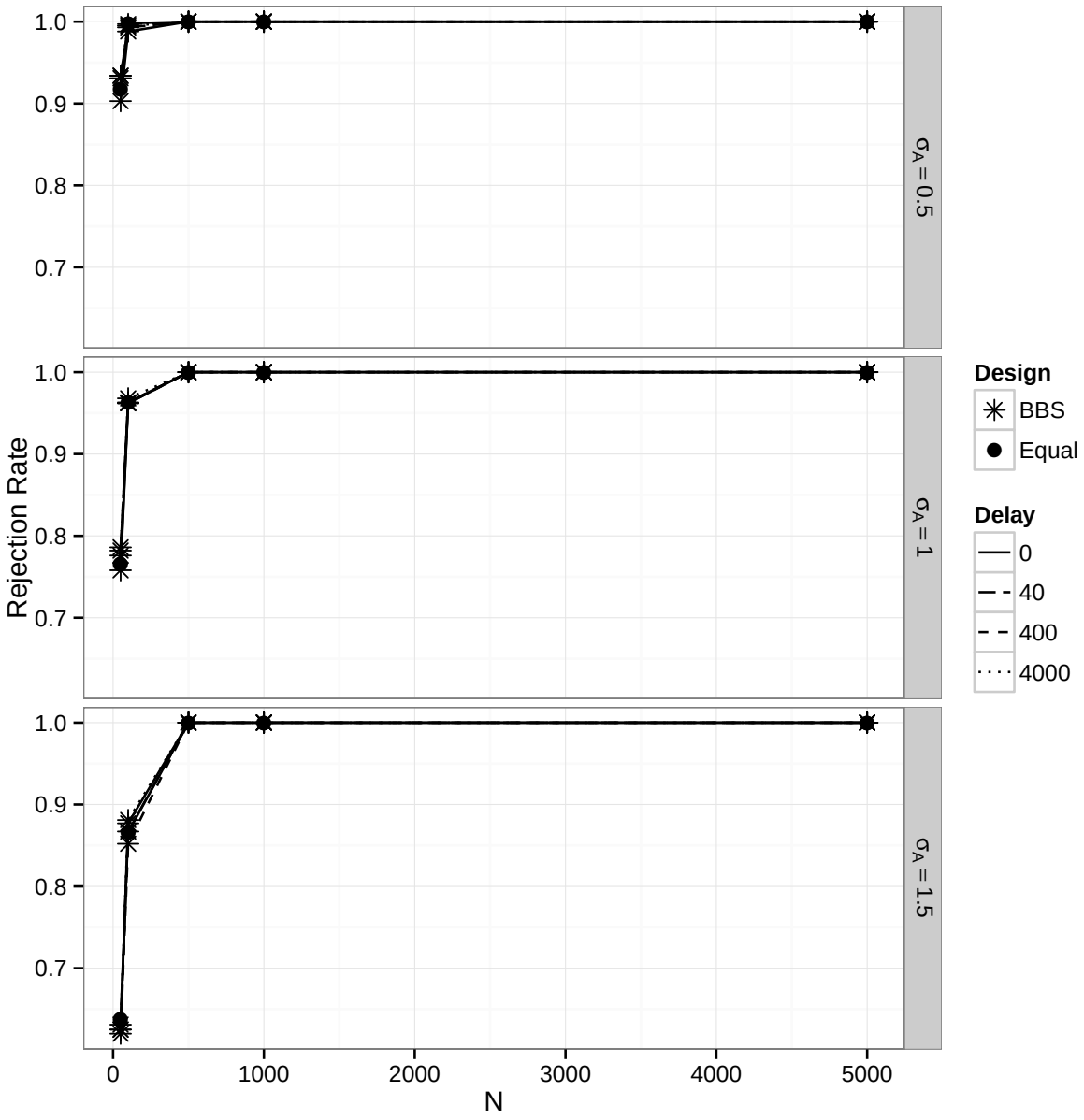


Figure 4.8: Proportion of trials which correctly reject the null hypothesis by design when $\mu_A = 1$, $\mu_B = 0$, σ_A varies, $\sigma_B = 1$, and N varies. Data points are grouped by design; BBS is represented by $*$ and equal allocation by $\bullet$. The time until a patient responds to treatment is an exponentially distributed delay. As delay in treatment observation increases, the line connecting delayed data points are increasingly dashed. No delay is represented by a solid line (—) and a mean delay of 4,000 is represented by a dotted line ($\cdots$).



Rejection rates are distinguished by design and delay parameter; $*$ for BBS and $\bullet$ for equal allocation connected by increasingly dashed lines as response delay increases.

As seen in Table 4.5 and Figure 4.8, for a fixed difference in treatments ($\mu_A = 1$, $\mu_B = 0$), rejection rates depend largely on clinical trial sample size and treatment variance. Power is similar between the two designs, including at different levels of delay, after accounting for N and σ_A. In fact, the rejection rate for all simulations with enrollment size at least 500 is 100% with standard deviation <0.005 for both BBS and Equal designs when treatment A has a positive effect and treatment B has no effect, regardless of treatment variance values. Moreover, after adjusting for treatment A variance, power within a clinical trial size differs by no more than 4% across both designs at all levels of delay, with lower power in the smaller clinical trials. Similarly, after adjusting for N, rejection rates at each level of treatment variance is consistent across both designs at all levels of delay, with lower rejection rates as total variance increases.

4.4.3 Delay

Both the BBS design and equal allocation initially randomize patients using a fixed allocation ratio of $\frac{1}{2}$. For equal allocation, this randomization proportion continues for the duration of the enrollment period. By contrast, the BBS design only retains a constant assignment probability until an adequate number m of patient responses are collected to inform an adaptive randomization probability. In particular, after ten responses are received for each treatment arm, the BBS leverages these observations to estimate a response-adaptive randomization probability, $\Phi\left(\frac{\widehat{\mu}_A - \widehat{\mu}_B}{S}\right)$. Patients assigned to treatment using a response-adaptive randomization ratio are considered to be adaptively randomized; patients assigned to treatment using the fixed allocation

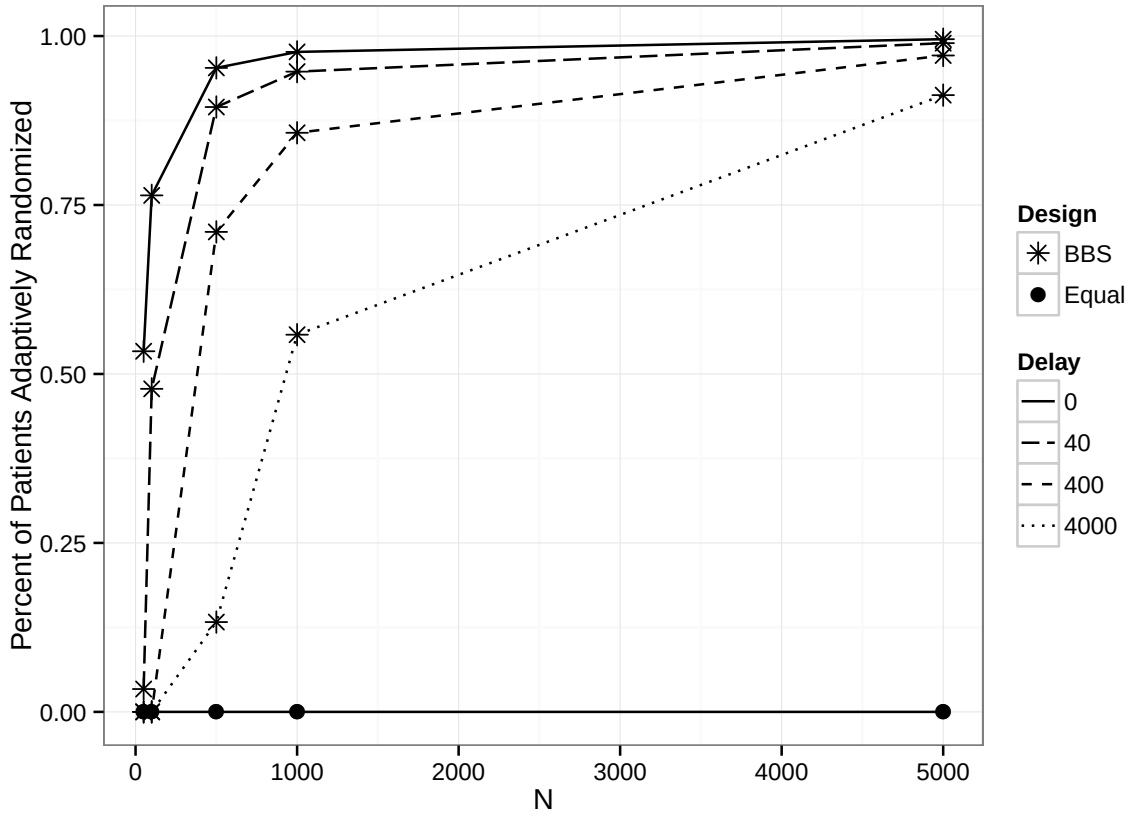
$\frac{1}{2}$ are not.

Table 4.6 describes the proportion of patients adaptively randomized under BBS for all combinations of delay and trial enrollment size. The first two columns indicate the design (BBS) and delay (0, 40, 400, and 4,000) parameters for the row. The remaining five columns are the mean (SD) proportion of patients adaptively randomized over 1,000 repetitions of each clinical trial size $N = 50$, 100, 500, 1,000, and 5,000. For example, the first row of data represents BBS adaptive randomization proportions when responses are immediately observable (Delay = 0). With a trial enrollment of $N = 50$, BBS randomizes 53% of patients adaptively with a standard deviation of 0.07 under conditions of no delay. With a sample size of $N = 500$, the proportion rises to 95% with a standard deviation of 0.01. When $N = 5{,}000$, nearly 100% of patients are adaptively randomized and variation in this proportion is negligible. While not shown, these values are the same across all treatment parameter values. That is, the size of the treatment mean or variance may impact *how* a patient is adaptively allocated, but has no influence on whether or not a patient *is* adaptively allocated. Equal allocation is not included in Table 4.6 as no patients are adaptively randomized under equal allocation regardless of trial parameters or treatment outcomes.

Figure 4.9 illustrates the trends in the proportion of patients adaptively randomized from Table 4.6 across the five simulated trial sizes. Adaptively randomized patient proportions are represented by $*$ for the BBS design and by $\bullet$ for equal allocation. In both designs, immediate responses (Delay = 0) are represented by solid lines and for BBS increasing delay in outcomes are represented by increasingly dashed lines.

As seen in Table 4.6 and Figure 4.9, the proportion of patients adaptively randomized is consistently high for large trials. Clinical trials with 5,000 patients see more than 90% of the patients adaptively randomized. Trials where $N = 1{,}000$ av-

Figure 4.9: Proportion of patients adaptively randomized by design as N varies. Data points are grouped by design; BBS is represented by $*$ and equal allocation by $\bullet$. The expected time until a patient responds to treatment is the exponentially distributed Delay. As delay in treatment observation increases, the line connecting delayed data points are increasingly dashed. No delay is represented by a solid line (—) and a mean delay of 4,000 is represented by a dotted line ($\cdots$).



erage at least 86% of patients adaptively randomized on the BBS design when delay does not exceed the total number of patients. Moreover, a trial with 1,000 patients still adaptively allocates more than half of its subjects when the expected time until a response is observed exceeds the enrollment period by a factor of four (Delay = 4,000).

On the other hand, for small trials the proportion of adaptively randomized patients barely surpasses 50% for $N = 50$ and 75% for $N = 100$ without any delay in observed outcomes. These rates are due to the requirement that at least 10 patients

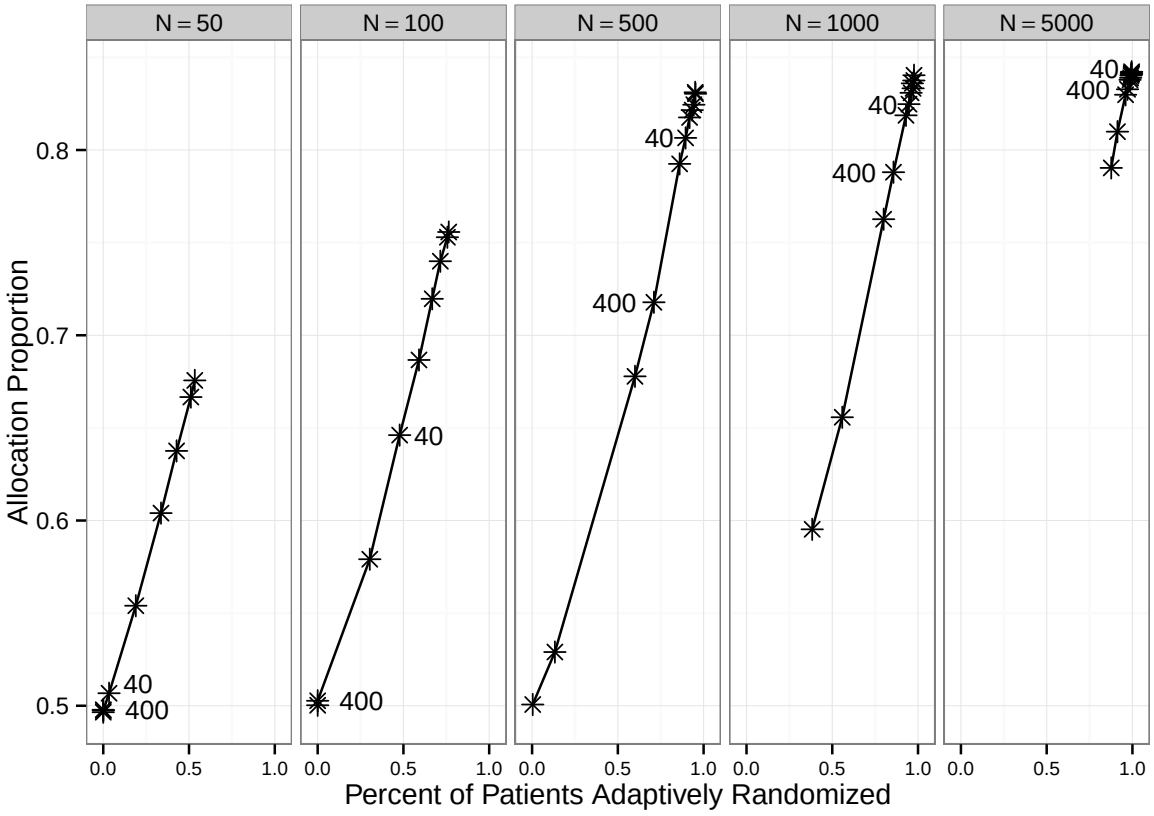
respond to each treatment prior to estimating an adaptive randomization probability. When the expected response delay is equivalent to the time it takes 40 patients to arrive, only 4% of 50 patients and 48% of 100 patients are adaptively randomized by the BBS design. Increasing the mean delay to 400, simulations confirm that no patients receive the benefit of an adaptive allocation for a small trial with relatively large delay.

Similarly, if expected delays greatly surpass the trial enrollment period, moderately sized trials will also lose their ability to adaptively randomize patients. For example, when $N = 500$ and the anticipated delay is eight times the anticipated the enrollment period, a trial may only adaptively randomize 14% of its participants—approximately 70 patients. Doubling the trial size or equivalently halving the relationship between expected delay and trial enrollment length allows 56% of enrolled subjects to be adaptively randomized for a clinical trial with 1,000 subjects. This compares favorably to the aforementioned trial of size 100 with no adaptive exposures for a similar delay to enrollment period relationship.

Figure 4.10 illustrates the nearly linear relationship between the proportion of patients adaptively randomized and the proportion of patients assigned to treatment A when treatment A is superior to treatment B and both treatment standard deviations equal one across all five simulated trial sizes. Equal allocation is excluded from this scatter plot. Additional levels of delay are included in Figure 4.10. Although specific delay values are not labeled, smaller delays correspond monotonically to increased allocation proportions and larger percents of patients adaptively randomized. Delays simulated and graphed include no delay, 1, 5, 10, 20, 40, 80, 400, 800, 4,000, and 8,000.

Figure 4.10 testifies to the strength of the relationship between a response-adaptive randomization's tendencies to expose patients to the better treatment and the pro-

Figure 4.10: Proportion of patients assigned to treatment A based on the proportion of patients adaptively randomized as N varies for $\mu_A = 1$, $\mu_B = 0$, and $\sigma_A = \sigma_B = 1$. All data points represent the BBS design at varying levels of exponentially distributed delay in treatment response. Delay values include immediate responses (0) as well as average delay times 1, 5, 10, 20, 40, 80, 400, 800, 4,000, and 8,000.



portion of enrolled patients who receive the adaptive randomization. Within each clinical trial size, greater delays in observing patient outcomes correspond monotonically to decreased allocation proportions. To illustrate, when $N = 50$, trials with no response delay adaptively allocated 53% of patients with a total of 68% of patients exposed to treatment A; trials with responses delayed by a mean rate of 20 additionally enrolled patients were able to adaptively allocate 19% of patients with a total of 55% of patients exposed to treatment A; and trials with responses delayed by a mean rate of 40 additionally enrolled patients were only able to adaptively allocate

3% of patients with a mere total of 51% of patients exposed to treatment A. All higher levels of delay resulted in negligible response-adaptive randomizations and a fixed equal allocation to both treatments.

Figure 4.10 also demonstrates the importance of trial enrollment size in this relationship. When a trial is unaffected by delayed responses, the larger the clinical trial, the higher the proportion of adaptively randomized patients will be. Similarly, for unaffected trials, higher the rates of exposure to the superior treatment correspond to larger trial size, with rates increasing rapidly from small to moderate trials and plateauing for large trials. Furthermore, for small enrollment populations, delays in treatment outcomes quickly reduce the BBS ethical allocation towards a fixed, even treatment split while, for large clinical trials, similar delays in treatment outcomes have relatively minor impacts to the adaptive and ethical nature of the design.

4.5 Application

In the past 25 years, Human Immunodeficiency Virus (HIV) infected an estimated 80 million people, half of whom are believed to have died from HIV-related causes [61, 26]. Newer medications such as antiretrovirals can help minimize viral transmission, extend life expectancies, and generally improve the quality of life of infected individuals—provided patients practice an adequate level of adherence to their prescribed therapy [13, 31]. Unfortunately, antiretroviral courses are associated with frequent non-compliance due to the timing and complexity of the medications as well as their side effects [13, 57]. Moreover, non-adherence may negatively impact the individual and the larger population via development and transmission of treatment-resistant viral strains. In a 2003 adherence intervention, [41] prospectively randomize 244 HIV patients on antiretroviral therapy to standard medication alone (treatment

M) or medication with additional educational and counseling therapy (treatment C). The authors examine the intervention's clinically and statistically significant effect on HIV RNA suppression (in RNA copies/mL) as a continuous measure, conditioned on baseline adherence as a prognostic factor (covariate Z).

Simulations in this section leverage findings from [41], including the underlying adherence covariate parameters and an improved reduction in HIV RNA by 0.36 copies/mL when medication is supplemented with educational and counseling therapy ($\mu_C - \mu_M$). In particular, baseline adherence is normally distributed throughout the patient population with mean $\mu_Z = 0.60$ and standard deviation $\sigma_Z = 0.49$. That is, the baseline adherence covariate is $Z \sim N(0.60, 0.49^2)$. Adherence impacts individual outcomes in both treatments, magnified by $\beta = 1.11$. Patients on standard medication therapy have a normally distributed treatment effect of $\mu_M = 0.22$ fewer HIV RNA copies/mL with a standard deviation of $\sigma_M = 0.54$ copies/mL. Hence the average medication-only outcome is a reduction in HIV RNA by $Y_M \approx \mu_M + \beta\mu_Z = 0.89$ copies/mL. Patients on medication therapy coupled with educational and counseling therapy have a treatment effect of $\mu_C = 0.58$ fewer HIV RNA copies/mL with a standard deviation of $\sigma_C = 0.47$ copies/mL, also normally distributed. Thus the average supplemented therapy outcome is a reduction in HIV RNA by $Y_C \approx \mu_C + \beta\mu_Z = 1.25$ copies/mL. The one-sided null hypothesis (combination treatment is no more effective than medication alone) is tested at significance $\alpha = 0.05$. For each scenario, 10,000 trials are simulated with 244 patients enrolled in each trial.

In [41], patients are assigned to treatment using a balanced design. As the randomization is not adaptive, delay in patient responses does not impact the allocation scheme. This is simulated via equal allocation. Simulations also consider how employing the BBS design might impact a clinical trial under similar conditions. For the BBS design, both patient enrollment (inter-arrival times) and patient outcomes

(response delay) are modeled via independent exponentially distributed random variables as described in Section 4.3.3. Patients are sequentially available for enrollment, arriving at an average rate of one person per unit of time. Patients respond at exponentially distributed times post-enrollment. In between a patient's enlistment and that same patient's response, 0, 61, 122, 183, or 244 additional patients would be expected to enroll in the clinical trial, based on the amount of delay incorporated in the specific simulation. Since the original trial's enrollment period is three months with expected response at six months, a delay of 488 is also simulated. All treated patient prognostic factors are considered, regardless of treatment outcome availability, for response-adaptive randomization estimates during the trial enrollment period.

Table 4.7 records patient assignment, power, and adaptive randomization for each combination of design and response delay over 10,000 simulated clinical trials. The first two columns indicate the design (BBS or Equal) and average delay (0, 61, 122, 183, 244, or 488) for the represented data. The next two columns contain the mean and standard deviation of the proportion of simulated patients assigned to the superior treatment C (medication supplemented with educational and counseling sessions) from a total enrollment of 244. The three subsequent columns represent the average number of simulated patients assigned to arm C (N_C), the average number of additional simulated patients assigned to the superior treatment than were assigned under equal allocation, and the average percent of additional simulated patients assigned to the superior arm for that row versus for equal allocation. The following two columns catalogue the mean proportion and proportion standard deviation of simulated trials which correctly reject the null hypothesis to conclude that educational and counseling therapy coupled with medication provide a greater reduction in HIV RNA copies/mL than does medication alone. The final two columns describe the average percent and standard deviation of simulated patients that were adaptively randomized during a

trial's enrollment period. This includes all patients who arrive after at least ten previously enrolled patients respond on each treatment arm.

For example, the first row of data represents BBS patient allocation when responses are immediately observable (Delay = 0). BBS randomizes 75% of simulated patients to medication supplemented with educational and counseling sessions with a standard deviation of 0.10. That means 182 patients would have received the superior treatment under BBS with no delay—60 more patients or an increase in 49% compared to the number of patients who would have received treatment C under equal allocation. Moreover, 95% of simulated BBS trials with no delay correctly reject the null hypothesis with a standard deviation of 0.21. Finally, 90% of simulated enrolling patients are randomized based on an estimated BBS adaptive randomization probability. By contrast, in the last row of data, equal allocation assigns only 50% of 244 simulated participants to the superior treatment. Hence only 122 patients would have received education and counseling sessions with no exposure benefits to leveraging this design. The point estimate of the simulated power of equal allocation is only slightly higher than that of the BBS designs at a 97% rejection rate and the standard deviation is also similar to but less than BBS at 0.18. No patients would have been randomized adaptively, regardless of whether or not responses were available to inform treatment mean and standard deviation estimates.

Figure 4.11 illustrates the trends in patient assignment from Table 4.7 across the six BBS delay values as well as equal allocation. Allocation means are represented by $*$ for the BBS design and by $\bullet$ for equal allocation. In both designs, immediate responses (Delay = 0) are represented by solid lines and for BBS increasing delay in outcomes are represented by increasingly dashed lines.

As seen in Table 4.7 and Figure 4.11, ethical patient allocation improves on the BBS design compared to balanced assignment. When delay is nonexistent, 75% of

Figure 4.11: Allocation proportions for HIV antiretroviral therapy adherence intervention by randomization design through 10,000 simulated trials. BBS allocation is represented by $*$ and equal allocation by $\bullet$. The points (bars) represent the proportion (standard deviation) of 244 patients assigned to the more effective combined therapy C of medication, education, and counseling versus the less effective treatment M of medication alone. The difference in treatment effects ($\mu_C - \mu_M$) is 0.36 HIV RNA copies/mL. The time until a patient responds to treatment is an exponentially distributed delay. As response delay increases, associated error bars are increasingly dashed. No delay is represented by a solid line (—) and a mean delay of 488 is represented by a dotted line ($\cdots$).

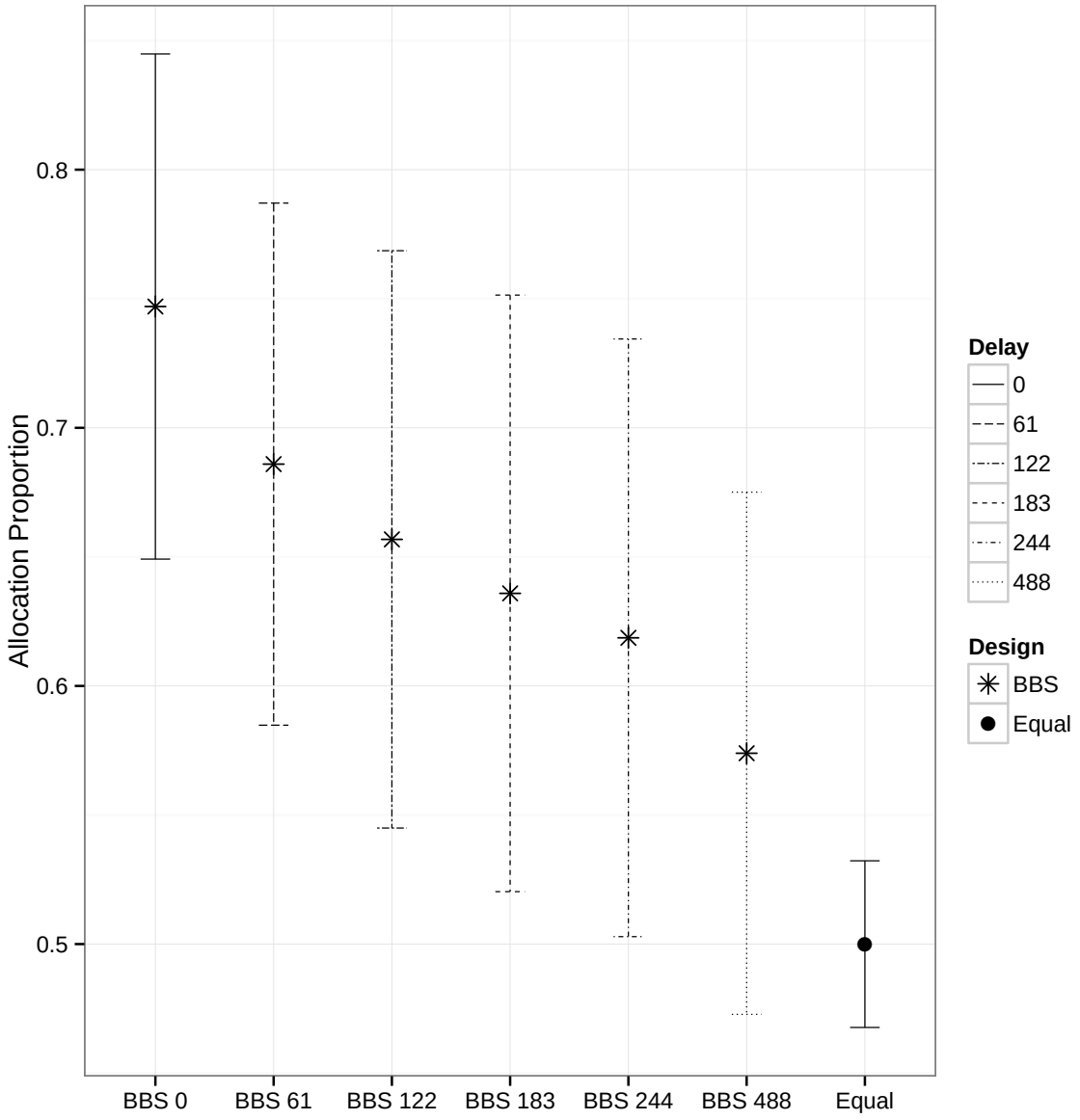
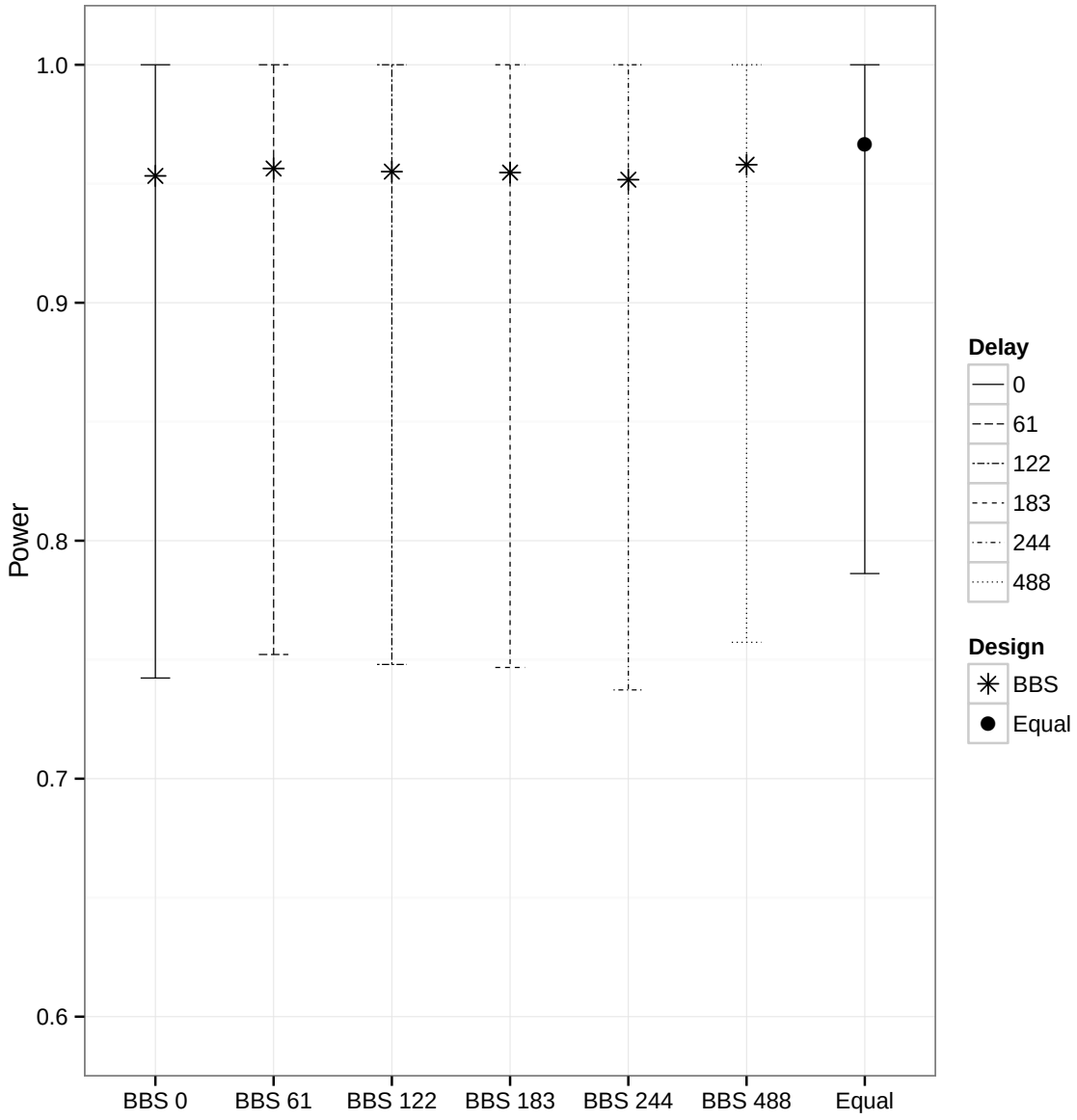


Figure 4.12: Power of HIV antiretroviral therapy adherence intervention by randomization design through 10,000 simulated trials. BBS allocation is represented by $*$ and equal allocation by $\bullet$. The point (bars) represent the proportion (standard deviation) of simulated antiretroviral therapy clinical trials which correctly reject the null hypothesis and conclude that an education and counseling intervention C in conduction with medication is more effective than medication treatment M alone. The difference in treatment effects ($\mu_C - \mu_M$) is 0.36 HIV RNA copies/mL. The time until a patient responds to treatment is an exponentially distributed delay. As response delay increases, associated error bars are increasingly dashed. No delay is represented by a solid line (—) and a mean delay of 488 is represented by a dotted line ($\cdots$).



patients in the trial are exposed to the superior treatment versus 50% on equal allocation. Even when the average response delay is equal to twice the enrollment period, 57% of patients still benefit from superior treatment assignment. In this case, 7% or 18 more patients are randomized to the better treatment with BBS than with equal allocation. Assignment variation, however, increases on the response-adaptive randomization scheme. The standard deviation of assignment proportions run from 0.10 to 0.12 for BBS but is a mere 0.03 for the nonadaptive equal allocation.

Figure 4.12 illustrates the trends in power from Table 4.7 across the six BBS delay values as well as equal allocation. Rejection rates are distinguished by design and delay parameter; $*$ for BBS and $\bullet$ for equal allocation with increasingly dashed lines as response delay increases. As seen in Table 4.7 and Figure 4.12, the mean proportion of clinical trials rejected differs by a mere 2% across all six scenarios. Moreover, the standard deviation of trials is also consistent at 0.21 in all but two of the BBS simulations and at 0.18 for equal allocation.

4.6 Discussion

4.6.1 Ethical Patient Allocation

Previous chapter results suggest the BBS design be considered as a viable alternative to equal allocation assuming all outcomes are immediately observable. This chapter investigates the impact to the design when treatment outcomes are not immediately observable. In particular, Sections 4.4.1 and 4.5 simulate the dampening effect delays in patient responses can have on the ethical nature of the BBS patient assignment.

In small trials ($N = 50$), an average delay in patient responses equal to 80% of the expected enrollment period essentially eliminates any benefit that could be

derived from using the BBS design. This is not surprising since only 4% of patients (8 individuals) were adaptively randomized in these scenarios. In a small trial, because at least 20 individuals must respond to treatment before the randomization proportions can deviate from $\frac{1}{2}$, an adequate number of responses are only available near the end of a trial.

For moderate sized trials, an 80% delay can still allocate an additional 20% of patients to the superior treatment. For example, in simulations with $N = 500$ and mean delay equal to 80% of the trial enrollment window, at least 10 responses are collected on each treatment arm after only 30% of patients have been randomized. In this case, approximately 70% of patients are exposed to the superior therapy. Moreover, for moderate trials, even when the expected delay exceeds the treatment window, simulations show that patients benefit from the adapted ratios based on the responses received. For example, with a 1,000 patient enrollment and a delay four times that of the enrollment window, half of the patients in a clinical trial may be adaptively randomized and, in fact, an additional 15% of patients can receive the better treatment compared to a similar cohort under equal allocation.

For large enough samples, treatment variance has a greater impact on patient assignment than delay in patient responses, provided an adequate proportion of patients are adaptively randomized. In particular, for large Phase III trials ($N = 5{,}000$ patients), exposure to the superior arm may increase by 10% as σ_A decreases by 0.5. Within a fixed level of σ_A, however, a moderate delay only decreases the proportion of ethical assignments by a few percentage points. When treatments are identical, the impact of delay on patient allocation is negligible at all trial size and variance combinations.

The simulation results presented in Sections 4.4.1 and 4.5 substantiate that the BBS design will assign more patents to the better treatment when a treatment dif-

ference exists, provided an adequate number of patient outcomes become available during the trial enrollment period. In the adaptive version of the small HIV intervention simulated, between 18 and 45 additional patients would receive the improved intervention even with delayed response. That is, the equal allocation clinical trial could expose at least 7% and up to 19% of the 244 enrollees to a better therapy by leveraging BBS randomization, even with significant delays in observed outcomes. The more patients left in the enrollment period after the design begins adaptively randomizing, the higher the proportion of ethically allocated patients could be. Researchers must weigh the benefits of improved patient exposure to treatment against the additional complication of implementing an adaptive randomization schema and the increased risk of a type I error.

4.6.2 Rejection Rates

Previous chapters determine that the BBS design maintains power levels relatively well compared to equal allocation; however, the variance in BBS rejection rates increase slightly versus equal allocation. Moreover, type I error rates also risk inflation, particularly for small trial sizes. This chapter considers the impact to the design's power and type I error rates when delayed patient responses are introduced to the design.

Simulations presented in Sections 4.4.2 and 4.5 further testify that for moderate to large clinical trials, the BBS design can sustain power levels similar to that of equal allocation. Redesigning the [41] clinical trial to leverage BBS, the power to detect the observed difference decreases by only 1% to 2% compared to equal allocation with standard deviation increase of 0.02 to 0.03. In general, the larger the delay in treatment response, the less adaptive the BBS randomization scheme is and the more

this design behaves like equal allocation. Power levels appear to rise as expected response delays increase. For smaller enrollment sizes, the power of the BBS design can drop slightly compared to equal allocation; however, with delayed observations BBS trial power can match and even exceed that of equal allocation. Likewise, type I error rates can be inflated by an additional four percent above equal allocation for the BBS design employed in smaller trials when outcomes are immediately available. This augmented risk also diminishes when responses are delayed and the design adaptivity decreases.

The BBS design can maintain relatively high power to detect a difference while simultaneously exposing more patients to the superior treatment. These qualities are desirable in an allocation algorithm for optimizing patient care while maintaining randomization in a controlled clinical trial. These trade-offs must be considered when selecting a randomization scheme, especially in the case of medical research.

4.6.3 Delay Dampening

Both Chapter 2 and 3 ignore the potential impact of delay by only considering simulations and situations where patient outcomes are immediately observable. Unfortunately, in many areas of medicine patients respond to treatment only after a period of time has passed from the initial therapy or exposure. Section 4.2.3 touches on three areas where delayed responses are prevalent, but the appearance of delayed outcomes in clinical trials is vast [7, 11, 28]. This chapter considers the impact to the properties of a clinical trial leveraging BBS when delayed patient responses are introduced to the design.

Section 4.6.1 summarizes the relationship between the response-adaptive randomization of the BBS design and the expected delay in observing treatment responses.

The adaptive nature of small trials is already constrained by the initial number of patients needed to estimate an adaptive assignment probability. Delayed responses further exacerbate this situation. Even delays which are modest relative to total enrollment can overwhelm a trial's ability to adaptively assign patients. Nevertheless, for moderately sized Phase III clinical trials, the ethical nature of the adaptive assignment endure except when the expected delay drastically exceeded the total enrollment period (for example, by a factor of 8).

In scenarios with extended periods of time elapsing before patient responses can be recorded, researchers may look to shorten this delay even for nonadaptive medical trials. Common approaches to circumvent this issue involve censoring data after a certain period, dichotomizing outcomes, and leveraging surrogate measures or alternative biomarkers [18, 22, 63]. While none of these solutions are simulated, two of these options may be suitable for the BBS design.

First, instead of waiting until a patient's full response is available, it may be possible to include a patient's response at a pre-specified time. For example, instead of waiting for a six month reading of a particular outcome, researchers may instead leverage the patient's progress of this measure at month three to help estimate the adaptive allocation proportion in the interim three months until the true response is ready. At month six, when the final patient endpoint becomes available, that value may replace the three month indicator in the randomization probability estimates, if patient enrollment is still ongoing.

Inclusion of the three-month measure could allow the redesigned trial in Section 4.5 to benefit from allocation levels similar to the BBS design with delay of 244 (equal to the three month enrollment period) rather than the delay of 488 (equivalent to the six month observed endpoint). That is, 5% of patients could be assigned to the superior treatment in addition to the 7% already benefiting from the BBS design

with a six month delay. Up to 29 people would be exposed to the better therapy above and beyond an equal allocation randomization.

If the three month observation is expected to differ drastically from the six month outcome—as may be the case in situations where a six month outcome is determined to be the earliest acceptable final indicator, then interim observations may skew the assignment probabilities away from the desired allocation ratio. The skewed proportion, however, would still be more ethical than balanced randomization provided that the skewed ratio falls between 0.5 and the true ratio. Simulations and a thorough understanding of the endpoint evolution over time are important factors in selecting an earlier estimate.

Second, instead of employing the true treatment outcome for the response-adaptive randomization probability, an alternative indicator for the outcome could be utilized instead. For instance, if the final outcome of interest is the time until an event occurs, a surrogate measure might be a predictive biomarker correlated with the future occurrence of an event. Of course, the alternative indicator is unlikely to be a perfect predictor. Moreover, the typical biomarker might be otherwise unsuitable: dichotomous or with known large variance. In fact, the reasons keeping the predictive indicator from being the final endpoint could also preclude its use in informing allocation proportions. As with many clinical trial choices, a substitution of this nature must be carefully considered.

Section 4.6.2 highlights how the BBS design's ability to preserve rejection rates relative to equal allocation is actually improved when delay is incorporated into a clinical trial. While the trend manifesting in Section 4.5 is more reminiscent of noise than of convergence to equal allocation, the power levels simulated in Section 4.4.2 are indeed increasing as delay increases. Furthermore, the impact of delay on both general simulations and a redesigned clinical trial via decreasingly ethical allocation

rates is unambiguous. When a response-adaptive randomization such as the BBS is able to maintain adequate levels of power, such a design should be considered in lieu of equal allocation. The risks associated with response-adaptive randomization and delayed responses must be balanced against the benefits to the medical community and to the individual patients involved in a study when selecting a study design.

4.6.4 Recommendations

As discussed in Chapters 2 and 3, the primary goal of many Phase III clinical trials is to determine the best care for future patients from a pair or group of similar therapies. The "gold standard" for such research includes assigning patients to treatments in equal proportions for the duration of the experiment, ignoring accumulating evidence about each treatment's effectiveness [10, 21, 36]. The importance of the individuals being treated clashes with this historically prevalent ideal [45, 46, 58]. Within a clinical trial such as [41], exposure to the superior treatment can be life changing. Balanced randomization studies bear the opportunity cost of introducing more of the population to the better therapy.

Through detailed simulations, the benefits and risks of the BBS design are clear. This response-adaptive randomization scheme leverages data garnered during a study to expose more patients to the better treatment arm throughout the course of the trial, rather than waiting for the trial conclusion. When the difference in treatment mean effects are large relative to treatment variance, the ethical gains can be substantial. There are three main downsides to utilizing BBS within a clinical trial. First and foremost, the patient assignment—while more ethical than equal allocation—has a higher variability than a nonadaptive randomization scheme. Second, the type I error rates can be inflated compared to the rates of incorrectly rejecting the null

hypothesis when employing equal allocation. This is particularly true for small trials and is mitigated for moderate to large enrollment levels ($N \geq 500$). The absolute increased risk is a few percentage points and, moreover, may not be clinically relevant as discussed in Chapter 2. Third, the power of the BBS design can be marginally less than the rates of correct rejection of the null hypothesis under equal allocation. Again, the impact is particularly prevalent in smaller sized studies.

The BBS design applies to situations where there are two competing treatments with continuous outcomes and researchers would prefer a more ethical patient assignment than equal allocation. To employ BBS randomization, the primary indication of a treatment's superiority should be the treatment mean effect, net of covariate impacts. Previous chapters demonstrate the design's relevance and ethicality when treatment variances are unknown or unequal as well as when covariate impacts differ by treatment—provided the covariate-treatment interaction does not contradict the primary indicator of treatment superiority. This chapter confirms that delays in patient responses are easily incorporated into the response-adaptive design. Furthermore, this chapter concludes that BBS is more ethical than equal allocation when delay in outcomes are modest relative to the enrollment period. When delays are extensive such that no patients are adaptively assigned throughout the trial enrollment, the BBS design is equivalent to equal allocation.

Redesigning a clinical trial to decrease HIV viral loads with only 244 patients and large delays in response, the BBS design would expose an extra 7% of patients to the better treatment with a mere 1% reduction in power versus the original balanced assignment. The potential benefit of the BBS design in treating patients during a trial is clear. Nevertheless, in any situation where adaptive randomization is being considered, the benefit of ethical patient exposure must be weighed against the increased complications of implementing an adaptive design. Simulations to increase researcher

understanding should included anticipated treatment parameters, trial design factors such as sample size and enrollment length, and delays to patient responses.

4.7 Appendix

Table 4.1: Representation of arrival and response times for a clinical trial with $N = 50$ patients with expected inter-arrival time of 1 unit of time and an expected delay of 40 units of time. Patient inter-arrival times are independently and identically distributed exp(1). Patients are assumed to be randomized and treated immediately upon arrival. The arrival, randomization, and treatment of the first patient is time zero. Patient response delays are independently and identically distributed exp(40). A patient's response time depends only on the time of that patient's randomization (treatment exposure) and the delay in that patient's response. In this scenario patient 3 would be the 20^{th} patient to respond to treatment at 44.2 units of time. All patients assigned after $m = 10$ responses have been observed on each treatment arm would be adaptively randomized under the BBS design.

Patient	Inter-arrival Time	Randomization Time	Response Delay	Response Time	Response Order
1	—	0.0	30.9	30.9	11
2	0.1	0.1	33.9	34.0	14
3	3.8	3.9	40.3	44.2	**20**[1]
4	2.6	6.5	8.0	14.5	2
5	0.2	6.7	24.1	30.8	10
6	1.0	7.6	46.7	54.3	28
7	0.6	8.2	7.5	15.7	3
8	1.7	10.0	18.8	28.8	8
9	0.7	10.7	51.6	62.3	32
10	0.6	11.3	24.0	35.3	18
11	0.5	11.8	8.1	20.0	4
12	0.1	11.9	10.7	22.6	6
13	0.7	12.6	22.0	34.7	17
14	0.1	12.7	111.4	124.1	44
15	0.6	13.3	9.9	23.2	7
16	0.4	13.7	0.3	14.0	1
17	0.4	14.1	21.4	35.5	19
18	2.0	16.1	18.3	34.3	15
19	0.3	16.4	15.8	32.2	13
20	2.4	18.8	45.9	64.7	33

Continued on next page

[1]Patient 3 is the 20th patient to respond and response occurs at 44.2 time units into the trial.

Table 4.1 Second header – *Cont.*

Patient	Inter-arrival Time	Randomization Time	Response Delay	Response Time	Response Order
21	0.5	19.2	2.6	21.9	5
22	2.2	21.5	9.9	31.4	12
23	0.6	22.0	132.1	154.2	48
24	1.2	23.2	11.2	34.4	16
25	0.8	24.1	55.2	79.3	36
26	1.3	25.3	65.9	91.2	40
27	0.5	25.8	24.4	50.2	25
28	0.2	26.0	61.9	87.9	39
29	1.1	27.1	152.0	179.1	50
30	0.1	27.1	24.7	51.9	26
31	0.1	27.2	2.8	30.0	9
32	0.2	27.4	28.0	55.5	30
33	1.7	29.2	136.6	165.8	49
34	1.7	30.9	43.4	74.3	34
35	0.6	31.5	23.7	55.2	29
36	0.0	31.6	55.2	86.8	37
37	1.0	32.6	16.0	48.6	24
38	0.9	33.5	114.1	147.7	47
39	0.8	34.4	44.5	78.9	35
40	1.5	35.8	51.4	87.2	38
41	2.8	38.7	63.7	102.4	42
42	0.7	39.4	100.8	140.2	46
43	1.3	40.6	86.7	127.4	45
44	1.3	42.0	3.5	45.5	21
45	0.8	42.8	16.8	59.6	31
46	1.2	44.0	8.2	52.2	27
47[2]	2.0	46.0	1.3	47.3	23
48	0.4	46.4	45.1	91.4	41
49	0.1	46.5	0.6	47.1	22
50	2.2	48.7	59.7	108.4	43

[2]Patients 47 – 50 would all be randomized after 20 patients had responded to treatment. Assuming $m = 10$ observations had been collected from each trial arm, the BBS design would allocate these remaining four patients to treatment using an adaptive assignment proportion.

Table 4.2: Proportion (SD) of patients assigned to Treatment A under BBS and equal allocation (Equal) with $\mu_A = \mu_B = 0$. Randomization design and response delay are given in the two leftmost columns. Allocation proportion and standard deviation across 1,000 simulations of trial size N given in remaining five columns for each standard deviation combination simulated.

Design	Delay	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
		$\mu_A = 0,\ \mu_B = 0,\ \sigma_A = 0.5,\ \sigma_B = 1.0$				
BBS	0	0.51 (0.12)	0.52 (0.13)	0.52 (0.09)	0.51 (0.07)	0.51 (0.04)
	40	0.50 (0.07)	0.50 (0.13)	0.51 (0.09)	0.51 (0.06)	0.50 (0.03)
	400	0.50 (0.07)	0.50 (0.05)	0.52 (0.15)	0.51 (0.12)	0.51 (0.04)
	4000	0.50 (0.07)	0.50 (0.05)	0.50 (0.06)	0.51 (0.16)	0.51 (0.10)
Equal	0	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
		$\mu_A = 0,\ \mu_B = 0,\ \sigma_A = 1.0,\ \sigma_B = 1.0$				
BBS	0	0.50 (0.12)	0.50 (0.13)	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)
	40	0.50 (0.07)	0.51 (0.12)	0.50 (0.08)	0.50 (0.06)	0.50 (0.02)
	400	0.50 (0.07)	0.50 (0.05)	0.50 (0.14)	0.50 (0.11)	0.50 (0.03)
	4000	0.50 (0.07)	0.50 (0.05)	0.50 (0.06)	0.50 (0.15)	0.50 (0.09)
Equal	0	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
		$\mu_A = 0,\ \mu_B = 0,\ \sigma_A = 1.5,\ \sigma_B = 1.0$				
BBS	0	0.49 (0.13)	0.49 (0.13)	0.49 (0.08)	0.50 (0.05)	0.50 (0.03)
	40	0.50 (0.07)	0.50 (0.12)	0.50 (0.08)	0.50 (0.06)	0.50 (0.02)
	400	0.50 (0.07)	0.50 (0.05)	0.49 (0.12)	0.49 (0.09)	0.50 (0.03)
	4000	0.50 (0.07)	0.50 (0.05)	0.50 (0.06)	0.49 (0.14)	0.50 (0.07)
Equal	0	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)

Table 4.3: Proportion (SD) of patients assigned to Treatment A under BBS and equal allocation (Equal) with $\mu_A = 1$ and $\mu_B = 0$. Randomization design and response delay are given in the two leftmost columns. Allocation proportion and standard deviation across 1,000 simulations of trial size N given in remaining five columns for each standard deviation combination simulated.

Design	Delay	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
		$\mu_A = 1,\ \mu_B = 0,\ \sigma_A = 0.5,\ \sigma_B = 1.0$				
BBS	0	0.70 (0.09)	0.83 (0.07)	0.93 (0.05)	0.95 (0.04)	0.97 (0.02)
	40	0.51 (0.08)	0.66 (0.09)	0.87 (0.05)	0.92 (0.04)	0.96 (0.02)
	400	0.50 (0.07)	0.50 (0.05)	0.75 (0.10)	0.83 (0.07)	0.93 (0.03)
	4000	0.50 (0.07)	0.50 (0.05)	0.53 (0.06)	0.67 (0.11)	0.87 (0.06)
Equal	0	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
		$\mu_A = 1,\ \mu_B = 0,\ \sigma_A = 1.0,\ \sigma_B = 1.0$				
BBS	0	0.68 (0.09)	0.76 (0.09)	0.83 (0.06)	0.84 (0.04)	0.84 (0.02)
	40	0.51 (0.07)	0.65 (0.10)	0.81 (0.06)	0.82 (0.04)	0.84 (0.02)
	400	0.50 (0.07)	0.50 (0.05)	0.72 (0.10)	0.79 (0.07)	0.83 (0.02)
	4000	0.50 (0.07)	0.50 (0.05)	0.53 (0.06)	0.66 (0.11)	0.81 (0.06)
Equal	0	0.49 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)
		$\mu_A = 1,\ \mu_B = 0,\ \sigma_A = 1.5,\ \sigma_B = 1.0$				
BBS	0	0.64 (0.10)	0.70 (0.09)	0.75 (0.05)	0.76 (0.04)	0.76 (0.02)
	40	0.51 (0.07)	0.62 (0.10)	0.74 (0.05)	0.75 (0.04)	0.76 (0.02)
	400	0.50 (0.07)	0.50 (0.05)	0.69 (0.10)	0.73 (0.06)	0.76 (0.02)
	4000	0.50 (0.07)	0.50 (0.05)	0.53 (0.05)	0.64 (0.11)	0.75 (0.05)
Equal	0	0.50 (0.07)	0.50 (0.05)	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)

Table 4.4: Proportion (SD) of trials where the null hypothesis is incorrectly rejected under BBS and equal allocation (Equal) with $\mu_A = \mu_B = 0$. Randomization design and response delay are given in the two leftmost columns. Rejection rate and standard deviation across 1,000 simulations of trial size N given in remaining five columns for each standard deviation combination simulated.

Design	Delay	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
		$\mu_A = 0,\ \mu_B = 0,\ \sigma_A = 0.5,\ \sigma_B = 1.0$				
BBS	0	0.10 (0.29)	0.10 (0.30)	0.09 (0.29)	0.06 (0.24)	0.07 (0.25)
	40	0.07 (0.25)	0.08 (0.27)	0.07 (0.26)	0.06 (0.24)	0.06 (0.24)
	400	0.07 (0.25)	0.06 (0.24)	0.08 (0.27)	0.08 (0.27)	0.07 (0.26)
	4000	0.08 (0.27)	0.05 (0.22)	0.07 (0.26)	0.07 (0.25)	0.07 (0.26)
Equal	0	0.06 (0.24)	0.07 (0.26)	0.07 (0.25)	0.05 (0.21)	0.06 (0.24)
		$\mu_A = 0,\ \mu_B = 0,\ \sigma_A = 1.0,\ \sigma_B = 1.0$				
BBS	0	0.07 (0.26)	0.06 (0.23)	0.06 (0.24)	0.05 (0.23)	0.05 (0.22)
	40	0.05 (0.22)	0.07 (0.25)	0.06 (0.24)	0.06 (0.23)	0.05 (0.22)
	400	0.07 (0.25)	0.07 (0.25)	0.06 (0.23)	0.06 (0.23)	0.07 (0.25)
	4000	0.08 (0.26)	0.06 (0.24)	0.05 (0.22)	0.06 (0.23)	0.05 (0.23)
Equal	0	0.07 (0.25)	0.06 (0.24)	0.05 (0.22)	0.06 (0.24)	0.04 (0.20)
		$\mu_A = 0,\ \mu_B = 0,\ \sigma_A = 1.5,\ \sigma_B = 1.0$				
BBS	0	0.06 (0.23)	0.06 (0.23)	0.05 (0.22)	0.05 (0.22)	0.05 (0.21)
	40	0.07 (0.25)	0.05 (0.23)	0.05 (0.22)	0.06 (0.23)	0.06 (0.23)
	400	0.06 (0.24)	0.06 (0.23)	0.05 (0.22)	0.05 (0.22)	0.05 (0.22)
	4000	0.07 (0.25)	0.07 (0.25)	0.06 (0.23)	0.07 (0.25)	0.06 (0.24)
Equal	0	0.06 (0.24)	0.07 (0.25)	0.05 (0.21)	0.04 (0.19)	0.05 (0.22)

Table 4.5: Proportion (SD) of trials where the null hypothesis is correctly rejected under BBS and equal allocation (Equal) with $\mu_A = 1$ and $\mu_B = 0$. Randomization design and response delay are given in the two leftmost columns. Rejection rate and standard deviation across 1,000 simulations of trial size N given in remaining five columns for each standard deviation combination simulated.

Design	Delay	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
		$\mu_A = 1,\ \mu_B = 0,\ \sigma_A = 0.5,\ \sigma_B = 1.0$				
BBS	0	0.90 (0.30)	0.99 (0.11)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	40	0.93 (0.25)	0.99 (0.08)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	400	0.93 (0.25)	1.00 (0.05)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	4000	0.93 (0.25)	1.00 (0.07)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
Equal	0	0.92 (0.27)	1.00 (0.04)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
		$\mu_A = 1,\ \mu_B = 0,\ \sigma_A = 1.0,\ \sigma_B = 1.0$				
BBS	0	0.76 (0.43)	0.96 (0.19)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	40	0.79 (0.41)	0.96 (0.19)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	400	0.78 (0.41)	0.96 (0.19)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	4000	0.78 (0.42)	0.97 (0.18)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
Equal	0	0.77 (0.42)	0.96 (0.19)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
		$\mu_A = 1,\ \mu_B = 0,\ \sigma_A = 1.5,\ \sigma_B = 1.0$				
BBS	0	0.63 (0.48)	0.88 (0.33)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	40	0.63 (0.48)	0.87 (0.34)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	400	0.62 (0.49)	0.85 (0.36)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	4000	0.63 (0.48)	0.88 (0.32)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
Equal	0	0.64 (0.48)	0.87 (0.34)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)

Table 4.6: Proportion (SD) of patients adaptively randomized under BBS. Randomization design and response delay are given in the two leftmost columns. The proportion and standard deviation of patients adaptively randomized across 1,000 simulations of trial size N given in remaining five columns.

Design	Delay	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
BBS	0	0.53 (0.07)	0.76 (0.03)	0.95 (0.01)	0.98 (0.00)	1.00 (0.00)
	40	0.04 (0.06)	0.48 (0.07)	0.89 (0.01)	0.95 (0.01)	0.99 (0.00)
	400	0.00 (0.00)	0.00 (0.00)	0.71 (0.04)	0.86 (0.02)	0.97 (0.00)
	4000	0.00 (0.00)	0.00 (0.00)	0.14 (0.10)	0.56 (0.05)	0.91 (0.01)

Table 4.7: HIV adherence intervention summary characteristics by randomization design and response delay over 10,000 simulated trials. P_C (SD) is the proportion (standard deviation) of patients assigned to the superior treatment of combined medication, education, and counseling (C). N_C is the number of patients assigned to treatment C. # is the number of additional patients randomized to the superior treatment via the current design compared to equal allocation assignment (Equal). % is the percent of additional patients assigned to the superior treatment under the current design versus randomization under equal allocation. Power (SD) is the proportion (standard deviation) of trials which correctly reject the null hypothesis that the addition of education and counseling does not improve the reduction in HIV RNA copies/mL over medication alone. AR (SD) is the proportion (standard deviation) of patients which were adaptively randomized within a trial.

Design	Delay	P_C (SD)	N_C	#	%	Power (SD)	AR (SD)
BBS	0	0.75 (0.10)	182	60	49%	0.95 (0.21)	0.90 (0.01)
	61	0.69 (0.10)	167	45	37%	0.96 (0.20)	0.74 (0.03)
	122	0.66 (0.11)	160	38	31%	0.96 (0.21)	0.66 (0.05)
	183	0.64 (0.12)	155	33	27%	0.95 (0.21)	0.59 (0.06)
	244	0.62 (0.12)	151	29	24%	0.95 (0.21)	0.53 (0.06)
	488	0.57 (0.10)	140	18	15%	0.96 (0.20)	0.35 (0.08)
Equal	0	0.50 (0.03)	122	0	0%	0.97 (0.18)	0.00 (0.00)

4.8 References

[1] Actelion. *Study of Macitentan (ACT-064992) on Morbidity and Mortality in Patients With Symptomatic Pulmonary Arterial Hypertension (SERAPHIN).* URL: https://clinicaltrials.gov/ct2/show/NCT00660179.

[2] Aziz, N. et al. "Time to viral load suppression in antiretroviral-naive and -experienced HIV-infected pregnant women on highly active antiretroviral therapy: implications for pregnant women presenting late in gestation". In: *British Journal of Obstetrics and Gynaecology* 120.12 (2013), pp. 1534–1547.

[3] Bai, Z.D., Hu, F., and Rosenberger, W.F. "Asymptotic properties of adaptive designs for clinical trials with delayed response". In: *The Annals of Statistics* 30.1 (2002), pp. 122–139.

[4] Bandyopadhyay, Uttam and Biswas, A. "Adaptive designs for normal response with prognostic factors". In: *Biometrika* 88 (2001), pp. 409–419.

[5] Bartfai, A. et al. "The protocol and design of a randomised controlled study on training of attention within the first year after acquired brain injury". In: *BMC Neurology* 14 (2014), pp. 102–110.

[6] Bayer. *A Study to Evaluate Efficacy and Safety of Oral BAY63-2521 in Patients With Pulmonary Arterial Hypertension (PAH) (PATENT-1).* URL: https://clinicaltrials.gov/ct2/show/NCT00810693.

[7] Berry, S.M. et al. *Bayesian Adaptive Methods for Clinical Trials.* 1st ed. Chapman & Hall/CRC Biostatistics Series, 2010.

[8] Biswas, A. "Generalized Delayed Response in Randomized Play-the-Winner Rule". In: *Communications in Statistics, Simulation and Computation* 32.1 (2003), pp. 259–274.

[9] Brolmann, F.E. et al. "Randomized clinical trial of donor-site wound dressings after split-skin grafting". In: *British Journal of Surgery* 100.5 (2013), pp. 619–627.

[10] Byar, D.P. "Ethics and statistics in randomized clinical trials: Comment". In: *Statistical Science* 6.1 (1991), pp. 65–68.

[11] Byar, D.P. et al. "Randomized Clinical Trials: Perspectives on Some Recent Ideas". In: *New England Journal of Medicine* 295.2 (1976), pp. 74–80.

[12] Calmy, A. et al. "HIV Viral Load Monitoring in Resource-Limited Regions: Optional or Necessary?" In: *Clinical Infectious Diseases* 44.1 (2007), pp. 128–134.

[13] M. A. Chesney, M. Morin, and L. Sherr. "Adherence to HIV combination therapy". In: *Soc Sci Med* 50.11 (2000), pp. 1599–1605.

[14] The Medicines Company. *Clevidipine in the Treatment of Patients With Acute Hypertension and Intracerebral Hemorrhage (ACCELERATE)*. URL: `https://clinicaltrials.gov/ct2/show/NCT00666328`.

[15] The Medicines Company. *Study of Clevidipine Assessing Its Preoperative Antihypertensive Effect in Cardiac Surgery (ESCAPE-1)*. URL: `https://clinicaltrials.gov/ct2/show/NCT00093249`.

[16] Omeros Corporation. *Efficacy and Safety of OMS103HP in Patients Undergoing Allograft Anterior Cruciate Ligament (ACL) Reconstruction*. URL: `https://clinicaltrials.gov/ct2/show/NCT00226759`.

[17] Instituto Bioclon S.A. de C.V. *A Phase III Multicenter Clinical Trial of Analatro® [Antivenin Latrodectus (Black Widow) Equine Immune F(ab)2] in Patients With Systemic Latrodectism*. URL: https://clinicaltrials.gov/show/NCT00657540.

[18] Fisher, Lloyd D. "Advances in clinical trials in the twentieth century". In: *Annual Review of Public Health* 20.1 (1999), pp. 109–125.

[19] Food and Drug Administration. *Critical Path Initiative: Transforming the way FDA-regulated products are developed, evaluated, and manufactured*. Report on Key Achievements in 2009. 2010.

[20] Food and Drug Administration. "Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics". Draft Document. 2010.

[21] Foster, C. *The Ethics of Medical Research on Humans*. 1st ed. University Press, Cambridge, UK: Cambridge University Press, 2001.

[22] Friedman, L.M., Furberg, C., and DeMets, D.L. *Fundamentals of Clinical Trials*. 3rd ed. Springer, 1998.

[23] Hasatsri, S, Angspatt, A., and Aramwit, P. "Randomized Clinical Trial of the Innovative Bilayered Wound Dressing Made of Silk and Gelatin: Safety and Efficacy Tests Using a Split-Thickness Skin Graft Model". In: *Evidence-Based Complementary and Alternative Medicine* 2015 (2015).

[24] Methodist Healthcare. *The Efficacy and Safety of Cleviprex in Ventriculostomy Patients Requiring IV Antihypertensive Therapy*. URL: https://clinicaltrials.gov/ct2/show/NCT01042574.

[25] Hinkelmann, K. *Design and Analysis of Experiments, Vol. 3: Special Designs and Applications*. 1st ed. Vol. 3. Wiley-Blackwell, 2011.

[26] Joint United Nations Programme on HIV and AIDS (UNAIDS). *Fact sheet*. URL: `http://www.unaids.org/en/resources/campaigns/globalreport2013/factsheet` (visited on 02/16/2015).

[27] Hu, F. and Rosenberger, W.F. *The Theory of Response-Adaptive Randomization in Clinical Trials*. 1st ed. Hoboken, NJ: John Wiley & Sons, Inc., 2006.

[28] Hu, F. et al. "Doubly-Adaptive Biased Coin Designs with Delayed Responses". In: *Canadian Journal of Statistics* 36.4 (2008), pp. 541–559.

[29] Hu, Z.C. et al. "Randomized clinical trial of autologous skin cell suspension combined with skin grafting for chronic wounds". In: *British Journal of Surgery* 102.2 (2015), pp. 117–123.

[30] Inova Health Care Services. *Pain Management in Response to Exparel vs. Standard Bupivicaine (VATS Exparel)*. URL: `https://clinicaltrials.gov/ct2/show/study/NCT02499159`.

[31] H. Knobel et al. "Validation of a simplified medication adherence questionnaire in a large cohort of HIV-infected patients: the GEEMA Study". In: *AIDS* 16.4 (2002), pp. 605–613.

[32] Lopes, R.T. et al. "Time to improve and recover from depressive symptoms and interpersonal problems in a clinical trial". In: *Clinical Psychology and Psychotherapy* 22.2 (2015), pp. 97–105.

[33] Macrocure Ltd. *Evaluation of CureXcell(R) in Treating Chronic Venous Leg Ulcers*. URL: `https://clinicaltrials.gov/ct2/show/NCT02130310`.

[34] McLaughlin, V.V. et al. "End Points and Clinical Trial Design in Pulmonary Arterial Hypertension". In: *Journal of the American College of Cardiology* 54.1 (2009), S97–S107.

[35] Inc. Medivation. *Safety and Efficacy Study of MDV3100 in Patients With Castration-Resistant Prostate Cancer Who Have Been Previously Treated With Docetaxel-based Chemotherapy (AFFIRM)*. URL: `https://clinicaltrials.gov/ct2/show/NCT00974311`.

[36] Palmer, C.R. "Ethics, data-dependent designs, and the strategy of clinical trials: time to start learning-as-we-go?" In: *Statistical Methods in Medical Research* 11 (2002), pp. 381–402.

[37] Phillis, A.N. et al. "HIV Viral Load Response to Antiretroviral Therapy According to the Baseline CD4 Cell Count and Viral Load". In: *Journal of the American Medical Association* 286.20 (2001), pp. 2560–2567.

[38] Pocock, S.J. "Allocation of Patients to Treatment in Clinical Trials". In: *Biometrics* 35.1 (1979), pp. 183–197.

[39] Pollack, C.V. et al. "Clevidipine, an intravenous dihydropyridine calcium channel blocker, is safe and effective for the treatment of patients with acute severe hypertension". In: *Annals of Emergency Medicine* 53.3 (2009), pp. 329–328.

[40] Pong, A. and Chow, S.C., eds. *Handbook of Adaptive Designs in Pharmaceutical and Clinical Development*. Boca Raton: CRC Press, 2011.

[41] C. Pradier et al. "Efficacy of an educational and counseling intervention on adherence to highly active antiretroviral therapy: French prospective controlled study". In: *HIV Clin Trials* 4.2 (2003), pp. 121–131.

[42] Rosenberger, W.F. "New Directions in Adaptive Designs". In: *Statistical Science* 11 (1996), pp. 137–149.

[43] Rosenberger, W.F. and Hu, F. "Maximizing power and minimizing treatment failures in clinical trials". In: *Clinical Trials* 1.2 (2004), pp. 141–147.

[44] Rosenberger, W.F. and Lachin, J. M. *Randomization in Clinical Trials Theory and Practice*. 1st ed. Wiley-Interscience, 2002.

[45] Royall, R.M. "Ethics and statistics in randomized clinical trials". In: *Statistical Science* 6.1 (1991), pp. 52–62.

[46] Royall, R.M. "Ethics and statistics in randomized clinical trials: Rejoinder". In: *Statistical Science* 6.1 (1991), pp. 83–88.

[47] Mexico Coordinación de Investigación en Salud. *Analgesic Efficacy of Two Concentrations of Bupivacaine in Women in Labor (EBWL)*. URL: https://clinicaltrials.gov/ct2/show/NCT02244086.

[48] SAS Institute Inc. *SAS/IML Software, Version 9.2*. Cary, NC, 2011. URL: http://www.sas.com/.

[49] F. E. Satterthwaite. "An Approximate Distribution of Estimates of Variance Components". In: *Biometrics Bulletin* 2.6 (Dec. 1946), pp. 110–114. ISSN: 00994987. DOI: 10.2307/3002019. URL: http://dx.doi.org/10.2307/3002019.

[50] Seoul Medical Center. *ED50 and ED95 of Intrathecal Bupivacaine With or Without Epinephrine for Total Knee Replacement Arthroplasty D50 and ED95 of Intrathecal Bupivacaine With or Without Epinephrine for Total Knee Replacement Arthroplasty*. URL: https://clinicaltrials.gov/ct2/show/NCT01264575.

[51] Simon, Richard. "Adaptive Treatment Assignment Methods and Clinical Trials". In: *Biometrics* 33.4 (1977), pp. 743–749.

[52] European Collaborative Study. "Time to Undetectable Viral Load after Highly Active Antiretroviral Therapy Initiation among HIV-Infected Pregnant Women". In: *Clinical Infectious Diseases* 44.12 (2007), pp. 1647–1656.

[53] Sverdlov, Oleksandr, Rosenberger, W.F., and Hu, F. "Adaptive Randomization for Clinical Trials". In: (2011).

[54] Tamura, R.N. et al. "A Case Study of an Adaptive Clinical Trial in the Treatment of Out-Patients with Depressive Disorder". In: *Journal of the American Statistical Association* 89.427 (1994), pp. 768–776.

[55] Uher, R. et al. "Early and delayed onset of response to antidepressants in individual trajectories of change during treatment of major depression: a secondary analysis of data from the Genome-Based Therapeutic Drugs for Depression (GENDEP) study". In: *Journal of Clinical Psychiatry* 72.11 (2011), pp. 1478–1484.

[56] U.S. Department of Health and Human Services and Food and Drug Administration. *Innovation/stagnation: Challenge and opportunity on the critical path to new medical products*. 2004.

[57] Viswanathan, S. et al. "Level of Adherence and HIV RNA Suppression in the Current Era of Highly Active Antiretroviral Therapy (HAART)". In: *AIDS Behav* (2014).

[58] Weinstein, M. C. "Allocation of subjects in medical experiments". In: *New England Journal of Medicine* 291.24 (1974), pp. 1278–1285.

[59] B. L. Welch. "The Generalization of 'Student's' Problem when Several Different Population Variances are Involved". In: *Biometrika* 34.1/2 (1947), pp. 28–35. ISSN: 00063444. DOI: 10.2307/2332510. URL: http://dx.doi.org/10.2307/2332510.

[60] World Health Organization (WHO). *Technical and Operational Considerations for Implementing HIV Viral Load Testing*. Iterim Technical Update. 2014.

[61] World Health Organization (WHO). *WHO Global Health Observatory: HIV/AIDS*. URL: `http://www.who.int/gho/hiv/en/` (visited on 02/16/2015).

[62] Zhang, L. and Rosenberger, W.F. "Response-Adaptive Randomization for Clinical Trials with Continuous Outcomes". In: *Biometrics* 62.2 (2006), pp. 562–569.

[63] Zhang, L. and Rosenberger, W.F. "Response-adaptive randomization for survival trials: the parametric approach". In: *Journal of Applied Statistics* 56.2 (2007), pp. 153–165.

[64] Zhang, L.-X. et al. "A generalized drop-the-loser urn for clinical trials with delayed responses". In: *Statistica Sinica* 17 (2007), pp. 387–409.

Chapter 5

Conclusions

5.1 Discussion

Response-adaptive randomization (RAR) is a growing class of clinical trial designs which leverage accumulating information from patients—in particular treatment outcomes—to alter the study's allocation probabilities throughout the trial in a prospectively planned manner [4, 18, 32, 36]. Typically, RAR designs' adaptivity focus on ethical aspirations such as minimizing treatment failures or maximizing exposures to the superior treatment [7, 17, 22, 31, 34]. These aims are important in medicine as researchers and clinicians attempt to provide each individual with the best available care. The difficulty with these intentions is exactly what necessitates an experiment in the first place: the lack of knowledge and/or agreement about which therapy is "best." This new breed of design continues to emphasize the practical importance of randomization while equipoise prevails, but insists on biasing the exposure probabilities as knowledge accrues. Hence RAR attempts to strike a compromise between the population need (determining the most appropriate treatment for the community benefit) and the personal obligation (providing the best treatment to each patient)

[5, 6, 16, 30, 37].

This dissertation proposes a new RAR biased coin design, the BBS design, intended to help phase III clinical trials assign more patients to the treatment with larger mean effect. While many RAR designs apply in cases where the outcome of interest is binary, fewer designs exist to modify studies with continuous outcomes [8, 9, 17, 31, 39]. The BBS design discussed in these papers pertains to treatments with continuous responses and furthermore allows for treatment outcomes to be impacted by patient covariates. Chapter 2 presents the BBS design, contrasting the newcomer with its predecessor and with equal allocation. In this chapter, simulations demonstrate that BBS allocates more patients to the superior treatment while mitigating the loss of power and increased bias inherent in many adaptive designs. Chapter 2 examines the expanded applicability of the BBS, illustrating the design's robustness to relaxing the assumption of known, equal variances. The utility and benefits of the BBS are illustrated through application to an HIV treatment adherence intervention, exposing more patients to the superior therapy than equal allocation. Chapter 3 further broadens the scope of the BBS design via simulation, confirming its applicability and ethicality when covariate-treatment interactions are present. Once again, an increased number of patients receive the more effective intervention in a redesigned HIV clinical trial even as covariate impacts on observed outcomes vary across treatments. In Chapter 4, simulations explore the impact of delayed treatment responses on the BBS design's ability to adaptively and ethically allocate patients. BBS randomization proves more ethical than equal allocation whenever a minimum threshold of observations occur prior to the end of the enrollment period. For the HIV intervention study where patient responses typically lag initial treatment by several months, BBS still assigned more than half of the recruited population to the more effective treatment arm.

In practice, implementing an adaptive design can be challenging [12, 14, 15, 33, 38]. In addition to the concerns associated with a nonadaptive clinical trial, there are three major components which should be carefully considered for RAR designs. The first factor gets at the heart of RAR: how ethical the patient allocation will be under the anticipated range of treatment parameters. The second aspect focuses on the preservation of the statistical properties of the trial as it undergoes RAR, namely power and type I error rates. The third facet is the potential impacts of delayed treatment responses on the design's abilities to adapt. For the BBS design, the overarching implications of these three concerns and the results from Chapters 2 through 4 are discussed below. Section 5.1.4 gives a brief treatment of the available asymptotic research as it applies to the BBS design. This dissertation concludes with recommendations based on the above findings as well as direction for future research.

5.1.1 Ethical Patient Allocation

Response-adaptive randomization designs depend on available outcomes to modify the randomization proportions in order to better allocate future subjects. If responses are scarce, an adaptive design may not have enough data to inform its assignment probabilities. In the case when information is scant or unreliable, an adaptive design—like the researchers of a trial—should not prefer one treatment assignment over another [11, 12, 26, 27]. As responses accrue, however, RAR designs can allocate patients to treatment in a more ethical manner. For example, the BBS design adapts to accumulating information by putting more patients on the treatment which is performing better, relative to the variability in patient responses.

Chapter 2 establishes that when treatment effects differ, the BBS design assigns more patients to the trial arm with larger treatment mean. Moreover, BBS does so in

similar or larger proportions than its predecessor, depending on the scaling parameter of the original design. A slight tradeoff is noted: as the expected number of patients exposed to the superior treatment rises, so does the allocation variance. This variance decreases as the trial enrollment size increases. Simulations show that the BBS design randomizes an additional 1–6% of patients to the superior treatment compared to equal allocation when the difference in treatment means is approximately one-tenth of the pooled standard deviation; an extra 7%–29% of patients when the difference is about half of the pooled standard deviation; and a total of 63%–97% of patients when the difference is comparable to the pooled standard deviation. Chapter 3 identifies the difference in treatment means, the treatment variances, and total enrollment as the major contributors to the BBS's ethical assignments. Covariate impacts on patient outcomes have negligible impact on the design's allocation in all simulations.

Chapter 4 finds that delays in patient responses impact the BBS design's ability to ethically randomize patients. When delay is small relative to the enrollment period, a majority of participants can be exposed to the superior treatment arm. As the delay in patient response times grows, however, that majority will decline. As logically expected, when simulated responses are delayed such that essentially no outcomes are observed prior to the conclusion of trial enrollment, BBS is nonadaptive. That is, the BBS performs equal allocation when sample sizes are small and recruitment is quick relative to the delay in patent responses. Nevertheless, as long as an adequate number of outcomes are observed before all subjects have been enrolled, the BBS design is still preferable to equal allocation in terms of ethical patient randomization. The degree of ethicality follows a nearly linear pattern in relation to the proportion of patients adaptively randomized within the trial.

Overall, the BBS design is an ethically advantageous RAR procedure compared to the balanced assignment of equal allocation. Simulations will help researchers

and clinicians understand anticipated ethical randomization proportions. Factors to consider include percent of patients adaptively randomized, variance in allocation proportion, and impact of delay based on likely values of treatment difference, treatment variances, average delay, enrollment size and period of time, and number of patient responses needed prior to adaptive allocation.

5.1.2 Rejection Rates

Maintaining adequate power to detect a difference in treatments is a common concern in response-adaptive clinical trials [17, 34]. Maximizing power via equal allocation for continuous outcomes is standard practice [23]. Nevertheless, exposing patients to experimental therapies in an ethical manner—especially as equipoise wanes throughout the course of a clinical trial—is gaining increasing attention, despite potential risks of decreased power and increased type I error rates[11, 12, 14, 15].

Chapter 2 determines that the BBS design maintains power levels relatively well compared to equal allocation, particularly when treatment variances are unknown or unequal. On the other hand, the variance in rejection rates increases slightly with the BBS design compared to nonadaptive equal allocation. Moreover, type I error rates also risk inflation under BBS RAR, particularly for small trial sizes. Chapter 3 concludes that the aforementioned rejection rates are not impacted when covariate impacts differ by treatment. Chapter 4 also supports these findings; BBS rejection rates depend on the difference in treatments, treatment variances, and clinical trial total enrollment levels. In fact, power levels appear to rise as simulated response delays grow and the BBS design allocates an increased number of patients non-adaptively. Similarly, type I error rates decrease to be on a par with equal allocation when BBS outcome observations are delayed.

The power to correctly reject a null hypothesis depends on the difference in treatment effects relative to treatment variance [18, 32] and the BBS design is no exception to this rule. Maintaining adequate power without drastically extending a trial or unduly exposing patients to less effective treatment requires thorough inspection from resourcing and ethical perspectives. Conversely, the incorrectly rejecting a null hypothesis may not incur detrimental impacts to the future patient population, particularly in comparison with failing to reject a false null hypothesis and denying future patients a superior treatment. The risks of type I and type II errors must be considered—along with possible treatment and covariate parameters, enrolled sample size, and anticipated delays in response—before implementing any clinical trial, especially an adaptive design [15].

5.1.3 Delay Dampening

Response-adaptive randomization designs such as the BBS leverage accumulating information to ethically assign patients to the treatment with superior outcomes. When an adaptive design is able to maintain adequate levels of power, such a design should be considered in lieu of equal allocation. Delayed responses are sometimes thought to preclude implementation of adaptive allocation [10, 35]; however, many RAR designs are easily revised so that assignment proportions depend only on observed outcomes. Instead of postponing subsequent patient exposures until previous patients respond to treatment, patients are assigned to trial arms as they enroll based on the information available at that time. Delayed outcomes are incorporated into the RAR as they are observed. This particular modification is commonly recommended due to its practicality, ease of implementation, and ethicality [2, 20, 39, 42]. Unfortunately, while the aforementioned articles prove that delay has minimal impact on the asymptotic char-

acteristics of select adaptive designs—namely randomization ratios and power—small to moderate trials may not behave similarly under delayed response conditions. For example, both allocation and rejection rates for the BBS design have higher variance than do equal allocation exposure, power, and type I error rates for similar trial sizes.

Both Chapters 2 and 3 ignore the potential impacts of delayed responses, simulating instead only instantaneously observed outcomes. Simulations throughout Chapter 4, by contrast, focus on the effects of such delays. The BBS design initially randomizes patients to treatment A with fixed probability $\frac{1}{2}$ until a minimum number of outcomes are observed in each treatment. Thereafter, BBS leverages the accumulated estimates to adaptively allocate patients to treatment. The longer the delay in responses relative to the trial enrollment period, the smaller the proportion of adaptively randomized patients will be. In the same vein, the larger a trial is for a given delay, the higher the proportion of adaptively assigned subjects will be.

Small trials are disproportionally impacted due to the size of the minimum necessary outcomes relative to the total trial size. For example, in a trial of 50 patients, 20 instantaneously observed responses must accrue prior to adaptive randomization. A mere 60% of patients remain who can be adaptively allocated—and this number shrinks rapidly as responses lag. Compare this with a moderately sized study enrolling 500 subjects with 20 outcomes immediately observed: up to 96% of patients will be adaptively exposed to treatment. Delayed responses still reduce the proportion of recruits adaptively assigned to treatment, but the impact is not as sever as it is in small trials.

It may be possible to counteract this early reduction in adaptive randomizations by lowering the number of individuals required to respond before leveraging the adaptive estimates. Of course, decreasing the initial burn-in requirements may not adequately offset a delay in responses as the estimated variance will be larger for smaller initial

estimates. That is, the ratio of treatment mean difference to pooled treatment standard deviation may not differ enough from a balanced assignment to benefit the few additional patients who are adaptively randomized. Even so, the minimum number of responses required for adaptive allocation should be carefully considered, especially for smaller trials and/or larger delays.

There are myriad applications within healthcare alone where delays in patient responses are modest relative to the enrollment period of an experiment. For example, adaptive designs are recommended when testing new cures for aggressive diseases such as ebola [25, 28, 29]. Adaptive designs are also recommended for emergency medicine trials [13, 24]. In both of these cases, there may be a delay in patient responses but final outcomes are observed and recorded within a matter of days—a short length of time relative to the trial duration of weeks or months. Incorporating BBS into a researcher's toolbox of ethical randomization schemes in scenarios where the impact of delay is moderate relative to the enrollment period is encouraged. As clinical trial behavior differs across values of treatment parameters, sample size, allocation probabilities, and delay, design simulations are imperative.

5.1.4 Asymptotic Properties

In the early 2000's authors Hu and Zhang pioneered asymptotic research in RAR designs. Their 2004 paper [20] opened the floodgates by deriving the asymptotic properties of the allocation produced by a generalized adaptive biased coin design. Subsequently, these results were leveraged towards a variety of outcomes. Hu, Rosenbergerm and Zhang [19] derive a lower bound on the asymptotic variance of RAR procedures with binomial outcomes. Zhang and Rosenberger [39] determine the asymptotic distribution of five continuous RAR allocations, including the BB design. Zhang

and Rosenberger [40] also apply the asymptotic properties to RAR designs with exponential survival outcomes. Hu, Zhang, Cheung, and Chan [21] demonstrate that the RAR asymptotic properties are "relatively insensitive" to the effects of exponentially distributed delayed responses, although small sample applications can be impacted by delay. Additionally, research in the sister field of covariate-adjusted response adaptive design employs similar asymptotic approaches [41, 43].

In [20], Hu and Zhang find that, for RAR procedures which meet a set of conditions, two asymptotic properties hold. First, as the number of patient responses increases, the estimated allocation proportion based on patient responses is asymptotically normally distributed with mean equal to the true allocation proportion based on treatment parameters. Second, as the number of patient responses increases, the proportion of patients assigned to each treatment is also asymptotically normally distributed with mean equal to the true allocation proportion. That is,

$$\sqrt{n}\left(\frac{N_A(n)}{n} - \rho, \widehat{\rho}(n) - \rho\right) \to N\left(0, \begin{bmatrix} \rho(1-\rho) + 2\theta & \theta \\ \theta & \theta \end{bmatrix}\right),$$

where n is the total number of patients randomized, $N_A(n)$ is the number of patients assigned to treatment A out of n total patients randomized, ρ is the true allocation proportion for treatment A based on the actual treatment parameters of the design, $\widehat{\rho}(n)$ is the estimated allocation proportion for treatment A based on patient responses from n randomized patients, and

$$\theta = \nabla(\rho)^T \begin{bmatrix} \mathrm{Var}(X_A^2) & \mathrm{Cov}(X_A^2, X_A) & 0 & 0 \\ \mathrm{Cov}(X_A^2, X_A) & \mathrm{Var}(X_A) & 0 & 0 \\ 0 & 0 & \mathrm{Var}(X_B^2) & \mathrm{Cov}(X_B^2, X_B) \\ 0 & 0 & \mathrm{Cov}(X_B^2, X_B) & \mathrm{Var}(X_B) \end{bmatrix} \nabla(\rho)$$

with

$$\nabla(\rho)^T = \left[\frac{\partial\rho}{\partial(\sigma_A^2+\mu_A^2)} \quad \frac{\partial\rho}{\partial\mu_A} \quad \frac{\partial\rho}{\partial(\sigma_B^2+\mu_B^2)} \quad \frac{\partial\rho}{\partial\mu_B}\right]$$

and treatment effects $X_k \sim N(\mu_k, \sigma_k^2)$ for $k = A, B$.

In [39], Zhang and Rosenberger determine that the BB design is, in fact, a special case of the generalized RAR procedure from [20]. Moreover, [39] explicitly calculates the asymptotic distributions for the BB design, finding that

$$\sqrt{n}\left(\frac{N_A(n)}{n} - \rho_T, \widehat{\rho_T}(n) - \rho_T\right) \to N\left(0, \begin{bmatrix} \rho_T(1-\rho_T)+2\theta_T & \theta_T \\ \theta_T & \theta_T \end{bmatrix}\right),$$

where

$$\rho_T = \Phi\left(\frac{\mu_A - \mu_B}{T}\right)$$

is the true BB allocation proportion for treatment A based on the actual (unknown) treatment parameters with $\Phi(\cdot)$ as the cumulative distribution function of a standard normal variable, $\widehat{\rho_T}(n)$ is the estimated BB allocation proportion for treatment A based on patient responses from n randomized patients, and

$$\theta_T = \left[\Phi'\left(\frac{\mu_A-\mu_B}{T}\right)\cdot\frac{1}{T}\right]^2 \frac{\sigma^2}{\rho_T(1-\rho_T)}$$

with σ^2 as the variance of treatment A and of treatment B.

The BBS design is a modification of the BB design, replacing the fixed scaling constant T with

$$S = \sqrt{\frac{(N_A-1)s_A^2 + (N_B-1)s_B^2}{N_A+N_B-2}},$$

an estimate of the pooled treatment standard deviation. When the treatment variances σ_A^2 and σ_B^2 are equal, then S^2 is an unbiased estimator of this single treatment

variance σ^2. Otherwise, S is a weighted average of the two treatments where the weighting depends on the number of patients assigned to each treatment. In the case that $\sigma_A^2 = \sigma_B^2 = \sigma^2$, the BBS design is also a special case of the generalized RAR procedure in [20]. Thus the conditions for asymptotic normality are met. Hence the techniques described in [39] can be applied to compute the asymptotic distribution of the BBS allocation proportions. In particular, for the BBS design

$$\sqrt{n}\left(\frac{N_A(n)}{n} - \rho_S, \widehat{\rho_S}(n) - \rho_S\right) \to N\left(0, \begin{bmatrix} \rho_S(1-\rho_S) + 2\theta_S & \theta_S \\ \theta_S & \theta_S \end{bmatrix}\right),$$

where $\rho_S = \Phi\left(\frac{\mu_A - \mu_B}{\sigma}\right)$ is the true BBS allocation proportion for treatment A based on the unknown treatment parameters, $\widehat{\rho_S}(n)$ is the estimated BBS allocation proportion for treatment A based on patient responses from n randomized patients, and

$$\theta_S = \left[\Phi'\left(\frac{\mu_A - \mu_B}{\sigma}\right)\right]^2 \left[\frac{1}{2} \cdot \left(\frac{\mu_A - \mu_B}{\sigma}\right)^2 + 1\right] \frac{1}{\rho_S(1-\rho_S)}$$

with σ^2 as the variance of treatment A and of treatment B.

These asymptotic results can be compared against the simulated study outcomes in Chapter 2. Table 5.1 contains patient allocation standard deviations from BBS simulations and from the BBS asymptotic normal distribution, scaled by $\frac{1}{\sqrt{N}}$ where N is the relevant clinical trial size. Outcomes are given for treatment parameters $\mu_A = 0.0$, 0.1, 0.5, and 1.0; $\mu_B = 0.0$; $\sigma_A = \sigma_B = 1.0$ and clinical trial sizes $N = 50$; 100; 500; 1,000; and 5,000. In Table 5.1, the first column on the left specifies whether the rows' values originate from Chapter 2 BBS simulations or from the BBS asymptotic distribution above. The second column identifies which values of μ_A are represented for a given row. The remaining five columns present either the simulated BBS treatment

A allocation standard deviation or the scaled BBS asymptotic treatment A allocation standard deviation

$$\sqrt{\frac{1}{N}\left(\rho_S(1-\rho_S)+2\left[\Phi'\left(\frac{\mu_A-\mu_B}{\sigma}\right)\right]^2\left[\left(\frac{\mu_A-\mu_B}{\sigma}\right)^2+2\right]\frac{1}{2\rho_S(1-\rho_S)}\right)}$$

for each of the five values of N.

Table 5.1: Patient allocation standard deviations for the BBS design from simulated and asymptotic results where μ_A varies, $\mu_B = 0.0$, and $\sigma_A = \sigma_B = 1.0$ for trial size N.

Design	μ_A	$N = 50$	$N = 100$	$N = 500$	$N = 1{,}000$	$N = 5{,}000$
BBS Simulation	0.0	0.12	0.13	0.07	0.05	0.02
	0.1	0.12	0.13	0.07	0.05	0.02
	0.5	0.11	0.11	0.06	0.05	0.02
	1.0	0.09	0.09	0.06	0.04	0.02
BBS Asymptotic	0.0	0.17	0.12	0.06	0.04	0.02
	0.1	0.17	0.12	0.06	0.04	0.02
	0.5	0.17	0.12	0.06	0.04	0.02
	1.0	0.17	0.12	0.05	0.04	0.02

For example, the first row of Table 5.1 indicates that the BBS simulation for $\mu_A = \mu_B = 0.0$ and $\sigma_A = \sigma_B = 1.0$ produced treatment allocations with standard deviation 0.12 when $N = 50$, 0.07 when $N = 500$, and 0.02 when $N = 5{,}000$. Similarly, the fifth row of Table 5.1 reveals that the BBS treatment allocation asymptotic standard deviation for $\mu_A = \mu_B = 0.0$ and $\sigma_A = \sigma_B = 1.0$ is 0.17 when $N = 50$, 0.06 when $N = 500$, and 0.02 when $N = 5{,}000$. For each value of μ_A, the treatment allocation standard deviation decreases as N increases, regardless of whether the row represents simulated or asymptotic results. Likewise, for a given study sample size, the standard deviations decrease as μ_A increases in the BBS simulated outcomes. BBS asymptotic calculations of standard deviation appear consistent for varying values of μ_A given

N.

When $N = 50$, the simulated standard deviations are smaller than the asymptotic standard deviations for the BBS design. This reduction is due to the fact that an initial $m = 10$ patients must respond to each treatment prior to adaptive allocation. For $N = 50$, delaying the adaptive randomization until nearly half of patients are randomized using the fixed assignment probability $\frac{1}{2}$ lowers the allocation standard deviation. When $N \geq 100$, the initial 20 patient burn-in requirements are less impactful; the simulated allocation standard deviations are typically slightly larger than the associated asymptotic results.

The asymptotic findings in this section support the conclusions derived from previous chapters' simulations. The simulated allocation variances for BBS are larger than those of equal allocation. Nevertheless, simulated BBS allocation standard deviation are comparable to their small-sample adjusted asymptotic standard deviations. The additional patients assigned to the superior treatment under BBS should be weighed against the increased allocation variability when considering a clinical trial randomization design.

5.1.5 Recommendations

If an affliction merits the resources necessary for a large scale clinical trial, it follows that treating a patient with the best possible care during the investigation is worth the additional resources of weighing the benefit of a RAR against the risks of implementing an adaptive trial. While there are potential additional complications associated with RAR studies [12, 14, 15, 33, 38], the impactful factors are becoming more well known [1, 3, 19, 30, 37, 39]. Moreover, advances in modern computing power facilitate performing a diverse array of simulations to better understand a RAR

trial and potential implications of various treatment and design parameters. The BBS design is a new RAR which preserves rejection rates while allowing more patients to be exposed to the study arm with superior mean effects, even as treatment outcomes are delayed. The BBS should be added to medical researchers' design toolkits and should be considered in lieu of equal allocation for phase III clinical trials.

While this dissertation contributes a new design to the area of RAR, knowledge about the BBS is only foundational at this stage. Future study on the asymptotic properties of the BBS design as well as convergence to these properties is important, particularly when treatment variances are not assumed to be equal. Additional questions arise as other modifications are considered. How is the BBS performance altered when tackling non-gaussian outcomes? What are the optimal burn-in requirements for initial adaptive allocation estimates? How do other variable delay models impact the proportion of patients adaptively randomized? Is there a more appropriate variance estimate that should replace the pooled standard deviation in the BBS design and how much supplemental bias does such an estimator generate? How do interim analyses and early stopping rules fit into this RAR schema? Continued research is imperative to providing the medical research community with adequate designs to cover the vast array of clinical domains in an ethical manner.

5.2 References

[1] Bai, Z.D. and Hu, F. "Asymptotics in randomized urn models". In: *Annals of Applied Probability* 15.1b (2005), pp. 914–940.

[2] Bai, Z.D., Hu, F., and Rosenberger, W.F. "Asymptotic properties of adaptive designs for clinical trials with delayed response". In: *The Annals of Statistics* 30.1 (2002), pp. 122–139.

[3] Baldi Antognini, A. "A theoretical analysis of the power of biased coin designs". In: *Journal of Statistical Planning and Inference* 138.6 (2008), pp. 1792–1798.

[4] Baldi Antognini, A. and Giovagnoli, A., eds. *Adaptive Designs for Sequential Treatment Allocation*. 1st ed. Chapman & Hall/CRC Biostatistics Series, 2015.

[5] Baldi Antognini, A. and Giovagnoli, A. "Compound optimal allocation for individual and collective ethics in binary clinical trials". In: *Biometrika* 97.4 (2010), pp. 935–946.

[6] Baldi Antognini, A. and Giovagnoli, A. "Ethical Adaptive Designs for Binary Clinical Trials". In: *Societa Italiana di Statistica* (2007).

[7] Bandyopadhyay, Uttam and Biswas, A. "Adaptive designs for normal response with prognostic factors". In: *Biometrika* 88 (2001), pp. 409–419.

[8] Biswas, A. and Bhattacharya, Rahul. "Optimal response-adaptive designs for normal responses". In: *Biometrical Journal* 51.1 (2009), pp. 193–202.

[9] Biswas, A., Bhattacharya, Rahul, and Zhang, L. "Optimal Response-Adaptive Designs for Continuous Responses in Phase III Trials". In: *Biometrical Journal* 49.6 (2007), pp. 928–940.

[10] Byar, D.P. et al. "Randomized Clinical Trials: Perspectives on Some Recent Ideas". In: *New England Journal of Medicine* 295.2 (1976), pp. 74–80.

[11] Ezekiel J. Emanuel et al., eds. *Ethical and Regulatory Aspects of Clinical Research: Readings and Commentary*. Baltimore, MD: The Johns Hopkins University Press, 2004.

[12] Fisher, Lloyd D. "Advances in clinical trials in the twentieth century". In: *Annual Review of Public Health* 20.1 (1999), pp. 109–125.

[13] Flight, L., Julious, S., and Goodacre, S. "Assessing the Current and Potential Use of Adaptive Study Designs in Emergency Medicine Clinical Trials". In: *Clinical Trial Design II*. Ed. by Ko, C.W. 259. American Statistical Association. 2015.

[14] Food and Drug Administration. *Critical Path Initiative: Transforming the way FDA-regulated products are developed, evaluated, and manufactured*. Report on Key Achievements in 2009. 2010.

[15] Food and Drug Administration. "Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics". Draft Document. 2010.

[16] Geraldes, M. et al. "The doubly adaptive weighted difference design". In: *Journal of Statistical Planning and Inference* 136.6 (2006), pp. 1923–1939.

[17] Hu, F. and Rosenberger, W.F. "Optimality, Variability, Power: Evaluating Response-Adaptive Randomization Procedures for Treatment Comparisons". In: *Journal of the American Statistical Association* 98.463 (2003), pp. 671–678.

[18] Hu, F. and Rosenberger, W.F. *The Theory of Response-Adaptive Randomization in Clinical Trials*. 1st ed. Hoboken, NJ: John Wiley & Sons, Inc., 2006.

[19] Hu, F., Rosenberger, W.F., and Zhang, L.-X. "Asymptotically best response-adaptive randomization procedures". In: *Journal of Statistical Planning and Inference* 136 (2006), pp. 1911–1922.

[20] Hu, F. and Zhang, L.-X. "Asymptotic Properties of Doubly Adaptive Biased Coin Designs for Multitreatment Clinical Trials". In: *The Annals of Statistics* 32.1 (2004), pp. 268–301.

[21] Hu, F. et al. "Doubly-Adaptive Biased Coin Designs with Delayed Responses". In: *Canadian Journal of Statistics* 36.4 (2008), pp. 541–559.

[22] Ivanova, A. V., Biswas, A., and Lurie, A. "Response-adaptive designs for continuous outcomes". In: *Journal of Statistical Planning and Inference* 136.6 (2006), pp. 1845–1852.

[23] Lachin, J. M. "Statistical properties of randomization in clinical trials". In: *Controlled Clinical Trials* 9 (1988), pp. 289–311.

[24] Lewis, R.J. and Bessen, H.A. "Sequential clinical trials in emergency medicine". In: *Annals of Emergency Medicine* 19.12 (1990), p. 1435.

[25] Longini, I.M. and Dean, N. "Design, Analysis, and Interpretation of Ebola Vaccine Efficacy Trials in West Africa". In: *Meeting the Challenges of a Pandemic: The Statistical Aspects of Dealing with Ebola*. Ed. by Russek-Cohen, E. 462. American Statistical Association. 2015.

[26] Palmer, C.R. "Ethics, data-dependent designs, and the strategy of clinical trials: time to start learning-as-we-go?" In: *Statistical Methods in Medical Research* 11 (2002), pp. 381–402.

[27] Pong, A. and Chow, S.C., eds. *Handbook of Adaptive Designs in Pharmaceutical and Clinical Development*. Boca Raton: CRC Press, 2011.

[28] Proschan, M. "Novel Vaccine Designs". In: *Statistical Challenges in Clinical Trials of Novel Vaccines*. Ed. by Solomon, G. 118. American Statistical Association. 2015.

[29] Proschan, M. and Dodd, L. "A Flexible Randomized Clinical Trial for Evaluating Therapeutics for Ebola Disease". In: *Meeting the Challenges of a Pandemic: The Statistical Aspects of Dealing with Ebola*. Ed. by Russek-Cohen, E. 462. American Statistical Association. 2015.

[30] Rosenberger, W.F. "New Directions in Adaptive Designs". In: *Statistical Science* 11 (1996), pp. 137–149.

[31] Rosenberger, W.F. and Hu, F. "Maximizing power and minimizing treatment failures in clinical trials". In: *Clinical Trials* 1.2 (2004), pp. 141–147.

[32] Rosenberger, W.F. and Lachin, J. M. *Randomization in Clinical Trials Theory and Practice*. 1st ed. Wiley-Interscience, 2002.

[33] Rosenberger, W.F. and Lachin, J. M. "The use of response-adaptive designs in clinical trials". In: *Controlled Clinical Trials* 14.6 (1993), pp. 471–484.

[34] Rosenberger, W.F. et al. "Optimal Adaptive Designs for Binary Response Trials". In: *Biometrics* 57 (2001), pp. 909–913.

[35] Spiegelhalter, D.J., Abrams, K.R., and Myles, J.P. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. New York, NY: Wiley & Sons, 2004.

[36] Sverdlov, Oleksandr, ed. *Modern Adaptive Randomized Clinical Trials: Statistical and Practical Aspects*. 1st ed. Chapman & Hall/CRC Biostatistics Series, 2015.

[37] Sverdlov, Oleksandr, Rosenberger, W.F., and Hu, F. "Adaptive Randomization for Clinical Trials". In: (2011).

[38] Tamura, R.N. et al. "A Case Study of an Adaptive Clinical Trial in the Treatment of Out-Patients with Depressive Disorder". In: *Journal of the American Statistical Association* 89.427 (1994), pp. 768–776.

[39] Zhang, L. and Rosenberger, W.F. "Response-Adaptive Randomization for Clinical Trials with Continuous Outcomes". In: *Biometrics* 62.2 (2006), pp. 562–569.

[40] Zhang, L. and Rosenberger, W.F. "Response-adaptive randomization for survival trials: the parametric approach". In: *Journal of Applied Statistics* 56.2 (2007), pp. 153–165.

[41] Zhang, L.-X. and Hu, F. "A new family of covariate-adjusted response-adaptive designs and their properties". In: *Applied Mathematics - A Journal of Chinese Universities* 24.1 (2009), pp. 1–13.

[42] Zhang, L.-X. et al. "A generalized drop-the-loser urn for clinical trials with delayed responses". In: *Statistica Sinica* 17 (2007), pp. 387–409.

[43] Zhang, L.-X. et al. "Asymptotic properties of covariate-adjusted response-adaptive designs". In: *The Annals of Statistics* 35.3 (2007), pp. 1166–1182.

Appendix A

SAS IML Simulation Code Written to Support the Work in this Dissertation

```
/**********BBS Design with delay**********/
    /**********'N' patients total ('m' of which give initial estimates) and 'Reps' replications**********/
    proc iml;
    /*****This simulation incorporates delay*****/
    /*****Overarching strategy is as follows*****/
    /*****Begin by creating an ordered vector of N patients*****/
    /*****Assign each patient a delay in response times*****/
    /*****Random exponential iid delays form a Poisson Process*****/
    /*****For this we'll create a Responded matrix*****/
    /*****At each new patient arrival, we know which responses have been received*****/
    /*****This matrix will also be used to help with estimates at that point in time–
```

```
not just the number of responses in each estimate*****/
    /*****Assign patients using a 50-50 allocation scheme until m patient responses
have been observed in each treatment*****/
    /*****Thereafter, assign based on RAR design using the available estimates at
that point in time*****/
    /*****Once all patients are assigned, can skip ahead to final estimates*****/
    /*****Major components to consider are the typical ones, including*****/
    /*****Probability of assignment to better treatment, power, and bias (alpha,
variance in assignment probability)*****/
    /*****Variance, Covariate Slope, and Delay differing by treatment assignment
(above plus efficiency)*****/
    /*****Which covariates are being used in estimation–All arrived (including cur-
rent, un-randomized patient), All assigned, or All responded*****/
    /*****Other major components not included in this simulation*****/
    /*****Early stopping rules–possible but unlikely to be desirable given power/bias/
efficiency issues*****/
    /*****Cara-izing design–really changes what is being considered here and would
prefer not to include in scope*****/
   start Simulate(MuA, MuB, StdA, StdB, CovMu, CovStd, CovSlopeA, CovSlopeB,
IncCurCov, DelayA, DelayB, Alpha, Sides, N, m);
    /*****Initialize Simulation Parameters*****/
    /*****MuA = Mean Effect Size of Treatment A*****/
    /*****MuB = Mean Effect Size of Treatment B*****/
    /*****StdA = Standard Deviation of Effect Size of Treatment A*****/
    /*****StdB = Standard Deviation of Effect Size of Treatment B*****/
    /*****CovMu = Mean of Covariate*****/
```

/*****CovStd = Standard Deviation of Covariate*****/

/*****CovSlopeA = Slope of Covariate for Treatment A*****/

/*****CovSlopeB = Slope of Covariate for Treatment B*****/

/*****IncCurCov = Covariate inclusion scheme for estimating randomization parameters*****/

/*****Options are 0=N, 1=Y, and 9=Resp*****/

/*****0 = N = No, do NOT include current patient's covariate in estimation–include all previous patients but not this one*****/

/*****1 = Y = Yes, include current patient's covariate in estimation–include all previous and current patients*****/

/*****9 = Resp = Responded, include only responded patients' covariates in estimation*****/

/*****DelayA = Exponential mean (and variance) for Treatment A response times*****/

/*****DelayB = Exponential mean (and variance) for Treatment B response times*****/

/*****Delays in response times can be negligible (0) for immediate responses*****/

/*****Delays in response times should be considered relative to time between arrivals (which is set at 1 unit of time)*****/

/*****Alpha = Alpha level to implement*****/

/*****Sides = Specifies whether the test is one-sided or two-sided*****/

/*****N = Number of Patients*****/

/*****m = Initial Number of Patients to assign to each Treatment to get Initial Estimates*****/

/*****There must be m responses from each Treatment in order to begin RAR*****/

SeedArrive=0; *Seed for Interarrival Time–0 for system clock; SeedDelay=0; *Seed for Response Delay Time regardless of treatment–0 for system clock; SeedCov=0; *Seed for Covariate error–0 for system clock; SeedA=0; *Seed for Trt A error–0 for system clock; SeedB=0; *Seed for Trt B error–0 for system clock; SeedRand=0; *Seed for patient randomization–0 for system clock;

/*****Create Interarrival Time Vector for time until the next patient arrives*****/

/*****Interarrival times follow a Poisson Process–with IID exponential interarrival times of mean/variance 1*****/

InterarrivalTime = J(N-1,1,SeedArrive); *J(r,c,v) creates a matrix of size (rc) filled with element v;

*We are creating an (N-11) column vector prepped for random number generation;

*Since we have N patients, we only need N-1 interarrival times–N-1 fences between N fence posts;

call randgen(InterarrivalTime,'EXPONENTIAL'); *Fills the vector with exponential(1) values;

InterarrivalTime = InterarrivalTime*1; *multiplying by C turns the values into exponential(C) values–for arrivals C=1;

/*****Create Actual Arrival Time Vector for time each patient actually arrives*****/

/*****Calculate actual arrival times–should be strictly increasing since exponential is continuous and positive*****/

ActualArrivalTime = 0//CuSum(InterarrivalTime); *Calculate the Actual arrival times of all N patients;

*First patient arrival time is set to 0, hence Actual Arrival Time Vector is (N1) column vector;

/*****Create Response Delay Vectors for time until the patient response is obtained–and which may differ by treatment*****/

/*****Response Delay times follow a Poisson Process–with IID exponential delay times of mean/variance DelayA or DelayB based on treatment assignment*****/

DelayTime = J(N-1,1,SeedDelay); *J(r,c,v) creates a matrix of size (rc) filled with element v;

*We are creating an (N-11) column vector prepped for random number generation;

*Since we have N patients, we only need N-1 response delay times to randomly assign all patients to treatment–N-1 fences between N fence posts;

call randgen(DelayTime,'EXPONENTIAL'); *Fills the vector with exponential(1) values;

DelayATime = DelayTime*DelayA; *multiplying by C turns the values into exponential(C) values–here C=DelayA;

DelayBTime = DelayTime*DelayB; *multiplying by C turns the values into exponential(C) values–here C=DelayB;

*Confirmed that these to vectors are indeed identical when DelayA = DelayB and different otherwise, yay!;

/*****Create Actual Response Time Vector for time each patient actually responds*****/

/*****Calculate actual response times–not necessarily increasing: patients may arrive faster than treated patients respond*****/

ActualResponseTimeA = ActualArrivalTime+(DelayATime//0); *Calculate the Actual response times of all N patients if assigned to Trt A;

ActualResponseTimeB = ActualArrivalTime+(DelayBTime//0); *Calculate the Actual response times of all N patients if assigned to Trt A;

*Last patient delay time is set to 0 because no further patients to be randomized, hence Actual response time vectors are (N1) column vectors;

/*****Create Treatment Assignment Vectors to track treatment assignments by patient*****/

/*****Begin with two (N1) vectors full of zeros*****/

/*****then fill them with 1's for each patient based on treatment assignment*****/

AssignedA = J(N,1,0); *J(r,c,v) creates a matrix of size (rc) filled with element v;

AssignedB = J(N,1,0); *J(r,c,v) creates a matrix of size (rc) filled with element v;

*We are creating two (N1) column vectors full of only zeros–will add ones as patients are assigned;

/*****Create Treatment Responded Vectors to track response by treatment for assigned patients*****/

/*****Begin with two (N1) vectors full of zeros*****/

/*****then fill them with 1's for each patient based on treatment assignment and response*****/

RespondedA = J(N,1,0); *J(r,c,v) creates a matrix of size (rc) filled with element v;

RespondedB = J(N,1,0); *J(r,c,v) creates a matrix of size (rc) filled with element v;

*We are creating two (N1) column vectors full of only zeros–will add ones as patients are assigned;

/*****Create Treatment Outcome Vector(s)*****/

/*****First we need the patient covariates which do not vary by treatment*****/

/*****This is done by creating an (N1) column vector for Covariate Effect*****/

/*****Randomly generate numbers from a normal distribution (Z N(0,1))*****/

/*****multiply them by the covariate's standard deviation*****/

/*****then add this to the covariate's mean (X=Z*std+mu N(mu,stdˆ2))*****/

Covariates=CovMu+(normal(J(N,1,SeedCov)))#CovStd; *Create all patient covariates;

/*****Then we need the patient response which include the covariates whose slopes may vary by treatment*****/

/*****This is done by creating two (N1) column vectors for Treatment Outcome*****/

/*****Randomly generate numbers from a normal distribution (Z N(0,1))*****/

/*****multiply them by the treatment's standard deviation*****/

/*****then add this to the treatment's mean (Y=Z*STD+MU N(MU,STDˆ2))*****/

/*****and finally add this to the covariate's effect–multiplied by trt-dependent slope (Y=Z*STD+MU+Slope*X)*****/

TrtAOutcomes=(MuA+(normal(J(N,1,SeedA)))#StdA)+(CovSlopeA*Covariates); *Get outcomes: responses + slopes*covariates;

TrtBOutcomes=(MuB+(normal(J(N,1,SeedB)))#StdB)+(CovSlopeB*Covariates); *Get outcomes: responses + slopes*covariates;

/*****Start Do Loop for patients 1 through N*****/

do i=1 to N;

*i_vector = I(N)[,i]; *I(N) creates an identity matrix of size N (NxN);

*I(N)[,i] just keeps the i-th column of that matrix–a column vector with all zeros except for a 1 in the i-th row;

*****Don't think we're needing/using this right now...;

/*****The first thing to do is to reset the time and update the Responded vectors*****/

CurrentTime = ActualArrivalTime[i,1]; *This sets the current time to the arrival time of the next patient;

CurrentTimeVec = J(N,1,CurrentTime); *This makes a whole Nx1 column vector full of current time;

RespondedA = (ActualResponseTimeA ¡ CurrentTimeVec)#AssignedA;

*First creates vector of zeros and ones, zero where response time is larger than current time (has not occured yet)

and 1 where response time is smaller than current time (has occurred prior to new patient arrival);

*Then zeros out all patients that have not been assigned to Trt A (or have not yet been assigned) where # is element-wise multiplication;

RespondedB = (ActualResponseTimeB ¡ CurrentTimeVec)#AssignedB;

/*****The second thing to do is to check and see if there are at least 2m patients responded, m from each treatment*****/

ARespTot = RespondedA[+,]; *Matrix[+,columns] returns a row vector of the sum of all columns specified;

BRespTot = RespondedB[+,]; *Matrix[+,columns] returns a row vector of the sum of all columns specified;

*Since there are no columns specified, this sums down each of the columns;

*Since there is only one column, these row vectors are actually scalars;

/*****If there are fewer than m patient responses in either treatment*****/

/*****assign patient i to treatment using a 50-50 probability*****/

if (ARespTot ¡ m) | (BRespTot ¡ m) then do; *must use |, cannot use 'or';

Phi=1/2; *Arbitrarily select first Phi–should be reset before any further randomization takes place;

/*****Randomly assign patient i based on Phi and a uniformly randomly generated number*****/

*If uniformly randomly generated number is less than Phi:

*assign patient to Trt A

*update Trt A assignment vector;

*Else assign patient to Trt B and update Trt B assignment vector;

if uniform(SeedRand)¡Phi then do;

AssignedA[i,1]=1; *Adds patient i to list of patients assigned to Trt A;

end;

else do;

AssignedB[i,1]=1; *Adds patient i to list of patients assigned to Trt B;

end;

NotAdRand = i; *Sets NotAdRand to THIS patient just assigned using Phi = 1/2;

*The LAST time this gets updated will be when the last patient randomized using 50-50 probability is allocated;

*Hence it will be set to the LAST PATIENT who does NOT get adaptively randomized;

end; *ends case where estimating for patient i with i ¿= N but fewer than m responses in at least one treatment;

/*****If there are at least m patient responses in each treatment*****/

/*****estimate parameters that go into Phi: MuA, MuB, and sigma_hat–and their antecedents*****/

/*****and then randomly assign patient i to treatment using a Phi/(1-Phi) probability*****/

*if (ARespTot ¿= m) & (BRespTot ¿= m) then do; *must use &, cannot use 'and';

else do;

/*****Estimate Covariate Mean, Sum of Squares (of errors), and Variance*****/

/*****Determine whether or not current patient covariate is included in covariate estimate*****/

/*****and truncate recorded covariate vector accordingly so mean can be taken*****/

/*****Only one covariate mean–does not differ by treatment*****/

if IncCurCov = 0 then do; *if 0 = N, then exclude current covariate–use only i-1 randomized patient covariates;

CovariateSub = Covariates[1:i-1,]; *truncate covariate vector to include all patients up through previous patient, not including this patient;

*X_Bar = CovariateSub[:];

*S_XX_Bar = (CovariateSub-X_Bar)[##]; *Calculate sum of squares of Covariate (technically SS of difference between individual observations and mean of observations);

*Var_X_Hat = S_XX_Bar/((i-1)-1); *Estimate Variance of Covariate (X) using sums of squares;

Cov_A_Sub = CovariateSub[loc((AssignedA[1:i-1,]) = 1),1]; *truncate covariate vector to include only patients who have responded to Trt A;

Cov_A_Mean = Cov_A_Sub[:];

Cov_A_Var = (Cov_A_Sub-Cov_A_Mean)[##]/((AssignedA[1:i-1,])[+]-1);

Cov_B_Sub = CovariateSub[loc((AssignedB[1:i-1,]) = 1),1]; *truncate covariate vector to include only patients who have responded to Trt A;

Cov_B_Mean = Cov_B_Sub[:];

Cov_B_Var = (Cov_B_Sub-Cov_B_Mean)[##]/((AssignedB[1:i-1,])[+]-1);

end;

if IncCurCov = 1 then do; *if 1 = Y, then include current covariate–use all i patient covariates, including this patient's covariate even though not yet randomized;

CovariateSub = Covariates[1:i,]; *truncate covariate vector to include all patients up through and including this patient;

*X_Bar = CovariateSub[:];

*S_XX_Bar = (CovariateSub-X_Bar)[##]; *Calculate sum of squares of Covariate (technically SS of difference between individual observations and mean of observations);

*Var_X_Hat = S_XX_Bar/(i-1); *Estimate Variance of Covariate (X) using sums of squares;

Cov_A_Sub = CovariateSub[loc((AssignedA[1:i,]) = 1),1]; *truncate covariate vector to include only patients who have responded to Trt A;

Cov_A_Mean = Cov_A_Sub[:];

Cov_A_Var = (Cov_A_Sub-Cov_A_Mean)[##]/((AssignedA[1:i,])[+]-1);

Cov_B_Sub = CovariateSub[loc((AssignedB[1:i,]) = 1),1]; *truncate covariate vector to include only patients who have responded to Trt A;

Cov_B_Mean = Cov_B_Sub[:];

Cov_B_Var = (Cov_B_Sub-Cov_B_Mean)[##]/((AssignedB[1:i,])[+]-1);

end;

if IncCurCov = 9 then do; *if 9 = Resp, then include only covariate of patients who have responded;

CovariateSub = Covariates[loc((RespondedA+RespondedB) = 1),1]; *truncate covariate vector to include only patients who have responded to treatment;

*X_Bar = (Covariates[loc((RespondedA+RespondedB) = 1),1])[:];

*Average of Responded Covariates: calculates mean of all elements in Covariates corresponding to patients that have responded to either treatment;

*loc saves the location where RespondedA+RespondedB = 1

–patients who have responded before/though/by arrival of patient i;

*Covariates[loc,1] turns all the elements in Covariates corresponding to saved location into a new vector

–covariates of patients with outcomes/responses recorded before/though/by arrival of patient i;

*[:] takes the mean of all elements in that new, covariates-of-those-who-have-responded vector;

*S_XX_Bar = (CovariateSub-X_Bar)[##]; *Calculate sum of squares of Covariate (technically SS of difference between individual observations and mean of observations);

*Var_X_Hat = S_XX_Bar/((ARespTot+BRespTot)-1); *Estimate Variance of Covariate (X) using sums of squares;

Cov_A_Sub = Covariates[loc((RespondedA) = 1),1]; *truncate covariate vector to include only patients who have responded to Trt A;

Cov_A_Mean = Cov_A_Sub[:];

Cov_A_Var = (Cov_A_Sub-Cov_A_Mean)[##]/(ARespTot-1);

Cov_B_Sub = Covariates[loc((RespondedB) = 1),1]; *truncate covariate vector to include only patients who have responded to Trt A;

Cov_B_Mean = Cov_B_Sub[:];

Cov_B_Var = (Cov_B_Sub-Cov_B_Mean)[##]/(BRespTot-1);

end;

/*****Estimate Treatment Outcome Means for each treatment*****/

TrtASub = TrtAOutcomes[loc(RespondedA = 1),1];

*loc saves the location where RespondedA = 1–patients who have responded before/though/by arrival of patient i;

*TrtAOutcomes[loc,1] turns all the elements in TrtAOutcomes corresponding to saved location into a new vector

–outcomes/responses recorded before/though/by arrival of patient i;

Y_A_Bar = TrtASub[:];

*Average of Outcome A: calculates mean of all elements in TrtAOutcomes corresponding to patients that have responded;

*[:] takes the mean of all elements in that new, outcomes-of-those-who-have-responded vector;

TrtBSub = TrtBOutcomes[loc(RespondedB = 1),1];

*loc saves the location where RespondedB = 1–patients who have responded before/though/by arrival of patient i;

*TrtBOutcomes[loc,1] turns all the elements in TrtBOutcomes corresponding to saved location into a new vector

–outcomes/responses recorded before/though/by arrival of patient i;

Y_B_Bar = TrtBSub[:];

*Average of Outcome B: calculates mean of all elements in TrtBOutcomes corresponding to patients that have responded;

*[:] takes the mean of all elements in that new, outcomes-of-those-who-have-responded vector;

/*****Estimate Covariance between Outcome and Covariate for each treatment–see above for vector creation/manipulation/calculation details*****/

Cov_A_Sub = Covariates[loc((RespondedA) = 1),1]; *truncate covariate vector to include only patients who have responded to Trt A;

*S_A_XY_Bar = (Cov_A_Sub#TrtASub)[+]-ARespTot*Y_A_Bar*X_Bar;

S_A_XY_Bar = (Cov_A_Sub#TrtASub)[+]-ARespTot*Y_A_Bar*Cov_A_Mean;

Cov_A_XY_Hat = S_A_XY_Bar/(ARespTot-1);

Cov_B_Sub = Covariates[loc((RespondedB) = 1),1]; *truncate covariate vector to include only patients who have responded to Trt A;

*S_B_XY_Bar = (Cov_B_Sub#TrtBSub)[+]-BRespTot*Y_B_Bar*X_Bar;

S_B_XY_Bar = (Cov_B_Sub#TrtBSub)[+]-BRespTot*Y_B_Bar*Cov_B_Mean;

Cov_B_XY_Hat = S_B_XY_Bar/(BRespTot-1);

*Estimates using sums of products (Y*x), minus double-counting products;

/*****Estimate Slopes (BetaA & BetaB) of Covariate (X) for each treatment*****/

Slope_A_Hat = (Cov_A_XY_Hat)/(Cov_A_Var);

Slope_B_Hat = (Cov_B_XY_Hat)/(Cov_B_Var);

*Estimates using ratio of covariance/variance–Covariance between Trt and Outcome over Variance of Covariate;

/*****Estimate Treatment Means and Difference in Treatment Effects*****/

MuA_Hat = Y_A_Bar - Cov_A_Mean*Slope_A_Hat; *Estimate Trt A response less covariate effect;

MuB_Hat = Y_B_Bar - Cov_B_Mean*Slope_B_Hat; *Estimate Trt B response less covariate effect;

Trt_Diff_Est = MuA_Hat - MuB_Hat; *Estimate of Trt Diff using all previous estimates;

/*****Estimate Treatment Variances–variances in treatment effect without accounting for covariate effect*****/

VarA_Hat_Est = (TrtASub-Y_A_Bar)[##]/(ARespTot-1) - (Slope_A_Hat**2)*Cov_A_Var - 2*Slope_A_Hat*0;

VarB_Hat_Est = (TrtBSub-Y_B_Bar)[##]/(BRespTot-1) - (Slope_B_Hat**2)*Cov_B_Var - 2*Slope_B_Hat*0;

*Estimate of Treatment Effect Variances: Use Property of Variance Addition for Dependent Variables;

/* Var(A+B) = Var(A) + Var(B) + 2Cov(A,B) or equivalently, Var(A) = Var(A+B) - Var(B) - 2Cov(A,B)

Outcome = TrtEffect + TrtErr + Slope*(Covariate + CovErr)

= TrtErr + Slope*CovErr + TrtEffect + Slope*Covariate –where latter 2 are constants with 0 variance

Var(Outcome) = Var(TrtErr) + Var(Slope*CovErr) + 2Cov(TrtErr,Slope*CovErr) + 0 + 0

= Var(TrtErr) + Slopeˆ2*Var(CovErr) + Slope*2Cov(TrtErr,CovErr)

Var(TrtErr) = Var(Outcome) - Slopeˆ2*Var(CovErr) - 2*Slope*Cov(TrtErr,CovErr)

Note: 2*Slope*Cov(TrtErr,CovErr) is Covariance of Treatment Error and Covariate Error, NOT of Outcome Error and Covariate Error

Note: Furthermore, Treatment and Covariate errors are independent, uncorrelated, and have no covariance, i.e., last term is ZERO

= Var_Y - Slopeˆ2*Var_X - 0 –with Var_Y estimated in the usual manner */

VarA_Hat = max(VarA_Hat_Est,0.00001); *Need to ensure positive (non-zero, non-negative) variances, so take max just in case;

VarB_Hat = max(VarB_Hat_Est,0.00001); *Need to ensure positive (non-zero, non-negative) variances, so take max just in case;

/*****Estimate Phi([MuA - MuB]/Sigma) and randomized based on it*****/

/*****Sigma is estimated using a pooled variance estimate that is unbiased when the variances are equal*****/

Ratio = (MuA_Hat - MuB_Hat)/sqrt(((ARespTot-1)*VarA_Hat+(BRespTot-1)*VarB_Hat)/(ARespTot+BRespTot-2));

Phi = CDF('NORMAL',Ratio,0,1);

/*****Now randomly assign patient i based on Phi and a uniformly randomly generated number*****/

*If uniformly randomly generated number is less than Phi:

*assign patient to Trt A

*update Trt A assignment vector;

*Else assign patient to Trt B and update Trt B assignment vector;

if uniform(SeedRand)¡Phi then do;

AssignedA[i,1]=1; *Adds patient i to list of patients assigned to Trt A;

end;

else do;

AssignedB[i,1]=1; *Adds patient i to list of patients assigned to Trt B;

end;

end; *ends case where estimating for patient i with i ¡= N but at least m responses in each treatment;

end; *ends loop for patient 1 through N, inclusive;

/*****All N patients have been assigned*****/

/*****Since there are no more randomizations performed, we can skip ahead to time when all patients have responded*****/

/*****Gather final estimates to conclude study*****/

if i = N+1 then do;

CurrentTime = max(ActualResponseTimeA[¡¿,1],ActualResponseTimeB[¡¿,1]); *This sets the current time to the maximum arrival time of all patients;

CurrentTimeVec = J(N,1,CurrentTime); *This makes a whole Nx1 column vector full of current tie;

RespondedA = (ActualResponseTimeA ¡= CurrentTimeVec)#AssignedA;

*First creates vector of zeros and ones, zero where response time is larger than current time (has not occurred yet)

and 1 where response time is smaller than or equal to current time;

*Added equals here because we are explicitly choosing the current time to be equal to a response time

but we still want to include that response;

*In previous versions of this, did not include equal sign

because times should have been continuous and equal should not happen

but for arguments' sake if equal, assume would not have had time to adequately process recorded response, I guess;

*Then zeros out all patients that have not been assigned to Trt A (or have not yet been assigned) where # is element-wise multiplication;

RespondedB = (ActualResponseTimeB ¡= CurrentTimeVec)#AssignedB;

/*****Find total number of patients assigned to each treatment*****/

ARespTot = RespondedA[+,]; *Matrix[+,columns] returns a row vector of the sum of all columns specified;

BRespTot = RespondedB[+,]; *Matrix[+,columns] returns a row vector of the sum of all columns specified;

*Since there are no columns specified, this sums down each of the columns;

*Since there is only one column, these row vectors are actually scalars;

CheckN = ARespTot + BRespTot;

if CheckN ^= N then do;

print "Patients are not all randomized!", "i= ", i, "N= ", N, "Check= ", CheckN;

end;

CheckA = (AssignedA = RespondedA)[+];

if CheckA ¡ N then do;

print "A Patients have not all responded!", "AssignedA= ", AssignedA, "RespondedA= ", RespondedA, "CheckA= ", CheckA;

end;

CheckB = (AssignedB = RespondedB)[+];

if CheckB ¡ N then do;

print "B Patients have not all responded!", "AssignedB= ", AssignedB, "RespondedB= ", RespondedB, "CheckB= ", CheckB;

end;

/*****Estimate Covariate Mean, Sum of Squares (of errors), and Variance*****/

* X_Bar = Covariates[:];

* S_XX_Bar = (Covariates-X_Bar)[##]; *Calculate sum of squares of Covariate (technically SS of difference between individual observations and mean of observations);

* Var_X_Hat = S_XX_Bar/(i-1); *Estimate Variance of Covariate (X) using sums of squares;

/*****Estimate Treatment Outcome Means for each treatment*****/

TrtASub = TrtAOutcomes[loc(RespondedA = 1),1];

*loc saves the location where RespondedA = 1–patients who have responded before/though/by arrival of patient i;

*TrtAOutcomes[loc,1] turns all the elements in TrtAOutcomes corresponding to saved location into a new vector

–outcomes/responses recorded before/though/by arrival of patient i;

Y_A_Bar = TrtASub[:];

*Average of Outcome A: calculates mean of all elements in TrtAOutcomes corresponding to patients that have responded;

*[:] takes the mean of all elements in that new, outcomes-of-those-who-have-responded vector;

TrtBSub = TrtBOutcomes[loc(RespondedB = 1),1];

*loc saves the location where RespondedB = 1–patients who have responded before/though/by arrival of patient i;

*TrtBOutcomes[loc,1] turns all the elements in TrtBOutcomes corresponding to saved location into a new vector

–outcomes/responses recorded before/though/by arrival of patient i;

Y_B_Bar = TrtBSub[:];

*Average of Outcome B: calculates mean of all elements in TrtBOutcomes corresponding to patients that have responded;

*[:] takes the mean of all elements in that new, outcomes-of-those-who-have-responded vector;

/*****Estimate Covariance between Outcome and Covariate for each treatment–see above for vector creation/manipulation/calculation details*****/

Cov_A_Sub = Covariates[loc((RespondedA) = 1),1]; *truncate covariate vector to include only patients who have responded to Trt A;

Cov_A_Mean = Cov_A_Sub[:];

Cov_A_Var = (Cov_A_Sub-Cov_A_Mean)[##]/(ARespTot-1);

* S_A_XY_Bar = (Cov_A_Sub#TrtASub)[+]-ARespTot*Y_A_Bar*X_Bar;

S_A_XY_Bar = (Cov_A_Sub#TrtASub)[+]-ARespTot*Y_A_Bar*Cov_A_Mean;

Cov_A_XY_Hat = S_A_XY_Bar/(ARespTot-1);

Cov_B_Sub = Covariates[loc((RespondedB) = 1),1]; *truncate covariate vector to include only patients who have responded to Trt B;

Cov_B_Mean = Cov_B_Sub[:];

Cov_B_Var = (Cov_B_Sub-Cov_B_Mean)[##]/(BRespTot-1);

* S_B_XY_Bar = (Cov_B_Sub#TrtBSub)[+]-BRespTot*Y_B_Bar*X_Bar;

S_B_XY_Bar = (Cov_B_Sub#TrtBSub)[+]-BRespTot*Y_B_Bar*Cov_B_Mean;

Cov_B_XY_Hat = S_B_XY_Bar/(BRespTot-1);

*Estimates using sums of products (Y*x), minus double-counting products;

/*****Estimate Slopes (BetaA & BetaB) of Covariate (X) for each treatment*****/

Slope_A_Hat = (Cov_A_XY_Hat)/(Cov_A_Var);

Slope_B_Hat = (Cov_B_XY_Hat)/(Cov_B_Var);

*Estimates using ratio of covariance/variance–Covariance between Trt and Outcome over Variance of Covariate;

/*****Estimate Treatment Means and Difference in Treatment Effects*****/

MuA_Hat = Y_A_Bar - Cov_A_Mean*Slope_A_Hat; *Estimate Trt A response less covariate effect;

MuB_Hat = Y_B_Bar - Cov_B_Mean*Slope_B_Hat; *Estimate Trt B response less covariate effect;

Trt_Diff_Est = MuA_Hat - MuB_Hat; *Estimate of Trt Diff using all previous estimates;

/*****Estimate Treatment Variances–variances in treatment effect without accounting for covariate effect*****/

VarA_Hat_Est = (TrtASub-Y_A_Bar)[##]/(ARespTot-1) - (Slope_A_Hat**2)*Cov_A_Var - 2*Slope_A_Hat*0;

VarB_Hat_Est = (TrtBSub-Y_B_Bar)[##]/(BRespTot-1) - (Slope_B_Hat**2)*Cov_B_Var - 2*Slope_B_Hat*0;

*Estimate of Treatment Effect Variances: Use Property of Variance Addition for Dependent Variables;

/* Var(A+B) = Var(A) + Var(B) + 2Cov(A,B) or equivalently, Var(A) = Var(A+B) - Var(B) - 2Cov(A,B)

Outcome = TrtEffect + TrtErr + Slope*(Covariate + CovErr)

= TrtErr + Slope*CovErr + TrtEffect + Slope*Covariate –where latter 2 are constants with 0 variance

Var(Outcome) = Var(TrtErr) + Var(Slope*CovErr) + 2Cov(TrtErr,Slope*CovErr) + 0 + 0

= Var(TrtErr) + Slopeˆ2*Var(CovErr) + Slope*2Cov(TrtErr,CovErr)

Var(TrtErr) = Var(Outcome) - Slopeˆ2*Var(CovErr) - 2*Slope*Cov(TrtErr,CovErr)

Note: 2*Slope*Cov(TrtErr,CovErr) is Covariance of Treatment Error and Covariate Error, NOT of Outcome Error and Covariate Error

Note: Furthermore, Treatment and Covariate errors are independent, uncorrelated, and have no covariance, i.e., last term is ZERO

= Var_Y - Slope2*Var_X - 0 –with Var_Y estimated in the usual manner */

VarA_Hat = max(VarA_Hat_Est,0.00001); *Need to ensure positive (non-zero, non-negative) variances, so take max just in case;

VarB_Hat = max(VarB_Hat_Est,0.00001); *Need to ensure positive (non-zero, non-negative) variances, so take max just in case;

/*****Estimate Phi([MuA - MuB]/Sigma) and randomized based on it*****/

/*****Sigma is estimated using a pooled variance estimate that is unbiased when the variances are equal*****/

Ratio = (MuA_Hat - MuB_Hat)/sqrt(((ARespTot-1)*VarA_Hat+(BRespTot-1)* VarB_Hat)/(ARespTot+BRespTot-2));

Phi = CDF('NORMAL',Ratio,0,1);

end; *ends gathering final estimates for study;

/*****Gather overall study stats*****/

/*****Number and proportion of patients assigned to each treatment*****/

n_A = ARespTot;

n_B = BRespTot;

P_A = n_A/N;

P_B = n_B/N;

/*****Number and proportion of patients assigned using adaptive randomization*****/

```
    NumAdRand = N - NotAdRand;
    PctAdRand = NumAdRand/N;
    *MuA_Hat = MuA_Hat;
    *MuB_Hat = MuB_Hat;
    *StdA_Hat = sqrt(VarA_Hat);
    *StdB_Hat = sqrt(VarB_Hat);
    /*****Actual and estimated pooled variances*****/
    Sigma = sqrt( ((n_A-1)*(StdA**2) + (n_B-1)*(StdB**2)) / (N-2) );
    Sigma_Hat = sqrt(((ARespTot-1)*VarA_Hat+(BRespTot-1)*VarB_Hat)/
    (ARespTot+BRespTot-2));
    *Ratio = Ratio;
    *Phi_Hat = Phi;
    Phi_Real = CDF('NORMAL',(MuA - MuB)/sqrt(Sigma),0,1);
    /*****Trial acceptance (0) or rejection (1) of the null hypothesis*****/
    /*****Use this when under the assumption that the variances are equal*****/
    DF_Eq = N - 2;
    SP_Eq = ((ARespTot-1)*VarA_Hat+(BRespTot-1)*VarB_Hat)/(ARespTot+BRespTot-
2);
    *WT = (1/ARespTot + 1/BRespTot);
    WT_Eq = 1/ARespTot + 1/BRespTot + (Cov_A_Mean + Cov_B_Mean)**2/((Cov_A_Sub
- Cov_A_Mean)[##]+(Cov_B_Sub - Cov_B_Mean)[##]);
    SE_Eq = sqrt(SP_Eq*WT_Eq);
    *Use this when using a pooled standard error for equal-variances t-test;
    /*****Use this when under the assumption that the variances are unequal–generally
safer assumption*****/
    W_A = VarA_Hat/n_A;
```

W_B = VarB_Hat/n_B;

DF_Un = (W_A+W_B)**2/(W_A**2/(n_A-1)+W_B**2/(n_B-1));

*SE_Unequal = sqrt((VarA_Hat/ARespTot)+(VarB_Hat/BRespTot));

SE_Un = sqrt((VarA_Hat/ARespTot)+(VarB_Hat/BRespTot)+

sqrt(VarA_Hat*VarB_Hat)*(Cov_A_Mean + Cov_B_Mean)**2/

((Cov_A_Sub - Cov_A_Mean)[##]+ (Cov_B_Sub - Cov_B_Mean)[##]));

*Use this when using an unpooled Welsh-Satterthwait standard error for separate-variances t-test;

/*

Satterthwaite's DF Approximation

The degrees of freedom are adjusted for unequal group variances as follows

df= (w1 + w2)**2/(w1**2/(n1-1) + w2**2/(n2-1))

where

w1 = s1**2/n1 , w2 = s2**2/n2, s1**2 and s2**2 are the sample variances for groups 1 and 2, respectively,

and n1 and n2 are the number of observations for groups 1 and 2, respectively.

*/

/*****Define Rejection based on assumption of Equal or Unequal variances for t-test*****/

/*****Define t-test test statistic based on whether the t-test is one-sided or two-sided*****/

if Sides = 1 then Reject_Equal=(1-probt((MuA_Hat-MuB_Hat)/SE_Eq,DF_Eq) ¡ Alpha/Sides);

else if Sides = 2 then Reject_Equal=(1-probt(abs(MuA_Hat-MuB_Hat)/SE_Eq,DF_Eq) ¡ Alpha/Sides);

if Sides = 1 then Reject_Unequal=(1-probt((MuA_Hat-MuB_Hat)/SE_Un,DF_Un) ¡ Alpha/Sides);

else if Sides = 2 then Reject_Unequal=(1-probt(abs(MuA_Hat-MuB_Hat)/SE_Un,DF_Un) ¡ Alpha/Sides);

*RejectTrue=(3 ¿ 1.96); * = 1 when True–i.e., when Rejecting the Null Hypothesis;

*RejectFalse=(0 ¿ 1.96); * = 0 when False–i.e., when Failing to reject the Null Hypothesis;

Dist1 = sqrt(N)*(P_A - CDF('NORMAL',((MuA - MuB)/Sigma),0,1));

Dist2 = sqrt(N)*(P_B - (1-CDF('NORMAL',((MuA - MuB)/Sigma),0,1)));

Vec1 = sqrt(N)*(MuA_Hat - MuA);

Vec2 = sqrt(N)*(Slope_A_Hat - CovSlopeA);

Vec3 = sqrt(N)*(MuB_Hat - MuB);

Vec4 = sqrt(N)*(Slope_B_Hat - CovSlopeB);

Ans = n_A||n_B||P_A||P_B

||NumAdRand||PctAdRand

||MuA_Hat||MuB_Hat

||VarA_Hat||VarB_Hat

||Sigma||Sigma_Hat

||Cov_A_Mean||Cov_A_Var

||Cov_B_Mean||Cov_B_Var

||Slope_A_Hat||Slope_B_Hat

||Cov_A_XY_Hat||Cov_B_XY_Hat

||Ratio||Phi||Phi_Real

||DF_Eq||SE_Eq||Reject_Equal

||DF_Un||SE_Un||Reject_Unequal

```
||Dist1||Dist2||Vec1||Vec2||Vec3||Vec4;
*print Ans;
return(Ans);
finish Simulate;
/*****Iterate the Simulation using varying parameters and allocations*****/
start Iter(MuA, MuB, StdA, StdB, CovMu, CovStd, CovSlopeA, CovSlopeB, IncCurCov, DelayA, DelayB, Alpha, Sides, N, m, Reps);
i=1;
Sim = Simulate(MuA, MuB, StdA, StdB, CovMu, CovStd, CovSlopeA, CovSlopeB, IncCurCov, DelayA, DelayB, Alpha, Sides, N, m);
do i=2 to Reps;
Sim=Sim//Simulate(MuA, MuB, StdA, StdB, CovMu, CovStd, CovSlopeA, CovSlopeB, IncCurCov, DelayA, DelayB, Alpha, Sides, N, m);
end;
n_A = sum(Sim[,1]);
n_A_Hat = sum(Sim[,1])/Reps;
n_B = sum(Sim[,2]);
n_B_Hat = sum(Sim[,2])/Reps;
P_A_Hat = sum(Sim[,3])/Reps;
P_B_Hat = sum(Sim[,4])/Reps;
P_A_Var = ((Sim[,3]-P_A_Hat)[##])/(Reps-1);
P_B_Var = ((Sim[,4]-P_B_Hat)[##])/(Reps-1);
NumAdRand = sum(Sim[,5]);
NumAdRand_Hat = sum(Sim[,5])/Reps;
PctAdRand_Hat = sum(Sim[,6])/Reps;
MuA_Hat = sum(Sim[,7])/Reps;
```

```
MuB_Hat = sum(Sim[,8])/Reps;
VarA_Hat = sum(Sim[,9])/Reps;
VarB_Hat = sum(Sim[,10])/Reps;
Sigma = sum(Sim[,11])/Reps;
Sigma_Hat = sum(Sim[,12])/Reps;
Cov_A_Mean = sum(Sim[,13])/Reps;
Cov_A_Var = sum(Sim[,14])/Reps;
Cov_B_Mean = sum(Sim[,15])/Reps;
Cov_B_Var = sum(Sim[,16])/Reps;
Slope_A_Hat = sum(Sim[,17])/Reps;
Slope_B_Hat = sum(Sim[,18])/Reps;
Cov_A_XY_Hat = sum(Sim[,19])/Reps;
Cov_B_XY_Hat = sum(Sim[,20])/Reps;
Ratio = sum(Sim[,21])/Reps;
Phi = sum(Sim[,22])/Reps;
Phi_Real = sum(Sim[,23])/Reps;
DF_Eq = sum(Sim[,24])/Reps;
SE_Eq = sum(Sim[,25])/Reps;
RejectEqPct = sum(Sim[,26])/Reps;
RejectEqVar = ((Sim[,26]-RejectEqPct)[##])/(Reps-1);
DF_Un = sum(Sim[,27])/Reps;
SE_Un = sum(Sim[,28])/Reps;
RejectUnPct = sum(Sim[,29])/Reps;
RejectUnVar = ((Sim[,29]-RejectUnPct)[##])/(Reps-1);
CI_Eq_L = (MuA_Hat-MuB_Hat) - tinv(1-Alpha/2,DF_Eq) * SE_Eq;
CI_Eq_U = (MuA_Hat-MuB_Hat) + tinv(1-Alpha/2,DF_Eq) * SE_Eq;
```

```
    CI_Un_L = (MuA_Hat-MuB_Hat) - tinv(1-Alpha/2,DF_Un) * SE_Un;
    CI_Un_U = (MuA_Hat-MuB_Hat) + tinv(1-Alpha/2,DF_Un) * SE_Un;
    Dist1_Mu_Hat = sum(Sim[,30])/Reps;
    Dist2_Mu_Hat = sum(Sim[,31])/Reps;
    Dist1_Var_Hat = ((Sim[,30]-Dist1_Mu_Hat)[##])/(Reps-1);
    Dist2_Var_Hat = ((Sim[,31]-Dist2_Mu_Hat)[##])/(Reps-1);
    Vec1_Mu_Hat = sum(Sim[,32])/Reps;
    Vec2_Mu_Hat = sum(Sim[,33])/Reps;
    Vec3_Mu_Hat = sum(Sim[,34])/Reps;
    Vec4_Mu_Hat = sum(Sim[,35])/Reps;
    Vec1_Var_Hat = ((Sim[,32]-Vec1_Mu_Hat)[##])/(Reps-1);
    Vec2_Var_Hat = ((Sim[,33]-Vec2_Mu_Hat)[##])/(Reps-1);
    Vec3_Var_Hat = ((Sim[,34]-Vec3_Mu_Hat)[##])/(Reps-1);
    Vec4_Var_Hat = ((Sim[,35]-Vec4_Mu_Hat)[##])/(Reps-1);
    Vec12_Cov_Hat = ((Sim[,32]-Vec1_Mu_Hat)'*(Sim[,33]-Vec2_Mu_Hat))/(Reps-1);
/* dot product (scalar product, inner product) */
    Vec13_Cov_Hat = ((Sim[,32]-Vec1_Mu_Hat)'*(Sim[,34]-Vec3_Mu_Hat))/(Reps-1);
/* dot product (scalar product, inner product) */
    Vec14_Cov_Hat = ((Sim[,32]-Vec1_Mu_Hat)'*(Sim[,35]-Vec4_Mu_Hat))/(Reps-1);
/* dot product (scalar product, inner product) */
    Vec23_Cov_Hat = ((Sim[,33]-Vec2_Mu_Hat)'*(Sim[,34]-Vec3_Mu_Hat))/(Reps-1);
/* dot product (scalar product, inner product) */
    Vec24_Cov_Hat = ((Sim[,33]-Vec2_Mu_Hat)'*(Sim[,35]-Vec4_Mu_Hat))/(Reps-1);
/* dot product (scalar product, inner product) */
    Vec34_Cov_Hat = ((Sim[,34]-Vec3_Mu_Hat)'*(Sim[,35]-Vec4_Mu_Hat))/(Reps-1);
/* dot product (scalar product, inner product) */
```

```
    Ans=MuA||MuB||StdA||StdB||CovMu||CovStd||CovSlopeA||CovSlopeB
    ||IncCurCov||DelayA||DelayB||Sides||m
    ||N||n_A_Hat||n_B_Hat||P_A_Hat||P_B_Hat||P_A_Var||P_B_Var||PctAdRand_Hat
    ||MuA_Hat||MuB_Hat
    ||VarA_Hat||VarB_Hat
    ||Sigma||Sigma_Hat
    ||Cov_A_Mean||Cov_A_Var
    ||Cov_B_Mean||Cov_B_Var
    ||Slope_A_Hat||Slope_B_Hat
    ||Cov_A_XY_Hat||Cov_B_XY_Hat
    ||Ratio||Phi||Phi_Real
    ||DF_Eq||SE_Eq||RejectEqPct||RejectEqVar
    ||DF_Un||SE_Un||RejectUnPct||RejectUnVar
    ||CI_Eq_L||CI_Eq_U||CI_Un_L||CI_Un_U
    ||Dist1_Mu_Hat||Dist2_Mu_Hat||Dist1_Var_Hat||Dist2_Var_Hat
    ||Vec1_Mu_Hat||Vec2_Mu_Hat||Vec3_Mu_Hat||Vec4_Mu_Hat
    ||Vec1_Var_Hat||Vec2_Var_Hat||Vec3_Var_Hat||Vec4_Var_Hat
    ||Vec12_Cov_Hat||Vec13_Cov_Hat||Vec14_Cov_Hat
    ||Vec23_Cov_Hat||Vec24_Cov_Hat||Vec34_Cov_Hat;
    *print Ans;
    return(Ans);
    finish Iter;
    *start Iter(MuA, MuB, StdA, StdB, CovMu, CovStd, CovSlopeA, CovSlopeB,
IncCurCov, DelayA, DelayB, Alpha, Sides, N, m, Reps);
    WorkAround = 50 ,100 ,500 ,1000, 5000, 10000;
    /*Run Sim over several different intervals*/
```

```
    do Row=1 to 5;
    CovMu = 1;
    CovStd = 1;
    CovSlopeA = 2;
    CovSlopeB = 2;
    IncCurCov = 0; *0=N, 1=Y, 9=Resp;
    DelayA = 40;
    DelayB = 40;
    Sides = 1; *1 or 2;
    Alpha = 0.05;
    Beta = 0.20;
    N = WorkAround[Row];
    m = 10;
    Reps = 1000;
    a = iter(0,0,0.5,1,CovMu,CovStd,CovSlopeA,CovSlopeB,IncCurCov,
DelayA,DelayB,Alpha,Sides,N,m,Reps);
    a = a//iter(0,0,0.7,1,CovMu,CovStd,CovSlopeA,CovSlopeB,IncCurCov,
DelayA,DelayB,Alpha,Sides,N,m,Reps);
    a = a//iter(0,0,0.9,1,CovMu,CovStd,CovSlopeA,CovSlopeB,IncCurCov,
DelayA,DelayB,Alpha,Sides,N,m,Reps);
    a = a//iter(0,0,1.0,1,CovMu,CovStd,CovSlopeA,CovSlopeB,IncCurCov,
DelayA,DelayB,Alpha,Sides,N,m,Reps);
    a = a//iter(0,0,1.1,1,CovMu,CovStd,CovSlopeA,CovSlopeB,IncCurCov,
DelayA,DelayB,Alpha,Sides,N,m,Reps);
    a = a//iter(0,0,1.3,1,CovMu,CovStd,CovSlopeA,CovSlopeB,IncCurCov,
DelayA,DelayB,Alpha,Sides,N,m,Reps);
```

```
a = a//iter(0,0,1.5,1,CovMu,CovStd,CovSlopeA,CovSlopeB,IncCurCov,
DelayA,DelayB,Alpha,Sides,N,m,Reps);
a = a//iter(1,0,0.5,1,CovMu,CovStd,CovSlopeA,CovSlopeB,IncCurCov,
DelayA,DelayB,Alpha,Sides,N,m,Reps);
a = a//iter(1,0,0.7,1,CovMu,CovStd,CovSlopeA,CovSlopeB,IncCurCov,
DelayA,DelayB,Alpha,Sides,N,m,Reps);
a = a//iter(1,0,0.9,1,CovMu,CovStd,CovSlopeA,CovSlopeB,IncCurCov,
DelayA,DelayB,Alpha,Sides,N,m,Reps);
a = a//iter(1,0,1.0,1,CovMu,CovStd,CovSlopeA,CovSlopeB,IncCurCov,
DelayA,DelayB,Alpha,Sides,N,m,Reps);
a = a//iter(1,0,1.1,1,CovMu,CovStd,CovSlopeA,CovSlopeB,IncCurCov,
DelayA,DelayB,Alpha,Sides,N,m,Reps);
a = a//iter(1,0,1.3,1,CovMu,CovStd,CovSlopeA,CovSlopeB,IncCurCov,
DelayA,DelayB,Alpha,Sides,N,m,Reps);
a = a//iter(1,0,1.5,1,CovMu,CovStd,CovSlopeA,CovSlopeB,IncCurCov,
DelayA,DelayB,Alpha,Sides,N,m,Reps);
a = a//iter(0,1,0.5,1,CovMu,CovStd,CovSlopeA,CovSlopeB,IncCurCov,
DelayA,DelayB,Alpha,Sides,N,m,Reps);
a = a//iter(0,1,0.7,1,CovMu,CovStd,CovSlopeA,CovSlopeB,IncCurCov,
DelayA,DelayB,Alpha,Sides,N,m,Reps);
a = a//iter(0,1,0.9,1,CovMu,CovStd,CovSlopeA,CovSlopeB,IncCurCov,
DelayA,DelayB,Alpha,Sides,N,m,Reps);
a = a//iter(0,1,1.0,1,CovMu,CovStd,CovSlopeA,CovSlopeB,IncCurCov,
DelayA,DelayB,Alpha,Sides,N,m,Reps);
a = a//iter(0,1,1.1,1,CovMu,CovStd,CovSlopeA,CovSlopeB,IncCurCov,
DelayA,DelayB,Alpha,Sides,N,m,Reps);
```

```
a = a//iter(0,1,1.3,1,CovMu,CovStd,CovSlopeA,CovSlopeB,IncCurCov,
DelayA,DelayB,Alpha,Sides,N,m,Reps);
a = a//iter(0,1,1.5,1,CovMu,CovStd,CovSlopeA,CovSlopeB,IncCurCov,
DelayA,DelayB,Alpha,Sides,N,m,Reps);
print "Simulations for delayed RAR procedure with large response desirable";
name=repeat("Delay",21,1); *3x7=21 rows in table now;
print a[colname="MuA" "MuB" "StdA" "StdB" "CovMu" "CovStd" "CovSlo-
peA" "CovSlopeB"
"IncCurCov" "DelayA" "DelayB" "Sides" "m"
"N" "nAHat" "nBHat" "PAHat" "PBHat" "PAVar" "PBVar" "PctAdRandHat"
"MuAHat" "MuBHat" "VarAHat" "VarBHat" "Sigma" "SigmaHat"
"CovAMu" "CovAVar" "CovBMu" "CovBVar"
"SlopeAHat" "SlopeBHat"
"CovAXYHat" "CovBXYHat"
"Ratio" "Phi" "PhiReal"
"DFEq" "SEEq" "RejectEqPct" "RejectEqVar"
"DFUn" "SEUn" "RejectUnPct" "RejectUnVar"
"CI_Eq_L" "CI_Eq_U" "CI_Un_L" "CI_Un_U"
"Dist1_Mu_Hat" "Dist2_Mu_Hat" "Dist1_Var_Hat" "Dist2_Var_Hat"
"Vec1_Mu_Hat" "Vec2_Mu_Hat" "Vec3_Mu_Hat" "Vec4_Mu_Hat"
"Vec1_Var_Hat" "Vec2_Var_Hat" "Vec3_Var_Hat" "Vec4_Var_Hat"
"Vec12_Cov_Hat" "Vec13_Cov_Hat" "Vec14_Cov_Hat"
"Vec23_Cov_Hat" "Vec24_Cov_Hat" "Vec34_Cov_Hat" rowname=name];
end; *ends iteration of work-around to run through all sims for various sample
sizes;
quit;
```