The Texas Medical Center Library

DigitalCommons@TMC

The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access)

The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences

12-2019

A SIGNATURE ENRICHMENT DESIGN WITH BAYESIAN ADAPTIVE RANDOMIZATION FOR CANCER CLINICAL TRIALS

Fang Xia

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

 Part of the Biostatistics Commons, Clinical Trials Commons, and the Medicine and Health Sciences Commons

Recommended Citation

Xia, Fang, "A SIGNATURE ENRICHMENT DESIGN WITH BAYESIAN ADAPTIVE RANDOMIZATION FOR CANCER CLINICAL TRIALS" (2019). *The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access)*. 985.
https://digitalcommons.library.tmc.edu/utgsbs_dissertations/985

This Dissertation (PhD) is brought to you for free and open access by the The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences at DigitalCommons@TMC. It has been accepted for inclusion in The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact digitalcommons@library.tmc.edu.

A SIGNATURE ENRICHMENT DESIGN WITH BAYESIAN ADAPTIVE RANDOMIZATION FOR CANCER CLINICAL TRIALS

by

Fang Xia, M.S.

APPROVED:

_____

Xuelin Huang, Ph.D.
Advisory Professor

_____

Jing Ning, Ph.D.

_____

Liang Li, Ph.D.

_____

Yu Shen, Ph.D.

_____

Courtney DiNardo, M.D.

APPROVED:

_____

Dean, The University of Texas
MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences

A SIGNATURE ENRICHMENT DESIGN WITH BAYESIAN ADAPTIVE RANDOMIZATION FOR CANCER CLINICAL TRIALS

A

DISSERTATION

Presented to the Faculty of

The University of Texas

MD Anderson Cancer Center UTHealth

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

by

Fang Xia, M.S.

Houston, Texas

December, 2019

To my dear family

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Xuelin Huang, for his continued guidance, support, and patience throughout my graduate study at the University of Texas MD Anderson UTHealth Graduate School of Biomedical Sciences.

I would like to thank my advisory committee: Dr. Jing Ning, Dr. Liang Li, Dr. Yu Shen, and Dr. Courtney DiNardo, for their comments and suggestions. I thank them for their valuable insights and efforts that helped shape my dissertation project.

I would like to give a special thanks to my mentor, Dr. Stephen L. George, for introducing me to the world of clinical trial design and for the tremendous support and encouragement throughout my ups and downs.

ABSTRACT

A SIGNATURE ENRICHMENT DESIGN WITH BAYESIAN ADAPTIVE RANDOMIZATION FOR CANCER CLINICAL TRIALS

Fang Xia, M.S.

Advisory Professor: Xuelin Huang, Ph.D.

Clinical trials in the era of precision medicine demand more flexible and efficient trial designs. Adaptive clinical trial designs allow pre-specified modifications of an on-going clinical trial and could shorten the trial duration. We reviewed five common types of adaptive clinical trials based on adaptation methods. In particular, outcome-randomization becomes more popular as it can assign more patients to the promising treatments based on the accumulated trial data. This data-driven allocation allows more patients to benefit from the trial, which is especially important for cancer patients. We compared different Bayesian outcome-adaptive randomization methods and discussed them from both methodological and ethical aspects.

When the group of patients who are likely to benefit from the test treatment is known, the clinical trial should focus on that sensitive subpopulation. The use of biomarkers in adaptive clinical trials can guide the assignment of individually optimal treatments to patients. To address this objective, we proposed a cross-validated signature enrichment design combined with Bayesian response-adaptive randomization. We evaluated the performance of this design using four criteria based on the benefits and losses for individuals inside and outside of the clinical trial. The proposed design allows more patients to receive optimal personalized

treatments, thus yielding a higher response rate. This design can identify therapies that are globally beneficial as well as treatments that are effective only in a sensitive subset. Simulation studies demonstrate the advantages of the proposed design over alternative designs. The approach is also illustrated by an example based on an actual clinical trial in non-small-cell lung cancer.

In contrast to the traditional two-arm trial, an umbrella trial is a master protocol that studies multiple drugs within a single indication, which can result in more accurate and efficient drug development. Based on SEDAR, we proposed a SEDAR-U design that is suitable for umbrella trials. We extended SEDAR to allow multiple arms and incorporate an early termination rule. Simulation studies showed that the SEDAR-U design can allow more enrolled patients to benefit from the promising treatment and achieve favorable operating characteristics.

TABLE OF CONTENTS

List of Figures

List of Tables

CHAPTER 1

Introduction

As the understanding of molecular, genomic, and phenotypic properties of cancer emerge, novel clinical trial designs are pressingly needed to improve the efficiency of drug development. Adaptive clinical trials allow pre-specified modification to the trial as it proceeds, which may shorten the trial duration and require fewer patients [4, 5]. Some common adaptation features in oncology trials include sample size re-estimation, adaptive randomization, and interim monitoring. Adaptive design can be used in all phases of clinical development. Phase I clinical trials aim to estimate the maximum tolerable dose (MTD) for the new drug, which is the highest dose with acceptable side effects. The continual reassessment method (CRM) [6, 7, 8] is an adaptive phase I trial that estimates the MTD based on a model instead of the traditionally rule-based designs [9, 10].

Phase II clinical trials aim to identify the efficient new treatment that can then be tested further in a phase III trial. The primary endpoint for a phase II trial is usually tumor response (response/no response). Phase II oncology trials are often single-arm, but may also be randomized [11, 12, 13]. Most adaptations have been applied to phase II trials. Some common adaptations include adaptive randomization, biomarker adaptive, sample size re-estimation, et al. Other than the frequentist

framework designs [14, 15, 16, 17], Bayesian approaches permit modifications based on accumulating data [18, 19].

Phase III clinical trials aim to evaluate the efficacy of the experimental treatments compared with the existing therapies or standard of care. They are large-scale randomized trials, typically consist of two or more treatment arms with hundreds to thousands of patients. The primary endpoints for phase III oncology trials usually are the time-to-event outcome (progression-free survival, overall survival, and disease-free survival) or binary outcome (response/no response). Phase III trials often include an interim analysis that allows early termination due to futility [5] and follows intention-to-treat principle [20]. At the end of a positive phase III trial, the experimental treatments will be submitted to the FDA for approval. The seamless phase II/III trial is an adaptive design that promotes a more rapid transition between phases II and III by combining effectiveness assessment and confirmatory into one trial [21, 22].

The rapid advancement of cancer biomarker development allows a better fit between patients and treatment options [23, 24, 25]. There are two types of biomarkers based on usage: prognostic and predictive biomarkers. Prognostic biomarkers provide clinical outcome information independent of treatments, while predictive biomarkers directly link the clinical outcome information to a specific treatment. If a treatment that targets a particular biomarker, such information can be used to categorize patients into either "sensitive" or "non-sensitive" patient subgroups to further assess the patients' responses. When the group of "sensitive" patients who are likely to benefit from the experimental treatment is known, the clinical trial should focus on that sensitive subpopulation. Some predictive cancer biomarkers that have been used to guide patient screening include human epidermal growth factor receptor

2 (HER2) for breast cancer [26, 27], BRAF for colorectal cancer and melanoma [28], estimated glomerular filtration rate (EGFR) for non-small-cell lung cancer and colorectal cancer [29, 23]. Furthermore, in 2017, the FDA approves Pembrolizumab for any solid tumor with the microsatellite instability-high (MSI-H) or mismatch repair deficient (dMMR) biomarker marked the agency's first approval based on biomarkers rather than tumor location [30].

The emergence of translating biomarker information into clinical use calls for more efficient and accelerated drug development. Master protocols is an overarching protocol which concurrently operates multiple substudies to simultaneously evaluate one or more interventions in one or more disease types, where each substudy may be conducted based on tumor type or biomarker [31, 32, 33, 25]. Based on the characteristics of the study, a master protocol is classified into basket, umbrella, and platform trials. The basket trial evaluates one treatment for multiple diseases with a common biomarker; the umbrella trial evaluates multiple treatments with each sub-study targeting one biomarker for one disease type. The platform trials assess perpetually several treatments for one disease and permit flexible addition and/or exclusion of new treatments during the trial. If basket and umbrella trials allow the addition of new therapies and exclusion of inferior treatments during the trial, they can also be considered as platform trials [33].

This dissertation is organized as follows. In chapter 2, we provide an overview of adaptive clinical trial designs. In chapter 3, we compare different Bayesian response-adaptive randomization methods and discuss their application from methodological and ethical considerations. In chapter 4, we propose a signature enrichment design with Bayesian adaptive randomization (SEDAR) for cancer clinical trials. In chapter 5, we propose an extension to the SEDAR that incorporates an early termination

rule and allows multiple treatments for umbrella trials. In chapter 6, we conclude with a brief discussion and future work.

CHAPTER 2

Review of Adaptive Clinical Trial Designs

Traditional randomized clinical trial uses a fixed ratio to assign patients to each treatment arm and analyze the treatment effect at the end of the study. In recent years, adaptive designs have become increasingly popular because of its flexibility as an alternative to traditional randomized clinical trials. Adaptive designs allow modification of the on-going clinical trials according to the information collected during the trial without undermining its integrity and validity. The use of adaptive design is attractive to investigators, sponsors, and even patients because it may reduce the trial duration, limit the number of patients exposing to the ineffective/inferior treatments and increase the probability of success in later phase drug development. Due to the modifications applied, some common types of adaptive designs include a) adaptive dose-finding, b) response-adaptive randomization, c) adaptive group sequential, d) biomarker adaptive, e) seamless phase II/III. Details of each design are described below.

2.1 Adaptive Dose-finding Design

The main objective of phase I clinical trials is to establish the maximum tolerated dose (MTD) for phase II trials. One fundamental assumption is that the efficacy of the drug monotonically increases with dose levels. However, as the doseage

gets higher, the toxicity increases at the same time. If the dose level is too low, the drug may not be efficient; if its too high, the drug may be too toxic. The MTD is considered the optimal dose level. A well-designed dose-finding should achieve a high probability of finding the MTD while limiting the number of patients being exposed to the dose-limiting toxicity (DLT).

Dose-finding designs can be classified as rule-based design or model-based design. The most widely used rule-based design is called "3+3" dose-escalation design [9, 10]. Below is a simple example of the "3+3" design:

Suppose three patients enter the trial at the initial dose level *M*. If no patient experience the DLT, then escalate to the next dose. If more than two patients experience the DLT, then stop the trial and claim dose M is the MTD. If only one patient has the DLT, then enter an additional three patients at the same dose level *M*. If one out of the six patients has a toxic response, then escalate to the next dose level *M+1*; Otherwise, stop the trial and claim dose M is the MTD. In summary, the estimated MTD is the highest dose level with observed DLT less than 0.33.

Some alternative rule-based designs to the traditional "3+3" design include the "2+4", "3+1+1" and "3+3+3" designs [10].The main advantage of the rule-based designs is easy implementation. However, it only utilizes the data at the current dose to make the decision and may enroll many patients to the ineffective low dose levels and expose some patients to the unsafe toxic dose levels [34, 35]. Some published performance of the "3+3" design also shows that it has a low probability of selecting the true MTD [36] and is highly variable in estimating MTD [35].

An adaptive dose-finding design based on a statistical model that proposed by O'Quigley et al. [6, 7], the continual reassessment method (CRM), is a common alternative to the "3+3" design. It uses a statistical model to estimate the relation-

ship between the dose and the risk of experiencing a DLT. It is a Bayesian procedure based on a one-parameter model. The CRM starts with a predefined set of dosages and the target DLT that defines the MTD, and then uses Bayes theorem to update the distribution of MTD continuously. After observing each patient's response, the posterior distribution of model parameters and predicted probabilities of a toxic reaction at each dose level are updated. The dose level for the next patient is chosen as the one with the predicted toxicity response probability closest to the target level of response. This procedure is repeated until the requisite N patients are enrolled. The final estimate of the MTD is the dose with the closest posterior probability to the target DLT. Additional model-based designs based on CRM includes dose escalation with overdose control EWOC [37], Bayesian model averaging CRM [38], time-to-event CRM [39] and Bayesian logistic regression model (BLRM) [40] etc. Although CRM has shown more desirable operating characteristics than the rule-based designs including a higher probability of determining the true MTD, its application is still limited due to conceptual and operational complexities [41].

Recently, another type of adaptive dose-finding design that combines key elements of the rule-based and model-based designs named model-assisted designs. Similar to rule-based designs, their dose escalation/de-escalation rules can be determined before the start of the trial. It also borrows the strength from model-based designs by using a statistical model to aid decision making. Some examples include the modified toxicity probability interval (mTPI) design [42] and its variation mTPI-2 [43] and the Bayesian optimal interval (BOIN) design [44].

2.2 Response-adaptive Randomization Design

Traditional clinical trials use a fixed ratio to allocate patients among treatment arms. Adaptive randomization allows modification of the allocation probabilities based on accumulated patient responses (outcome). This adjustment enables more patients to receive superior treatments at that moment, limits patients exposing to the subtherapeutic therapies, and potentially results in more accurate and faster drug development. Response-adaptive randomization may be conducted continuously throughout the trial, or as a one-time adaptation in the allocation ratio at the time of an interim futility analysis. Response-adaptive randomization can also be approached from the Frequentist and the Bayesian framework.

A Frequentist approach named doubly-adaptive biased coin design was first proposed by Eisele and Woodroofea [45] and further modified by Hu and Zhang [46]. This design depends on the current proportion of patients from each treatment arm and the current estimate of the desired allocation proportion. The advantages of this design include flexibilities to target any given allocation and allowing multiple types of responses while attaining the lower bound of variance.

Figure 2.1 presents a diagram of the doubly-adaptive biased coin design. In general, it consists of three steps: first, assign a small proportion of patients using equal randomization as a burn-in period; second, calculate the optimal allocation probability based on current patient responses. third, apply the allocation function to calculate the assignment probability for the next new patient, which targets the optimal allocation probability in step 2.

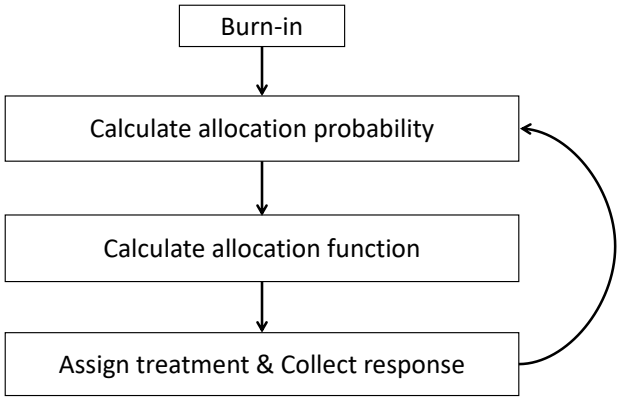

Figure 2.1: Diagram of the Doubly-adaptive Biased Coin Design.

To calculate the optimal allocation probability, the Neyman allocation [47] is a popular choice. For two treatment scenarios, one may also consider the RSIHR proportion [48] to minimize the number of failures. Some discussion of generalizing the RSIHR proportion for K treatments has been discussed in Tymofyeyev et al. [49].

Let k denote the treatment arm index, $k = C, E$, n_k denote the intermediate sample size, $n = 1, ..., N$, and π_k refers to the response probability for arm k. Suppose the allocation probability after step 2 is ρ_k and the allocation function $g(r, s)$ which is defined on $[0, 1] \times [0, 1]$ is as follows [50, 46]:

$$g_k(r, s) = \frac{s_k(s_k/r_k)^\gamma}{\sum_{j=1}^{K} s_j(s_j/r_j)^\gamma}, k = 1, \cdots, K, \tag{2.1}$$

where $\gamma \geq 0$ is a tuning parameter that controls the randomness.

Then the assignment probability of treatment k for the next incoming patient is $g(n_k/N, \hat{\rho}_k)$, where n_k/N is the current proportion of patients in arm k.

In contrast to the Frequentist approach, Bayesian response-adaptive randomization assumes the randomization probability $r_{k,n}$ is proportional to the posterior

probability [51, 52, 18, 53]:

$$Pr(\pi_k = max\{\pi_1, ..., \pi_k\}|\ data_n),\ k = 1, ..., K,$$

When n is small, the posterior probabilities can be highly variable. To stabilize the randomization probability, we add a tuning parameter c such that

$$r_{k,n}^c = \frac{(Pr(\pi_k = max\{\pi_1, ..., \pi_k\}|\ data_n))^c}{\sum_{j=1}^K (Pr(\pi_j = max\{\pi_1, ..., \pi_j\}|\ data_n))^c}.$$

The value of c should be between 0 and 1. When c is equal to 0, it gives equal randomization. When c is equal to 1, the $r_{k,n}^c$ is the same as $\hat{r}_{k,n} = \hat{Pr}(\pi_k = max\{\pi_1, ..., \pi_k\}|\ data_n)$. In practice, c is usually set at 1/2.

Another method to control the randomization magnitude is to let the posterior probability be the randomization probability $r_{k,n}$, and restrict it within a reasonable range. More details of Bayesian response-adaptive randomization will be discussed in chapter 2.

2.3 Adaptive Group Sequential Design

Classical group sequential designs repeatitively apply significance testing at pre-specified stages based on accumulating data. At each stage, it offers a possibility of stopping early without the need of setting a fixed number of sample size. At each interim stage, a decision may be made to support or reject the null hypothesis based on a stopping criterion. A classical group sequential design requires to pre-specify the total number and the timeing of the analyses. The trial will be stopped until the maximum information is obtained or when crossing a stopping boundary. Let δ denote the treatment difference between experimental and control groups. Suppose a total number of K analyses, $k = 1, ..., K$, are planned with a maximum trial duration allowed T. Let t_k be the time of the k^{th} analysis, where $0 < t_1 < t_2 < ... < t_k = T$.

Thus, there is $k-1$ interim analyses and one final analysis. Let τ_k denote the fraction of the expected information observed at time t_k. τ_k can be approximately by

$$\tau_k = \begin{cases} \frac{n_k}{N_{max}} & \text{for quantitative endpoints} \\ \frac{d_k}{D_{max}} & \text{for time-to-event endpoints} \end{cases} \tag{2.2}$$

where N_{max} is the maximum requisite sample size for the trial and D_{max} is the maximum number of events required for the trial. They are often referred to as the maximum information.

Due to multiple testings in the group sequential design trial, Type I error rate will be inflated. To preserve the Type I error rate at α, we need to determine the stopping boundaries $c_1, c_2, ..., c_k$ such that the trial will be stopped under H_0 at the first τ_k when

$$|Z(\tau_k)| \geq c_k.$$

That is, to select a c_k such that $P_0(\bigcap_{j=1}^{K} |Z(\tau_k)| < c_k) = 1 - \alpha$. Examples of the stopping boundaries include the Pocock [54], the O'Brien & Fleming (OBF) [55], the Wang & Tsiatis [56] and the Lan & DeMets [57]. The Pocock uses constant boundaries [54], while the O'Brien & Fleming utilizes increasing boundaries [55]. The Wang & Tsiatis boundaries depend on a shape parameter and the pre-specified information fraction [56]. The Lan & DeMets proposed a flexible boundary approach called the α spending function without requiring a pre-specified number of interim analyses and the timing of those stops [57].

The α spending function $\alpha^*(\tau)$ is a monotone increasing function of τ which defines how much α is spent at each τ [58]. For τ_0, $\alpha^*(0) = 0$; for τ_K, $\alpha^*(\tau_K) = \alpha$. The stopping boundaries c_k is solved recursively from

$$P_0(|Z(\tau_1)| \geq c_1) = \alpha(\tau_1) \text{ for } k = 1;$$

$$\alpha(\tau_{k-1}) + P_0(|Z(\tau_1)| < c_1, ..., |Z(\tau_{j-1})| < c_{j-1}, |Z(\tau_j)| \geq c_j) = \alpha(\tau_j) \text{ for } k = 2, ..., K.$$

The α spending function can also approsimates the O'Brien & Fleming [55] and the Pocock group sequential boundaries [54].

To determine the requisite sample size for a group sequential design, we first need to specify α, the spacing, the number of analyses, desired power β and $\alpha^*(\tau)$ to compute the stopping boundaries. Then compute the required maximum information I_{max} in order to reach the desired power based on the stopping boundaries. The final step is to convert I_{max} into N_{max} for quantitative endpoints or into D_{max} for time-to-event endpoints.

Adaptive group sequential design allows adaptations or modifications at interim analyses [59]. Those adaptations may include but not limited to sample size re-estimation, adding/dropping treatment arm, randomization scheme modification, etc. Although the classical group sequential design is well understood and the Type I error rates can be well controlled under spending functions, adding changes at interim analysis may inflate Type I error rate. Statistical procedures are still needed to control study-wise Type I error rate [60].

2.4 Biomarker Adaptive Design

A biomarker is a biological molecule that is objectively measured and evaluated as an indicator of normal biological, pathogenic processes, or pharmacologic responses to a treatment[61]. The emergence of biomarker development can guide treatment selection for patients according to their biological characteristics or genomic profiles. Based on the type of usage, biomarkers can be classified into prognostic or predictive biomarkers. Prognostic biomarkers assess patient responses independent of treatment, while predictive biomarkers provide information regarding the

most likely response to a given treatment. Prognostic biomarkers can separate potential patients into good or poor prognosis at diagnosis, while predictive biomarkers can separate patients into sensitive or non-sensitive for a given treatment [62]. Based on the characteristic of predictive biomarkers, they can be used to guide treatment selection to allow a better fit between patients and treatment options.

Many molecular cancer treatments only benefit a subset of patients. Biomarker adaptive design identifies a subgroup of patients who are most likely to respond to the test treatment based on their biomarker profiles. Biomarker adaptive design can allow more patients to benefit from the trial compared to the traditional all-comers design, which may increas the trial efficiency and limit non-sensitive patients to experience the toxicity of the test treatment.

Xu et al. propose a subgroup-based adaptive (SUBA) design for multi-arm biomarker trials [63]. The SUBA design simultaneously searches for prognostic subgroups using a random partition model and assigns patients to the best subgroup-specific treatment based on posterior predictive probabilities.

If a biomarker is known at patient screening, this trial should only enroll sensitive patients. If the biomarker information is unknown, the adaptive signature design [1] may be used. The adaptive signature design tests the overall treatment effect at a significance level of α_1. If the overall treatment effect is statistically significant, then conclude that the experimental treatment may benefit all treated patients. Otherwise, test the treatment effect in a sensitive patient subset at a significance level of α_2. This controls the overall type I error rate at α with $\alpha = \alpha_1 + \alpha_2$.

Jiang et al. proposed a biomarker-adaptive threshold design procedure to evaluate treatment with possible biomarker defined subset effect [64]. Their proce-

dure combines both overall and subset tests by incorporating a correlation structure of the two-stage test statistics. Their simulation and real trial data application show that the biomarker-adaptive design substantially improves the efficacy compared with the traditional design. Specifically, their design consists of two procedures: procedure A which is similar to the adaptive signature design [1] and procedure B which combines testing treatment effect for both overall and sensitive patients. Procedure B incorporates a correlation structure of the two-stage test statistics from procedure A. It only chooses the maximum statistic T from all the cut-off models, each corresponding to a potential biomarker cut-ff value. To ensure the power is adequate when the test treatment is indeed beneficial to all treated patients, a positive constant R is added to the overall effect statistic S, whose value is calculated when cut-off value $= 0$. Then the permutation distribution of the test statistic T is used to calculate the P value.

Gu et al. proposed a Bayesian two-stage biomarker-based adaptive randomization design (BATTLE-2) [65]. This design has three main goals: 1) to test the treatment effect 2) to identify prognostic and predictive biomarkers for targeted agents 3) to provide better treatment for patients enrolled in the trial. Bayesian adaptive randomization approach applies to both the biomarker development (stage I) and validation stages (stage II). In stage I, adaptive randomization is applied based on a known biomarker. At the end of the first stage, a Go or No-Go decision is made from testing the overall treatment effect. If a Go decision is made at the end of stage I, a two-step Bayesian Lasso strategy will be used to select additional prognostic or predictive biomarkers based on the accumulated data: group lasso will be used to select important biomarkers, and adaptive lasso will be used to refine the selection and estimation. Patients enrolled in the second stage will receive treatment based

on the refined adaptive randomization scheme.

2.5 Seamless Phase II/III Design

During traditional drug development, phase II trials select the best treatment/dose, and phase III trials further confirm the treatment efficacy. The seamless phase II/III design (also called pick-a-winner design, drop-the-loser design, adaptive seamless design) is a multi-stage design that combines phases II and III into a single trial. It consists of two stages: a selection stage (phase IIb) and a confirmation stage (phase III) [66]. For the selection stage, the objective is to select the best dose/treatment by employing a parallel design with multiple treatment arms and a control arm. The inclusion of a control arm at the selection stage allows the data to be easily combined with that from the confirmation stage for final analysis. The confirmation stage evaluates the efficacy of the selected treatments versus the control treatment.

The general framework for a seamless phase II/III design is as follows: patients entering the trial are randomized between the experimental arms and the control arm. At the end of the first stage, the selected dose/treatment will move to the confirmation stage. Newly enrolled patients will be randomized to receive the selected dose/treatment or the control treatment. If multiple arms are selected at the end of the first stage, or only the inferior treatments are dropped, each selected arm will be compared with the control arm at the confirmation stage.

Thall, Simon, and Ellenberg proposed a two-stage design with binary endpoints to avoid continuous monitoring [67]. Schaid, Wieand, and Therneau extended their design to the time-to-event endpoint and allowed early stopping for efficacy after the selection stage [68]. Stallard and Todd generalized their designs to allow more

than two stages using error spending function and suitable for all types of endpoints by using an efficient score for test statistics [69]. Huang, Liu, and Hsiao presented a seamless design that allows pre-specifying the probabilities of rejecting an arm at each stage [70]. Bischoff and Miller proposed a two-stage adaptive design based on the means of the normal distribution for treatment selection [71]. Their design requires a minimal number of patients to reach the desired power while controlling the type I error rate and limiting the probability of selecting an inferior treatment after the first stage. Todd and Stallard presented a group sequential design that allows a change of endpoint for the confirmation stage [72]. Chow and Tu derived formulas for sample size calculation, assuming the same study objectives at different stages [73]. Posch, Maurer, and Bretz presented two methods to control the type I error rate when applying sample size reassessment and/or treatment selection: one uses simulation to adjust the critical value, and the other uses an adaptive Bonferroni-Holm procedure [74]. Stallard described a method to control the type I error rate by adjusting the usual group sequential boundaries when using short-term endpoint for the treatment selection at the first interim analysis for a confirmatory seamless phase II/III trial[75].

Instead of running multiple experimental arms vs. control arm phase II and III designs, the seamless phase II/III design requires a smaller sample size by utilizing data from both stages and can expedite the drug development process [76, 66]. Since the results of a seamless II/III trial is sensitive to the timing of the analysis for the selection stage, simulations should be done to find an optimal time for the interim analysis [66].

CHAPTER 3

Comparison of Bayesian Response-adaptive Randomization Methods

When designing a clinical trial, equal randomization is commonly used. However, many argue that it is not ethical to assign the same number of patients to poor treatment and to more efficacious treatment. Response/Outcome-adaptive randomization aims to assign more patients to the superior arm based on the information of current patients' performance accumulated as the trial moves along. Although response-adaptive randomization offer many advantages, it has raised some concerns over its complexity in design and coordination, and the difficulty in informing patients and its validity. In this chapter, we compare two Bayesian response-adaptive randomization approaches and discuss them from the methodological and ethical considerations.

3.1 Comparison of Different Bayesian Response-adaptive Randomization Methods with Fixed Randomization

When designing a clinical trial for comparing two or more treatments, equal randomization is commonly used. However, in many cancer clinical trials, patients need to enroll in trials to receive the most advanced therapies. Therefore, it is essential for the trial to offer maximum treatment benefits to the participants. In early

oncology, we may find response-adaptive randomization [77, 78, 50, 79] attractive as it assigns more patients to receive the superior treatment based on current data evidence, which offers patients a higher chance to benefit from enrolling in the trial.

Figure 3.1 presents the diagram of a general Bayesian response-adaptive randomization procedure. A key feature of Bayesian response-adaptive randomization is that it repetitively updates the assignment probability based on accumulated data.

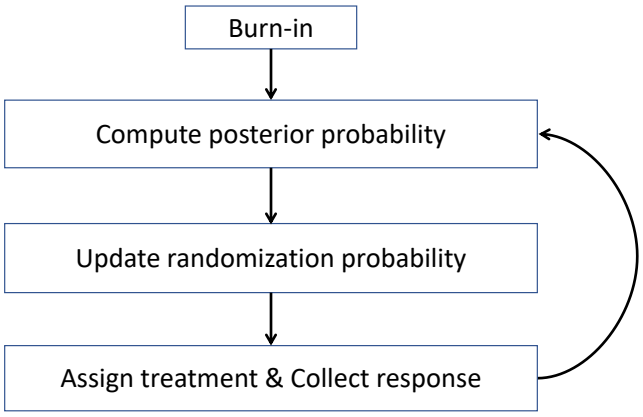

Figure 3.1: Diagram of the Bayesian adaptive randomization.

Consider the design of a two-arm clinical trial to compare treatment A to treatment B to study the efficacy. Assuming X_i is the operating characteristics/biomarkers for patient i. For the Adaptive Randomization Using Response Rate approach (ARRR), the probability of assigning patient i to each treatment arm is computed using the estimated response rates if he/she were received treatment A or B.

Let $r(A, X_i)$ and $r(B, X_i)$ denote the response rate which if patient i receive treatment A or B, respectively. Then probability of patient i receiving A when using ARRR, $P_{i,A,AR-RR}$, is

$$P_{i,A,AR-RR} = \frac{\hat{r}(A, X_i)^w}{\hat{r}(A, X_i)^w + \hat{r}(B, X_i)^w},$$

where w is a tuning parameter that controls the magnitude of the randomizaion.

Another method called Adaptive Randomization Using Superiority Confidence (ARSC) uses the posterior probability $P(r(A, X_i) > r(B, X_i)|data)$ to calculate the probability of treatment assignment. ARSC assumes the allocation probability for treatment A is proportional to the posterior probability $P(r(A, X_i) > r(B, X_i)|data)$. The probability of receiving treatment A when using AR-SC, $P_{i,A,AR-SC}$, is defined as

$$P_{i,A,AR-SC} = \frac{\hat{P}(r(A, X_i) > r(B, X_i)|data)^w}{\hat{P}(r(A, X_i) > r(B, X_i)|data)^w + \hat{P}(r(A, X_i) \leq r(B, X_i)|data)^w}. \quad (3.1)$$

Here we use a simple example to illustrate the main difference between ARRR and ARSC (Table 3.1). At the beginning of the trial, suppose each arm currently has 10 patients and the response rate is 0.6 for arm A and 0.3 for arm B. The probability of being assigned to arm A when using ARRR is 0.67 and when using ARSC is approximately 0.52. Later, suppose each arm reaches 100 patients and the response rate for each arm remains the same. The probability for ARRR still stays at 0.67, while that for ARSC increases to 0.70.

Table 3.1: A simple example: ARRR vs. ARSC

$n_A = n_B = 10$	$r_A = 0.6$	$r_B = 0.3$	$P_{A,ARRR} = 0.67$ $P_{A,ARSC} \approx 0.52$
$n_A = n_B = 100$	$r_A = 0.6$	$r_B = 0.3$	$P_{A,ARRR} = 0.67$ $P_{A,ARSC} \approx 0.70$

According to this example, with only 10 patients, the superiority of treatment A is unclear. Aggressive patient allocation to treatment A is not appropriate. However, with 100 patients, the superiority of treatment A is clear. Aggressive patient allocation to treatment A is desired. This simple example shows that $P_{A,ARRR}$ remains the same regardless of sample size; $P_{A,ARSC}$ increases as the sample size gets larger.

We further compare the performance of ARRR and ARSC in a clinical trial setting through simulation. We use the following logistic regression model for patient responses:

$$logit(r_i) = \beta_0 + \beta_1 T_i + \beta_2 X_i + \beta_3 T_i X_i$$

$$k = A, B; i = 1, \cdot, N.$$

where X_i is the patient biomarkers. We write $X \sim Normal(\mu, \sigma^2)$ to indicate that the random variable X has an normal distribution with mean μ and variance σ^2. Let T_i denote the treatment indicator which

$$T_i = \begin{cases} 0.5 & \text{Treatment} = \text{A} \\ -0.5 & \text{Treatment} = \text{B} \end{cases}, i = 1, ..., N.$$

Let equation $logit(r_i) = \mathbb{BX}_i$. Then response rate r_i is calculated from:

$$r_i = \frac{exp(\mathbb{BX}_i)}{1 + exp(\mathbb{BX}_i)}, i = 1, ..., N,$$

where each patient's response follows a Bernoulli distribution with probability r_i.

The required design parameters and the simulation settings are presented in Table 3.2: the sample size N; the type I error rate α; the initial number of equally randomly assignment patients n_{eq}; the adaptive randomization block size bl; the tuning parameter w and patient biomarker X. The simulation scenarios are listed in Table 3.3. For this simulation study, we focus on the overall response rates and power. The results are presented in Table 3.4.

Table 3.2: Required design parameters

N	200
n_{eq}	80
bl	30
w	0.5
α	0.05
X_i	i.i.d. N(0, 1)

Table 3.3: Simulation scenarios

	β_0	β_1	β_2	β_3
NULL	-0.20	0.00	0.00	0.00
Scenario 1	-0.20	1.20	0.15	1.60
Scenario 2	-0.20	1.20	0.15	0.50
Scenario 3	-0.20	0.90	0.15	1.60
Scenario 4	-0.20	0.90	0.15	0.50

Table 3.4: Simulation results

	Randomization	Power	ORR
NULL	ER	0.036	0.450
	ARRR	0.038	0.449
	ARSC	0.033	0.450
Scenario 1	ER	0.953	0.458
	ARRR	0.944	0.492
	ARSC	0.920	0.519
Scenario 2	ER	0.982	0.457
	ARRR	0.980	0.472
	ARSC	0.980	0.490
Scenario 3	ER	0.767	0.458
	ARRR	0.755	0.485
	ARSC	0.707	0.510
Scenario 4	ER	0.856	0.453
	ARRR	0.840	0.463
	ARSC	0.831	0.475

For the given values of the design parameters, ARSC clearly shows advantages from the increased response rates compared to ER and ARRR. The power, however, is the highest with the trial using ER. This results agrees with the finding from Lee, Chen and Yin [81]: using response-adaptive randomization may often result in some power loss. However, an example in status epilepticus has shown that applying adaptive randomization can achieve a higher power than using fixed randomization [82].

3.2 Methodological Considerations

Yin, Chen and Lee [81] proposed a phase II trial design: Bayesian adaptive randomization and predictive probability (BARPP) which implements response/outcome-adaptive randomization. This design is motivated by a neoadjuvant lung cancer trial at the University of Texas M.D. Anderson Cancer Center. Since the sample size of each arm is unknown due to the response-adaptive randomization mechanism, but such information is required to compute the predictive probability, they proposed two methods to approximate the predictive probability. In general, the predictive probability is defined as:

$$\begin{aligned} PP &= E_{Y_1,Y_2}[I\{P|p_2 - p_1| > \delta | X_1 = x_1, X_2 = x_2, Y_1, Y_2) \geq \theta_T\}] \\ &= \sum_{y_1=0}^{N_1-n_1} \sum_{y_2=0}^{N_2-n_2} P(Y_1 = y_1|X_1 = x_1) P(Y_2 = y_2|X_2 = x_2) \\ &\times \ I\{P(|p_2 - p_1| > \delta | X_1 = x_1, X_2 = x_2, Y_1 = y_1, Y_2 = y_2) \geq \theta_T\}, \end{aligned} \tag{3.2}$$

where Y_k is the number of future responses in arm k, p_k is the corresponding response rate, the number of response observed X_k, the maximum number of patients N_k, and the number of patients currently in arm k, n_k, $k = 1, 2$. δ is the clinically meaningful treatment difference and θ_T is a threshold probability.

The decision rule based on PP is defined as[83]:

- Equivalence stopping: Stop the trial and claim the two arms are equivalent if $PP < \theta_L$;
- Superiority stopping: Stop the trial and claim a superior arm if $PP > \theta_U$.

Type I error rate and statistical power are maintained by calibrating $(N, \delta, \theta_T, \theta_L, \theta_U)$. In outcome-adaptive randomization clinical trials, N_1 and N_2 are not fixed. The authors proposed two methods to compute the PP.

Recall that outcome-adaptive randomization assigns each newly enrolled patient to a treatment group based on his/her randomization probability. Randomization probability is calculated based on cumulative information from currently enrolled patients in both arms. It can be denoted following the work of Thall and Wathen [80] as:

$$\pi = \frac{P(p_2 > p_1 | X_1 = x_1, X_2 = x_2)^\tau}{P(p_2 > p_1 | X_1 = x_1, X_2 = x_2)^\tau + \{1 - P(p_2 > p_1 | X_1 = x_1, X_2 = x_2)\}^\tau}, \tag{3.3}$$

where π is the probability of assigning to arm 2 and τ is the turning parameter which controls the rate of adaptive randomization. τ is recommended to take value between 0 and 1. When $\tau = 0$, it is equivalent to equal randomization and when $\tau = 1$, it is equivalent to "play the winner" [84]. The results can be generalized to multiple arms setting with one standard treatment arm and k-1 experimental treatment arms.

Method 1 calculates PP through averaging Y_1 and Y_2 conditional on the number of patients who would be assigned to arm 2. Let m denote the fixed number of remaining patients in the trial and Z denote the number of patients who would be assigned to arm 2, then $Z \sim bin(m, \pi)$. The PP can be calculated as:

(3.4)

$$PP = \sum_{z=0}^{m} \sum_{y_1=0}^{m-z} \sum_{y_2=0}^{z} P_Z(z | X_1 = x_1, X_2 = x_2) \times P(Y_1 = y_1 | X_1 = x_1) P(Y_2 = y_2 | X_2 = x_2)$$
$$\times \; I\{P(|p_2 - p_1| > \delta | X_1 = x_1, X_2 = x_2, Y_1 = y_1, Y_2 = y_2) \geq \theta_T\}.$$

Method 2 directly approximates $N_k - n_k$ based on observed data. For example, $N_1 - n_1 = m(1 - \pi)$ for arm 1 and $N_2 - n_2 = m\pi$ for arm 2.

The authors perform simulations to further investigate the performance of Bayesian adaptive randomization and predictive probability under various p_1 and p_2 combinations while fixing N, δ, θ_T and θ_U. Since no information is available at the beginning of the trial, equal randomization is applied. As the trial proceeds, more

patients are assigned to the better performing arm and lead to higher randomization probabilities.

For the given values of the design parameters, the Bayesian adaptive randomization trial and predictive probability design and group sequential design produce almost identical power curves. Under a certain scenario, the power when using Bayesian adaptive randomization and predictive probability design is lower than using group sequential design. The authors explain that this is due to the conservative approach group sequential taken to control type I error. The simulations also show that when the difference between the two response rates is substantially large, the response percentage in the superior arm when using outcome adaptive design is higher since more patients are assigned to that arm compared to fixed randomization. However, the required sample size for a response-adaptive randomization trial is also larger than the trial using fixed randomization given the same type I error rate and the statistical power.

When early termination is allowed, trials using outcome-adaptive randomization does not necessarily result in a larger sample size than the ones using fixed randomization. If the treatment difference is large, early termination rule kicks in and terminates the trial early. Thus, the sample size during the adaptive randomization stage may be small, which may lead to a smaller total sample size.

The authors also note that since Bayesian adaptive randomization and predictive probability design assign more patients to the superior arm, the imbalance between two arms increases accordingly. Thus, more patients are needed to make up for the loss of statistical power, leading to a larger sample size needed for Bayesian adaptive randomization and predictive probability design compared to fixed randomization.

Lee, Chen, and Yin [81] also exam the operating characteristics of outcome-adaptive randomization and compare it with fixed randomization for binary outcome through simulations. For binary outcomes, patients either respond to the treatment or not. In the content of a two-arm trial, let p_1 denote the response rate of the control arm and p_2 denote the response rate of the experimental arm. The experimental arm is declared as superior to the control arm if $Pr(p_2 > p_1|D) > \theta_T$ conditional on observed data D and pre-specified cut-off value θ_T. The randomization probability of assigning a patient to arm 2 is

$$\frac{Pr(p_2 > p_1|D)^c}{Pr(p_2 > p_1|D)^c + Pr(p_1 > p_2|D)^c}, \tag{3.5}$$

where c is the turning parameter.

For their simulation studies, the authors aim to control 10% type I error rate and to reach 90% statistical power. They consider two turning parameter c values: $c = n/(2N)$ and $c = (n/N)^{0.1}$, n is the current patients in the trial and N is the total sample size. When $c = n/2N$, at the beginning of the trial c = 0 and at the end of the trial c = 0.5. When $c = (n/N)^{0.1}$, this reflects a larger contrast: c = 0 at the beginning of the trial, c = 0.87 when 25% of patients in the trial, c = 0.93 when 50% of patients in the trial and c = 0.97 when 75% of patients in the trial. p_1 is fixed at 0.2 and p_2 ranges from 0.05 to 0.95. Beta(1, 1) is the prior distribution. The null hypothesis is defined as $p_1 = p_2 = 0.2$. The alternative hypothesis is defined as $p_1 = 0.2, p_2 = 0.4$. The randomization probability is restricted between 0.1 and 0.9 to avoid extreme patient allocation.

The simulation results show that the number of the sample size required to achieve the requisite power is the smallest for equal randomization and the largest for outcome adaptive randomization with $c = (n/N)^{0.1}$. Outcome adaptive randomization with $c = (n/N)^{0.1}$ yields the highest overall response rate for all p_2 value,

especially when the difference between the two response rates is large. The number of nonresponders is the lowest for equal randomization when $p_2 < 0.6$, but for outcome adaptive randomization with $c = (n/N)^{0.1}$ when $p_2 \geq 0.6$.

When early termination rule is incorporated, the early efficacy and futility stopping rates are similar between outcome adaptive randomization and equal randomization. On average, the sample size for outcome adaptive randomization with $c = (n/N)^{0.1}$ is larger than those of equal randomization, but those of outcome adaptive randomization with $c = n/2N$ is similar to equal randomization.

Their simulations show that the trial using equal randomization requires a smaller sample size and produces smaller non-responders while controlling type I error rate and maintaining the desired statistical power than that using outcome-adaptive randomization. While outcome-adaptive randomization often requires a larger sample size, it allocates more patients to the superior treatment arm. Given the same total sample size, outcome-adaptive randomization tends to yield a higher number of responders than equal randomization. When the response difference is substantially large, outcome-adaptive randomization may require a smaller sample size and yield a higher response rate than equal randomization when early termination rule applies.

3.3 Ethical Considerations

Randomized clinical trials have always been treated as the golden standard of conducting clinical trials. Thus, it is natural to compare outcome-adaptive clinical trials with fixed randomized clinical trials. Many literature has discussed the advantages and disadvantages of outcome-adaptive clinical trials and fixed randomized clinical trials from theoretical and methodological considerations, but it is also

important to consider from the ethical perspective.

Hey and Kimmelman stimulates a discussion between outcome-adaptive allocation trials and fixed randomized clinical trials in ethical considerations [85]. The authors argue that outcome-allocation adaptive trials are no better than fixed randomized clinical trials regarding mainly minimizing patient burden, informed consent, and validity in the setting of two-arm trials, but may be more appropriate in the multi-arm context. The authors claim that outcome-adaptive randomization fails to minimize patients' burden since only 50% success rate of achieving regulatory license after reaching phase III which implies that patients in outcome-adaptive randomization trials still only have a 50% probability of receiving the superior treatment in best case scenario. They also point out that most new treatments only deliver minimal improvement over standard treatments, therefore one must justify the disadvantage of the need for a larger sample size. The authors continue their objection that since outcome-adaptive randomization is complicated in nature, which increases patients' burden in understanding the informed consent. For validity, the authors highlight that the dynamic trial environment and treatment may also induce bias. They also point out since treatment allocation is based on currently enrolled patients, the latter enrolled patients will have more advantage of being assigned to a better performance arm than earlier enrolled patients. This may introduce bias too. However, the authors believe outcome-adaptive randomization provides more benefits than a fixed randomized clinical trial in the context of multi-arm trials since it can evaluate many different hypotheses at once. In summary, Hey and Kimmelman found that outcome-adaptive randomization shows no advantage than a fixed randomized clinical trial regarding patients' burden, informed consent, and validity, but maybe more appropriate in the multi-arm trial setting.

The article by Hey and Kimmelman stimulates discussions over the ethical consideration of outcome-adaptive randomization clinical trials. Many statisticians have expressed their opinions corresponding to the arguments that Hey and Kimmelman brought up to.

Joffe and Ellenberg express their opinions on outcome-adaptive randomization from three ethical research criteria: scientific validity, favorable benefit/risk ratio, and informed consent, with a focus on the data monitoring perspective [86]. They note that researches need to justify participants' risk in producing generalizable knowledge. In outcome-adaptive randomization trials, latter enrolled patients may differ from earlier enrolled patients systematically. This results in imbalanced allocation, and therefore introduces bias. They mention that outcome-adaptive randomization improves benefit/risk balance for enrolled patients, particularly if the superior treatment effect can be observed in a short amount of time. The authors highlight that fixed randomization with sequential design may be the optimal approach when the goal of the trial is to minimize the number of patients enrolled in the poor performance arm, especially with large difference outcomes. The authors particularly provide three plausible disclosure approaches in open-label adaptive trials: first, investigators only reveal randomization without emerging data plan; Second, investigators may disclose the dynamic treatment allocation probabilities based on interim data, but withhold the prospective randomization probabilities for newly enrolled patients; lastly, investigators may inform patients both the adaptive randomization design and the actual randomization probabilities, which presents a difficult communication challenge. In all scenarios, investigators need to address the fact to patients that it is still possible that they will be randomized to receive the worse performing arm.

Korn and Freidlin [87] see no evidence that outcome-adaptive randomization is useful, especially compared to equal randomization with interim monitoring. Interim monitoring shares the same goal as outcome-adaptive randomization: to minimize patients being assigned to the inferior arm. They mainly consider the operating characteristics of trial designs with interim monitoring. They point out that although outcome-adaptive randomization may provide a slightly larger proportion of positive responders than fixed randomization design, outcome-adaptive randomization tends to results in a larger proportion of non-responders due to the larger sample size required. If the null hypothesis is true, outcome-adaptive randomization trial shows no benefit since it requires a larger sample size, which will expose more patients to the inferior treatment and result in a longer trial.

Although many statisticians question the use of outcome-adaptive randomization, Berry [88] strongly disagrees with Hey and Kimmelman's opinions about the unethical news of outcome-adaptive randomization. Berry argues that the authors failed to understand the fundamental goal of clinical trials. The major goal of conducting clinical trials is for scientific efficiency, not for better treatment for participants. Since the conflict between clinical practice and science is inevitable, the Belmont Report provides a guideline on how to deal with this situation. Adaptive clinical trials provide a better option than fixed randomization especially for rare diseases, and they can also make an adjustment to multiple objectives under a two-arm trial setting in a more efficient and timely manner. Berry also points out that fixed randomization will be outdated with rapid advancement in oncology and being adaptive can treat participants more efficiently.

Lee comments on the ethical issues from three categories: statistical properties, practical considerations, and real-life performance [89]. Lee states that equal

randomization is in collective ethics, adaptive randomization is in individual ethics. While keeping the desired statistical power, equal randomization is designed to have the smallest trial to test treatment differences, but adaptive randomization focuses more on providing the best treatment to enrolled patients. Lee argues that short term efficacy often fails to translate to long term efficacy which may be due to poor choice of surrogate endpoint regardless of randomization methods used. The low success rate of phase III trials is mainly because of the inefficiency of the drug development process. Lee also points out that adaptive randomization may utilize more resources than equal randomization, but it takes patient benefit into account. Regarding the informed consent concern raised by Hey and Kimmelman, Lee believes this is rather a communication challenge instead of a randomization issue. Lee agrees that population and treatment changes during the trial may induce bias, solutions such as block randomization may be applied. Additionally, Lee encourages putting more effort in educating adaptive randomization and in utilizing adaptive randomization.

Saxman shares some opinions with Lee that there is no sufficient evidence to support Hey and Kimmelman's argument that outcome-adaptive randomization has no advantage over fixed randomization due to the low success rate of phase III clinical trials [90]. Saxman objects Hey and Kimmelman's argument that the complicated to plan and coordinate the nature of outcome-adaptive randomization is an ethical predicament for outcome adaptive randomization. On the patient autonomy issue, particularly informed consent, Saxman agrees with Hey and Kimmelman that it is difficult for patients to comprehend outcome-adaptive randomization due to its complexity and the outcome-adaptive randomization supporters need to propose a more effective solution to this problem. In continue with Hey and Kimmelman's argument on informed consent, Saxman also supports that outcome-adaptive randomization

may have an increasing bias due to the dynamic entities.

CHAPTER 4

A Signature Enrichment Design with Bayesian Adaptive Randomization

4.1 Introduction

Most human cancers are heterogeneous with respect to their molecular, genomic and phenotypic properties and treatments that target specific molecules may only benefit a subset of patients [91, 92, 93, 94, 95, 96, 97]. Thus, it is important to design efficient trials that offer optimal treatments for as many patients as possible based on their biomarker profiles.

In contrast to traditional non-adaptive design methods, adaptive designs have recently gained popularity due to their flexibility and efficiency. In 2006, the Pharmaceutical Research and Manufactures of America (PhRMA) Working Group defined an adaptive design as a design that allows modification of the on-going study based on accumulating data, without undermining the validity and integrity of the trial [98]. Examples of adaptive designs include those that use adaptive randomization [99, 100, 48, 101, 102, 103, 104, 105], sample size re-estimation [106, 107, 108, 109, 110], changes to eligibility criteria [3, 111, 112], and early stopping rules for safety, futility or efficacy [113, 114, 115, 116, 117, 105, 118, 119]. Adaptive randomization procedures can use a frequentist framework [84, 120, 50] or a Bayesian framework [121, 122]. Commonly used adaptive randomization methods include treatment-

adaptive randomization, covariate adjustment randomization and response adaptive randomization [99, 100, 48, 123, 124, 125, 126, 127]. Bayesian adaptive randomization methods allow the combination of prior knowledge and observed data to learn about parameters of interest [128]. Bayesian adaptive randomization procedures apply Bayes theorem repetitively based on accumulating data to adjust the randomization probability for each newly enrolled patient. Such procedures may be considered more ethical by assigning more patients to the more effective treatment arm. Several clinical trials have adopted Bayesian adaptive randomization, including BATTLE and I-SPY2 [129, 124, 130, 131, 132, 133, 134, 135, 136, 137, 138].

Recently, the rapid advancement of biomarker studies in oncology has promoted the development and application of precision medicine, previously known as "personalized medicine" [139, 140]. Precision medicine targets a subpopulation of patients who are most likely to respond to the treatment based on their characteristics or biomarker profile. Prognostic biomarkers provide information on clinical outcome independently of the treatment received, while predictive biomarkers provide information on clinical outcome for a particular treatment [141]. Prognostic biomarkers can be used to screen good and bad prognostic patients, while predictive biomarkers can be used to measure how likely the patient will respond to a particular treatment. Before using biomarker information in clinical practice, it is essential to test the biomarkers for analytical and clinical validity and clinical utility [142, 143]. For the designs described in this paper, we focus on predictive biomarkers.

A biomarker adaptive design utilizes patient biomarker information measured at baseline to allow adaptation based on previous patients' responses. When the biomarkers used to categorize the sensitive status for potential patients are known before the start of the clinical trial, a biomarker adaptive design is particularly useful.

If the biomarker is unknown, one may consider an adaptive signature design [1] or the cross-validated variation of this design [2]. These designs use an equal randomization approach but include a development stage and a validation stage to determine a sensitive subset classifier. An adaptive enrichment design [3] allows changes to the enrollment criteria as the trial proceeds. Restricting enrollment to patients who are most likely to benefit from the treatment based on a classifier improves the efficiency of the clinical trial but may require a longer time to accrue the targeted patients.

In this chapter, we propose a cross-validated signature enrichment design with Bayesian response-adaptive randomization. This design combines the elements of cross-validated adaptive signature designs and adaptive enrichment designs. During patient enrollment, our design uses an enrichment strategy which oversamples the sensitive patient cohort and undersamples the non-sensitive patients. It also applies Bayesian response-adaptive randomization, adaptively adjusting future allocation probabilities based on the accumulated data. As a result, this design has the advantages of Bayesian adaptive randomization as well as an enrichment strategy, yielding a higher overall response rate on the trial and other advantages described below and with comparable power to the other two designs. At the end of the trial, we test the treatment effect in the overall patient group, and in the sensitive patient subgroup with the control of type I error rates.

In addition to the usual evaluation of statistical power, we propose four trial evaluation criteria to assess the performance of the designs: current individual loss (CIL), future individual loss (FIL), the probability of an individual in the trial receiving optimal treatment (PCO) and the probability of a future individual receiving personalized optimal treatment (PFO). These criteria evaluate the benefits for both trial participants and future patients from the perspective of precision medicine, pro-

viding a more comprehensive assessment of trial performance. Results are demonstrated in extensive simulations.

The rest of this chapter is organized as follows. In section 2, we briefly review the cross-validated adaptive signature design with respect to individual patient criteria, the adaptive enrichment design, and our Bayesian response-adaptive randomization scheme. In section 3, we describe our cross-validated signature enrichment design with response-adaptive randomization. In section 4, we define the trial performance evaluation criteria. In section 5, we report the results of simulation studies to compare our proposed design with the cross-validated adaptive signature design and the adaptive enrichment design. In section 6, we present an example based on the Iressa Pan-Asia Study (IPASS) [144]. In section 7, we conclude with a brief discussion.

4.2 Review of Adaptive Signature Designs, Adaptive Enrichment Designs and Bayesian Adaptive Randomization

As the name suggests, our proposed cross-validated signature enrichment design combines with Bayesian response-adaptive randomization. In this section, we briefly review the cross-validated adaptive signature design, the adaptive enrichment design, and the Bayesian response-adaptive randomization method. For all the designs mentioned in this article, we consider a two-arm clinical trial to compare the efficacy of an experimental treatment E and a control treatment C.

4.2.1 Adaptive signature design (ASD)

When a reliable classifier for identifying sensitive patients is not available at the start of the trial, an adaptive signature design [1, 145] may be useful. If the overall test for treatment effect is not significant, this design defines a development

stage and a validation stage for a sensitive patient subpopulation classifier. Since only a portion of patients contributes to each stage, the cross-validated extension of the adaptive signature design [2] adopts a K-fold cross-validation method to classify patients from the entire population and to adjust the P value using a permutation test [146].

As illustrated in Figure 1, this design uses equal randomization and if the overall test is not significant, divides patients into two cohorts: a development cohort and a validation cohort for determining a sensitive subset classifier. The design controls the overall type I error rate at α, where $\alpha = \alpha_1 + \alpha_2$. If the overall treatment effect is statistically significant at some pre-specified level α_1, then the experimental treatment is considered globally beneficial. Otherwise, the adaptive signature design develops a sensitive patient classifier and tests whether the classifier is useful for treatment selection at type I error rate α_2.

The classification of a sensitive patient subgroup is based on machine learning voting methods in two stages. The first stage identifies predictive biomarkers. The second stage identifies sensitive patients based on the chosen predictive biomarkers from stage I. A patient in the validation set is classified as sensitive if the predicted odds ratio of experimental arm versus standard arm exceeds some pre-specified value. The treatment effect only in the sensitive patients in the validation cohort is then assessed at a pre-specified reduced type I error rate α_2 to test whether the classifier is useful for treatment selection.

The cross-validated adaptive signature design utilizes K-fold cross-validation to obtain a final sensitive patient subset from the entire trial population. Since the sensitive patient subset is obtained using cross-validation, a permutation test may be used to adjust the P value [2, 146].

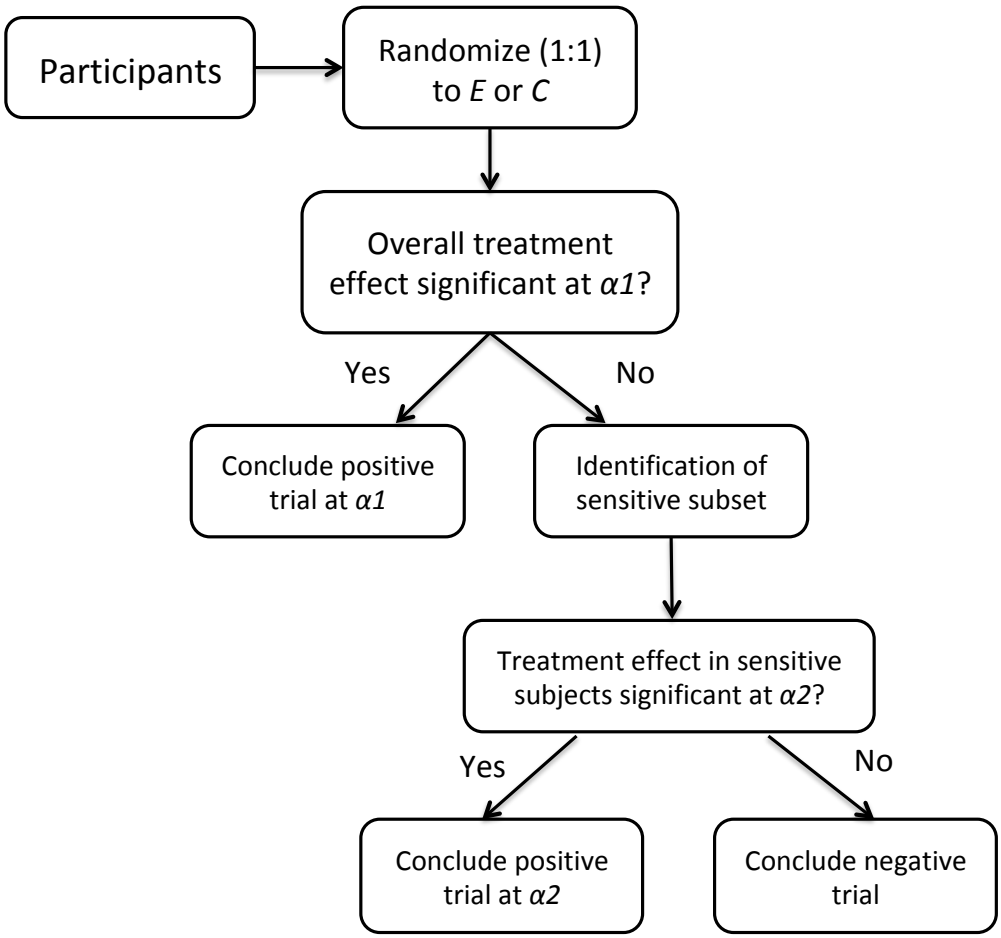

Figure 4.1: Diagram for Adaptive Signature Design (ASD) [1, 2].

4.2.2 Adaptive enrichment design (AED)

The adaptive enrichment design [3, 147, 148, 149] restricts enrollment to sensitive patients, as defined by an adaptive classifier function, and excludes all nonsensitive patients. Patients are grouped into blocks according to their entry time. For the initial patient block, equal randomization is used to allocate patients to treatment arms. For subsequent patient blocks, as illustrated in Figure 2, the probability of response is estimated for patient i with each treatment E and C, based on the cumulative estimated response rates from patients in the previous blocks. Let $\hat{r}_{i,E}$ and $\hat{r}_{i,C}$ denote the response rate of patient i when receiving E or C, respectively. The classifier function determines the sensitive status, which is defined as

$$CF(\hat{r}_{i,E}, \hat{r}_{i,C}) = I(\hat{r}_{i,E} - \hat{r}_{i,C} > \delta), \tag{4.1}$$

where δ denotes the minimal clinical meaningful difference. Let $\hat{CF}_i$ refer to the latest estimate of the classifier function at the time of the entry of the i^{th} patient. The classifier function restricts entry into the trial to patients with $\hat{CF}_i = 1$, the sensitive patients, and excludes patients whose $\hat{CF}_i = 0$, the non-sensitive patients. Newly enrolled patients will be assigned to one of the treatment arms using equal randomization. The classifier will be updated once a new block data becomes available. The process is repeated until the desired total number of patients is reached.

As the adaptive enrichment design will include mostly sensitive patients (except in the initial patient block), it can improve the efficiency of the clinical trial and protect non-sensitive patients from exposure to ineffective treatments. However, it can be difficult to estimate treatment effect size and choose a strategy to update enrollment criteria and requires a longer time to accrue sufficient number of sensitive patients than the traditional all-comers designs. Also, if the treatment is effective

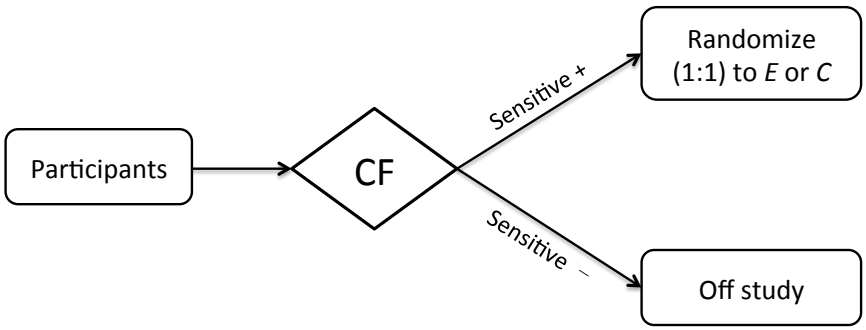

Figure 4.2: Diagram for Adaptive Enrichment Design (AED) [3]. *CF* denotes the sensitive status classifier function.

in all patients, restricting enrollment may exclude patients who would benefite from participating in the trial [150, 4, 5].

4.2.3 Bayesian adaptive randomization

Bayesian adaptive randomization updates allocation probabilities based on currently enrolled patient responses, assigning more patients to the more effective treatments. Compared with equal randomization, response-adaptive randomization places more emphasis on individual benefit and advantages, especially with complicated designs [53, 81]. Some objections to adaptive randomization include potential bias due to inhomogeneous patient population throughout the trial, statistical inefficiency due to unbalanced patient allocation and the requirement that responses be collected shortly after treatment [5, 151, 152, 90, 153]. Generally, the advantages of response-adaptive randomization are more evident with larger sample sizes and with larger differences between treatments.

Bayesian adaptive randomization assumes that the allocation probability for treatment E is proportional to the posterior probability $\hat{P}(r(i,E) > r(i,C)|data)$.

The probability of patient i receiving treatment E, $P_{i,E}$, is given by [80, 154, 79] as

$$P_{i,E} = \frac{\hat{P}(r(i,E) > r(i,C)|data)^w}{\hat{P}(r(i,E) > r(i,C)|data)^w + \hat{P}(r(i,E) \leq r(i,C)|data)^w} \tag{4.2}$$

where w is a tuning parameter between 0 and 1. When w is equal to 0, there is equal randomization. When w is equal to 1, $P_{i,E}$ is the same as $\hat{P}(r(i,E) > r(i,C)|data)$. Common practice is to set w equal to $1/2$ or to use $n/2N$ where n is the cumulative sample size [80].

4.3 Signature Enrichment Design with Adaptive Randomization (SEDAR)

We propose a cross-validated signature enrichment design with response-adaptive randomization. For the initial block, we use equal randomization and let the initial block size be twice that of the subsequent adaptive randomization blocks [155]. The framework for subsequent patient blocks of the design is presented in Figure 3. We assume patient responses can be observed shortly after receiving treatments. The enrollment procedure for future patients is as follows:

1. For each patient, apply the current sensitive patients classifier function (CF) derived on data from previous blocks. This separates patients into one of the two categories: sensitive and non-sensitive.
2. If the patient is determined to be sensitive, randomize to E or C based on the Bayesian adaptive randomization scheme.
3. If the patient is determined to be non-sensitive, calculate the enrollment function (EF = 1 with probability P; EF=0 with probability 1-P). If EF=1, randomize based on the Bayesian randomization scheme as above; if EF = 0, the patient is off study (i.e., not randomized). The details of determining the enrollment probability are given in section 3.1.

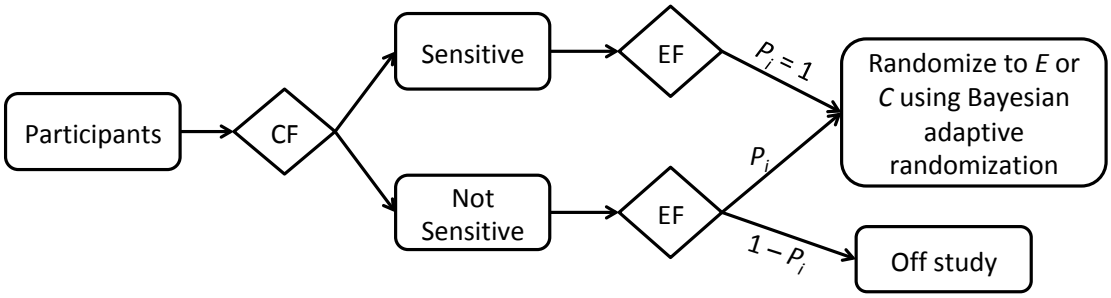

Figure 4.3: Diagram for Signature Enrichment Design with Adaptive Randomization (SEDAR). *CF* is the classifier for determining patient sensitive status; *EF* is the enrollment function; P_i is the probability of enrolling the i^{th} patient.

4. After the current block is completely accrued, refine the classifier function based on the additional results from the current block and repeat the steps above for the next block.

As is common with all enrichment designs, it is difficult to calculate the required sample size a priori. To estimate the total required sample size for our design, we first calculate the sufficient sample size to reach the desired power using traditional methods for non-enrichment designs [156, 157]. Then we use simulations to modify the sample size for the given design parameters.

4.3.1 Sensitive-patient classifier function and enrollment function

After each block of data becomes available, we develop a sensitive patient classifier to screen for future patients. Many classification algorithms have been studied for biomarker classifier development [158, 159, 160, 161, 142]. Some popular choices include logistic regression, random forest [162], support vector machines [163] and diagonal linear discriminant analysis [164]. To estimate the probabilities of response for a patient with the experimental or control treatment, many statistical

models or algorithmic methods, e.g., logistic regression and random forests, may be used [165, 166]. For our proposed design, we use a random forest approach. In practice, a different classification algorithm may be used depending on trial-specific consideration. A brief description of the random forest technique is summarized below. More details can be found elsewhere [162, 167].

Random forests grow a large collection of de-correlated decision trees [168]. Each tree starts with drawing a bootstrap sample from the original dataset. At each node, a best split is determined among a randomly selected subset of the total input variables. When a random forests approach is used for classification purpose, the class prediction is the majority vote from the total number of decision trees. The class probabilities are the proportion of votes from the ensemble of trees.

At an early stage of the clinical trial, the available patient data to build a classifier is limited. In contrast to the adaptive enrichment designs that restrict entry into the trial to sensitive patients, we use an enrichment strategy which will include both sensitive and non-sensitive patients but with different enrollment probabilities. Let δ denote the minimal clinical meaningful difference and $\hat{r}_{i,x,E}$ and $\hat{r}_{i,x,C}$ denote the estimated response rates for potential patient i with signatures x if receiving treatment E or C, respectively. The sensitive patient status is determined in the same way as AED, which is:

$$CF(\hat{r}_{i,E}, \hat{r}_{i,C}) = I(\hat{r}_{i,x,E} - \hat{r}_{i,x,C} > \delta), \tag{4.3}$$

Let $\hat{CF}_i$ refer to the latest estimate of the classifier function at the time of entry of the i^{th} patient. If $\hat{CF}_i = 1$, then patient i is sensitive; otherwise, the patient i is non-sensitive. For sensitive patients, the probability of enrollment is 1. For non-sensitive patients, we use a continuous function as the enrollment function to determine the enrollment probability for patient i with signatures x. The enrollment probabilities

are calculated as

$$P_i = min\{max(\frac{\hat{r}_{\{i,x,E\}} - \hat{r}_{\{i,x,C\}}}{\delta}, 0.10), 1\} \tag{4.4}$$

Thus, the probability for sensitive patients is 1 (i.e., all sensitive patients are enrolled) and for non-sensitive patients the enrollment probability is between 0.1 and 1. When all patient data are available, we build a final sensitive patient subset classifier. In our design, we use random forests as the algorithm for sensitive status classification.

During the trial process, we screen sensitive patients to enroll in the next block. This sensitive patient classifier is built on the previous patient response, treatment received and biomarker information. At the final analysis, the classifier is developed by cross-validation based on complete patient data, as described in the next subsection. The sensitive patient subset selected is used to test the treatment effect difference in the sensitive subgroup.

4.3.2 Bayesian decision rule

For the proposed design, we estimate the power to detect the difference of the treatment effect between E and C in the overall patient group and in the sensitive patient subset using Bayesian decision rules.

To test the treatment effect in the sensitive population, we first identify the overall sensitive patient subset from the entire trial population using 10-fold cross-validation. This technique is also used in the cross-validated extension of the adaptive signature design [2]. For K-fold cross-validation, the entire trial population is randomly partitioned into K nonoverlapping equal-sized subpopulations. In the final analysis, one of the K folds is used as the validation cohort, and the rest $K-1$ folds are treated as the development cohort and a sensitive patient subset S_k $(k = 1, ..., K)$ is developed. This procedure is repeated K times. Each patient appears once in the

validation cohort and $K-1$ times in the development cohort. Then the total number of the sensitive patient subset from the entire trial patients is $S = \bigcup_{k=1}^{K} S_k$.

At final analysis, arm E is claim efficacious if

$$\hat{Pr}(r_E > r_c \mid data) > a_U \tag{4.5}$$

where a_U is a cut-off parameter whose value is tuned to control the type I error rates. The type I error rate for testing treatment effect in the overall patient group is α_1, and for testing treatment effect in the sensitive subset it is α_2, so that the overall type I error rate $\alpha = \alpha_1 + \alpha_2$. That is, we calibrate different a_U values for the overall analysis and for the sensitive subset analysis to control their type I error rates.

4.3.3 Required design parameters

The required design parameters (Table 4.1) are the sample size for the initial equal randomization block (N_0) and for the adaptive randomization blocks (N_{bar}). Then the total sample size $N = N_0 + m * N_{bar}$, where $m + 1$ is the total number of blocks. Additionally, the proposed design requires one to specify the tuning parameter w; minimal clinical difference δ for sensitive status classification; the type I error rate for testing the overall treatment effect (α_1) and the type I error rate for testing treatment effect in the sensitive subset (α_2).

Table 4.1: Design parameters required for SEDAR.

N_0	Sample size for the initial equal randomization block
N_{bar}	Sample size for adaptive randomization blocks
N	Total sample size
w	Tuning parameter
δ	Minimal clinical difference for sensitive status classification
α_1	Type I error rate for testing treatment effect in overall patient group
α_2	Type I error rate for testing treatment effect in sensitive subset

4.4 Trial performance evaluation criteria for simulations

In addition to commonly used operating characteristics such as the average response rate and power, we propose the following trial performance evaluation criteria for simulation studies of a signature clinical trial:

- Current individual loss (CIL)
- Future individual loss (FIL)
- Probability of a current individual in the trial receiving personalized optimal treatment (PCO)
- Probability of a future individual receiving personalized optimal treatment (PFO)

For CIL and FIL, we first define a *match* for enrolled and future patients. For currently enrolled patients, a *match* occurs when the patient's actual treatment received is the same as the best treatment from the true model. For future patients, a *match* occurs when the superior treatment selected based on the final fitted model is the same as the best treatment from the true model. For a currently enrolled patient i with signature x, let $\hat{P}_i(Y = 1|T, x)$ denote the probability of responding to the received treatment T. Similarly, for a potential future patient, $\hat{P}_i(Y = 1|T, x)$ refers to the probability of responding to treatment T as determined by the fitted model. Let $\hat{P}_i(Y = 1|T = OPT, x)$ denote the probability of responding to the *optimal* treatment determined by the true model. Then we define the personalized loss function as the follows:

(4.6)

$$Loss(i) = \begin{cases} 0 & \text{if it is a match} \\ \hat{P}_i(Y = 1|T, x) - \hat{P}_i(Y = 1|T = OPT, x) & \text{if there is not a match} \end{cases}$$

Then CIL and FIL are the average values of the individual loss for trial partic-

ipants and for future patients, respectively. A small value of CIL indicates that most currently enrolled patients have received the treatment that they will respond at least similarly as the optimal treatments according to the true model. The interpretation is similar for FIL.

The PCO and PFO quantities are the probabilities of receiving the personalized optimal treatment for trial participants and future patients, respectively. PCO can be estimated as the average fraction of matches between the optimal treatment selection by the true model and by the patients' actual treatment received.

PFO is the probability of a future patient receiving the personalized optimal treatment if the treatment recommendation from the trial were followed. It can be estimated as the average fraction of matches for prospective patients between the optimal treatment selections by the true model and by the final fitted model. A large value in PCO and PFO indicates that most patients have received and will receive their optimal treatments determined by the true model, respectively.

The main differences between CIL & FIL and PCO & PFO are that for CIL and FIL, the penalties for non-matches depend on the individual loss; for PCO and PFO, all non-matches receive the same penalty. Small values of CIL and FIL, and large values of PCO and PFO are desirable.

4.5 Simulation Studies

We performed a series of simulations to evaluate our design and compared it with the adaptive signature design and the adaptive enrichment design. For all the designs, we implemented the cross-validated procedures to search for final sensitive subsets from the entire patient population. To make ASD and AED comparable to SEDAR, all the designs used the Bayesian decision rules described above at final

analyses.

We assume a two-treatment clinical trial designed to compare a control treatment (C) with an experimental treatment (E). We include situations where the experimental treatment is effective only in a sensitive subset. Let T_i denote the treatment indicator which

$$T_i = \begin{cases} 0.5 & \text{Treatment} = E \\ -0.5 & \text{Treatment} = C \end{cases}, \ i = 1, \cdots, n. \tag{4.7}$$

4.5.1 Simulation setting

The total sample size N is set at 300; the size of the adaptive randomization blocks N_{bar} is set at 50. We choose the size N_0 of the initial block, for which equal randomization is used, to be $N_0 = 100$. Hence, we have a total of 5 blocks of sizes 100, 50, 50, 50 and 50. The tuning parameter for Bayesian adaptive randomization w is set at 0.5. The minimal clinical meaningful difference for sensitive patient status qualification is set at 0.1. The type I error rate α_1 for testing overall treatment effect is controlled at 0.04, and the type I error rate α_2 for testing treatment effect in the sensitive subset is controlled at 0.01.

We generate the i^{th} patient's response from the following logistic regression model with ten biomarkers:

$$logit(r_i) = \beta_0 + \beta_1 T_i + \sum_{k=1}^{10} \gamma_k X_{ki} + \sum_{k=1}^{10} \eta_k T_i X_{ki}, \ i = 1, \cdots, 300; k = 1, \cdots, 10. \tag{4.8}$$

where X_{ki} is the k^{th} biomarker/signature for patient i. Each biomarker X_{ki} is assumed to follow a normal/multivariate normal distribution with mean of 0.5 and standard deviation of 1.0.

For the Bayesian adaptive randomization, we assume the prior is

$$\beta_0, \beta_1, \gamma_k, \eta_k \sim Normal(\mu = 0, \sigma^2 = 100).$$

At the final analysis, for each treatment arm, we use a non-informative prior $Beta(1, 1)$.

We consider three categories of biomarker distributions. For each type, we include four scenarios and a null scenario to compare the performance of SEDAR with ASD and AED (Table 4.2). The biomarker or biomarker-treatment interaction effects not presented in the table are assumed to be zero.

For the first category, we assume all biomarkers are independently distributed following a normal distribution with mean 0.5 and variance 1.0. For scenarios 1 and 2, we consider only X_1 and the corresponding treatment interaction affect the response with different magnitude. For scenarios 3 and 4, we assume X_1 and X_6, and their interactions with treatment will affect the patient response with different scales.

For the second category, we assume the first five biomarkers have a multivariate normal distribution $X \sim NVM(0.5, \Sigma)$ with covariance matrix which has 1 for the variance and 0.25 for ρ. The other five variables are still independently distributed with a normal distribution with mean 0.5 and variance 1.0. For scenarios 7 and 8, we consider only X_1 and the corresponding treatment interaction affect the response. For scenarios 9 and 10, we assume one biomarker from the first five correlated biomarkers, X_1 and its interaction with treatment and another biomarker from the independent biomarker group, X_6 and its interaction with treatment, affect the response.

For the third category, we assume the first five biomarkers have a multivariate normal distribution with mean 0.5 and covariance matrix with $\rho = 0.75$, and the remaining five biomarkers have a multivariate normal distribution with mean 0.5

Table 4.2: Simulation scenarios

		β_0	β_1	γ_1	η_1	γ_6	η_6
x_1 to x_{10} are continuous with $\rho = 0.00$	Scenario 1	-0.8	0.0	0.0	0.0	0.0	0.0
	Scenario 2	-0.8	0.0	0.8	0.8	0.0	0.0
	Scenario 3	-0.8	0.0	0.3	1.1	0.0	0.0
	Scenario 4	-0.8	0.0	0.3	0.3	0.7	0.7
	Scenario 5	-0.8	0.0	0.4	0.4	1.1	1.1
x_1 to x_5 are continuous with $\rho = 0.25$	Scenario 6	-0.8	0.0	0.0	0.0	0.0	0.0
x_6 to x_{10} are continuous with $\rho = 0.00$	Scenario 7	-0.8	0.0	0.8	0.8	0.0	0.0
	Scenario 8	-0.8	0.0	0.3	1.1	0.0	0.0
	Scenario 9	-0.8	0.0	0.3	0.3	0.7	0.7
	Scenario 10	-0.8	0.0	0.4	0.4	1.1	1.1
x_1 to x_5 are continuous with $\rho = 0.75$	Scenario 11	-0.8	0.0	0.0	0.0	0.0	0.0
x_6 to x_{10} are continuous with $\rho = 0.25$	Scenario 12	-0.8	0.0	0.8	0.8	0.0	0.0
	Scenario 13	-0.8	0.0	0.3	1.1	0.0	0.0
	Scenario 14	-0.8	0.0	0.3	0.3	0.7	0.7
	Scenario 15	-0.8	0.0	0.4	0.4	1.1	1.1

and covariance matrix with $\rho = 0.25$. The simulation scenarios are similar to the other two categories.

All the categories and scenarios are listed in Table 4.2. The null scenarios for each category numbers 1, 6 and 11, respectively. For each scenario, we generated 5,000 simulations.

4.5.2 Simulation results

We compare the results obtained using our proposed design with those obtained using ASD and AED. The results for the estimated power are presented in Table 4.3. The type I error rates for the treatment effect in the overall patient group and the sensitive subgroup are controlled at 0.04 and 0.01, respectively.

For the rest scenarios, ASD always has the lowest power, while AED or SEDAR has the highest power. In scenario 3, 8, 13, AED has a higher power than SEDAR. For example, in scenario 3, the power in the overall patient group for ASD, AED and SEDAR are 0.750, 0.987 and 0.964, respectively. In the sensitive

subgroup, the power is 0.721 for ASD, 0.920 for AED and 0.823 for SEDAR. For the other scenarios, SEDAR achieves the highest power in both overall group and sensitive subset. For scenario 7, the power in the overall patient group is 0.539 for ASD, 0.830 for AED and 0.892 for SEDAR.

The results of the four trial performance evaluation criteria we proposed in section 4 are presented in Table 4.3. While both ASD and AED adopt an equal randomization approach when assigning patients to treatment arm, resulting in equal numbers of patients per treatment arm, SEDAR assigns more patients to the superior treatment based on baseline biomarker profiles and data accumulation during the trial. As a result, SEDAR always achieves a higher objective response rate (ORR) than ASD and AED. For example, in scenario 15, 55.6% of patients respond to the assigned treatment for SEDAR compared to 47.7% for ASD and 53.7% for AED. AED results in a higher proportion of patients who respond to the assigned treatment than ASD. This demonstrates that the enrichment strategy allows more trial participants to achieve responses than the all-comers designs.

Table 4.3: Overall analysis: operational characteristics

SCEN	Parameters	Design	Ovl Power	Sens Power	ORR	CIL	FIL	PCO	PFO
1	NULL	ASD	0.010	0.010	0.010	0.310	0.000	1.000	1.000
		AED	0.037	0.010	0.011	0.310	0.000	1.000	1.000
		SEDAR	0.038	0.011	0.008	0.310	0.000	1.000	1.000
2	$\gamma_1 = \eta_1 = 0.8$	ASD	0.510	0.300	0.410	0.073	0.031	0.501	0.715
		AED	0.815	0.560	0.461	0.088	0.014	0.500	0.902
		SEDAR	0.888	0.577	0.466	0.057	0.033	0.606	0.695
3	$\gamma_1 = 0.3$	ASD	0.744	0.721	0.353	0.107	0.023	0.500	0.766
	$\eta_1 = 1.1$	AED	0.987	0.920	0.374	0.127	0.009	0.500	0.909
		SEDAR	0.968	0.823	0.414	0.072	0.024	0.637	0.751
4	$\gamma_1 = \eta_1 = 0.3$	ASD	0.635	0.348	0.430	0.077	0.032	0.500	0.723
	$\gamma_6 = \eta_6 = 0.7$	AED	0.854	0.590	0.472	0.090	0.014	0.500	0.911
		SEDAR	0.911	0.614	0.483	0.059	0.034	0.610	0.701
5	$\gamma_1 = \eta_1 = 0.4$	ASD	0.794	0.602	0.478	0.097	0.028	0.500	0.771

To be continued

Table 4.3 (continued)

SCEN	Parameters	Design	Ovl Power	Sens Power	ORR	CIL	FIL	PCO	PFO
	$\gamma_6 = \eta_6 = 1.1$	AED	0.964	0.833	0.537	0.111	0.012	0.500	0.927
		SEDAR	0.993	0.892	0.557	0.067	0.030	0.639	0.760
6	NULL	ASD	0.040	0.008	0.310	0.000	0.000	1.000	1.000
		AED	0.040	0.008	0.310	0.000	0.000	1.000	1.000
		SEDAR	0.040	0.010	0.310	0.000	0.000	1.000	1.000
7	$\gamma_1 = \eta_1 = 0.8$	ASD	0.523	0.327	0.410	0.073	0.031	0.501	0.718
		AED	0.807	0.587	0.463	0.088	0.013	0.499	0.908
		SEDAR	0.894	0.594	0.468	0.057	0.033	0.607	0.697
8	$\gamma_1 = 0.3$	ASD	0.744	0.693	0.353	0.107	0.023	0.500	0.767
	$\eta_1 = 1.1$	AED	0.990	0.931	0.377	0.128	0.009	0.500	0.917
		SEDAR	0.974	0.842	0.415	0.073	0.024	0.636	0.752
9	$\gamma_1 = \eta_1 = 0.3$	ASD	0.653	0.362	0.430	0.077	0.032	0.500	0.725
	$\gamma_6 = \eta_6 = 0.7$	AED	0.855	0.587	0.471	0.090	0.014	0.500	0.911
		SEDAR	0.917	0.638	0.482	0.058	0.034	0.612	0.706
10	$\gamma_1 = \eta_1 = 0.4$	ASD	0.808	0.615	0.478	0.097	0.028	0.501	0.771
	$\gamma_6 = \eta_6 = 1.1$	AED	0.971	0.826	0.537	0.111	0.013	0.500	0.926
		SEDAR	0.991	0.895	0.555	0.068	0.030	0.637	0.756
11	NULL	ASD	0.037	0.010	0.310	0.000	0.000	1.000	1.000
		AED	0.039	0.010	0.310	0.000	0.000	1.000	1.000
		SEDAR	0.038	0.009	0.310	0.000	0.000	1.000	1.000
12	$\gamma_1 = \eta_1 = 0.8$	ASD	0.514	0.329	0.412	0.073	0.029	0.500	0.732
		AED	0.851	0.625	0.468	0.090	0.010	0.500	0.922
		SEDAR	0.898	0.620	0.473	0.058	0.031	0.608	0.708
13	$\gamma_1 = 0.3$	ASD	0.737	0.729	0.353	0.107	0.022	0.501	0.785
	$\eta_1 = 1.1$	AED	0.992	0.945	0.379	0.131	0.007	0.500	0.938
		SEDAR	0.980	0.885	0.419	0.074	0.024	0.640	0.768
14	$\gamma_1 = \eta_1 = 0.3$	ASD	0.637	0.351	0.428	0.077	0.032	0.500	0.723
	$\gamma_6 = \eta_6 = 0.7$	AED	0.852	0.585	0.472	0.090	0.013	0.500	0.914
		SEDAR	0.909	0.631	0.485	0.059	0.035	0.610	0.696
15	$\gamma_1 = \eta_1 = 0.4$	ASD	0.798	0.606	0.478	0.097	0.029	0.501	0.770
	$\gamma_6 = \eta_6 = 1.1$	AED	0.972	0.849	0.539	0.112	0.012	0.500	0.928
		SEDAR	0.991	0.880	0.557	0.068	0.031	0.639	0.753

CIL and FIL measure the loss for trial participants and future patients. ASD and AED have similar values of CIL, while SEDAR has much smaller values than the other designs. Using scenario 5 as an example, CIL for SEDAR is 0.067 compared to 0.097 for ASD and 0.111 for AED. This difference in CIL shows that SEDAR results in more trial participants receiving optimal treatments. For FIL, the results

are similar for ASD and SEDAR, but AED has the smallest values in all scenarios. This show that a complete enrichment strategy may allow more future patients to receive their optimal treatments.

SEDAR also has higher PCO values than the other designs. That is, SEDAR results in a higher proportion of enrolled patients receiving their optimal treatments as defined by the true model. Similar to the relative performance among the three designs, SEDAR and ASD show similar results for PFO, which refers to the probability for a future patient receiving the personalized optimal treatment if the treatment recommendation from the trial were followed.

The results for true and the estimated treatment effect for the entire trial patient population and for sensitive patient subset are presented in the supplemental materials. The results show that the difference of the average response rate among all patients in a trial between E and C mainly determines the power of a trial. The results of the proportion of truly sensitive patients in the overall patient group and in the sensitive patient subset are also included in the supplemental materials. According to the results, AED trials have the highest proportion of truly sensitive patients, following by SEDAR and ASD, demonstrating that enrichment strategies include more patients who are likely to respond to the treatment than all comers designs, one of the key reasons for considering enrichment designs.

4.6 Example

To illustrate our approach, consider the design of a trial in non-small-cell lung cancer (NSCLC) comparing two treatment arms, one using a tyrosine kinase inhibitor (gefitinib) and the other using standard chemotherapy (carboplatin plus paclitaxel), similar to the Iressa Pan-Asia Study (IPASS)[144]. The single binary

biomarker in this case is the EGFR mutation status, either positive or negative, with a prevalence of EGFR positive patients of approximately 50% in Asian countries and approximately 20% in North America [169]. For the purpose of illustrating the application in the setting of this paper (binary response assessed soon after the start of treatment), we consider the endpoint to be objective response rate. Unfortunately, since the assessment of EGFR mutation status was not an eligibility requirement, only 36% of the patients (437 of 1217) had a known status. To conduct simulations using the IPASS study as a guide, we assume the prevalence of EGFR+ to be 0.50 (commonly observed in an Asian population) and the following response rates in the various treatment arms and EGFR mutation status groups (no missing values): 0.45 in the EGFR positive chemotherapy arm; 0.70 in the EGFR positive gefitinib arm; 0.15 in the EGFR negative chemotherapy arm; and 0.10 in the EGFR negative gefitinib arm.

Although this example is presented as an illustration of the setting considered here, the final endpoints in such a study would ordinarily be overall progression-free survival (PFS) or overall survival (OS), not the response rates. In the initial results reported for the IPASS study, there was improved PFS for gefitinib in the EGFR mutation-positive group but worse PFS for gefitinib in the EGFR mutation-negative group. These overall results are suggested in the early results, but not definitively so except perhaps in the EGFR negative group because of the dramatic early difference.

To apply our design to the IPASS study, we assume a total sample size of 440; an initial block size of 176 patients, and subsequent adaptive randomization block size of 88 patients. The overall type I error rate is controlled at 0.05 with 0.04 for the global treatment effect test and 0.01 for the sensitive subset test. The clinical meaningful difference for enrolling sensitive patients is set at 0.1.

Following the approach used in the IPASS paper for progression-free survival [144], we use the following covariates in a logistic regression model for ORR, all as binary covariates: Age ($<$ 65 years vs. $\geq$ 65 years), Sex (Male, Female), Smoking history (non-smoker vs. former light smoker), WHO performance status (0 or 1 vs. 2). We assume only the EGFR mutation status and its interaction have a treatment effect. In the true model, no treatment main effect is included. The prevalence of each biomarker stratum is presented in Table 4.4.

Table 4.4: IPASS example: biomarker settings

Biomarkers	Gefitinib	Chemotherapy
EGFR status		
EGFR +	0.50	0.50
EGFR -	0.50	0.50
Age		
$<$ 65 years	0.72	0.76
$\geq$ 65 years	0.28	0.24
Gender		
Male	0.20	0.21
Female	0.80	0.79
Smoking history		
Non-smoker	0.938	0.936
Former light smoker	0.062	0.064
WHO performance status		
0 or 1	0.90	0.89
2	0.10	0.11

For the given design parameters, the results are presented in Table 4.5. Generally, the results confirm the conclusion drawn from the previous simulation studies. The power of the overall population is 0.738 for ASD; 0.995 for AED; and 0.999 for SEDAR, and of the sensitive subset is 0.946 for ASD, 0.993 for AED and 0.982 for SEDAR. The objective response rates among all trial participants are 0.361 by ASD; 0.442 by AED; and 0.446 by SEDAR. In addition, SEDAR also has the lowest CIL index and high values for PCO. Thus, our proposed design can assign more patients

to receive their optimal treatments.

Table 4.5: Simulation results: IPASS example

Design	Ovl Power	Sens Power	ORR	CIL	FIL	PCO	PFO
ASD	0.738	0.946	0.361	0.096	0.089	0.501	0.605
AED	0.995	0.993	0.442	0.121	0.035	0.449	0.838
SEDAR	0.999	0.982	0.446	0.063	0.051	0.662	0.768

4.7 Discussion

In this article, we have proposed a cross-validated signature enrichment design with adaptive randomization for cancer clinical trials. Our proposed design builds on the advantages of the adaptive signature design and the adaptive enrichment design. Assuming that patient responses can be observed shortly after receiving treatment, our design adopts an adaptive enrichment approach by developing a sensitive patient classifier to screen sensitive and non-sensitive patients for enrollment. Once a patient enters the trial, we use response-adaptive randomization to adjust treatment allocation probabilities in a Bayesian framework. At the final analysis, we use an adaptation of the adaptive signature designs. For the sensitive patient subset test, this design efficiently identifies sensitive patients from the entire trial participant population using a cross-validated procedure.

Our results have shown that ASD always has the lowest power, while SEDAR has the highest power in most scenarios. The main difference between AED or SEDAR and ASD is that both AED and SEDAR are "enrichment" designs, they enrolling more "sensitive" patients than ASD, which enrolls all patients.

To help clarify the relative performance of power between AED and SEDAR, denote by $r|E$ the response rate among all patients in a trial who receive the experimental treatment (E), and $r|C$ the response rate among those who receive the control

treatment (C). The main reason the power of SEDAR is consistently higher than that of AED in our simulation scenarios is that $r|E - r|C$ is larger in the SEDAR trials than in the AED trials. This is caused by the covariate-based outcome-adaptive randomization adopted by SEDAR. A toy example is given below.

Suppose the population consists of two groups of patients: 2/3 sensitive and 1/3 non-sensitive. For sensitive patients, the response rates by E and C are 0.6 and 0.3 respectively. For non-sensitive patients, the response rate is 0.3 regardless of treatment. Assume the total sample size is 300, and by the AED design, 80 non-sensitive patients and 220 sensitive patients are enrolled. Due to the equal randomization, among each group (sensitive or non-sensitive), 50% patients have received E, and 50% received C. Then in this setting, $r|E = 0.52$, $r|C = 0.30$, thus $r|E - r|C = 0.22$.

For SEDAR, assume 100 non-sensitive and 200 sensitive patients are enrolled. By SEDAR, non-sensitive patients are equally randomized between E and C, whereas sensitive patients are more likely to be assigned to E. Assume 80% of sensitive patients have received E, and 20% received C. In this setting, $r|E = 0.54$, $r|C = 0.30$, thus $r|E - r|C = 0.24$. SEDAR results in a higher value of $r|E - r|C$ than AED, thus higher power.

For clinical trials that do not consider patients' biomarkers X or in which the biomarker is not predictive, the underlying assumption is that all patients are homogeneous and the patient heterogeneity is ignored. In this situation, it is true that, when compared with equal randomization, adaptive randomization will lose power with some exceptions (e.g., some scenarios with binary outcomes). However, in this paper, we account for patient heterogeneity and use regression models of response on X to predict future patients response rates to E or C. Although both

AED and SEDAR use these results to select patients to enroll, only SEDAR uses these results to adjust patients' probability to receive E or C. The consequence is that, as seen from above, SEDAR trials have larger values of $r|E - r|C$, and thus higher power levels than AED.

Another feature in our designs affects the power. In the simulations, for AED and SEDAR, we have used slightly different enrollment functions for non-sensitive patients. While AED gives a zero chance for non-sensitive patients to enroll, SEDAR gives a chance between 0.1 and 1 for them. This difference favors AED for power (i.e., should make AED have higher power than SEDAR). The final power is the combined effect of this enrollment function and the above effect on $r|E - r|C$ caused by the adaptive randomization.

Our results have also shown that adopting an enrichment strategy - oversampling the sensitive patients and undersampling the nonsensitive patients will offer benefits to trial participants. This is particularly important for late-stage cancer patients since many will need to enroll in clinical trials to have access to novel treatments. The enrichment strategy described here can be applied to other biomarker adaptive designs [170, 157, 140, 171]. For example, the recently-proposed enriched biomarker stratified design and the auxiliary-variable-enriched biomarker stratified design also show that applying an enrichment strategy can result in a more cost-efficient design in terms of power [111, 112]. For both of these designs, the issue is how to choose the optimal enrichment proportion. Our proposed design uses a continuous loss function based on the response rate estimates for each treatment for a given patient using a classification algorithm. In addition to using an enrichment strategy to increase trial efficiency, our proposed design also utilizes Bayesian adaptive randomization approach to sequentially update treatment assignment probabil-

ities based on patients' biomarker signatures measured at baseline. The simulation results imply that our design has more promising trial operating characteristics in most scenarios.

To address the *personalized* aspect of precision medicine, we also proposed four trial performance evaluation criteria in addition to the traditional operating characteristics. The criteria include CIL, FIL, PCO and PFO for both currently enrolled patients and future patients. These criteria numerically present how well the trial design will offer patients their own personalized optimal treatment.

Although our design has distinct advantages, there are some caveats. The complexity of utilizing an enrichment strategy as well as a Bayesian adaptive randomization scheme, yields operational complexities. In addition, in common with other types of adaptive enrichment trials, there is a potential for the trial duration to increase and raise issues of interpretation of the final tests and treatment effect estimates arising from the adaptive enrichment process.

Compared to the all-comers, equal randomization designs, one limitation of an adaptive enrichment strategy is that the final sensitive subset may not represent the true subgroup in the real world. However, if the true proportions of patients in the various subcategories were known, one could reconstruct the "real world" setting by post-hoc weighting to give some idea of the potential bias. Our proposed design allows more patients to receive optimal personalized treatments, thus yielding a higher overall response rate on the trial. This design can identify therapies that are globally beneficial as well as treatments that are effective only in a sensitive subset. The proposed cross-validated signature enrichment design with adaptive randomization outperforms the all-comers equal randomization trial designs in many scenarios.

CHAPTER 5

A Signature Enrichment Design with Bayesian Adaptive Randomization for Umbrella Trials

5.1 Introduction

Traditional "one-size-fits-all" cancer trial assumes all patients are homogeneous. However, in the modern era of cancer research, tumor molecular profiles are now better understood. This advent of precision medicine promotes a shift in treatment development from "one-size-fits-all" to biomarker-driven therapies that are tailored based on each patient's biomarker profile [172, 173, 25]. Novel clinical trial designs are needed to answer more clinical questions more efficiently and in less time.

A master protocol is one overarching protocol which allows simultaneous evaluation of multiple treatments or disease population in multiple sub-studies [88, 174, 175, 32]. Based on the characteristics, a master protocol is categorized into three types: basket, umbrella, and platform trials. Each type includes one or more components that are more efficient than running a single trial.

Basket trials evaluate single treatment within a biomarker positive subgroup but for multiple histologies. The goal is to identify unambiguous signals of treatment based on molecular features rather than histology [32, 33, 176]. Basket trials are usually early phase, single-arm studies, and only need to develop one assay. They

are more efficient than separate studies and are particularly useful to test if an already approved treatment can work on other disease indications. However, the choice of endpoints may be limited due to prognostic heterogeneity across disease types. It can also be challenging to define historical controls across different diseases.

Platform trial refers to a randomized trial design with multiple experimental arms but shares a common control arm. One of the critical features of the platform trials is the flexibility that allows treatments to enter or exit during the course of the trial [32, 33, 176]. New treatment can be added, and existing ones can be declared for superiority or dropped early due to futility, often based on some Bayesian decision rule throughout a study. The Bayesian hierarchical model is commonly used to model the various treatment effects by using independent parameters across subtypes. Adaptive randomization can also be used to allocate more patients to the most promising arm based on the accumulated data [51, 52].

Different from basket and platform trials, umbrella trials target single histology but evaluate multiple treatments each match to a biomarker. Since all patients are from the same tumor group, any observed benefits can be directly linked to the biomarker [33, 176]. Umbrella trials are generally mid-to-late phase studies aiming to identify large treatment effects within a single disease type [176]. Umbrella trials are flexible, new treatment arm can be added in the process, and inferior arms can also be terminated early. Umbrella trials also have limitations: it usually requires a longer duration and larger sample size[176]. The enrollment can be difficult when the prevalence of a biomarker is low [33, 176]. Umbrella trials can also be combined with the basket trials to open to multiple histologies.

In this chapter, we aim to extend the Signature Enrichment Design with Bayesian Adaptive Randomization (SEDAR) to a multi-arm setting, which is suitable

for umbrella trials. The rest of the paper is organized as follows. In section 2, we briefly review the umbrella trial design with examples. In section 3, we describe our signature enrichment design with Bayesian adaptive randomization for umbrella trials. In section 4, we present the results of simulation studies to compare our proposed design with the adaptive signature design and the adaptive enrichment design. In section 5, we conclude with a summary.

5.2 Review of Umbrella Trials

Umbrella trials evaluate multiple treatments within a single disease group. To determine a patient's eligibility, a multiplex assay is used, and only biomarker positive patients will be enrolled. The diagram for umbrella trials is in Figure 5.1. After screening, patients will be assigned to a sub-study based on their molecular makeup. Each sub-study has its own treatment targeting the corresponding molecular agents. For scenarios in which a patient is eligible for multiple arms, inference and hypothesis testing should be used to account for the overlap. In this section, we review three on-going or recent umbrella trials: LUNG-MAP, ALCHEMIST, and FOCUS4.

5.2.1 Examples

Lung-MAP

The Lung Cancer Master Protocol (Lung-MAP) is the first precision medicine trial for patients with previously treated advanced non-small cell lung cancer from the National Cancer Institute [177, 178]. It is an umbrella study in which the patient treatment sub-study assignment is based on her/his biomarkers. If a patient has biomarkers that are targeted by one of the treatments studied in the trial, then he/she will be enrolled in the corresponding sub-study. For patients with multiple matched biomarkers, they will be randomized to a cohort using a weighted algorithm

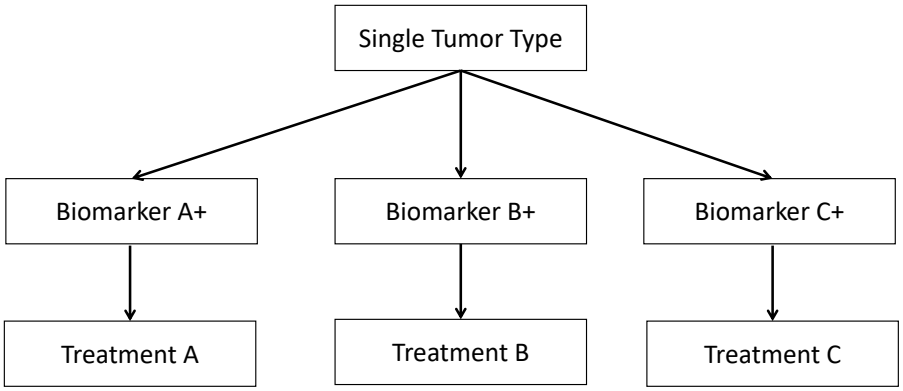

Figure 5.1: General Schema for Umbrella Trials.

with a priority given to low prevalence biomarker cohorts. For non-matched patients, they will be assigned to the non-match group to receive immunotherapy drug combinations. Each treatment sub-study can close after enough number of patients have enrolled. New treatments can be added, and treatments can be dropped from the trial.

The Lung-MAP trial is a phase II/III trial [179]. The endpoint for the phase II study is progression-free-survival, and for the phase III endpoint is progression-free-survival and overall survival. The phase II patients will contribute to the phase III study if the progression-free survival exceeds an efficacy boundary. In October 2017, 1,400 patients registered. One initial cohort - c-MET-positive was terminated early due to toxicity [179]. In March 2015, the FDA approved nivolumab in the same patient population [180]. As a result, the initial control treatment, docetaxel, is no longer the standard of care. This promotes the Lung-MAP to re-open for modifications: phase II trials no longer have a control treatment and now become

single arm with only experimental treatment [176].

ALCHEMIST

The Adjuvant Lung Cancer Enrichment Marker Identification and Sequencing Trials (ALCHEMIST) is an umbrella trial for patients with early-stage non-small cell lung cancer (NSCLC) whose tumors have been completely removed by surgery [181, 182]. The ALCHEMIST trials aim to test if adding erlotinib, crizotinib, or nivolumab can prolong overall survival for NSCLC patients who have completed standard therapy.

The ALCHEMIST trials include a screening trial and 3 phase III adjuvant trials: ALCHEMIST-Screening, ALCHEMIST-EGFR, ALCHEMIST-ALK, and ALCHEMIST-Immunotherapy. The ALCHEMIST-Screening trials exam the patients' EGFR and the ALK molecular profiles. For patients who test positive for the EGFR mutation, they will be enrolled in the ALCHEMIST-EGFR trial to be randomly assigned erlotinib or placebo. For patients who test positive for the ALK rearrangement, they will be enrolled in the ALCHEMIST-ALK trial to be randomly assigned crizotinib or placebo. For patients who test negative for neither mutations, they may be enrolled in an ALCHEMIST trial that compares an immunotherapy drug, nivolumab, with observation. The nivolumab is a checkpoint inhibitor that blocks programmed cell death protein 1 (PD-1) to allow T-cells attack cancer cells.

The ALCHEMIST trials aim to screen at least 8,000 patients. All patients screened will be monitored for five years. For the ALCHEMIST-EGFR and the ALCHEMIST-ALK trials, patients will be assigned to receive the experimental treatment or placebo for two years and will be monitored for up to 10 years after treatment. For the ALCHEMIST-Immunotherapy trials, patients will be randomly assigned to receive nivolumab or just observation for up to 1 year.

FOCUS4

FOCUS4 is a phase 2-3 randomized trial that uses the umbrella design to evaluate different new therapies for patients with varying subtypes of advanced colorectal (bowel) cancer [183, 184]. The sub-types include mutations in the BRAF gene (FOCUS4-A trial), in the PIK3CA gene (FOCUS4-B trial), in the KRAS or NRAS gene (FOCUS4-C trial), no mutation in the above genes or all wild type (FOCUS4-D trial) and unclassified biomarker results (FOCUS4-N trial). The primary outcome measure is progression-free survival and overall survival as the secondary outcome measure. multiarm multistage (MAMS) trial design methodology

Patients first will be followed after 16 weeks of standard first-line chemotherapy. Once completed the chemo, if the tumor has not progressed, patients will be screened for the sub-types. Within each sub-type trial, patients will be randomly allocated to receive either a new treatment or a placebo or active monitoring.

FOCUS4-D patients will receive either AZD8931 or active monitoring. FOCUS4-N patients will receive either capecitabine or active monitoring. Multiple interim analyses are planned for early termination.

Between July 7, 2014, and March 7, 2016, the FOCUS4-D trial randomly assigned 16 patients to receive AZD8931 and 16 patients to placebo. At first interim analysis (March 2016), the independent data monitoring committee (IDMC) recommended early termination due to lack of activity. At final analysis (August 2016), 15 with AZD8931 and 16 with place reached the primary endpoint. Median progression-free survival was 2.96 months in the AZD8931 group and 3.48 groups in the placebo group. No significant benefit of AZD8931 has shown over the placebo. A new treatment replaces the current one for the FOCUS4-D trial.

5.3 Signature Enrichment Design with Bayesian Adaptive Randomization for Umbrella Trials

We extend the proposed cross-validated signature enrichment design with Bayesian adaptive randomization to allow multiple experimental treatment arms and add an early stopping rule that is suitable for umbrella trials (SEDAR-U). Similar to SEDAR, we assume patient responses can be observed shortly after receiving treatment. For the initial burn-in period, identical to SEDAR, we use equal randomization. The diagram for SEDAR-U is presented in Figure 5.2.

After the initial block, we enter an interim analysis to check the futility of each experimental arm. If all arms fail, we terminate the study. If at least one experimental arm passes, the trial continues. The first step is to derive a sensitive patient classifier function (CF) to classify incoming patients into one of the two categorize: sensitive and non-sensitive. Next, an enrollment function (EF) will be applied to determine the probability of enrolling for each patient. Once the block size of adaptive randomization is reached, patients are allocated among treatment arms using Bayesian adaptive randomization. At the final analysis, we check the efficacy of each experimental arm.

5.3.1 Sensitive-patient classifier function (CF)

Let $\hat{r}_{i,x,E}$ and $\hat{r}_{i,x,C}$ denote the estimated response rates for potential patient i with signatures x if receiving treatment E or C, respectively. δ is the minimal clinical meaningful difference for determining patient sensitive status. Then the sensitive patient status is determined as follows:

- If $\hat{r}_{\{i,x,E\}} - \hat{r}_{\{i,x,C\}} > \delta$, then patient i is sensitive;
- If $\hat{r}_{\{i,x,E\}} - \hat{r}_{\{i,x,C\}} \leq \delta$, then patient i is non-sensitive.

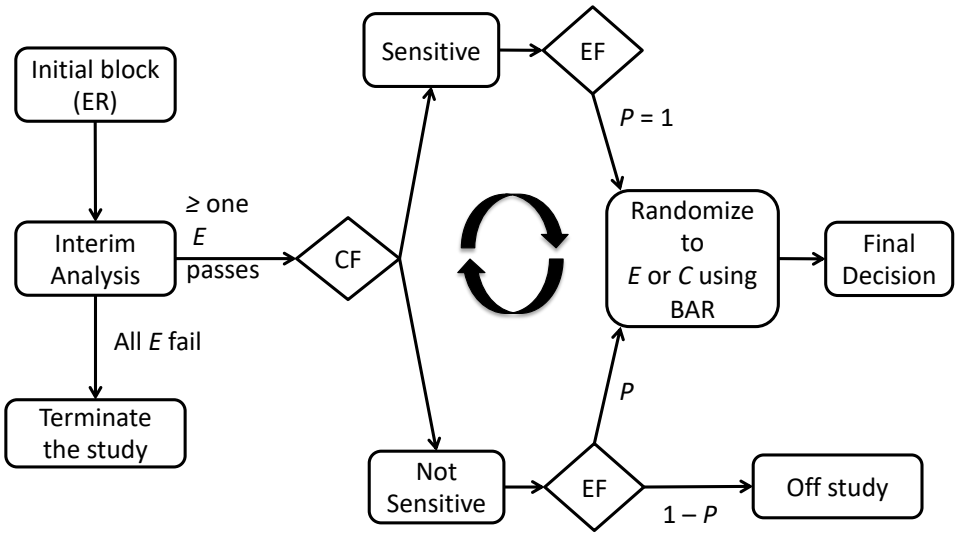

Figure 5.2: Diagram for Signature Enrichment Design with Bayesian Adaptive Randomization for Umbrella Trials (SEDAR-U). *CF* is the classifier for determining patient sensitive status; *EF* is the enrollment function; P_i is the probability of enrolling the i^{th} patient.

5.3.2 Enrollment function (EF)

Once a new patient's sensitive status has been determined, the next step is to calculate the enrollment probability. We use the following function for patient i with signatures x:

$$P_i = \frac{max(\hat{r}_{\{i,x,E\}} - \hat{r}_{\{i,x,C\}},\ \delta/10)}{\delta}.$$

According to this function, the probability for sensitive patients is 1 and that for non-sensitive patients is between 0.1 and 1.

5.3.3 Bayesian adaptive randomization for multi-arms

Let k denote the treatment arm index, $k = 1, ..., K$; E for experimental arms or C for control arm; n for the intermediate sample size, $n = 1, ..., N$; and π_k for the response probability for arm k. Bayesian adaptive randomization assumes the

randomization probability $r_{k,n}$ is proportional to the posterior probability [155]:

$$Pr(\pi_k = max\{\pi_1, ..., \pi_k\}|\ data_n),\ k = 1, ..., K,$$

When n is small, the posterior probabilities can be highly variable. To stabilize the randomization probability, we add a tuning parameter w such that

$$r_{k,n}^w = \frac{(Pr(\pi_k = max\{\pi_1, ..., \pi_k\}|\ data_n))^w}{\sum_{j=1}^K (Pr(\pi_j = max\{\pi_1, ..., \pi_k\}|\ data_n))^w}.$$

The value of w should be between 0 and 1. When w is equal to 0, it gives equal randomization. When w is equal to 1, the $r_{k,n}^w$ is the same as $\hat{r}_{k,n} = \hat{Pr}(\pi_k = max\{\pi_1, ..., \pi_k\}|\ data_n)$. In practice, w is usually set at 1/2.

Another method to control the randomization magnitude is to let the posterior probability be the randomization probability $r_{k,n}$, and restrict it between $[\sigma, 1-\sigma]$ where σ is a pre-specified value between 0 and 0.5.

5.3.4 Final sensitive subset

Similar to the cross-validated adaptive signature design, we use K-fold cross-validation to obtain a final sensitive subset from the entire trial patients. First, the entire trial population is randomly partitioned into M nonoverlapping equal-sized subpopulations; Then, one of the M folds is used as the validation cohort and the rest $M-1$ folds are treated as the development cohort. Then for each fold, we develop a sensitive patient subset S_m $(m = 1, ..., M)$. Then the total number of the sensitive patient subset from the entire trial patients is $S = \bigcup_{m=1}^K S_m$.

5.3.5 Early futility stopping rule and final selection rule

For each $\pi_k, k = 1, ..., E$, we use a non-informative prior $Beta(1, 1)$.

We apply an early stopping rule for experimental arms at interim analysis (after the initial block of patients):

Experimental arm k is terminated due to futility if:

$$\hat{Pr}(\pi_k > \pi_c + \delta \mid data_{INT}) < 0.01,\ k = 1, ..., K.$$

At final analysis, arm k is claim efficacious/selected if

$$\hat{Pr}(\pi_k > \pi_c + \delta \mid data_{FNL}) > a_U,\ k = 1, ..., K,$$

where a_U is a cut-off parameter whose value is determined/tuned to control the type I error rate under the null scenario.

5.4 Trial Performance Evaluation Criteria

Similar to SEDAR, in addition to the most commonly used response rates and power, we propose the following four criteria to further evaluate the trial performance for precision medicine settings:

- Current individual loss (CIL)
- Future individual loss (FIL)
- Probability of a current individual in the trial receiving personalized optimal treatment (PCO)
- Probability of a future individual receiving personalized optimal treatment (PFO)

The details can be found in chapter 4 section 4.4. To summarize, the penalties for non-matches depend on the individual loss for CIL and FIL. While for PCO and PFO, all non-matches receive the same penalty. Good clinical trial designs should result in small values for CIL and FIL and large values for PCO and PFO.

5.5 Required Design Parameters

The required design parameters for SEDAR-U is listed in Table 5.1. We require the sample size for the initial equal randomization block (N_0) for the adaptive

randomization blocks (N_{bar}); tuning parameter w for allocation probabilities; minimal clinical difference for sensitive status classification δ; minimal clinical difference for early stopping rule and final selection rule γ; the type I error rate for testing the overall treatment effect (α_1) and that for testing treatment effect in the sensitive subset (α_2), and the prior for final decision rule and for BAR procedure.

Table 5.1: Design parameters required for SEDAR-U.

N_0	Sample size for the initial equal randomization block
N_{bar}	Sample size for adaptive randomization blocks
N	Total sample size
w	Tuning parameter for allocation probabilities
δ	Minimal clinical difference for sensitive status classification
γ	Minimal clinical difference for early stopping rule and final selection rule
α_1	Type I error rate for testing treatment effect in overall patient group
α_2	Type I error rate for testing treatment effect in sensitive subset
α	Type I error rate for testing treatment effect in either overall patient group or sensitive subset
Prior	for BAR procedure and for r_k

5.6 Simulation Studies for Multiple Treatment Arms with Control

We performed a series of simulation studies to explore the performance of our proposed design that are suitable for umbrella trials. Similar to the previously signature Enrichment Design with Bayesian Adaptive Randomization for two treatments, we compare our proposed design with the cross-validated adaptive signature design and the adaptive enrichment design. For the adaptive enrichment design. At the end of the trial, for all the designs, we use cross validation to seek sensitive patient subset utilizing all the patient data. We are also vary the number of biomarkers to study its effect on the trial performances. Here we investigate three simulation categories: 1) with one control arm and ten biomarkers; 2) with one control arm and two biomarkers; 3) without control arm and two biomarkers. For all scearios, we

consider three treatments. The bivariate treatment indicators are

$$(T_{1i}, T_{2i}) = \begin{cases} (1,0) & \text{if patient } i \text{ receives } A \\ (0,1) & \text{if patient } i \text{ receives } B \\ (0,0) & \text{if patient } i \text{ receives } C \textit{ (control)} \end{cases} \tag{5.1}$$

We consider a three-arm clinical trial designed to compare two experimental treatments (A, B) with a control treatment (C). We generate the i^{th} patient's response from the following logistic regression model with ten biomarkers:

$$logit(r_i) = \beta_0 + \beta_1 T_{1i} + \beta_2 T_{2i} + \eta_1 T_{1i} X_{1i} + \eta_2 T_{2i} X_{1i}, i = 1, \cdots, N.$$

where X_{ki} is the k^{th} biomarker/signature for patient i. Each biomarker X_{ki} is assumed to follow a multivariate normal distribution $(0.5, \Sigma)$ with correlation coefficient ρ. The prior for each regression coeficient is $\beta_0, \beta_1, \beta_2, \eta_1, \eta_2 \sim Normal(0, 0.01)$, where 0.01 refers to the precision.

Simulation setting

The total sample size N is set at 300; the size of each adaptive randomization block N_{bar} is set at 50. We choose the size N_0 of the initial block, for which equal randomization is used, to be $N_0 = 150$. Hence, we have a total of 4 blocks of sizes 150, 50, 50, and 50. The tuning parameter for Bayesian adaptive randomization w is set at 0.5; the minimal clinical difference for sensitive status classification is 0.15, and for early stopping rule and final selection rule is 0; the type I error rate α_1 for testing overall treatment effect is controlled at 0.03; the type I error rate α_2 for testing treatment effect in each sensitive subset is controlled at 0.01. Thus, the total type I error rate for testing treatment effect in either overall patient group or sensitive subset α is controlled at 0.05. The prior distribution of the response rate for each treatment, r_k, is assumed to be non-informative with $Beta(1, 1)$.

Similar to the simulation scenarios in chapter 4, we consider three biomarker distributions. For each type, we include two scenarios and a null scenario to compare the performance of SEDAR with ASD and AED (Table 5.2). The biomarker or biomarker-treatment interaction effects not presented in the table are assumed to be zero.

For the first category, we assume all biomarkers are independently distributed following a normal distribution with mean 0.5 and variance 1.0. For the second category, we assume the first five biomarkers have a multivariate normal distribution $X \sim NVM(0.5, \Sigma)$ with the covariance matrix, which has 1 for the variance and 0.25 for ρ. The other five variables are still independently distributed with a normal distribution with mean 0.5 and variance 1.0. For the third category, we assume the first five biomarkers have a multivariate normal distribution with mean 0.5 and covariance matrix with $\rho = 0.25$, and the remaining five biomarkers have a multivariate normal distribution with mean 0.5 and covariance matrix with $\rho = 0.75$.

All the categories and scenarios are listed in Table 5.2. The null scenarios for each category numbers 1, 4, and 7, respectively. For scenarios 2, 5 or 8, we consider the main treatment effect only for T_1 and the corresponding treatment interaction between T_1 and X_1, T_2 and X_1 will affect the patient response with different scales. For scenarios 3, 6 or 9, we assume both the main treatment effect of T_1 and T_2, and each of their biomarker-treatment interactions with X_1 will affect the patient response with different scales. For each scenario, we generated 2,000 simulations.

Simulation results

We compare the results obtained using our proposed design with those obtained using ASD and AED. The results for the estimated power are presented in Table 5.3. The type I error rates for the treatment effect in the overall patient group

Table 5.2: Simulation scenarios - SEDAR-U.

		β_0	β_1	β_2	η_1	η_2
x_1 to x_{10} with $\rho = 0.00$	Scenario 1	-0.8	0.0	0.0	0.0	0.0
	Scenario 2	-0.8	0.2	0.0	1.6	1.2
	Scenario 3	-0.8	0.5	0.3	0.8	0.2
x_1 to x_5 with $\rho = 0.25$	Scenario 4	-0.8	0.0	0.0	0.0	0.0
x_6 to x_{10} with $\rho = 0.00$	Scenario 5	-0.8	0.2	0.0	1.6	1.2
	Scenario 6	-0.8	0.5	0.3	0.8	0.2
x_1 to x_5 with $\rho = 0.25$	Scenario 7	-0.8	0.0	0.0	0.0	0.0
x_6 to x_{10} with $\rho = 0.75$	Scenario 8	-0.8	0.2	0.0	1.6	1.2
	Scenario 9	-0.8	0.5	0.3	0.8	0.2

and each sensitive subgroup are controlled at 0.03 and 0.01, respectively. The composite type I error rate, which measures the probability of claiming that there is a significant treatment effect found in either overall group or at least one of the sensitive subsets when there is indeed no difference, is controlled at 0.05. The control of the type I error rates are shown under the null scenarios in Table 5.3.

For the alternative scenarios, where there is indeed some treatment effect, ASD has the highest overall power, following by AED and SEDAR. Though for the composite power, the differences among all three designs become minimal. For example, in scenario 6, the composite power for ASD, AED and SEDAR are 0.944, 0.940 and 0.937, respectively. In the sensitive subgroup, the relative performance would depend on the underlying true treatment effects in the setting. When the effect of treatment A is substantial and much larger than that of treatment B, the power in sensitive patient subset to treatment A is similar among all designs. Also note that the power of the treatment effect among patients who are sensitive to treatment A is higher than that among the overall patient group. This reflects the scenarios of precision medicine. When the effect of treatment A is moderate, its power in the sensitive subset of ASD and SEDAR may be higher than AED. For example, in scenario 9, the power of treatment effect among patients who are

sensitive to treatment A is 0.752 for ASD, 0.744 for AED and 0.762 for SEDAR.

Table 5.3: Simulation results: Type I error rate/Power - SEDAR-U.

SCEN	Parameters	Design	Overall	Sensitive (A)	Sensitive (B)	Composite
1	NULL	ASD	0.031	0.011	0.010	0.052
	(Type I error rate)	AED	0.027	0.009	0.009	0.045
		SEDAR	0.028	0.011	0.009	0.048
2	$\beta_1 = 0.2, \beta_2 = 0.0$	ASD	0.939	0.972	0.600	0.988
	$\gamma_1 = 1.6, \gamma_2 = 1.2$	AED	0.939	0.971	0.610	0.983
		SEDAR	0.920	0.950	0.600	0.974
3	$\beta_1 = 0.5, \beta_2 = 0.3$	ASD	0.932	0.760	0.269	0.948
	$\gamma_1 = 0.8, \gamma_2 = 0.2$	AED	0.926	0.742	0.263	0.942
		SEDAR	0.916	0.764	0.279	0.931
4	NULL	ASD	0.028	0.010	0.012	0.049
	(Type I error rate)	AED	0.027	0.010	0.012	0.048
		SEDAR	0.029	0.010	0.009	0.048
5	$\beta_1 = 0.2, \beta_2 = 0.0$	ASD	0.950	0.978	0.636	0.989
	$\gamma_1 = 1.6, \gamma_2 = 1.2$	AED	0.933	0.967	0.601	0.986
		SEDAR	0.924	0.958	0.608	0.977
6	$\beta_1 = 0.5, \beta_2 = 0.3$	ASD	0.933	0.765	0.293	0.944
	$\gamma_1 = 0.8, \gamma_2 = 0.2$	AED	0.929	0.735	0.267	0.940
		SEDAR	0.921	0.762	0.282	0.937
7	NULL	ASD	0.027	0.009	0.010	0.046
	(Type I error rate)	AED	0.030	0.008	0.011	0.049
		SEDAR	0.031	0.012	0.011	0.053
8	$\beta_1 = 0.2, \beta_2 = 0.0$	ASD	0.951	0.970	0.636	0.989
	$\gamma_1 = 1.6, \gamma_2 = 1.2$	AED	0.944	0.960	0.601	0.986
		SEDAR	0.925	0.956	0.608	0.977
9	$\beta_1 = 0.5, \beta_2 = 0.3$	ASD	0.942	0.752	0.267	0.955
	$\gamma_1 = 0.8, \gamma_2 = 0.2$	AED	0.936	0.744	0.275	0.949
		SEDAR	0.932	0.762	0.292	0.948

The results of the four trial performance evaluation criteria are presented in Table 5.4. Since both ASD and AED adopt equal randomization, the number of patients is similar in each treatment arm. Thus, the ORR is the same for ASD and AED. However, SEDAR uses adaptive randomization to assign more patients to the superior treatment as trial proceeds, unlike ASD or AED which uses equal randomization, SEDAR always achieves a higher objective response rate (ORR) than the others. For example, in scenario 6, 45.1% of patients respond to the assigned treatment for SEDAR compared to 41.0% for ASD and 40.9% for AED.

The CIL and FIL show the average values of the individual loss for trial participants and for future patients, respectively. For CIL, SEDAR has the smallest values following by ASD and then AED. Using scenario 3 as an example, the CIL for SEDAR is 0.284 compared to 0.337 for ASD and 0.342 for AED. This difference in CIL shows that SEDAR results in more trial participants receiving their optimal personalized treatments. For FIL, the results are similar for all three designs. Those results suggest that the situation is more complicated when there are multiple treatments and early stopping rules involved.

For the PCO values, SEDAR demonstrates that it offers the highest probability for trial participants to receive their personalized optimal treatments. This is achieved by assigning more patients to superior treatment instead of using a fixed ratio assignment method. The SEDAR also shows promising results for PFO, which refers to the probability for a future patient to receive their personalized optimal treatment if following the treatment recommendation from the trial.

To further investigate the patient allocation from applying Bayesian adaptive randomization, we take a look at the number of patients by treatment arm (Table 5.5). Since there is virtually no treatment effect under the null scenarios, Bayesian adaptive randomization acts the same as equal randomization by assigning the same number of patients to each treatment arm. For alternative scenarios, Bayesian adaptive randomization manages to assign many more patients to the truly superior treatment and much fewer patients to the control treatment compared to equal randomization. For example, in scenario 6, the number of patients assigns to treatment A is 143 for SEDAR and 100 for ASD and AED. This clearly demonstrates the advantages of Bayesian adaptive randomization, especially when the new treatment could bring substantial benefits to patients.

Table 5.4: Simulation results: Trial performance evaluation criteria - SEDAR-U.

SCEN		Design	ORR	CIL	FIL	PCO	PFO
1	NULL	ASD	0.310	0.000	0.000	1.000	1.000
		AED	0.310	0.000	0.000	1.000	1.000
		SEDAR	0.310	0.000	0.000	1.000	1.000
2	$\beta_1 = 0.2, \beta_2 = 0.0$	ASD	0.420	0.314	0.057	0.333	0.703
	$\gamma_1 = 1.6, \gamma_2 = 1.2$	AED	0.420	0.329	0.053	0.333	0.788
		SEDAR	0.455	0.274	0.060	0.437	0.764
3	$\beta_1 = 0.5, \beta_2 = 0.3$	ASD	0.410	0.337	0.037	0.333	0.710
	$\gamma_1 = 0.8, \gamma_2 = 0.2$	AED	0.410	0.342	0.042	0.333	0.774
		SEDAR	0.451	0.284	0.044	0.445	0.764
4	NULL	ASD	0.310	0.000	0.000	1.000	1.000
		AED	0.310	0.000	0.000	1.000	1.000
		SEDAR	0.310	0.000	0.000	1.000	1.000
5	$\beta_1 = 0.2, \beta_2 = 0.0$	ASD	0.420	0.314	0.057	0.333	0.701
	$\gamma_1 = 1.6, \gamma_2 = 1.2$	AED	0.420	0.329	0.052	0.333	0.789
		SEDAR	0.455	0.274	0.059	0.438	0.767
6	$\beta_1 = 0.5, \beta_2 = 0.3$	ASD	0.410	0.337	0.037	0.333	0.710
	$\gamma_1 = 0.8, \gamma_2 = 0.2$	AED	0.409	0.342	0.040	0.333	0.778
		SEDAR	0.451	0.284	0.043	0.445	0.769
7	NULL	ASD	0.310	0.000	0.000	1.000	1.000
		AED	0.310	0.000	0.000	1.000	1.000
		SEDAR	0.310	0.000	0.000	1.000	1.000
8	$\beta_1 = 0.2, \beta_2 = 0.0$	ASD	0.420	0.314	0.059	0.333	0.696
	$\gamma_1 = 1.6, \gamma_2 = 1.2$	AED	0.420	0.329	0.054	0.333	0.781
		SEDAR	0.455	0.273	0.060	0.439	0.760
9	$\beta_1 = 0.5, \beta_2 = 0.3$	ASD	0.409	0.338	0.040	0.333	0.702
	$\gamma_1 = 0.8, \gamma_2 = 0.2$	AED	0.409	0.343	0.043	0.333	0.769
		SEDAR	0.450	0.285	0.046	0.444	0.757

5.7 Summary

In the modern era of cancer research, the evolution of clinical trial designs calls for novel trial designs with improved efficiency and flexibility without undermining the validity and integrity. Among the newly proposed concepts, umbrella trials target multiple genetic mutations with different treatments for a single cancer type. Umbrella trials screen for patient molecular profiles and then assign treatments accordingly. This procedure provides patients more trial options and allows a better "fit" between the studies and patients.

Table 5.5: Simulation results: Number of patients by treatment arm - SEDAR-U.

SCEN	Parameters	Design	# in arm A	# in arm B	# in arm C
1	NULL	ASD	100.0	100.0	100.0
		AED	100.0	100.0	100.0
		SEDAR	100.0	99.0	101.0
2	$\beta_1 = 0.2, \beta_2 = 0.0$	ASD	100.0	100.0	100.0
	$\gamma_1 = 1.6, \gamma_2 = 1.2$	AED	100.0	99.9	100.2
		SEDAR	135.7	94.3	69.9
3	$\beta_1 = 0.5, \beta_2 = 0.3$	ASD	100.0	100.0	100.0
	$\gamma_1 = 0.8, \gamma_2 = 0.2$	AED	99.9	99.9	100.2
		SEDAR	142.4	89.1	68.5
4	NULL	ASD	100.0	100.0	100.0
		AED	100.0	100.0	100.0
		SEDAR	100.0	99.0	101.0
5	$\beta_1 = 0.2, \beta_2 = 0.0$	ASD	100.0	100.0	100.0
	$\gamma_1 = 1.6, \gamma_2 = 1.2$	AED	100.1	100.1	100.0
		SEDAR	135.8	94.4	69.8
6	$\beta_1 = 0.5, \beta_2 = 0.3$	ASD	100.0	100.0	100.0
	$\gamma_1 = 0.8, \gamma_2 = 0.2$	AED	99.9	100.1	100.0
		SEDAR	142.7	88.9	68.4
7	NULL	ASD	100.0	100.0	100.0
		AED	100.0	100.0	100.0
		SEDAR	100.0	99.0	101.0
8	$\beta_1 = 0.2, \beta_2 = 0.0$	ASD	100.0	100.0	100.0
	$\gamma_1 = 1.6, \gamma_2 = 1.2$	AED	100.1	100.0	99.9
		SEDAR	136.1	94.0	70.0
9	$\beta_1 = 0.5, \beta_2 = 0.3$	ASD	100.0	100.0	100.0
	$\gamma_1 = 0.8, \gamma_2 = 0.2$	AED	99.9	100.1	100.0
		SEDAR	141.9	90.2	68.0

In this chapter, we have extended the cross-validated signature enrichment design with Bayesian adaptive randomization to allow multiple treatment settings with an early stopping rule which is suitable for umbrella trials. In addition to the advantages of the cross-validated adaptive signature design and the adaptive enrichment design, our design applies the Bayesian adaptive randomization method for multi-arm and includes an early futility stopping rule. After the initial block of patients data became available, at the interim analysis, we apply an early stopping rule which allows early termination of the study if all the experimental arms show futility. If at least one experimental treatment passed the interim analysis, the proceeding

steps are similar to the cross-validated signature enrichment design with Bayesian adaptive randomization in chapter 4. That is, we enroll patients using an adaptive enrichment approach by developing a sensitive patient classifier to oversample sensitive patients and undersample non-sensitive patients. Once a patient enters the trial, we use Bayesian response-adaptive randomization for multi-arms to adjust allocation probabilities among the remaining arms. At the final analysis, we declare if each of the remaining arms is efficacious based on a Bayesian decision rule.

CHAPTER 6

Conclusion and Future Work

Clinical trial in the era of precision medicine promotes innovation in trial design to answer more questions in less time. Adaptive clinical trials allow pre-specified modifications on one or more trial components in the process so that more enrolled patients could benefit from participating in the trial and result in expedited drug development. After determining a safe and effective dose in phase I clinical trial, phase II clinical trial assesses if the new treatment works that can then be further evaluated in phase III clinical trial. In this dissertation, we mainly focus on developing a novel design for phase II clinical trials.

In chapter 3, we compared different Bayesian response-adaptive randomization methods and discussed their applications from both methodological and ethical aspects. Bayesian response-adaptive randomization allocates more patients to the superior treatment arms based on posterior probabilities, which is repetitively updated based on accumulating trial data. Our simulation studies demonstrate that Bayesian response-adaptive randomization results in more responders than ER, therefore display higher patient response rates.

In chapter 4, we proposed a signature enrichment design with Bayesian response-adaptive randomization (SEDAR) focusing on cancer clinical trials. In addition to

elements of ASD and AED, the proposed design uses an enrichment strategy to enroll patients and applies Bayesian response-adaptive randomization for patient allocation. As a result, SEDAR offers more advantages, in addition to the ones of ASD and AED. Also, we proposed four trial evaluation criteria that can be used to better evaluate the performance of trial design for precision medicine. Extensive simulation studies show that SEDAR presents more favorable characteristics than the other two designs. We also present an example based on the IPASS trial.

In chapter 5, we extended SEDAR to incorporate multiple treatments and add an early stopping for futility. Those elements are suitable for umbrella trials. Umbrella trial is a type of master protocols, which studies multiple treatments with each in a biomarker enriched sub-study for one disease type under one overarching protocol. Our simulation studies demonstrate that the proposed signature enrichment design with Bayesian adaptive randomization for umbrella trials (SEDAR-U) offers more benefits to the trial participants than the alternative designs.

In contrast to umbrella trials, basket trials evaluate the test treatment for multiple disease types but with common molecular alteration(s). Basket trials often require a fewer number of patients and result in a shorter duration than the traditional phase II trial. Also, the basket trial can be especially useful for rare molecular aberrations as the data may be pooled across substudies for analysis. Many designs for basket trials have been proposed that allow borrowing information across cancer types to improve the statistical power while controlling the type I error rate [185, 186, 187, 188, 189]. One future work could be to extend SEDAR for basket trial.

In this dissertation, we focus on a single primary endpoint. Trial designs that use more than one primary endpoint may offer a more comprehensive evaluation of

the treatment effect than ones that are using the single endpoint. In 2017, the US Food and Drug Administration (FDA) issued guidance regarding the use of multiple endpoints in clinical trials [190]. When designing a trial that uses co-primary endpoints to evaluate the joint effect of all the endpoints, the type II error needs to be controlled as it inflates with an increasing number of endpoints [191, 192]. For hypothesis testing, co-primary endpoints use the intersection-union principle [193]. The null hypothesis is rejected if and only if all hypotheses with each corresponding to each endpoint are rejected. Many designs and analyses have been proposed for co-primary endpoint trials [194, 195, 196, 197]. Extention of the SEDAR design to allow co-primary endpoints still warrants future research works.

Bibliography

[1] B. Freidlin and R. Simon. Adaptive signature design: an adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients. *Clinical cancer research*, 11(12):7872–7878, 2005.

[2] B. Freidlin, W. Jiang, and R. Simon. The cross-validated adaptive signature design. *Clinical cancer research*, 16(2):691–698, 2010.

[3] N. Simon and R. Simon. Adaptive enrichment designs for clinical trials. *Biostatistics (Oxford, England)*, 14(4):613–625, 2013.

[4] S-C. Chow and R. Corey. Benefits, challenges and obstacles of adaptive clinical trial designs. *Orphanet Journal of Rare Diseases*, 6(79):1186/1750–1172, 2011.

[5] EL. Korn and Boris. Freidlin. Adaptive clinical trials: Advantages and disadvantages of various adaptive design elements. *JNCI: Journal of the National Cancer Institute*, 109(6), 2017.

[6] J. O'Quigley, M. Pepe, and L. Fisher. Continual reassessment method: A practical design for phase 1 clinical trials in cancer. *Biometrics*, 46:33–48, 1990.

[7] J. O'Quigley and LZ. Shen. Continual reassessment method: A likelihood approach. *Biometrics*, 52:673–684, 1996.

[8] SN. Goodman, ML. Zahurak, and S. Piantadosi. Some practical improvements in the continual reassessment method for phase i studies. *Statistics in Medicine*, 14:1149–1161, 1995.

[9] BE. Storer. Design and analysis of phase I clinical trials. *Biometrics*, 45(3):925–937, 1989.

[10] BE. Storer. An evaluation of phase i clinical trial designs in the continuous dose-response setting. *Statistics in Medicine*, 20(16):2399–2408, 2001.

[11] JMG. Taylor, TM. Braun, and Z. Li. Comparing an experimental agent to a standard agent: relative merits of a one-arm or randomized two-arm phase II design. *Clinical Trials*, 3:335–338, 2006.

[12] GR. Pond and S. Abbasi. Quantitative evaluation of single-arm versus randomized phase II cancer clinical trials. *Clinical Trials*, 8(3):260–269, 2011.

[13] H Tang, NR. Foster, A. Grothey, SM. Ansell, RM. Goldberg, and DJ. Sargent. Comparison of error rates in single-arm versus randomized phase II cancer clinical trials. *Journal of Clinical Oncology*, 28(11):1936–1941, 2010.

[14] R. Simon. Optimal two-stage designs for phase II clinical trials. *Control Clinical Trials*, 10(1):1–10, 1989.

[15] TR. Fleming. One-sample multiple testing procedure for phase II clinical trials. *Biometrics*, 38:143–151, 1982.

[16] SH. Jung, TY. Lee, KM. Kim, and S. George. Admissible two-stage designs for phase II cancer clinical trials. *Statistics in Medicine*, 23:561–569, 2004.

[17] A. Banerjee and AA. Tsiatis. Adaptive two-stage designs in phase II clinical trials. *Statistics in Medicine*, 25(19):3382–3395, 2006.

[18] G. Yin, N. Chen, and JJ. Lee. Phase II trial design with bayesian adaptive randomization and predictive probability. *Journal of the Royal Statistical Society Series C, Applied statistics*, 61(2):1467–9876, 2012.

[19] A. Satlin, J. Wang, V. Logovinsky, S. Berry, C. Swanson, S. Dhadda, and D. A. Berry. Design of a bayesian adaptive phase 2 proof-of-concept trial for ban2401, a putative disease-modifying monoclonal antibody for the treatment of alzheimer's disease. *Alzheimer's & Dementia*, 2:1–12, 2016.

[20] LD. Fisher, DO. Dixon, J. Herson, RK. Frankowski, MS. Hearron, and KE. Peace. *Intention to treat in clinical trials*. Marcel Dekker, 1990.

[21] N. Stallard and S. Todd. Seamless phase II/III designs. *Statistical Methods in Medical Research*, 20(6):623–634, 2011.

[22] M. Wang, J. J. Dignam, Q. E. Zhang, J. F. DeGroot, M. P. Mehta, and S. Hunsberger. Integrated phase II/III clinical trials in oncology: a case study. *Clinical Trials*, 9(6):741–747, 2012.

[23] N. Goossens, S. Nakagawa, X. Sun, and Y. Hoshida. Cancer biomarker discovery and validation. *Translational cancer research*, 4(3):256–269, 2011.

[24] Sauter E. R. Reliable biomarkers to identify new and recurrent cancer. *European journal of breast health*, 13(4):162–167, 2017.

[25] P. Janiaud, S. Serghiou, and JPA. Loannidis. New clinical trial designs in the era of precision medicine: An overview of definitions, strengths, weaknesses, and current use in oncology. *Cancer Treatment Reviews*, 73:20–30, 2019.

[26] JS. Ross, JA. Fletcher, GP. Linette, J. Stec, E. Clark, and M. Ayers. The

her-2/neu gene and protein in breast cancer 2003: biomarker and target of therapy. *The Oncologist*, 8:307–325, 2003.

[27] DJ. Slamon, B. Leyland-Jones, S. Shak, H. Fuchs, V. Paton, A. Bajamonde, T. Fleming, W. Eiermann, J. Wolter, M. Pegram, J. Baselga, and L. Norton. Use of chemotherapy plus a monoclonal antibody against her2 for metastatic breast cancer that overexpresses her2. *New England Journal of Medicine*, 344(11):783–792, 2001.

[28] L. Cheng, A. Lopez-Beltran, F. Massari, GT. MacLennan, and R. Montironi. Molecular testing for braf mutations to inform melanoma treatment decisions: a move toward precision medicine. *Modern pathology*, 31(1):24–38, 2018.

[29] R. Rosell, E. Carcereny, R. Gervais, A. Vergnenegre, B. Massuti, E. Felip, R. Palmero, R. Garcia-Gomez, C. Pallares, JM. Sanchez, R. Porta, M. Cobo, P. Garrido, F. Longo, T. Moran, A. Insa, F. De Marinis, R. Corre, I. Bover, A. Illiano, E. Dansin, J. de Castro, M. Milella, N. Reguart, G. Altavilla, U. Jimenez, M. Provencio, MA. Moreno, J. Terrasa, J. Muñoz-Langa, J. Valdivia, D. Isla, M. Domine, O. Molinier, J. Mazieres, N. Baize, R. Garcia-Campelo, G. Robinet, D. Rodriguez-Abreu, G. Lopez-Vivanco, V. Gebbia, L. Ferrera-Delgado, P. Bombaron, R. Bernabe, A. Bearz, A. Artal, E. Cortesi, C. Rolfo, M. Sanchez-Ronco, A. Drozdowskyj, C. Queralt, I. de Aguirre, JL. Ramirez, JJ. Sanchez, MA. Molina, M. Taron, L. Paz-Ares, Spanish Lung Cancer Group in collaboration with Groupe Français de Pneumo-Cancérologie, and Associazione Italiana Oncologia Toracica. Erlotinib versus standard chemotherapy as first-line treatment for european patients with advanced egfr mutation-

positive non-small-cell lung cancer (eurtac): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncology*, 13(3):239–246, 2012.

[30] L. Marcus, S.J. Lemery, P. Keegan, and R. Pazdur. Fda approval summary: Pembrolizumab for the treatment of microsatellite instability-high solid tumors. 25(13):3753–3758, 2019.

[31] U.S. Food and Drug Administration. Master protocols: Efficient clinical trial design strategies to expedite development of oncology drugs and biologics. *Guidance Document*, 2018.

[32] J. Woodcock and LM. LaVange. Master protocols to study multiple therapies, multiple diseases, or both. *New England Journal of Medicine*, 377(1):62–70, 2017.

[33] A. Hirakawa, J. Asano, H. Sato, and S. Teramukai. Master protocol trials in oncology: Review and new trial designs. *Contemporary clinical trials communications*, 12:1–8, 2018.

[34] EL. Korn, D. Midthune, TT. Chen, LV. Rubinstein, MC. Christian, and RM. Simon. A comparison of two phase i trial designs. *Statistics in Medicine*, 13:1799–1806, 1994.

[35] SN. Goodman, ML. Zahurak, and S. Piantadosi. Some practical improvements in the continual reassessment method for phase I studies. *Statistics in Medicine*, 14:1149–1161, 1995.

[36] P. F. Thall and S. J. Lee. Practical model-based dose-finding in phase I clinical trials: Methods based on toxicity. *International Journal of Gynecological Cancer*, 13(3):251–261, 2003.

[37] J. Babb, A. Rogatko, and S. Zacks. Cancer phase i clinical trials: efficient dose escalation with overdose control. *Sstatistics in medicine*, 17(10):1103–1020, 1998.

[38] G. Yin and Y. Yuan. Bayesian model averaging continual reassessment method in phase i clinical trials. *Journal of the American Statistical Association*, 104(487):954–968, 2009.

[39] Y. Cheung and R. Chappell. Sequential designs for phase i clinical trials with late-onset toxicities. *Biometrics*, 56:1177–1182, 2000.

[40] B. Neuenschwander, M. Branson, and T. Gsponer. Critical aspects of the bayesian approach to phase I cancer trials. *Statistics in Medicine*, 27(13):2420–2439, 2008.

[41] C. Le Tourneau, JJ. Lee, and LL. Siu. Dose escalation methods in phase I cancer clinical trials. *Journal of the National Cancer Institute*, 101(10):708–720, 2009.

[42] Y. Ji and S. Wang. Modified toxicity probability interval design: A safer and more reliable method than the 3 + 3 design for practical phase i trials. *Journal of Clinical Oncology*, 31(14):1785–1791, 2013.

[43] W. Guo, SJ. Wang, S. Yang, H. Lynn, and Y. Ji. A bayesian interval dose-finding design addressing ockham's razor: mtpi-2. *Contemporary Clinical Trials*, 58:23–33, 2017.

[44] Y. Yuan, K. R. Hess, S. G. Hilsenbeck, and M. R. Gilbert. Bayesian optimal interval design: A simple and well-performing design for phase i oncology trials. *Clinical Cancer Research*, 22:4291–4301, 2017.

[45] J.R. Eisele and M.B. Woodroofe. Central limit theorems for doubly adaptive biased coin designs. *Annals of Statistics*, 23:234–254, 1995.

[46] F. Hu and L.X. Zhang. Asymptotic properties of doubly adaptive biased coin designs for multi-treatment clinical trials. *Annals of Statistics*, 21:268–301, 2004.

[47] J. Neyman. On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97:558–606, 1934.

[48] WF. Rosenberger, N. Stallard, A. Ivanova, CN. Harper, and ML. Ricks. Optimal adaptive designs for binary response trials. *Biometrics*, 57:309–913, 2001.

[49] F. Hu, L. Zhang, S.H. Cheung, and W.S. Chan. Doubly adaptive biased coin designs with delayed responses. *Canadian Journal of Statistics*, 36:541–559, 2008.

[50] F. Hu and WF. Rosenberger. *The theory of response adaptive randomization in clinical trials*. John Wiley & Sons, 2006.

[51] P. Thall and J. K. Wathen. Statistical controversies in clinical research: scientific and ethical problems with adaptive randomization in comparative clinical trials. *Annals of Oncology*, 26:1621–1628, 2015.

[52] P. Thall and J. Wathen. A simulation study of outcome adaptive randomization in multi-arm clinical trials. *Clinical Trials*, 14:432–440, 2017.

[53] D.A. Berry. Adaptive clinical trials: The promise and the caution. *Journal of Clinical Oncology*, 29(6):606–609, 2011.

[54] S.J. Pocock. Group sequential methods in the design and analysis of clinical trials. *Biometrika*, 64(2):191–199, 1977.

[55] PC. O'Brien and TR. Fleming. A multiple testing procedure for clinical trials. *Biometrics*, 35(3):549–556, 1979.

[56] SK. Wang and AA. Tsiatis. Approximately optimal one-parameter boundaries for group sequential trials. *Biometrics*, 43(1):193–199, 1987.

[57] K. Lan and D. DeMets. Discrete sequential boundaries for clinical trials. *Biometrika*, 70(3):659–663, 1983.

[58] DL. DeMets and KK. Lan. Interim analysis: the alpha spending function approach. *Statistics in Medicine*, 13:1341–52, 1994.

[59] HH. Ḿuller and H. Sch́afer. Adaptive group sequential designs for clinical trials: combining the advantages of adaptive and of classical group sequential approaches. *Biometrics*, 57(3):886–891, 2001.

[60] U.S. Food and Drug Administration. Adaptive design clinical trials for drugs and biologics. *Guidance Document*, 2018.

[61] Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics*, 69(3):1–7, 2001.

[62] O'Neill R. T. Wang, S. and H. M. Hung. Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset. *Pharmaceutical Statistics*, 6:227–244, 2007.

[63] Y. Xu, L. Trippa, P. Müller, and Y. Ji. Subgroup-based adaptive (suba) designs for multi-arm biomarker trials. *Statistics in Biosciences*, 8(1):159–180, 2016.

[64] W. Jiang, B. Freidlin, and R. Simon. Biomarker-adaptive threshold design: A procedure for evaluating treatment with possible biomarker-defined subset effect. *Journal of the National Cancer Institute*, 99(13):1036–1043, 2007.

[65] X. Gu, N. Chen, C. Wei, S. Liu, VA. Papadimitrakopoulou, RS. Herbst, and JJ. Lee. Bayesian two-stage biomarker-based adaptive design for targeted therapy development. *Statistics in Biosciences*, 8(1):99–128, 2016.

[66] M. Chang. *Introductory Adaptive Trial Designs: A Practical Guide with R*. Chapman & Hall/Crc Biostatistics, 2015.

[67] PF. Thall, R. Simon, and SS. Ellenberg. A two-stage de- sign for choosing among several experimental treatments and a control in clinical trials. *Biometrics*, 45:537–547, 1989.

[68] DJ. Schaid, S. Wieand, and TM. Therneau. Optimal two stage screening designs for survival comparisons. *Biometrika*, 77:659–663, 1990.

[69] N. Stallard and S. Todd. Sequential designs for phase III clinical trials incorporating treatment selection. *Statistics in Medicine*, 22:689–703, 2003.

[70] W. Huang, J. Liu, and C. Hsiao. An alternative phase II/III design for continuous endpoints. *Pharmaceutical Statistics*, 10:105–114, 2011.

[71] W. Bischoff and F. Miller. Adaptive two-stage test procedures to find the best treatment in clinical trials. *Biometrika*, 92:197–212, 2005.

[72] S. Todd and N. Stallard. A new clinical trial design combining phases II and III: sequential designs with treatment selection and a change of endpoint. *Therapeutic Innovation & Regulatory Science*, 39:109–118, 2005.

[73] SC. Chow and YH. Tu. On two-stage seamless adaptive design in clinical trials. *Journal of the Formosan Medical Association*, 107:S52–S60, 2008.

[74] M. Posch, W. Maurer, and F. Bretz. Type i error rate control in adaptive designs for confirmatory clinical trials with treatment selection at interim. *Pharmaceutical Statistics*, 10:96–104, 2011.

[75] N. Stallard. A confirmatory seamless phase II/III clinical trial design incorporating short-term endpoint information. *Statistics in Medicine*, 29(9):959–971, 2010.

[76] J. Maca, S. Bhattacharya, V. Dragalin, P. Gallo, and M. Krams. Adaptive seamless phase II/III designs—background, operational aspects, and examples. *Drug Information Journal*, 40(4):463–473, 2006.

[77] WR. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of the two samples. *Biometrika*, 25:285–294, 1933.

[78] W. Rosenberger and J.M. Lachin. The use of response adaptive designs in clinical trials. *Controlled Clinical Trials*, 14:471–484, 1993.

[79] J. Ning and X. Huang. Response-adaptive randomization for clinical trials with adjustment for covariate imbalance. *Statistics in Medicine*, 29(17):1761–1768, 2010.

[80] PF. Thall and JK. Wathen. Practical bayesian adaptive randomisation in clinical trials. *European journal of cancer (Oxford, England: 1990)*, 43:859–866, 2007.

[81] JJ. Lee, N. Chen, and G. Yin. Worth adapting? revisiting the usefulness of outcome-adaptive randomization. *Clinical Cancer Research*, 18(17):4498–4507, 2012.

[82] JT. Connor, JJ. Elm, KR. Broglio, ESETT, and ADAPT-IT Investigators. Bayesian adaptive trials offer advantages in comparative effectiveness trials: an example in status epilepticus. *Journal of Clinical Epidemiology*, 66:130–137, 2013.

[83] JJ. Lee and D.D. Liu. A predictive probability design for phase II cancer clinical trials. *Clinical Trials*, 5(2):93–106, 2008.

[84] M. Zelen. A new design for randomized clinical trials. *New England Journal of Medicine*, 300:1242–1246, 1979.

[85] SP. Hey and J. Kimmelman. Are outcome-adaptive allocation trials ethical? *Clinical Trials*, 12(2):102–106, 2015.

[86] S. Joffe and SS. Ellenberg. Commentary on hey and kimmelman. *Clinical Trials*, 12(2):116–118, 2015.

[87] E. Korn and B. Freidlin. Commentary on hey and kimmelman. *Clinical Trials*, 12(2):122–124, 2015.

[88] Berry DA. The brave new world of clinical cancer research: adaptive biomarker-driven trials integrating clinical practice with clinical research. *Molecular Oncology*, 9(5):951–959, 2015.

[89] JJ. Lee. Commentary on hey and kimmelman. *Clinical Trials*, 12(2):110–112, 2015.

[90] SB. Saxman. Commentary on hey and kimmelman. *Clinical Trials*, 12(2):113–115, 2015.

[91] V. Grunwald and M. Hidalgo. Developing inhibitors of the epidermal growth factor receptor for cancer treatment. *Journal of National Cancer Institute*, 95:851–867, 2003.

[92] SJ. Diaz-Cano. Tumor heterogeneity: Mechanisms and bases for a reliable application of molecular marker design. *International Journal of Molecular Sciences*, 13(2):1951–2011, 2012.

[93] R. Fisher, L. Pusztai, and C. Swanton. Cancer heterogeneity: implications for targeted therapeutics. *British Journal of Cancer*, 108(3):479–485, 2013.

[94] CE. Meacham and SJ. Morrison. Tumor heterogeneity and cancer cell plasticity. *Nature*, 501(7467):328–337, 2013.

[95] AG. Rivenbark, SM. O'Connor, and WB. Coleman. Molecular and cellular heterogeneity in breast cancer: challenges for personalized medicine. *The American Jounal of Pathology*, 183(4):1113–1124, 2013.

[96] LG. Martelotto, CK. Ng, S. Piscuoglio, B. Weigelt, and JS. Reis-Filho. Breast cancer intra-tumor heterogeneity. *Breast Cancer Research*, 16(3):210, 2014.

[97] X. Sun and Q. Yu. Intra-tumor heterogeneity of cancer cells and its implications for cancer treatment. *Acta Pharmacologica Sinica*, 36(10):1219–1227, 2015.

[98] P. Gallo, C. Chuang-Stein, V. Dragalin, B. Gaydos, M. Krams, and J. Pinheiro. Adaptive design in clinical drug development - an executive summary of the phrma working group (with discussions). *Journal of Biopharmaceutical Statistics*, 16(3):275–283, 2006.

[99] B. Efron. Forcing a sequential experiment to be balanced. *Biometrika*, 58:403–417, 1971.

[100] JM. Lachin. Statistical properties of randomization in clinical trials. *Controlled Clinical Trials*, 9:289–311, 1988.

[101] W. Lehmacher and G. Wassmer. Adaptive sample size calculations in group sequential trials. *Biometrics*, 55:1286–1290, 1999.

[102] LD. Kaiser. Dynamic randomization and a randomization model for clinical trials data. *Statistics in Medicine*, 31(29):3858–3873, 2012.

[103] BP. Hobbs, BP. Carlin, and DJ. Sargent. Adaptive adjustment of the randomization ratio using historical control data. *Clinical trials (London, England)*, 10(3):430–440, 2013.

[104] ZS. Hoare, CJ. Whitaker, and R. Whitaker. Introduction to a generalized method for adaptive randomization in trials. *Trials*, 10:14–19, 2013.

[105] Y. Du, X. Wang, and JJ. Lee. Simulation study for evaluating the performance of response-adaptive randomization. *Contemporary clinical trials*, 40:15–25, 2015.

[106] L. Cui, HMJ. Hung, and SJ. Wang. Modification of sample size in group sequential trials. *Biometrics*, 55:853–857, 1999.

[107] WJ. Shih. Group sequential, sample size re-estimation and two-stage adaptive designs in clinical trials: a comparison. *Statistics in Medicine*, 25:933–941, 2006.

[108] E. Hade, D. Jarjoura, and L. Wei. Sample size re-estimation in a breast cancer trial. *Clinical trials (London, England)*, 7(3):219–226, 2010.

[109] W. Zhong, JS. Koopmeiners, and BP. Carlin. A two-stage bayesian design with sample size reestimation and subgroup analysis for phase II binary response trials. *Contemporary clinical trials*, 36(2):1010–1016, 2013.

[110] J. Bowden and A. Mander. A review and re-interpretation of a group-sequential approach to sample size re-estimation in two-stage trials. *Pharmaceutical Statistics*, 13(3):163–172, 2014.

[111] T. Wang, X. Wang, H. Zhou, J. Cai, and S. L. George. Auxiliary-variable-enriched biomarker-stratified design. *Statistics in Medicine*, 37:4610–4635, 2018.

[112] XF. Wang, JZ. Zhou, T. Wang, and S. L. George. On enrichment strategies for biomarker stratified clinical trials. *Journal of Biopharmaceutical Statistics*, 28:292–308, 2018.

[113] SK. Wang and AA. Tsiatis. Approximately optimal one-parameter boundaries for a sequential trials. *Biometrics*, 43(1):193–200, 1987.

[114] M. Posch and P. Bauer. Adaptive two-stage designs and the conditional error function. *Biometrical Journal*, 41:689–696, 1999.

[115] C. Jennison and BW. Turnbull. *Group Sequential Methods with Applications to Clinical Trials*. Chapman and Hall, 2000.

[116] Q. Liu, MA. Proschan, and GW. Pledger. A unified theory of two-stage adaptive designs. *Journal of American Statistical Association*, 97:1034–1041, 2002.

[117] C. Jennison and BW. Turnbull. Meta-analysis and adaptive group sequential design in the clinical development process. *Journal of Biopharmaceutical Statistics*, 15(4):537–558, 2005.

[118] S. Urach and M. Posch. Multi-arm group sequential designs with a simultaneous stopping rule. *Statistics in Medicine*, 35(30):5536–5550, 2016.

[119] S. Schüler, M. Kieser, and G. Rauch. Choice of futility boundaries for group sequential designs with two endpoints. *BMC Medical Research Methodology*, 17:119, 2017.

[120] TG. Karrison, D. Huo, and R. Chappell. A group sequential, response-adaptive design for randomized clinical trials. *Control Clinical Trials*, 24:506–522, 2003.

[121] DA. Berry and SG. Eick. Adaptive assignment versus balanced randomization in clinical trials: a decision analysis. *Statistics in Medicine*, 14:231–246, 1995.

[122] YK. Cheung, LYT. Inoue, JK. Wathen, and PF. Thall. Continuous bayesian adaptive randomization based on event times with covariates. *Statistics in Medicine*, 25:55–70, 2006.

[123] CJ. Weir and KR. Lees. Comparison of stratification and adaptive methods for treatment allocation in an acute stroke clinical trial. *Statistics in Medicine*, 22:705–726, 2003.

[124] PF. Thall and JK. Wathen. Covariate-adjusted adaptive randomization in a sarcoma trial with multi-stage treatments. *Statistics in Medicine*, 24:1947–1964, 2005.

[125] X. Gu and JJ. Lee. A simulation study for comparing testing statistics in response-adaptive randomization. *BMC Medical Research Methodology*, pages 10–48, 2010.

[126] Y. Yuan, X. Huang, and S. Liu. A bayesian response-adaptive covariate-

balanced randomization design with application to a leukemia clinical trial. *Statistics in Medicine*, 30(11):1218–1229, 2011.

[127] M-O. Kim, C. Liu, F. Hu, and JJ. Lee. Outcome-adaptive randomization for a delayed outcome with a short-term predictor: imputation-based designs. *Statistics in Medicine*, 33(23):4029–4042, 2014.

[128] A. Gelman, JB. Carlin, HSA. Stern, and DB. Rubin. *Bayesian data analysis, 2nd ed.* Chapman and Hall/CRC, 2004.

[129] PF. Thall, LYT. Inoue, and T. Martin. Adaptive decision making in a lymphocyte infusion trial. *Biometrics*, 58:560–568, 2002.

[130] RG. Maki, ML. Hensley, JK. Wathen, et al. A sarc multicenter phase III study of gemcitabine vs. gemcitabine and docetaxel in patients with metastatic soft tissue sarcomas. *Journal of Clinical Oncology*, 24(18S):9514–9514, 2006.

[131] SJ. Mandrekar and DJ. Sargent. Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges. *Journal of Clinical Oncology*, 27(24):4027–4034, 2009.

[132] JJ. Lee, Xuemin. G, and Suyu. Liu. Bayesian adaptive randomization designs for targeted agent development. *Clinical trials (London, England)*, 7(5):584–596, 2010.

[133] X. Zhou, S. Liu, ES. Kim, RS. Herbst, and JJ. Lee. Bayesian adaptive design for targeted therapy development in lung cancer - a step toward personalized medicine. *Clinical trials (London, England)*, 5(3):181–193, 2008.

[134] DA. Berry, RS. Herbst, and EH. Rubin. Reports from the 2010 clinical and translational cancer research think tank meeting: design strategies for per-

sonalized therapy trials. *Clinical cancer research: an official journal of the American Association for Cancer Research*, 18(3):638–644, 2012.

[135] KA. Gold, ES. Kim, JJ. Lee, II. Wistuba, CJ. Farhangfar, and WK. Hong. The battle to personalize lung cancer prevention through reverse migration. *Cancer prevention research (Philadelphia, PA)*, 4(7):962–972, 2011.

[136] TL. Lai, PW. Lavori, M-CI. Shih, and BI. Sikic. Clinical trial designs for testing biomarker-based personalized therapies. *Clinical trials (London, England)*, 9(2):141–154, 2012.

[137] M. Buyse, S. Michiels, DJ. Sargent, A. Grothey, A. Matheson, and A. de Gramont. Integrating biomarkers in clinical trials. *Expert review of molecular diagnostics*, 11(2):171–182, 2011.

[138] E. Galanis, W. Wu, J. Sarkaria, SM. Chang, H. Colman, D. Sargent, and others. Incorporation of biomarker assessment in novel clinical trial designs: personalizing brain tumor treatments. *Current oncology reports*, 13(1):42–49, 2011.

[139] S. L. George. Statistical issues in translational cancer research. *Clinical cancer research*, 14(19):5954–5958, 2008.

[140] S. Matsui, M. Buyse, Simon R., and Eds. *Design and analysis of clinical trials for predictive medicine*. New York, Chapman and Hall/CRC, 2015.

[141] Q. Shi, SJ. Mandrekar, and DJ. Sargent. Predictive biomarkers in colorectal cancer: usage, validation, and design in clinical trials. *Scandinavian journal of gastroenterology*, 47(3):356–362, 2012.

[142] JJ. Chen, T-P. Lu, D-T. Chen, and S-J. Wang. Biomarker adaptive designs in clinical trials. *Translational Cancer Research*, 3(3):279–292, 2014.

[143] R. Simon. Clinical trial designs for evaluating the medical utility of prognostic and predictive biomarkers in oncology. *Personalized medicine*, 7(1):33–47, 2010.

[144] TS. Mok, YL. Wu, S. Thongprasert, CH. Yang, DT. Chu, N. Saijo, P. Sunpaweravong, B. Han, B. Margono, Y. Ichinose, Y. Nishiwaki, Y. Ohe, JJ. Yang, B. Chewaskulyong, H. Jiang, EL. Duffield, CL. Watkins, AA. Armour, and M. Fukuoka. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *New England Journal of Medicine*, 361:947–957, 2009.

[145] Z.W. Zhang, M.J. Li, M. Lin, G. X. Soon, T. Greene, and C. Y. Shen. Subgroup selection in adaptive signature designs of confirmatory clinical trials. *Journal of the Royal Statistical Society Series C - Applied Statistics*, 66(2):345–361, 2017.

[146] RM. Simon, EL. Korn, LM. McShane, MD. Radmacher, GW. Wright, and Y. Zhao. *Design and Analysis of DNA Microarray Investigations*. Springer, 2004.

[147] Z. W. Zhang, R. Z. Chen, G. X. Soon, and H. Zhang. Treatment evaluation for a data-driven subgroup in adaptive enrichment designs of clinical trials. *Statistics in Medicine*, 37(1):1–11, 2018.

[148] T. C. Qian, E. Colantuoni, A. Fisher, and M. Rosenblum. Sensitivity of adaptive enrichment trial designs to accrual rates, time to outcome measurement, and prognostic variables. *Contemporary Clinical Trials Communications*, 8:39–48, 2017.

[149] S. Matsui and J. Crowley. Biomarker-stratified phase III clinical trials: En-

hancement with a subgroup-focused sequential design. *Clinical Cancer Research*, 24(5):994–1001, 2018.

[150] N. Simon. Adaptive enrichment designs: applications and challenges. *Clinical Investigation (London)*, 5(4):383–391, 2015.

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

[152] J.J. Lee, X. Gu, and S. Liu. Bayesian adaptive randomization designs for targeted agent development. *Clinical trials (London, England)*, 7(5):584–596, 2010.

[153] J. Sim. Outcome-adaptive randomization in clinical trials: issues of participant welfare and autonomy. *Theoretical Medicine and Bioethics*, 40(2):83–101, 2019.

[154] X. Huang, S. Biswas, Y. Oki, J. Issa, and DA. Berry. A parallel phase I/II clinical trial design for combination therapies. *Biometrics*, 63:429–436, 2007.

[155] J.K. Wathen and P. Thall. A simulation study of outcome adaptive randomization in multi-arm clinical trials. *Clinical Trials*, 14(5):432 – 440, 2017.

[156] S-C. Chow and J-P. Liu. *Design and analysis of clinical trials, 2nd ed.* John Wiley & Sons, Inc., 2004.

[157] S. L. George, X. Wang, H. Pang, and Eds. *Cancer clinical trials*. New York, Chapman and Hall/CRC, 2016.

[158] JW. Lee, JB. Lee, M. Park, and others. An extensive comparison of recent classification tools applied to microarray data. *Computational Statistics & Data Analysis*, 48:869–885, 2005.

[159] H. Moon, H. Ahn, RL. Kodell, and others. Ensemble methods for classification of patients for personalized medicine with high-dimensional data. *Artificial Intelligence in Medicine*, 41:197–207, 2007.

[160] R. Simon. Development and validation of biomarker classifiers for treatment selection. *Journal of statistical planning and inference*, 138(2):308–320, 2008.

[161] S. Baek, CA. Tsai, and JJ. Chen. Development of biomarker classifiers from high-dimensional data. *Brief Bioinform*, 10:537–546, 2009.

[162] L. Breiman. Random forest. *Machine Learning*, 45:5–32, 2001.

[163] V. Vladimir. *The nature of statistical learning theory*. Springer, 1995.

[164] S Dudoit, J Fridlyand, and T. Speed. Comparison of discrimination methods for the classification of tumors using gene expression data. *Journal of the American Statistical Association*, 97:77–87, 2002.

[165] Jochen Kruppa, Yufeng Liu, Gèrard Biau, Michael Kohler, Inke R. König, James D. Malley, and Andreas Ziegler. Probability estimation with machine learning methods for dichotomous and multicategory outcome: Theory. *Biometrical Journal*, 56(4):534–563, 2014.

[166] Jochen Kruppa, Yufeng Liu, Gèrard Biau, Michael Kohler, Inke R. König, James D. Malley, and Andreas Ziegler. Probability estimation with machine learning methods for dichotomous and multicategory outcome: Applications. *Biometrical Journal*, 56(4):564–583, 2014.

[167] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning, second edition*. Springer, 2009.

[168] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regressin Trees*. Wadsworth, 1984.

[169] A. Midha, S. Dearden, and R. McCormack. Egfr mutation incidence in non-small-cell lung cancer of adenocarcinoma histology: a systematic review and global map by ethnicity (mutmapII). *American Journal of Cancer Research*, 5:2892–2911, 2015.

[170] B. Freidlin, LM. McShane, and EL. Korn. Randomized clinical trials with biomarkers: design issues. *Journal of the National Cancer Institute*, 102:152–160, 2010.

[171] R. Uozumi and C. Hamada. Interim decision-making strategies in adaptive designs for population selection using time-to-event endpoints. *Journal of Biopharmaceutical Statistics*, 27(1):84–100, 2017.

[172] J. Fridlyand, RM. Simon, JC. Walrath, N. Roach, R. Buller, DP. Schenkein, KT. Flaherty, JD. Allen, EV. Sigal, and HI. Scher. Considerations for the successful co-development of targeted cancer therapies and companion diagnostics. *Nature Reviews Drug Discovery*, 12(10):743–755, 2013.

[173] E. Joffe, A. Lasonos, and A. Younes. Clinical trials in the genomic era. *Journal of Clinical Oncology*, 35(9):1011–1017, 2017.

[174] AJ. Redig and PA. Jänne. Basket trials and the evolution of clinical trial design in an era of genomic medicine. *Journal of Clinical Oncology*, 33(9):975–977, 2015.

[175] BR. Saville and SM. Berry. Efficiencies of platform clinical trials: a vision of the future. *Clinical Trials*, 13(3):358–366, 2016.

[176] L. A. Renfro and D. J. Sargent. Statistical controversies in clinical research: basket trials, umbrella trials, and other master protocols: a review and examples. *Annals of Oncology*, 28(1):34–43, 2017.

[177] R. S. Herbst, D. R. Gandara, F. R. Hirsch, M. W. Redman, M. LeBlanc, P. C. Mack, L. H. Schwartz, E. Vokes, S. S. Ramalingam, J. D. Bradley, D. Sparks, Y. Zhou, C. Miwa, V. A. Miller, R. Yelensky, Y. Li, J. D. Allen, E. V. Sigal, D. Wholley, C. C. Sigman, G. M. Blumenthal, S. Malik, G. J. Kelloff, J. S. Abrams, C. D. Blanke, and V. A. Papadimitrakopoulou. Lung master protocol (lung-map)–biomarker-driven protocol for accelerating development of therapies for squamous cell lung cancer: Swog s1400. *Clinical Cancer Research*, 21(7):1514–1524, 2015.

[178] CE. Steuer, V. Papadimitrakopoulou, RS. Herbst, MW. Redman, FR. Hirsch, PC. Mack, SS. Ramalingam, and DR. Gandara. Innovative clinical trials: the lung-map study. *Clinical Pharmacology & Therapeutics*, 97(5):488–491, 2015.

[179] M. W. Redman and C. J. Allegra. The master protocol concept. *Seminars in Oncology*, 42(5):724–730, 2015.

[180] D. Kazandjian, DL. Suzman, G. Blumenthal, S. Mushti, K. He, M. Libeg, P. Keegan, and R. Pazdur. Fda approval summary: Nivolumab for the treatment of metastatic non-small cell lung cancer with progression on or after platinum-based chemotherapy. *Oncologist,*, 21(5):634–642, 2016.

[181] DE. Gerber, GR. Oxnard, and R. Govindan. Alchemist: bringing genomic discovery and targeted therapies to early-stage lung cancer. *Clinical Pharmacology & Therapeutics*, 97:447–450, 2015.

[182] B. Freidlin and EL. Korn. Biomarker enrichment strategies: matching trial design to biomarker credentials. *Nature reviews: Clinical oncology*, 11(2):81–90, 2014.

[183] RK. Kaplan, TM. Maughan, AC. Crook, DF. Fisher, RW. Wilson, LC. Brown, and MP. Parmar. Evaluation of many treatments and biomarkers in oncology: a new design. *Journal of Clinical Oncology*, 31(36):4562–4568, 2013.

[184] R. Adams, E. Brown, and et al. Brown, L. and. Inhibition of egfr, her2, and her3 signalling in patients with colorectal cancer wild-type for braf, pik3ca, kras, and nras (focus4-d): a phase 2-3 randomised trial. *Lancet Gastroenterology & Hepatology*, 3(3):162–171, 2018.

[185] K. M. Cunanan, M. Gonen, R. Shen, D. M. Hyman, G. J. Riely, C. B. Begg, and A. Iasonos. Basket trials in oncology: A trade-off between complexity and efficiency. *Journal of Clinical Oncology*, 35(3):271–273, 2017.

[186] Y. Chu and Y. Yuan. A bayesian basket trial design using a calibrated bayesian hierarchical model. *Clinical trials (London, England)*, 15(2):149–158, 2018.

[187] SM. Berry, KR. Broglio, S. Groshen, and DA. Berry. Bayesian hierarchical modeling of patient subpopulations: efficient designs of phase II oncology clinical trials. *Clinical Trials*, 10(5):149–158, 2013.

[188] R. Simon, S. Geyer, J. Subramanian, and S. Roychowdhury. The bayesian basket design for genomic variant-driven phase II trials. *Seminars in Oncology*, 43(1):13–18, 2016.

[189] C. Chen, N. Li, S. Yuan, Z. Antonijevic, R. Kalamegham, and Robert A. B. Statistical design and considerations of a phase 3 basket trial for simultaneous

investigation of multiple tumor types in one study. *Statistics in Biopharmaceutical Research*, 8:248–257, 2016.

[190] Food and Drug Administration. *Guidance for Industry: Multiple Endpoints in Clinical Trials*. U.S. Department of Health and Human Services Food and Drug Administration, 2017.

[191] HM. Hung and SJ. Wang. Some controversial multiple testing problems in regulatory applications. *journal of Biopharmaceutical Research*, 19(1):1–11, 2009.

[192] A. Dmitrienko, AC. Tamhane, and F. Bretz. *Multiple Testing Problems in Pharmaceutical Statistics*. Chapman & Hall/CRC, 2010.

[193] RL. Berger. Multiparameter hypothesis testing and acceptance sampling. *Technlabelmetrics*, 24:295–300, 1982.

[194] T. Hamasaki, K. Asakura, T. Ochiai, and SR. Evans. *Group-Sequential Clinical Trials with Multiple Co-Objectives*. Cham/Heidelberg/ New York: Springer, 2016.

[195] T. Sozu, T. Sugimoto, and T. Hamasaki. Sample size determination in clinical trials with multiple co-primary binary endpoints. *Statistics in Medicicine*, 29(21):2169–79, 2010.

[196] T. Sozu, T. Sugimoto, and T. Hamasaki. Sample size determination in clinical trials with multiple co-primary endpoints including mixed continuous and binary variables. *Biometrical Journal*, 54:716–729, 2012.

[197] T. Sugimoto, T. Sozu, T. Hamasaki, and SR. Evans. A logrank test-based

method for sizing clinical trials with two co-primary time-to-event endpoints. *Biostatistics*, 14(3):409–421, 2013.

VITA

Fang Xia was born in Chongqing Shi, China. After graduating from the Chongqing Yucai secondary school, she went to the United States to pursue her bachelor's degree in 2008. In May of 2012, she received a B.S. degree in Mathematics with a minor in Statistics from the University of Iowa. Then she joined Duke University and received a master's degree in Biostatistics in May of 2014. In August of 2015, she entered the University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences for the Ph.D. degree in Biostatistics.