2013

Response Adaptive Design using Auxiliary and Primary Outcomes

Shuxian Sinks
Virginia Commonwealth University

Response Adaptive Design using Auxiliary and Primary Outcomes

Dissertation submitted to the Committee of the

Virginia Commonwealth University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Biostatistics

Shuxian Z. Sinks

Roy T. Sabo, Ph.D.

Nitai D. Mukhopadhyayi, Ph.D.

R.K. Elswick, Jr., Ph.D.

Karl E. Peace, Ph.D.

Amir Toor, M.D.

Virginia Commonwealth University

Richmond, Virginia

November, 2013

Acknowledgements

I would like to express my sincere appreciation to my co-advisors, Dr. Roy Sabo and Dr. Nitai D. Mukhopadhyayi for their encouragement and guidance of this dissertation. I am grateful for all their precious time that they have contributed in weekly meeting to discuss and review the methodology and simulation results. I am also thankful to the member of my dissertation committee, Dr. R.K. Elswick, Dr. Karl Peace, and Dr. Amir Toor for their commitment to this dissertation. To Dr. R.K. Elswick, thanks for providing me the opportunity to work at School of Nursing, VCU. I have learned many skills and gained experience to work with clinicians in the past 4 years. I also appreciate the kindness, support, and encouragement from the rest of faculty and my classmates. To my husband Michael, the Sinks family, and my family in China, I am tremendously grateful for the patience, support, understanding, and love.

Response Adaptive Design Using Auxiliary and Primary Outcomes

Shuxian Z. Sinks

(ABSTRACT)

Response adaptive designs intend to allocate more patients to better treatments without undermining the validity and the integrity of the trial. The immediacy of the primary response (e.g. deaths, remission) determines the efficiency of the response adaptive design, which often requires outcomes to be quickly or immediately observed. This presents difficulties for survival studies, which may require long durations to observe the primary endpoint. Therefore, we introduce auxiliary endpoints to assist the adaptation with the primary endpoint, where an auxiliary endpoint is generally defined as any measurement that is positively associated with the primary endpoint. Our proposed design (referred to as bivariate adaptive design) is based on the classical response adaptive design framework. The connection of auxiliary and primary endpoints is established through Bayesian method. We extend parameter space from one dimension to two dimensions, say primary and auxiliary efficacies, by implementing a conditional weigh function on the loss function of the design. The allocation ratio is updated at each stage by optimization of the loss function subject to the information provided for both the auxiliary and primary outcomes. We demonstrate several methods of joint modeling the auxiliary and primary outcomes. Through simulation studies, we show that the bivariate adaptive design is more effective in assigning patients to better treatments as compared with univariate optimal and balanced designs. As hoped, this joint-approach also reduces the expected number of patient failures and preserves the comparable power as compared with other designs.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 What is adaptive design

Adaptive designs allow for the modification of clinical trials based on accumulating data, without undermining the validity and the integrity of the trial (Chow et al., 2005). The modifications include, but are not limited to: stopping a trial early due to safety, efficacy or futility based on interim analysis, sample size reestimation, and dropping inferior groups (Gallo et al., 2006). These designs are critical in the clinical development of new drugs in early phases, which involves long-term commitments in time and economic costs. The designs are flexible to be modified based on collected data, through such the modifications and adaptations have to be thoughtfully planned in randomization designs.

Adaptive design can be categorized based on four different rules: randomization rule, sampling rule, stopping rule, and decision rule (Mahajan and Gupta, 2010). Randomization rule intends to allocate more patients to better treatment by changing the probabilities of assigning patient in each arm. It consists of response-adaptive design and covariate-adaptive allocation. The sampling rule defines how many patients will be sampled at

the next stage to detect treatment difference. It includes sample-size reestimation and drop-the-loser design. The stopping rule intends to stop the trial when efficacy or futility of the treatment is obvious. It comprises group sequential design and adaptive treatment-switching design. The decision rule refers to the changes different from the other three rules, such as hypothesis-adaptive design, change the primary end-point design etc.

1.1.1 Response adaptive design

Response-adaptive designs are a common adaptive randomization procedure, which allows modification of the allocation rate and schemes for placing more patients on the better treatment based on patient responses. It may be used to allocate more patients to a better treatment for Phase III clinical trials when the sample size is large. Thompson (1933) has introduced the idea of assigning patients to the better treatment by adapting the allocation rate, if there exits a preference between the two treatments. He also uses bayes probabilities to judge between the rival two treatments, where non-informative priors are applied and their interval lie in the possible range $(0, 1)$. Several response-adaptive designs have subsequently been developed for binary responses in controlled clinical trials, such as the play-the-winner rule, the randomized play-the-winner rule, the drop-the-loser rule, optimal adaptive rule etc. Our response-adaptive design is an extension of the optimal response-adaptive design proposed by Rosenberger, Stallard, Ivanova, Harper and Ricks (RSIHR) (Rosenberger et al., 2001). The optimal allocation rate changes in order to meet certain objectives, such as to minimize number of patients to inferior treatment or to maximize the power and ensure sufficient sample size to the trial. The sample size is predetermined by the design according to the objectives, not resulting from an *ad hoc*

basis. Neyman allocation maximizes the power of the test of comparing two binomial probabilities, where its allocation rate of assigning to treatment A takes the form of $\frac{\sqrt{p_A q_A}}{\sqrt{p_A q_A}+\sqrt{p_B q_B}}$, the ratio of standard deviations. Optimal response adaptive designs should consider the statistical power and the number of patients assigned to inferior treatment. Neyman allocation does not perform well out of ethical consideration when $p_A + p_B > 1$, because it assigns more patient to the inferior treatment. However, RSIHR allocation (called optimal allocation) minimizes the expected number of patient deaths (or treatment failures) and maintains comparable statistical power as balanced randomization and Neyman allocation. Its allocation rate, instead, takes the form of $\frac{\sqrt{p_A}}{\sqrt{p_A}+\sqrt{p_B}}$.

1.2 Motivations

Treatment success rates, p_A and p_B both are unknown parameters and estimated based on the data. Therefore, the optimal procedure becomes effective when the data can be rapidly collected and do not require long-term follow-up. Unfortunately, quick availability of patient responses is one obstacle holding researchers from applying these methods in practice. It is common that no observations are observed at the early stage of a clinical trial. Then, the allocation rate will not be updated and stays at the specified initial leading probability, which undermines the purpose of the design. Ultimately, the slow adaptation due to slowly observed responses does not benefit the randomization. To alleviate this drawback, Rosenberger and Lachin (1993) suggested to adapt the allocation ratio after groups of patients instead of after each individual patient. The solution is not quite promising when the flow of enrollment is faster than the flow of data collection, particularly when the primary outcome is delayed and rarely observed.

We are motivated to introduce a response-adaptive allocation method that incorporates auxiliary outcome, which is hopefully correlated with the primary outcome. This is different from biomarker-adaptive designs, where the adaptation merely depends on the response of biomarker (eg. genomic markers). Also, the auxiliary endpoint is not a surrogate endpoint as defined by Prentice (1989) to directly substitute primary endpoint. The auxiliary endpoint is defined as a response variable that can strengthen primary endpoint analysis and frequently be observed (Fleming et al., 1994). Usually, we do not need to make any additional assumptions on the auxiliary outcome to use it in the analysis. The way of implementing the auxiliary outcome without ignoring the primary outcome information becomes critical to increase the flexibility of response-adaptive design on data collection. The essence of response adaptive design clearly states a goal or objective to minimize the total loss or failure of patients in each treatment group. Inspired by this characteristic, we extend the parameter space to dimensions, say primary and auxiliary efficacies, by implementing a conditional weight function to describe the loss. Bayesian analysis naturally suits for this purpose, as it easily incorporates observed responses and historical information about the parameters of interest in the adaptation.

Intuitively, there are two approaches to joint modeling auxiliary and primary outcomes using Bayesian methods. One approach is to construct joint likelihood and have independent priors for each parameter of interest. A bivariate binomial distribution would be naturally chosen for the joint likelihood of two binomial random variables. Several bivariate binomial distributions have been formulated and studied in the literature. Aitken and Gonin (1935) derived a bivariate binomial distribution with fixed equal marginals and different binomial probabilities, where the equal marginals limits its application. Hamdan (1972) derived a canonical form of bivariate binomial distribution with unequal

marginals but equal binomial probabilities, however, it is not applicable for our scenario. Marshall and Olkin (1985) discussed a bivariate binomial distribution based on a bivariate Bernoulli distribution, where the distribution is originally found by Teicher (1954). Hamdan and Jensen (1976) pointed out that this distribution is a bivariate binomial under certain conditions and can be illustrated as the case when the marginals of a 2*2 contingency table are observed but the cell counts are not available. Unfortunately, these distributions are in a complicated form and their properties have not been profoundly studied.

An alternative approach is to have independent likelihoods for each outcome and construct a joint prior for the parameters of interest. Therefore, we need a bivariate prior distribution to describe the association of the parameters. In one dimensional cases, we know that the beta distribution is the natural conjugate prior distribution for that binomial likelihood. Extending to two dimensional cases, a bivariate beta distribution might be an appropriate choice to consider. Gupta and Wong (1985) have studied three-parameter and five-parameter bivariate beta distributions. The three-parameter distribution is a bivariate Dirichlet distribution and its correlation is always negative. However, our aim is to find a bivariate distribution to model the positive correlation between two proportions. The five-parameter bivariate beta distribution is a Morgenstern type (Morgenstern, 1956) bivariate distribution. Unlike three-parameter distribution, it is not restricted to negative correlations. Also, the extra parameters allow more flexibility in application. However, it has been shown that the correlation of Morgenstern type bivariate distributions has a small maximal range from $-\frac{1}{3}$ and $\frac{1}{3}$ (Schucany et al., 1978). Olkin and Liu (2003) have constructed a three-parameter distribution based on gamma distributions with desirable properties. For instance, its expanded power series is related to the hypergeometric func-

tion, and the density is positively likelihood ratio dependent.

The purpose of presenting the distribution is to have it server as prior on the support $0 \leq x_i \leq 1$ for positively correlated parameters of the binomial distribution. In this trend, we want to construct a bayesian model where the association of the outcomes is solely through that of the corresponding parameters, assuming the two outcomes are independent given the parameters. This is the main model we will use during the entire research.

1.3 Prospectus

In the next chapters, we aim to solve the following problems:

1.3.1 Ranges of correlations using three- and five-parameter bivariate beta models in case of informative prior specification

Olkin and Liu have shown that the three-parameter bivariate distribution derived from gamma distributions is quite flexible on positive correlation when the marginals are not fixed. In bayesian analysis, we have it serve as the prior distribution for the correlated binomial random variables, and prior information usually places on the marginals and correlations. Therefore, the flexibility of the distribution with fixed marginals on correlation is critical to the analysis. In this chapter, we compare the performance of Olkin and Liu's bivariate beta distribution and Morgenstern type bivariate beta distribution on correlation when their marginals are fixed. We show that the correlation of Olkin and Liu's bivariate beta distribution with fixed marginals fails in a formulated bound-

ary. Using analytic methods, it has been shown that the correlation boundary changes corresponding to the marginals. Morgenstern type bivariate beta distributions with fixed marginals are always restricted in range of $[-\frac{1}{3}, \frac{1}{3}]$. For certain conditions, Olkin and Liu's bivariate beta distribution attains higher correlation than Morgenstern type bivariate beta distribution. For the sake of a relatively flexible correlation range, we have chosen the three-parameter bivariate beta distribution proposed by Olkin and Liu to be the prior distribution for our bivariate adaptive desgin model, even though its correlation has a limited range.

1.3.2 Response adaptive allocation using auxiliary and primary outcomes for binary outcomes in case of the two treatment groups

The bivariate optimal adaptive design framework is developed based on the optimal adaptive for binary outcomes (Rosenberger et al., 2001) in this chapter. We refer to the optimal adaptive proposed by Rosenberger et al. as the univariate optimal adaptive design in the following chapters. In order to implement the auxiliary outcome in the adaptation, we assume that auxiliary and primary outcomes follow binomial distribution, and the parameters of these distributions are also random variables and jointly distributed in the three-parameter bivariate beta distribution presented in Chapter 2. Also, we assume that the correlation of auxiliary and primary outcomes is predetermined regardless of treatment. Through some theoretical derivations, the bivariate optimal allocation depends on the ratio of the expectations of posterior conditional distribution of the primary parameter given the auxiliary parameter for both treatment groups, where the expectations

are related to the hypergeometric function. The univariate optimal design realizes the minimization of expected number of patient failures. The bivariate optimal design minimizes the risk of patient failures given auxiliary outcome information, which is different from the goal of the univariate optimal desgin. We perform simulation study to compare bivariate optimal adaptive designs to the univariate optimal adaptive and balanced randomization designs, in terms of expected number of patient failures, allocation ratios, power/error rate, and number of patients assigned to each treatment group. Then, we also have discussed how the correlation of auxiliary and primary outcomes affects the simulation results of the bivariate optimal adaptive design.

1.3.3 Applying copula method for constructing a bivariate distribution with beta marginals as prior distribution

We have noticed that the three-parameter bivariate beta distribution with fixed marginals is limited in its range of correlation. In order to overcome this obstacle, we use copula method to construct a bivariate distribution with beta marginal distributions as the prior distribution for the proposed bivariate optimal adaptive design. Sklar (1973) has stated that any multivariate joint distribution can be written as the combination of several univariate marginal distributions and a copula function which describe the dependence structure among random variables. Therefore, the copula method allows us to separately model the dependence of two random binomial variables and their marginal distributions. In this chapter, we disccuss the simulation results of the bivariate optimal method using copula-based prior and studied the performance of the adaptation using posterior variance and sample variance in the procedure. Different copula functions suit different

types of data, and two major copula functions are commonly used: elliptical copulas and Archimedean copulas. We have simulated the bivariate optimal design using Gaussian and Clayton copula-based prior distributions separately, which are representative of the two major copulas. By the mean of changing the dependence parameter of the copula, we intend to examine how the procedure is affected by the dependece of two random variables.

Chapter 2

Ranges of Correlations using Bivariate Beta Models for Informative Prior Specification (written as manuscript)

2.1 Introduction

The beta distribution is a natural-conjugate prior for the binomial likelihood, where the posterior expectation is a linear combination of data and prior parameters (Diaconis and Ylvisaker, 1979). The constructed correlated binomial distribution is complicated and not unique, because it can be formed from different directions to describe the association of two binomial distribution (Biswas and Hwang, 2002). Even though finding conjugate priors for two correlated binomial variables is challenging, the bivariate beta distribution can be used as prior for correlated binomial random variables (Pham and Turkkan, 1992). Even though it is not conjugate with correlated binomial variables, the computational problem can be handled through Bayesian sampling methods.

A few bivariate beta distributions have been proposed in the statistical literature, differing mainly in the number of parameters and their attainable correlation ranges. Gupta and Wong (1985) studied the five-parameter bivariate beta distribution, which is derived from the Morgenstern-system of curves. However, the allowable correlations resulting from this distribution are restricted (Schucany et al., 1978). Gupta and Wong also discussed a three-parameter bivariate beta distribution known as the Dirichlet distribution, where the correlations here are restricted to negative values. Jones (2002) first provided another three-parameter bivariate beta distribution through the links between beta and F distributions. Olkin and Liu (2003) proposed an alternative way of generating the exact formulation of this three-parameter bivariate beta distribution by using gamma distributions.

Though in their original contexts, these models allow for five and three "free" parameters to be specified, their use as informative priors with "known" marginal values reduces the numbers of "free" parameters. This has the unintended effect of reducing the range of

"allowable" correlation values. In this paper, we compare the attainable correlation range of Morgenstern-type five-parameter bivariate beta distribution with three-parameter bivariate beta distribution introduced by Liu and Olkin under fixed marginal means.

2.2 3 and 5- parameter bivariate beta distribution

Suppose we have two correlated clinical outcomes, say X_1 and X_2. $X_1|b_1$ follows Bin(n_1, b_1), where n_1 is number of patients and b_1 is the event rate in the study of X_1 ; $X_2|b_2$ follows Bin(n_2, b_2), where n_2 is number of patients and b_2 is the event rate in the study of X_2 . Both distributions of X_1 and X_2 depend on distributions of b_1 and b_2, where (b_1, b_2) jointly follow a bivariate beta distribution. Note that we could model b_1 and b_2 separately. Though we would then need a joint distribution for X_1 and X_2, the choice of the distribution is not unique for fixed marginal distribution and association between X_1 and X_2. Therefore, instead of directly modeling the correlation between X_1 and X_2, the correlation is modeled through the correlation between b_1 and b_2. We expect that two clinical outcomes are positively correlated.

Suppose b_j has marginal mean μ_j, where $j = 1, 2$. The degrees freedom (or free parameters) of the bivariate beta distribution is reduced from k to $k - 2$ by the imposition of fixed marginal means, where k is number of parameters of a bivariate beta distribution. Denote α_j as shape parameter and β_j as scale parameter for beta density of random variable b_j.The shape parameter is proportional to the scale parameter, such that $\alpha_j = \frac{\mu_j}{1-\mu_j}\beta_j = C_j\beta_j$, for $j = 1, 2$. The mean and variances of b_1 and b_2 can be also easily obtained:

$$
\begin{aligned}
E(b_1) = \mu_1 = \frac{\alpha_1}{\alpha_1 + \beta_1}, \quad Var(b_1) = \sigma_1^2 = \frac{\alpha_1 \beta_1}{(\alpha_1 + \beta_1)^2(\alpha_1 + \beta_1 + 1)} \\
E(b_2) = \mu_2 = \frac{\alpha_2}{\alpha_2 + \beta_2}, \quad Var(b_2) = \sigma_2^2 = \frac{\alpha_2 \beta_2}{(\alpha_2 + \beta_2)^2(\alpha_2 + \beta_2 + 1)}
\end{aligned} \tag{2.1}
$$

2.2.1 Correlation of three-parameter bivariate beta distribution

The density function for the three-parameter bivariate beta distribution proposed by Olkin and Liu (2003) is given by

$$
f(b_1, b_2) = \frac{\Gamma(\alpha_1 + \alpha_2 + \beta)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\beta)} \frac{b_1^{\alpha_1 - 1}(1 - b_1)^{\alpha_2 + \beta - 1} b_2^{\alpha_2 - 1}(1 - b_2)^{\alpha_1 + \beta - 1}}{(1 - b_1 b_2)^{\alpha_1 + \alpha_2 + \beta}}, \tag{2.2}
$$

where $0 < b_1 < 1$, $0 < b_2 < 1$. It can be shown that the marginal distributions of b_1 and b_2 follow univariate beta distributions. Each marginal distribution shares a common scale parameter β with each having different shape parameters (α_1 for b_1 and α_2 for b_2), thus giving the bivariate beta distribution three degrees freedom. As is known, the univariate beta distribution has a bell shape when both scale and shape parameters are greater than 1. The uniform distribution is achieved when both scale and shape parameters are equal to 1. The univariate beta distribution with alpha and beta less than 1 has a u-shaped curve which is not very useful for modeling in practice (Kelton and Law, 2000). Olkin and Liu has shown that the shape of the joint density function also varies with different combinations of $(\alpha_1, \alpha_2, \beta)$. For large values of α_1, α_2 and β, the density is similar to a bivariate normal density. For small values of α_1, α_2, β, the density is close to uniform. Therefore, we will only focus on the region of $\alpha_1 > 1$, $\alpha_2 > 1$, $\beta \geq 1$, which provides a bell shaped density for the prior information.

To calculate the correlation of the distribution, we utilize the following formula:

$$
\text{Corr}(b_1, b_2) = \frac{E(b_1 b_2) - E(b_1)E(b_2)}{\sqrt{Var(b_1)Var(b_2)}} \tag{2.3}
$$

The first moment $E(b_1 b_2)$ is integrated from the joint distribution of b_1 and b_2 :

$$E(b_1 b_2) = \int_0^1 \int_0^1 b_1 b_2 \frac{\Gamma(\alpha_1 + \alpha_2 + \beta)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\beta)} \frac{b_1^{\alpha_1 - 1}(1 - b_1)^{\alpha_2 + \beta - 1} b_2^{\alpha_2 - 1}(1 - b_2)^{\alpha_1 + \beta - 1}}{(1 - b_1 b_2)^{\alpha_1 + \alpha_2 + \beta}} db_1 db_2 \tag{2.4}$$

To simplify (2.4), the following results of integrals and formulas on the hypergeometric function are applied (Abramowitz and Stegun (1964), Bailey (1935)):

$${}_2F_1(a, b; c; z) = \Gamma(c)[\Gamma(b)\Gamma(c - b)]^{-1} \int_0^1 t^{b-1}(1 - t)^{c-b-1}(1 - tz)^{-a} dt \tag{2.5}$$

$${}_3F_2(a, b, c; e, f; z) = \frac{\Gamma(e)\Gamma(f)\Gamma(s)}{\Gamma(a)\Gamma(s + b)\Gamma(s + c)} {}_3F_2(e - a, f - a, s; s + b, s + c; z) \tag{2.6}$$

where $s = e + f - a - b - c$.

Therefore, $E(b_1 b_2)$ can be written as (see Appendix):

$$E(b_1 b_2) = \frac{\alpha_1 \alpha_2}{(\alpha_1 + \beta)(\alpha_2 + \beta)} *_3 F_2(1, 1, \beta; \beta + \alpha_1 + 1, \beta + \alpha_2 + 1; 1) \tag{2.7}$$

Substituting (2.3) with (2.7) and (2.1), the correlation between b_1 and b_2 in this 3-parameter model then becomes:

$$\begin{aligned} \text{Corr}(b_1, b_2) &= \sqrt{\frac{\alpha_1 \alpha_2 (\alpha_1 + \beta + 1)(\alpha_2 + \beta + 1)}{\beta^2}} \{ {}_3F_2(1, 1, \beta; \beta + \alpha_1 + 1, \beta + \alpha_2 + 1; 1) - 1 \} \\ &= \sqrt{\frac{\alpha_1 \alpha_2}{(\alpha_1 + \beta + 1)(\alpha_2 + \beta + 1)}} * \sum_{j=1}^{\infty} \frac{j! \frac{\Gamma(\beta + j)}{\Gamma(\beta + 1)}}{\frac{\Gamma(\alpha_1 + \beta + 1 + j)}{\Gamma(\alpha_1 + \beta + 2)} \frac{\Gamma(\alpha_2 + \beta + 1 + j)}{\Gamma(\alpha_2 + \beta + 2)}} \end{aligned} \tag{2.8}$$

With three degrees freedom, the correlation is free to range over [0,1]. In other words, we can reach any correlation ranging from zero to one by choosing α_1, α_2 and β. Olkin and Liu performed a simulation study to show how the correlation changes with different choices of α_1, α_2 and β. They found that the correlation is large when the shape parameters (α_1 and α_2) are large and the scale parameter (β) is small, while, small values of shape parameters (α_1 and α_2) and large value of scale parameters (β) yield small correlation. However, the degree freedom is reduced to one when we specify a value for each

of the marginal means (recall $\alpha_j = \frac{\mu_j}{1-\mu_j}\beta = C_j\beta$). Thus, the correlation only depends on the value of β when the marginals are fixed:

$$\text{Corr}(b_1, b_2|\beta) = \underbrace{\sqrt{\frac{C_1C_2\beta^2}{[(C_1+1)\beta+1][(C_2+1)\beta+1]}}}_{h(\beta)} \underbrace{\sum_{j=1}^{\infty} \frac{j!\frac{\Gamma(\beta+j)}{\Gamma(\beta+1)}}{\frac{\Gamma[(C_1+1)\beta+1+j]}{\Gamma[(C_1+1)\beta+2]}\frac{\Gamma[(C_2+1)\beta+1+j]}{\Gamma[(C_2+1)\beta+2]}}}_{g(\beta)} \tag{2.9}$$

We can see that the first part on the right of the correlation (2.9) is an increasing function of β, denoted as $h(\beta)$. $h(\beta)$ is then in the range $\left[\sqrt{\frac{C_1C_2}{(C_1+2)(C_2+2)}}, \sqrt{\frac{C_1C_2}{(C_1+1)(C_2+1)}}\right)$ given $\beta \geq 1$. Therefore, we have $\text{Corr}(b_1, b_2|\beta) < \sqrt{\frac{C_1C_2}{(C_1+1)(C_2+1)}}g(\beta)$ for any $\beta \geq 1$.

The second part is also a function of β, denoted as $g(\beta)$. The expansion of $g(\beta)$ function then becomes:

$$\begin{aligned} g(\beta) = 1 + \frac{2(\beta+1)}{((C_1+1)\beta+2)((C_2+1)\beta+2)} + \frac{2(\beta+1)}{((C_1+1)\beta+2)((C_2+1)\beta+2)} \\ * \frac{3(\beta+2)}{((C_1+1)\beta+3)((C_2+1)\beta+3)} + \cdots > 1 \end{aligned} \tag{2.10}$$

Each term in (2.10) is a decreasing function of β, therefore, $g(\beta)$ is also a decreasing function of β. Given $\beta \geq 1$, the maximum value of $g(\beta)$ is retained at $\beta = 1$. And as β approximates infinity, $g(\beta)$ approximates to 1. Therefore, $g(\beta)$ is in the range $(1, g(1)]$. It can be easily shown that the inequality $\text{Corr}(b_1, b_2|\beta) > h(\beta)$ holds for any $\beta \geq 1$.

The correlation coefficient (2.9) is then bounded by the sequences $h(\beta)$ and $Kg(\beta)$:

$$h(\beta) < \text{Corr}(b_1, b_2|\beta) < Kg(\beta), \text{where } K = \sqrt{\frac{C_1C_2}{(C_1+1)(C_2+1)}} \text{ and } \beta \geq 1 \tag{2.11}$$

Clearly, we also have an approximate boundary for the correlation from (2.11):

$$\min_{\beta\in[1,\infty)} h(\beta) < \text{Corr}(b_1, b_2|\beta) < K \max_{\beta\in[1,\infty)} g(\beta) \tag{2.12}$$

2.2.2 Correlation of Morgenstern-type bivariate beta distribution

The joint density distribution of Farlie-Gumbel-Morgenstern (FGM) distributions has a form of:

$$f(b_1, b_2) = [\prod_{j=1}^{2} f(b_j)][1 + \lambda \prod_{j=1}^{2} \{2F(b_j) - 1\}], \quad b_1, b_2 > 1, |\lambda| \leq 1 \tag{2.13}$$

where $f(b_j)$ is the marginal density of b_j, and $F(b_j)$ is the cumulative density function (c.d.f) of b_j. Integrating over the joint distribution,

$$E(b_1 b_2) = E(b_1)E(b_2) + \lambda \int_0^1 b_1\{2F(b_1) - 1\}f(b_1)db_1 \int_0^1 b_2\{2F(b_2) - 1\}f(b_2)db_2 \tag{2.14}$$

Let $g(b) = \int_0^1 b\{2F(b) - 1\}f(b)db$. Using the Cauchy-Schwarz inequality, it has been shown that $g(b) \leq \sqrt{\frac{1}{3}}\sigma$ by Schucany et al. (1978). Then, the general form of the correlation is composed of the product of two univariate integrals, λ, and the marginal variances as follows (see Appendix)

$$\text{Corr}(b_1, b_2) = \frac{\lambda g(b_1) g(b_2)}{\sigma_1 \sigma_2} \leq \frac{1}{3}\lambda \tag{2.15}$$

Since λ is in the range of [-1,1], the correlation in the FGM distribution is bounded between $-\frac{1}{3}$ and $\frac{1}{3}$ for any specified marginal distributions. The corresponding FGM bivariate beta distribution is obtained by having beta densities as the marginal distributions. The c.d.f of the beta distribution is given by

$$F(b) = \frac{1}{B(\alpha, \beta)} \int_0^b t^{\alpha-1}(1-t)^{\beta-1}dt, \quad \alpha, \beta > 0 \tag{2.16}$$

Note that (2.16) is the regularized incomplete beta function, and can be reformed in terms of the gauss hypergeometric function

$$
\begin{aligned}
F(b) &= \frac{1}{B(\alpha,\beta)} \int_0^b t^{\alpha-1} \sum_{k=0}^{\infty} \frac{\Gamma(1-\beta+k)t^k}{\Gamma(1-\beta)k!} dt \\
&= \frac{b^\alpha}{\alpha B(\alpha,\beta)} \sum_{k=0}^{\infty} \frac{(1-\beta)_k(\alpha)_k b^k}{(\alpha+1)_k k!} \\
&= \frac{b^\alpha}{\alpha B(\alpha,\beta)} {}_2F_1(\alpha, 1-\beta; \alpha+1; b)
\end{aligned} \tag{2.17}
$$

Through some calculations, we get

$$
\begin{aligned}
g(b) &= \frac{2}{\alpha B(\alpha,\beta)^2} \sum_{k=0}^{\infty} \left\{ \frac{(1-\beta)_k(\alpha)_k}{(\alpha+1)_k k!} \int_0^1 b^{2\alpha+k}(1-b)^{\beta-1} db \right\} - \int_0^1 b f(b) db \\
&= \frac{2\Gamma(2\alpha+1)\Gamma(\beta)}{\alpha\Gamma(1+2\alpha+\beta)B(\alpha,\beta)^2} \sum_{k=0}^{\infty} \frac{(1-\beta)_k(\alpha)_k(1+2\alpha)_k}{(\alpha+1)_k(1+2\alpha+\beta)_k k!} - E(b) \\
&= E(b)\{ \frac{2B(1+2\alpha,\beta)}{\alpha B(\alpha,\beta)B(1+\alpha,\beta)} {}_3F_2(1-\beta, \alpha, 1+2\alpha; 1+\alpha, 1+2\alpha+\beta; 1) - 1\}
\end{aligned}
$$

applying (2.6),

$$
= E(b)\{ \frac{2B(2\alpha,2\beta)}{\beta B(\alpha,\beta)B(\alpha,\beta)} {}_3F_2(1, 1+\alpha+\beta, 2\beta; 1+\beta, 1+2\alpha+2\beta; 1) - 1\} \tag{2.18}
$$

The correlation of b_1 and b_2 is

$$
\begin{aligned}
\mathrm{Corr}(b_1, b_2) = \frac{\lambda E(b_1)E(b_2)}{\sigma_1\sigma_2} \prod_{j=1}^{2} \{ &\frac{2B(2\alpha_j, 2\beta_j)}{\beta_j B(\alpha_j,\beta_j)^2} \\
&* {}_3F_2(1, 1+\alpha_j+\beta_j, 2\beta_j; 1+\beta_j, 1+2\alpha_j+2\beta_j; 1) - 1\}
\end{aligned} \tag{2.19}
$$

Under specified marginal means,the correlation coefficient (2.19) is a product of λ and a function of β_1 and β_2, denoted as $g(\beta_1, \beta_2)$. The range of possible correlations are determined by the range of $g(\beta_1, \beta_2)$, since $\lambda \in [-1, 1]$.

2.3 Calculation and comparison of correlation bounds

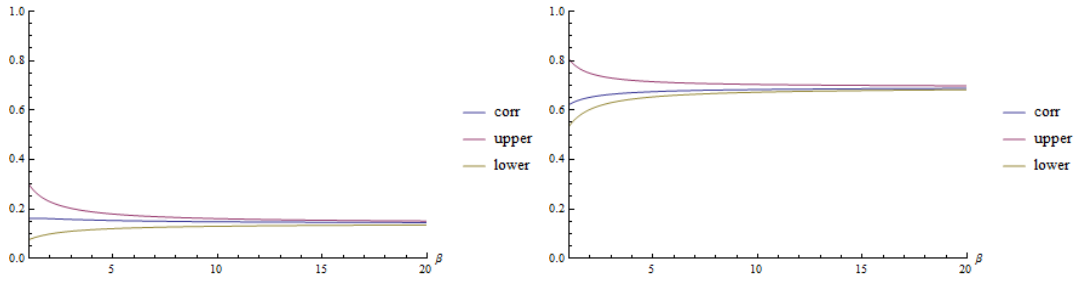
We selected different pairs of p_1 and p_2 to represent the proportions of success for two treatment outcomes .To reflect the commonly experienced reality of small clinical differences between two proportions, we assumed the difference to be either 0.1 or 0.2. Table

1 presents the boundary of the correlation for the three-parameter bivariate beta distribution. We note that this boundary is not the attainable boundary for the correlation, however, the boundary does provide some information on the performance of the distribution of the region, where β is as great as 1 and marginal means are fixed. The correlation realized at any value of β will be constrained by the bounds presented in Table 2.1. The correlation bounds depend on the value of p_1 and p_2. For small p_1 and p_2, the correlation will also be small. For instance, the correlations are less than 0.3 and the bounds are quite narrow, when p_1 and p_2 is out of any combination of (0.1, 0.2, 0.3). For large values of p_1 and p_2 (0.5 to 0.8), the correlation is able to reach large values, but again its range is restricted. The purpose of constructing these bounds is to show that the correlation value of the three parameter bivariate beta distribution is not free on $[0, 1]$. Moreover, the correlation is restricted to a narrow range given p_1 and p_2. In order to demonstrate it, we plot the correlation against sequence $h(\beta)$ (lower) and $Kg(\beta)$ (upper) given β . As presented in Figure 2.1, the maximum range between lower and upper corresponds to the correlation boundary in Table 2.1. Figure 2.1a is the case of $p_1 = 0.1$ and $p_2 = 0.2$, and the exact calculated correlation stays around 0.16. Figure 2.1b is the case of $p_1 = 0.6$ and $p_2 = 0.8$, and the exact calculated correlation ranges from 0.6 to 0.7. As β increases, the range between the lower and upper boundaries asymptotically decrease to a point.

Table 2.1: Approximate Correlation Boundary for Combinations of Efficacies

p_1	p_2	$\min h(\beta)$	$\max Kg(\beta)$
0.1	0.2	0.0765	0.1646
0.1	0.3	0.0964	0.2016
0.2	0.3	0.1400	0.2852
0.2	0.4	0.1667	0.3293
0.3	0.4	0.2100	0.4033
0.3	0.5	0.2425	0.4509
0.4	0.5	0.2887	0.5206
0.4	0.6	0.3273	0.5703
0.5	0.6	0.3780	0.6376
0.5	0.7	0.4237	0.6887
0.6	0.7	0.4804	0.7544
0.6	0.8	0.5345	0.8065

Figure 2.1: Correlation agianst the Bound given βs



(a) $p_1 = 0.1$, $p_2 = 0.2$

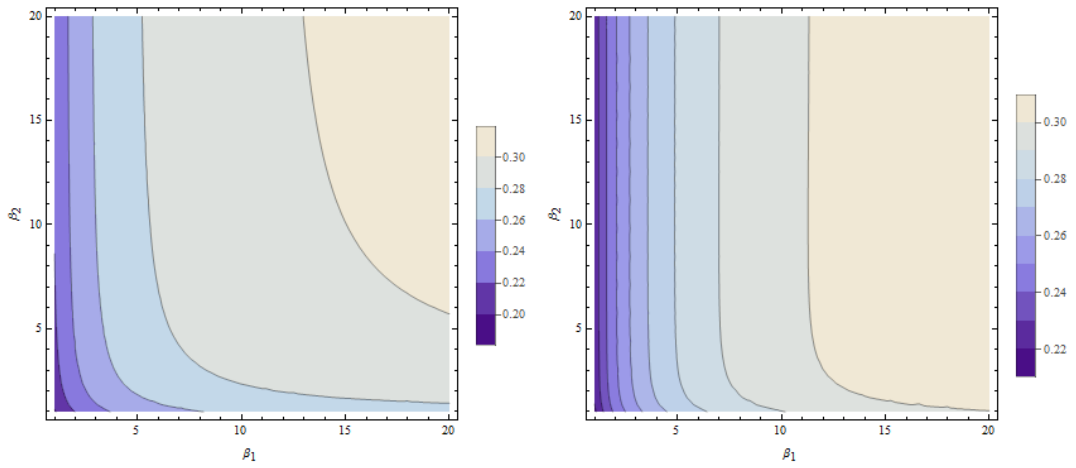
(b) $p_1 = 0.6$, $p_2 = 0.8$

(c) $p_1 = 0.3$, $p_2 = 0.5$

(d) $p_1 = 0.4$, $p_2 = 0.5$

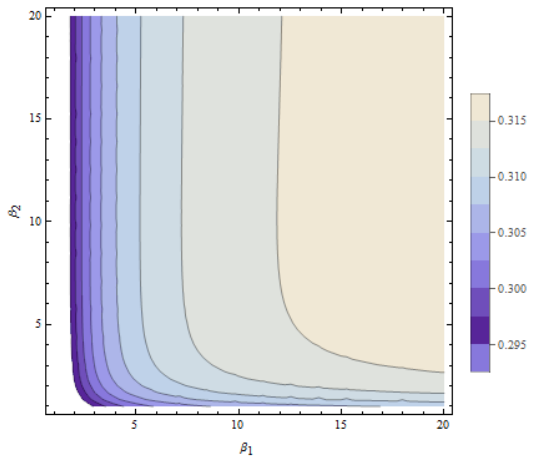
As we have discussed, the correlation of Morgenstern type bivariate distributions are bounded between $-\frac{1}{3}$ and $\frac{1}{3}$ in general. Figure 2.2 gives a contour plot of $g(\beta_1, \beta_2)$ given β_1, β_2 for the 5-parameter model. When we have specified beta marginals, we have noticed that the correlation of the distribution is always positive for the cases we are interested given $1 \leq \beta_1, \beta_2 \leq 20$. For small p_1 and p_2 , the correlation varies from 0.2 to 0.3. As p_1 and p_2 increases, the size of the correlation decreases and the correlation basically stays around 0.3 with minor variations. This property reduces the applicability of the distribution on Bayesian modeling. Meanwhile, the large correlation values are not allowed in the five parameter bivariate beta distribution, which is a serious disadvantage on modeling correlated binomial data.

Figure 2.2: Contour Plot for $g(\beta_1, \beta_2)$ given (p_1, p_2)

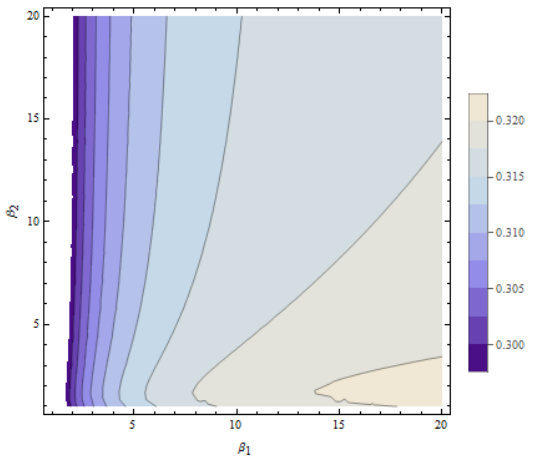


(a) $p_1 = 0.1$, $p_2 = 0.2$

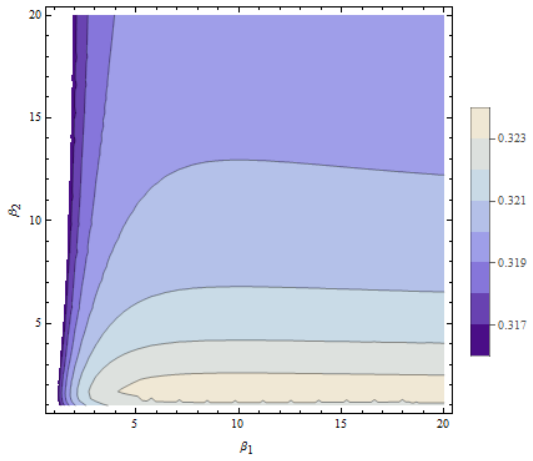
(b) $p_1 = 0.1$, $p_2 = 0.3$



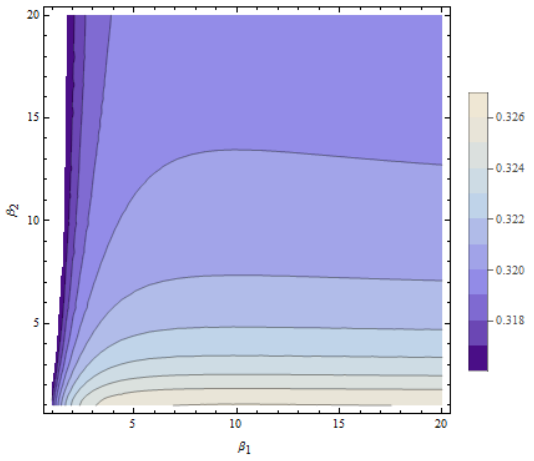
(c) $p_1 = 0.2$, $p_2 = 0.3$



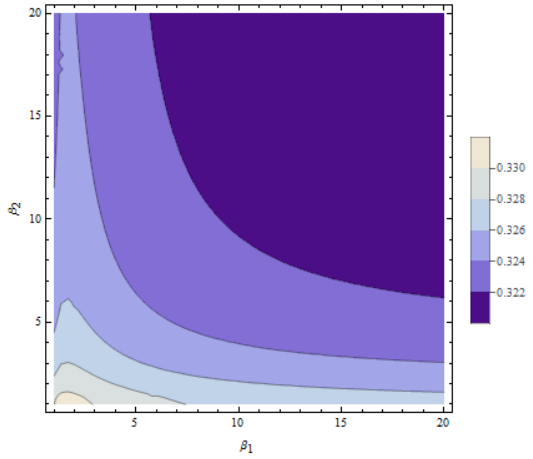
(d) $p_1 = 0.2$, $p_2 = 0.4$



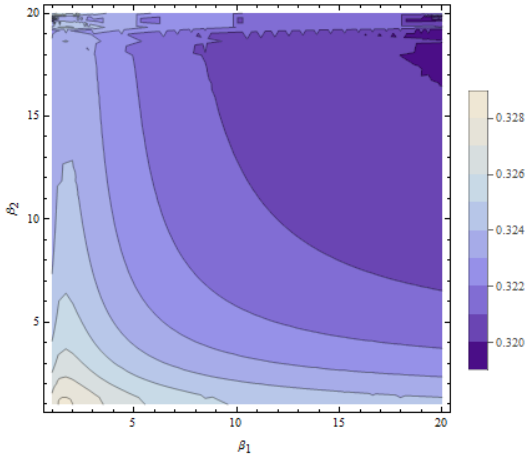
(e) $p_1 = 0.3$, $p_2 = 0.4$



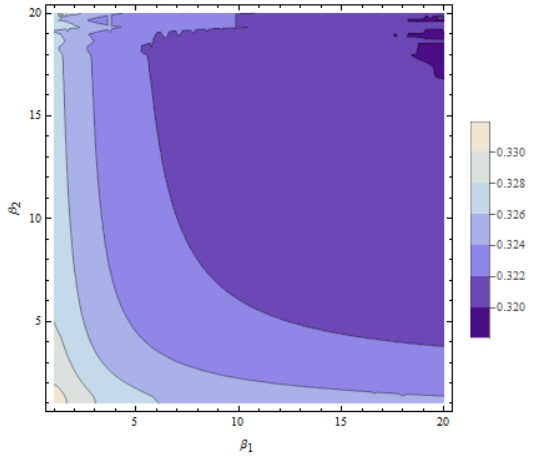
(f) $p_1 = 0.3$, $p_2 = 0.5$



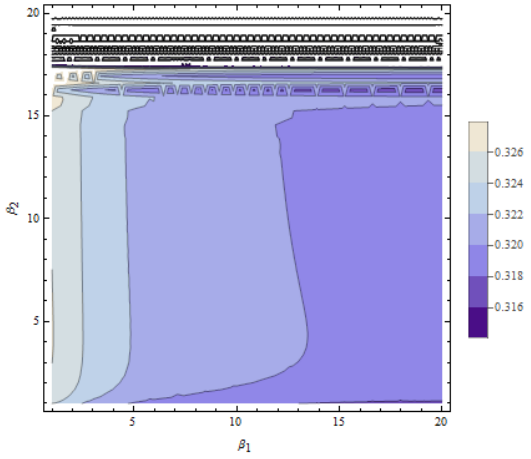
(g) $p_1 = 0.4$, $p_2 = 0.5$



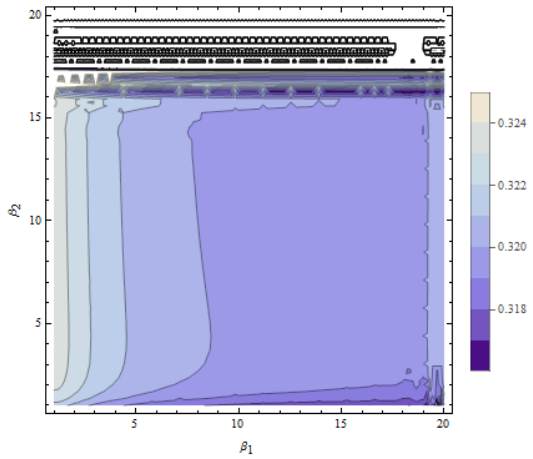
(h) $p_1 = 0.4$, $p_2 = 0.6$



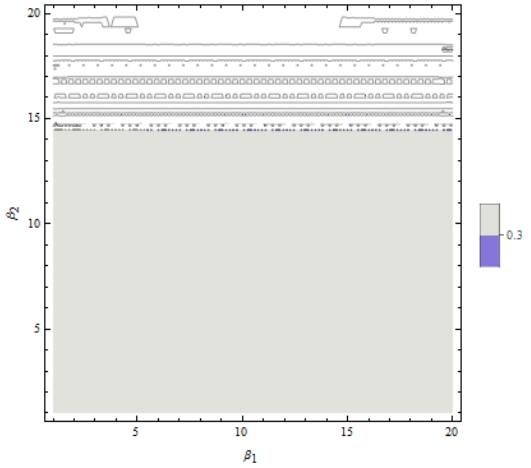
(i) $p_1 = 0.5$, $p_2 = 0.6$



(j) $p_1 = 0.5$, $p_2 = 0.7$



(k) $p_1 = 0.6$, $p_2 = 0.7$



(l) $p_1 = 0.6$, $p_2 = 0.8$

2.4 Discussion

According to Table 2.1, we notice that the range of correlation is narrow for reasonable values of marginal probabilities p_1 and p_2 for three parameter bivariate beta distribution. From this point, we are also interested in whether the five parameter bivariate beta distribution (α_1, β_1, α_2, β_2, λ) proposed by Gupta and Wong is free on [-1,1] under the same restriction for each of the marginal means. However, this five-parameter distribution is bounded from $-\frac{1}{3}$ to $\frac{1}{3}$. For lower values of p_1 and p_2, the five-parameter model allows more possible correlation values than the three- parameter distribution in terms of the maximum attainable correlation. However, the three-parameter allows larger correlations when p_1 and p_2 get larger, while the five-parameter model does not. In practice, we would appreciate the freedom to model the correlation as well as the marginal for binomial correlated data. Therefore, neither of these models offer flexibility, and thus seems unsuited for us as informative priors. It is possible that the copula functions may offer wider possible correlations in the presence of fixed marginal priors. We will study this possibility elsewhere.

2.5 Appendix

The Gaussian hypergeometric function ${}_2F_1(a, b; c; z)$ and the generalized hypergeometric function ${}_3F_2(a, b, c; e, f; z)$ are defined as:

$$ {}_2F_1(a, b; c; z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!} $$

$$ {}_3F_2(a, b, c; e, f; z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k (c)_k}{(e)_k (f)_k} \frac{z^k}{k!} $$

where $(a)_k$ takes the form:

$$(a)_k = \frac{\Gamma(a+k)}{\Gamma(a)} = (a+k-1)(a+k-2)(a+k-3)\cdots(a+2)(a+1)(a)$$

$$\begin{aligned}
E(b_1b_2) &= \int_0^1\int_0^1 b_1 * b_2 \frac{\Gamma(\alpha_1+\alpha_2+\beta)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\beta)} \frac{b_1^{\alpha_1-1}(1-b_1)^{\alpha_2+\beta-1} b_2^{\alpha_2-1}(1-b_2)^{\alpha_1+\beta-1}}{(1-p_1p_2)^{\alpha_1+\alpha_2+\beta}} db_1db_2 \\
&= \frac{\Gamma(\alpha_1+\alpha_2+\beta)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\beta)} \int_0^1\int_0^1 \frac{b_1^{\alpha_1}(1-b_1)^{\alpha_2+\beta-1} b_2^{\alpha_2}(1-b_2)^{\alpha_1+\beta-1}}{(1-p_1p_2)^{\alpha_1+\alpha_2+\beta}} db_1db_2 \\
&= \frac{\Gamma(\alpha_1+\alpha_2+\beta)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\beta)} \int_0^1\int_0^1 \sum_{j=0}^{\infty} \frac{\Gamma(\alpha_1+\alpha_2+\beta+j) * b_1^{\alpha_1+j}(1-b_1)^{\alpha_2+\beta-1}}{\Gamma(\alpha_1+\alpha_2+\beta)j!} \\
&* b_2^{\alpha_2+j}(1-b_2)^{\alpha_1+\beta-1} db_1db_2 \\
&= \frac{\Gamma(\alpha_1+\alpha_2+\beta)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\beta)} \sum_{j=0}^{\infty} \frac{\Gamma(\alpha_1+\alpha_2+\beta+j)}{\Gamma(\alpha_1+\alpha_2+\beta)} \frac{\Gamma(\alpha_1+1+j)\Gamma(\alpha_2+\beta)}{\Gamma(\alpha_1+\alpha_2+\beta+1+j)} * \frac{1}{j!} \\
&* \frac{\Gamma(\alpha_2+1+j)\Gamma(\alpha_1+\beta)}{\Gamma(\alpha_1+\alpha_2+\beta+1+j)} \\
&= \frac{\Gamma(\alpha_1+\alpha_2+\beta)\Gamma(\alpha_2+\beta)\Gamma(\alpha_1+\beta)}{\Gamma(\beta)\Gamma(\alpha_1+\alpha_2+\beta+1)\Gamma(\alpha_1+\alpha_2+\beta+1)} \sum_{j=0}^{\infty} \frac{(\alpha_1+\alpha_2+\beta)_n(\alpha_1+1)_n}{(\alpha_1+\alpha_2+\beta+1)_n} \\
&* \frac{(\alpha_2+1)_n}{(\alpha_1+\alpha_2+\beta+1)_n} \frac{1}{j!} \\
&= \frac{\alpha_2\alpha_1}{\alpha_1+\alpha_2+\beta} \frac{\Gamma(\alpha_2+\beta)\Gamma(\alpha_1+\beta)}{\Gamma(\beta)\Gamma(\alpha_1+\alpha_2+\beta+1)} * \\
&{}_3F_2(\alpha_1+\alpha_2+\beta, \alpha_1+1, \alpha_2+1; \alpha_1+\alpha_2+\beta+1, \alpha_1+\alpha_2+\beta+1; 1)
\end{aligned}$$

Define $v = \alpha_1+\alpha_2+\beta$,

$$\begin{aligned}
E(b_1b_2) &= \frac{\alpha_2\alpha_1}{v} \frac{\Gamma(\alpha_2+\beta)\Gamma(\alpha_1+\beta)}{\Gamma(\beta)\Gamma(v+1)} * \frac{\Gamma(v+1)\Gamma(v+1)\Gamma(\beta)}{\Gamma(v)\Gamma(\alpha_1+\beta+1)\Gamma(\alpha_2+\beta+1)} \\
&*_3 F_2(1,1,\beta; \beta+\alpha_1+1, \beta+\alpha_2+1; 1) \\
&= \frac{\alpha_1\alpha_2}{(\alpha_1+\beta)(\alpha_2+\beta)} *_3 F_2(1,1,\beta; \beta+\alpha_1+1, \beta+\alpha_2+1; 1)
\end{aligned}$$

The Cauchy-Schwarz inequality is stated as $| \int f(x)g(x)dx |^2 \leq \int f(x)^2 dx \int g(x)^2 dx$.

$$
\begin{aligned}
g(b) &= \int_0^1 b\{2F(b)-1\}f(b)db \\
&= \int_0^1 (b-E(b))\{2F(b)-1\}f(b)db \text{ , where } \int_0^1 \{2F(b)-1\}f(b)db = 0 \\
&\leq \sqrt{\int_0^1 (b-E(b))^2 f(b)db \int_0^1 (2F(b)-1)f(b)db} \\
&\leq \sqrt{\frac{1}{3}}\sigma
\end{aligned}
$$

Chapter 3

Bivariate Response Adaptive Design Part I (written as manuscript)

3.1 Background

Optimal response adaptive randomization designs are intended to minimize the risk of patients being assigned to an inferior treatment, given treatment outcomes (i.e. survival, remission) of previous patients. The average power of the procedure is a decreasing function of the variability of the randomization (Hu and Rosenberger, 2003). This leads to the trade-off between power and expected number of failures. The procedure is required to be the fully sequential, which means data need to be immediately observed, to balance the trade-off.

In practice, the primary clinical outcome may take a long term to be observed, especially in cancer trials (i.e. survival, remission). Therefore, there is usually a delay in the allocation rate update for the next patient or group of patients. However, the efficiency of the response adaptive design highly depends on the immediacy of observed data. If

few primary endpoints are observed at the early stage, little information is available to make the decision for the modification of the trial, which undermines the advantage of the design. A paper by Bai et al. (2002) has shown that moderate delay in response will not affect the asymptotic property of the adaptive procedure under certain delay mechanisms, however, the allocation rate through the trial is directly affected and there is a higher risk of assigning more patients to the inferior treatment.

3.2 Rationale

In the classical response adaptive procedure, the aim is to minimize the loss function given that the information level at each stage is constant, which has the form that contains the treatment efficacy difference ($\theta = p_A - p_B$, where p_A and p_B are the success rates for treatment A and B) and sample size ($n_i = n_{A,i} + n_{B,i}$) (Jennison and Turnbull):

$$L(\theta) = u(\theta)n_{A,i} + v(\theta)n_{B,i} \text{ , subject to } \frac{\sigma_A^2}{n_{A,i}} + \frac{\sigma_B^2}{n_{B,i}} = K \tag{3.1}$$

where $n_{A,i}$ and $n_{B,i}$ refer to cumulative number patients assigned to treatment A and treatment B at i^{th} stage, $u(\theta)$ is the loss for a patient allocated to treatment A, and $v(\theta)$ is the loss for a patient allocated to treatment B.

In randomization, patients are actually exposed to two risks: treatment failure and assigned to inferior treatment. Let $\theta < 0$ indicate treatment A is inferior ($p_A < p_B$) and $\theta > 0$ indicate treatment B is inferior ($p_A > p_B$). The treatment failure risk are described by $u(\theta)$ and $v(\theta)$, which are functions that depend on the value of θ. The function $u(\theta)$ increases as θ decreases and $v(\theta)$ increases as θ increases. The allocation ratio ($n_{A,i}/n_{B,i}$)

determines the probability of assigning patient to the inferior treatment. The loss function then integrates the two risks to which patients are exposed and our goal is to minimize this loss function subject to the constant variability at each stage. Assume σ_A^2 and σ_B^2 are the known variances of the population responses to treatment A and treatment B, respectively. The minimization of equation (3.1) can be solved for the allocation ratio using delta method (see Appendix), and the minimized allocation ratio is:

$$R = \frac{n_{A,i}}{n_{B,i}} = \frac{\sigma_A}{\sigma_B}\sqrt{\frac{v(\theta)}{u(\theta)}} \tag{3.2}$$

For Binary response trials, if $u(\theta) = v(\theta) = 1$, the allocation ratio $R = \frac{\sigma_A}{\sigma_B} = \sqrt{\frac{p_A q_A}{p_B q_B}}$ is Neyman allocation (Melfi and Page, 1998). Neyman allocation minimizes the total sample size given fixed variance. If $u(\theta) = 1-p_A$ and $v(\theta) = 1-p_B$, the allocation ratio $R = \sqrt{\frac{p_A}{p_B}}$ turns out to be RSIHR allocation (also called Optimal allocation), which minimizes the expected number of treatment failures (Rosenberger et al., 2001). Consequently, we need only to model $u(\theta)$ and $v(\theta)$ to realize a specific objective. $u(\theta)$ and $v(\theta)$ can also be treated as functions of unknown parameter p_A and p_B, which can be estimated based on patient responses using sequential estimation method. Since each adaptation depends on the previous estimates of $u(\theta)$ and $v(\theta)$, the accuracy of the estimates effects the whole adapative procedure. With delayed primary response, we have little information to make an accurate adaptation decision for the incoming patients.

In this paper, we are motivated to incorporate an auxiliary outcome that is easily observed and highly positively correlated with the primary outcome in the response adaptive randomization procedure. Such a procedure will need to take into account information from both the auxiliary and primary outcomes. Especially during early stages of randomization, we usually do not have enough information to update allocation ratio. A long lag time to observe primary outcome weakens the advantage of the response adaptive

design. Auxiliary outcomes provide additional information about the primary outcome and revive the adaptation procedure with delayed primary response. Then, the problem is how to appropriately incorporate the auxiliary outcome into the adaptation process along with the primary outcome. Bibby and Væth (2011) proposed a two-dimensional beta binomial distribution to fit the model for calculating the correlation of two count data. The distribution illuminates an approach to construct the bivariate variable model for the adaptive design. Based on the classical response adaptive design framework, we propose a new response adaptive design for binary data using both the auxiliary outcomes and primary outcomes. The goals of this paper are: 1) to introduce a response adaptive design framework that simultaneously use both primary and auxiliary outcomes, and 2) to utilize a bivariate beta distribution (Olkin and Liu, 2003) as prior distribution for correlated binomial data, to account for dependence between the two outcomes.

3.3 Method

3.3.1 New design using auxiliary and primary outcomes

For treatments j= A or B, suppose X_j is an auxiliary outcome for treatment j, Y_j is a primary outcome for treatment j, where X_j and Y_j both are binary variables. According to the observed outcome sequence, we denote $P_{1,j}$ as the auxiliary efficacy and $P_{2,j}$ as the primary efficacy for treatment j.

We make the following assumptions about the design that 1) $P_{1,j}, P_{2,j}$ are random variables and the joint distribution of $P_{1,j}, P_{2,j}$ follows a bivariate beta distribution, 2) The

conditional random variables $X_j|P_{1,j} \sim Bin(n_{1,j}, P_{1,j})$ and $Y_j|P_{2,j} \sim Bin(n_{2,j}, P_{2,j})$ are independent, where $n_{1,j}$ and $n_{2,j}$ are the sample size of auxiliary and primary outcomes, and 3) the correlation between X_j and Y_j is explained through the correlation of $P_{1,j}$ and $P_{2,j}$.

The posterior distribution of P_1 and P_2 can be expressed as:

$$f(P_1, P_2|X, Y) \propto f(X|P_1)f(Y|P_2)f(P_1, P_2) \tag{3.3}$$

As we have mentioned earlier, $u(\theta)$ and $v(\theta)$ are positive weight that measures the risk of assigning patients to treatment A and B given primary efficacies ($P_{2,A}$, $P_{2,B}$). In addition, we also have auxiliary efficacies $P_{1,A}$ and $P_{1,B}$, which offer the information about $P_{2,A}$ and $P_{2,B}$ respectively. Therefore, it is reasonable to average the $u(\theta)$ and $v(\theta)$ over all possible set of $P_{2,A}$ and $P_{2,B}$ given $(P_1, X, Y)_A$ and $(P_1, X, Y)_B$. Based on loss function (3.1) of the classical adaptive design framework, the loss function of the procedure using auxiliary and primary outcomes takes the following form:

$$L(\theta) = E[u(\theta)|(P_1, X, Y)_A, (P_1, X, Y)_B)]n_{A,i} + E[v(\theta)|(P_1, X, Y)_A, (P_1, X, Y)_B)]n_{B,i} \tag{3.4}$$

where $n_{A,i}$ and $n_{B,i}$ are the number of patients in treatment A and B at i^{th} stage. The two conditional expectations in (3.4) can be calculated through the conditional posterior distribution from (3.3).

The classical response adaptive design adopts equal increments of information at each stage, which means we update the allocation ratio after each subject or group of subjects are enrolled. Following in this trend, we will have $\frac{\sigma^2_{Y_A}}{n_{A,i}} + \frac{\sigma^2_{Y_B}}{n_{B,i}} = K$, where $\sigma^2_{Y_A}$ and $\sigma^2_{Y_B}$ are variances of primary outcomes and assumed to be known. The minimization of the function (3.4) is the same as that of the loss function (3.1) in classical response adaptive design framework, since the conditional expectations are assumed to be known. There-

fore, the allocation ratio is $R^* = \frac{\sigma_{Y_A}}{\sigma_{Y_B}} \sqrt{\frac{E[v(\theta)|(P_1,X,Y)_A,(P_1,X,Y)_B]}{E[u(\theta)|(P_1,X,Y)_A,(P_1,X,Y)_B]}}$.

3.3.2 Two-dimensional beta-binomial model

Olkin and Liu (2003) generated a bivariate beta distribution from three gamma distributed random variables, in order to provide a prior for correlated binomial distributions. As we discussed in Chapter 2, the bivariate beta distribution has limitations on the correlation of two variables. However, we still want to see the performance of the distribution serving as prior for $(P_{1,j}, P_{2,j})$. Given the assumptions about the design, the joint distribution of $(X_j, Y_j, P_{1,j}, P_{2.j})$ is the product of condition distributions of $X_j|P_{1,j}$ and $Y_j|P_{2.j}$, and prior distribution of $(P_{1,j}, P_{2,j})$.

To simplify our notation, the following distributions are generalized to any (X, Y, P_1, P_2) given a specific treatment.

$$\begin{aligned} f(X,Y,P_1,P_2) &= f(X,Y|P_1,P_2) * f(P_1,P_2;\alpha_1,\alpha_2,\beta) \\ &= \binom{n_1}{x} p_1^x (1-p_1)^{n_1-x} \binom{n_2}{y} p_2^y (1-p_2)^{n_2-y} \\ &* \frac{\Gamma(\alpha_1+\alpha_2+\beta)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\beta)} \frac{p_1^{\alpha_1-1}(1-p_1)^{\alpha_2+\beta-1} p_2^{\alpha_2-1}(1-p_2)^{\alpha_1+\beta-1}}{(1-p_1p_2)^{\alpha_1+\alpha_2+\beta}} \end{aligned} \tag{3.5}$$

Integrating with respect to P_2, the joint distribution of X, Y, P_1 is:

$$\begin{aligned} f(X,Y,P_1) &= \binom{n_1}{x}\binom{n_2}{y} \frac{\Gamma(\alpha_1+\alpha_2+\beta)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\beta)} p_1^{x+\alpha_1-1}(1-p_1)^{\alpha_2+\beta-1+n_1-x} \\ &* \frac{\Gamma(y+\alpha_2)\Gamma(\alpha_1+\beta+n_2-y)}{\Gamma(\alpha_1+\alpha_2+\beta+n_2)} *_2 F_1(\alpha_1+\alpha_2+\beta, y+\alpha_2; \\ &\alpha_1+\alpha_2+\beta+n_2; p_1) \end{aligned} \tag{3.6}$$

where ${}_2F_1$ is the Gaussian hypergeometric function. Therefore, the conditional distribution of P_2 given P_1 and the data X, Y is easily obtained:

$$\begin{aligned} f(P_2|X,Y,P_1) &= f(X,Y,P_1,P_2)/f(X,Y,P_1) \\ &= \frac{\Gamma(\alpha_1+\alpha_2+\beta+n_2)\,\frac{p_2^{y+\alpha_2-1}(1-p_2)^{\alpha_1+\beta-1+n_2-y}}{(1-p_1p_2)^{\alpha_1+\alpha_2+\beta}}}{\Gamma(y+\alpha_2)\,\Gamma(\alpha_1+\beta+n_2-y)} \\ &* \frac{1}{{}_2F_1(\alpha_1+\alpha_2+\beta, y+\alpha_2; \alpha_1+\alpha_2+\beta+n_2; p_1)} \end{aligned} \tag{3.7}$$

As presented in the defined loss function, $u(\theta)$ and $v(\theta)$ are function of $P_{2,A}$ and $P_{2,B}$. Also, we know that treatment A is independent from treatment B, which indicates $f(X,Y,P_1,P_2)_A$ and $f(X,Y,P_1,P_2)_B$ are independent. As long as we know the conditional distribution $f(P_2|X,Y,P_1)_A$ and $f(P_2|X,Y,P_1)_B$ for treatment A and B, we are able to calculate the conditional expectation from the loss function (3.4).

In this paper, we only focus on the case when $u(\theta) = 1 - P_{2,A}$ and $v(\theta) = 1 - P_{2,B}$, recalling that $P_{2,j}$ is the primary efficacy rate in the j^{th} treatment. Then, the loss function form is reduced to

$$L(\theta) = (1 - E[P_{2,A}|(P_1,X,Y)_A])\,n_{A,i} + (1 - E[P_{2,B}|(P_1,X,Y)_B])\,n_{B,i} \tag{3.8}$$

The optimal allocation ratio can be rewritten as $R^* = \frac{\sigma^2_{Y_A}}{\sigma^2_{Y_B}}\sqrt{\frac{E[1-P_{2,B}|(P_1,X,Y)_B]}{E[1-P_{2,A}|(P_1,X,Y)_A]}}$.

For a given treatment, the conditional expectation is a function of X, Y, P_1 with prior parameters $(\beta, \alpha_1, \alpha_2)$ (see Appendix).

$$E[P_2|P_1,X,Y] = \frac{y+\alpha_2}{\alpha_1+\alpha_2+\beta+n_2} * \frac{{}_2F_1(\alpha_1+\alpha_2+\beta, y+\alpha_2+1; \alpha_1+\alpha_2+\beta+n_2+1; p_1)}{{}_2F_1(\alpha_1+\alpha_2+\beta, y+\alpha_2; \alpha_1+\alpha_2+\beta+n_2; p_1)} \tag{3.9}$$

We may notice that the expression on the right side of the equation is the gauss continued function. The continued function of Gaussian hypergeometric function converges uniformly for $0 < p_1 < 1$. Therefore, $E[P_2|P_1,X,Y]$ is guaranteed to reside within the range (0,1). And the correlation of X and Y is proportional to the correlation of P_1 and

P_2 and takes the form (see Appendix):

$$\text{corr}(X, Y) = \sqrt{\frac{n_1 n_2}{(\alpha_1 + \beta + n_1)(\alpha_2 + \beta + n_2)}}\text{corr}(P_1, P_2) \tag{3.10}$$

3.3.3 Prior density choice

In the binomial-beta model, subject matter expertise can be used to provide some information to assess the probability of having a success event, which then determines the mean or mode of the beta distribution. The total (r) of α and β determines the variance of the beta distribution given the marginal. The larger the total r, the more compact will be the prior distribution. Basically, this r indicates how confident we are on the expert advice or literature information, and $r - 2$ is known as effective sample size. If we lack confidence on the prior belief of success probability, we will weight the data more by taking a wide unimodal beta density function (i.e. by selecting low r). For the bivariate beta distribution case, we still adopt the same logic to select marginal density, which follows the beta distribution. Recalling that the prior correlation of Olkin and Liu's distribution is bounded in a narrow range when the marginal means are given, the inflexibility of modeling correlation diminishes the reliability of the bivariate prior distribution. According to (3.10), the correlation of auxiliary and primary outcomes approximate to the correlation of auxiliary and primary efficacy as $(n_1, n_2) >> (\alpha_1, \alpha_2, \beta)$. Therefore, we intend to have a less informative prior by choosing r no greater than 15 when α_1, α_2 and β are greater than 1. As studied in Olkin and Liu's paper, the bivariate beta distribution tends to have a bivariate normal density when α_1, α_2 and β are large.

3.3.4 Estimation rule for allocation rate

Although the allocation rate depends on the unknown parameters, we will apply the sequential sampling rule following the trend of optimal adaptive design to update the allocation rate. The prior parameters $(\beta, \alpha_1, \alpha_2)$ reveal the knowledge about the correlation between the auxiliary and primary outcomes (X and Y) and efficacies of the outcomes (P_1 and P_2) for a specific treatment. Based on clinician experience or pilot study, we are able to determine an appropriate combination of $(\beta, \alpha_1, \alpha_2)$ that satisfy $\frac{\alpha_1}{\alpha_1+\beta} \approx$ prior P_1 and $\frac{\alpha_2}{\alpha_2+\beta} \approx$ prior P_2. Let paired (x_k, y_k) be the auxiliary and primary outcome for k^{th} subject. Let T_k be the treatment indicator for k^{th} subject. As mentioned earlier, the primary outcome might not be available immediately to update the allocation ratio for the next coming patient, though the auxiliary outcome should be selected so that it will be quickly available before the next accrued patient.

Let Iy_k indicate whether the primary response for the k^{th} patient become accessible when a new patient is enrolled in the study. Let $F\left((x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k), \ldots, (x_{i-2}, y_{i-2}), (x_{i-1}, y_{i-1}); I_{y_1}, I_{y_2}, \ldots, I_{y_k}, I_{y_{i-2}}, I_{y_{i-1}}; T_1, T_2, T_3, \cdots, T_{i-2}, T_{i-1}\right)$ be the history of the first $i-1$ patients, denoted as $F(\bullet)_{i-1}$. Based on $F(\bullet)_{i-1}$, we have following results for the i^{th} stage:

$$
\begin{aligned}
n_{1,A,i} &= \sum_{k=1}^{i-1} T_k, & n_{1,B,i} &= \sum_{k=1}^{i-1} (1 - T_k) \\
x_{A,i} &= \sum_{k=1}^{i-1} x_k T_k, & x_{B,i} &= \sum_{k=1}^{i-1} x_k (1 - T_k) \\
n_{2,A,i} &= \sum_{k=1}^{i-1} I_{y_k} T_k, & n_{2,B,i} &= \sum_{k=1}^{i-1} I_{y_k} (1 - T_k) \\
y_{A,i} &= \sum_{k=1}^{i-1} y_k I_{y_k} T_k, & y_{B,i} &= \sum_{k=1}^{i-1} y_k I_{y_k} (1 - T_k)
\end{aligned}
$$

3.3.5 Algorithm of the design

The algorithm for conducting the proposed adaptive design is:

(1) Set initial allocation rate to be 0.5 for the first patient.

(2) Update auxiliary efficacy for treatment A and B with estimates $\tilde{p}_{1,A} = \frac{\alpha_{1,A}+x_{A,i-1}}{\alpha_{1,A}+\beta_A+n_{A,i-1}}$ and $\tilde{p}_{1,B} = \frac{\alpha_{1,B}+x_{B,i-1}}{\alpha_{1,B}+\beta_B+n_{B,i-1}}$ for the i^{th} stage.

Even though the auxiliary outcome is obtainable immediately, it is always the case that the empirical estimate (sample proportion) of P_1 is either overestimated or underestimated due to its sensitivity on extremes values. The empirical estimate performs better as the sample size gets larger. Instead of using sample proportions, posterior means of auxiliary outcomes are an appropriate choice for the estimation of auxiliary efficacy. As it is known, posterior means are weighted averages of the sample proportion and prior mean. Notice that the posterior distribution of the auxiliary outcome is given by the beta-binomial distribution.

(3). Calculate R^* by substituting $E(P_{2,A} \mid \tilde{p}_{1,A}, F(\bullet)_{i-1})$, $E(P_{2,B} \mid \tilde{p}_{1,B}, F(\bullet)_{i-1})$, $\hat{\sigma}^2_{Y,A}$, and $\hat{\sigma}^2_{Y,B}$.

Sample variances would be a good choice for $\hat{\sigma}^2_{Y,A}$, and $\hat{\sigma}^2_{Y,B}$ given enough data, for the reason that they are maximum likelihood estimator of variances. Also, they are easy to compute and save computational time in the procedure. However, primary outcomes are binary data and their variances are associated with their means. We might not be able to calculate the sample variance for treatments at early stages, which could result from response-delay, no event observed, or patients clustered in one treatment. For $i \leq k$, the estimated variance for treatment j adopts the posterior conditional variance $Var(P_{2,j} \mid \tilde{p}_{1,j}, F(\bullet)_{i-1})$, for $i > k$, the empirical sample variance $\frac{y_{j,i-1}(n_{j,i-1}-y_{j,i-1})}{n^2_{j,i-1}}$ is substituted, where j=treatment A or B, and k is the stage where sample variance is available

for both treatment groups.

(4). Repeat (2) and (3).

(5). Terminate the randomization depending on the specified stopping rule (eg. final sample size, primary end-point or toxicity).

3.4 Simulation

3.4.1 Simulation targets

In the simulation study, we are interested in modeling different clinical scenarios to see the performance of our bivariate model allocation method compared with RSIHR allocation and balanced allocation: 1) how different primary outcomes between treatment A and B affect simulation results, in terms of allocation proportions, number of patients assigned to each treatment, error rate, number of treatment failures, 2) how different of auxiliary efficacies affect the simulation results, and 3) how a delay in primary response affect the simulation results.

3.4.2 Sampling method

We plan to compare our proposed method with RSIHR allocation and balanced allocation methods. Let us consider each simulation as a single trial. Intuitively, simulations of each method are done separately with large amount of trials. However, to model random events in this way may not approximate the real clinical trial setting, since the trial is actually an observed sequence of random variables. Then, it will be more realistic to

generate N_A random observations from treatment A population and N_B random observations from treatment B population for each single trial. Each method works on the same pool of subjects. Suppose N is the total sample size of the clinical trial, then N_A and N_B should both be greater than N. Within a trial, three allocation methods will actually share the same sample pool that simulate from populations of treatment A and B. Across trials, the sample pool is regenerated for each trial. Therefore, we are able to reduce variation between trials. The variation between the trials caused by sample pools randomly generating from the populations of treatment A and B.

The sample size of balanced allocation is fixed in advanced, while the sample size of the adaptive methods is allowed to adjust during the trial. In order to make allocation procedures comparable, the total sample size is selected to yield 0.90 power of two-sided t-test for balanced allocation in each case. Correlated binomial responses are sampled from a multinomial distribution given both auxiliary and primary efficacies with a specified correlation. Notice that this correlation is not free range to $[-1, 1]$, due to the joint probability of auxiliary and primary outcomes being calculated to obey the Frechet bounds. We assume that the correlation between auxiliary and primary outcomes is fixed regardless of treatment effect. In other words, the auxiliary outcome contains certain amount of information about primary outcome, no matter which treatment a patient receives.

3.4.3 Simulation settings

1) Lead-in sample

RSIHR allocation utilizes the primary outcome to update allocation ratio. The allocation ratio is calculated based on the sample proportion and sample variance. As we mentioned

earlier, these sample estimates are not estimable at the early stage of the trial, when no variability exits in treatment responses or no response is available by the time the next patient enrolled in the study. A lead-in is introduced to the simulation process, which means that patients in the lead-in phase are assigned to treatments with equal probability.

2) Prior parameter selection

Prior distributions take into account the uncertainty of P_1 and P_2 before observed data is considered. Recalling that auxiliary and primary efficacies (P_1 and P_2) follow beta distributions, α_1 and β are the shape and scale parameters for prior distribution on auxiliary efficacy (P_1), and α_2 and β are shape and scale parameters for prior distribution on primary efficacy (P_2). In the simulation study, we already know the true efficacies (p_1, p_2) of treatment A and treatment B, though in practice, we do not know(p_1, p_2). Therefore, the mean of prior distributions are set to some values around p_1 and p_2, which gives us the equations of $\frac{\alpha_1}{\alpha_1+\beta} \approx priorP_1$ and $\frac{\alpha_2}{\alpha_2+\beta} \approx priorP_2$. Given these two equations, the relationship among $(\alpha_1, \alpha_2, \beta)$ is determined. As long as we know the correlation between P_1 and P_2, the combination of $(\alpha_1, \alpha_2, \beta)$ can be resolved. The expectation of P_1 and P_2 varies for different treatment but the correlation between P_1 and P_2 stays the same regardless of treatments.

3) Cases

Our method is to target the scenario that the primary outcome has a rare event rate and the auxiliary outcome has a moderate event rate. Such as, the primary efficacy ranges from 0.1 to 0.3, and auxiliary efficacy ranges from 0.4 to 0.7. Due to the limitation of the correlation of bivariate beta distribution on this model, the correlation has a narrow window (see Chapter 2). We assume the correlation between auxiliary and primary out-

comes are 0.5 for all cases. Recall that our bivariate optimal method aims to alleviate the restriction of classic response design on response delay. In the following cases, we assume a scenario that 30 patients have delay in primary outcome in the trial, to reflect the realistic scenario that observation of the primary outcome is delayed. Alternatively, we assure that the auxiliary outcome is immediately observed.

3.4.4 Results

To make explanation easier, we refer to our proposed method as bivariate optimal method and RSIHR allocation method as the univariate optimal method. Table 3.1 presents the number of patients assigned to treatment B (the more effective treatment) given the sample size. When treatment B is better than treatment A (0.1 vs 0.2 or 0.1 vs 0.3), the bivariate optimal method assigns more patients to the beneficial treatment than dose the univariate optimal method. Also, the bivariate optimal method is more sensitive to treatment differences. Roughly, 64% (104/162) of patients are allocated to treatment B when there is 0.2 clinical difference, and 59.5% (313/526) patients are allocated to treatment B when the clinical difference is reduced to 0.1. However, the univariate optimal method results in 58% (304/526 or 95/162) of patients allocated to treatment B when the clinical difference is either 0.1 or 0.2. Both bivariate and univariate optimal methods intend to equally assign patients to treatments when no clinical difference presents. It seems that the bivariate optimal method depends slightly on the auxiliary efficacies, which only changes decimal places of the number of patients assigned to superior treatment. The effect of clinical difference overwhelms the effect of auxiliary efficacy difference and dominates the direction of allocation. For no clinical difference case (0.3 vs 0.3), the auxiliary

outcome difference protracts the allocation to be balanced. For instance, 97 out of 200 patients are assigned to treatment B while the univariate optimal and balance methods assigns 100 patients to the treatment.

The advantage of response adaptive designs is that they reduce treatment failures, as compared to traditional randomization designs. Therefore, Table 3.2 shows the expected number of patient failures using the three methods, corresponding to the cases presented in Table 3.1. If a clinical difference displays, the bivariate and univariate optimal method perform substantially better than the balance method in terms of reducing patient failures. Compared with the univariate optimal method, the results of bivariate optimal method are not statistically enhanced, though there is a small improvement in all cases considered. Generally, Balance method has slightly smaller standard deviation than the other two optimal methods, and the variability of the two adaptive methods are similar. Within paired primary outcomes, the number of patient failures slightly varies across different combination of auxiliary efficacies, which does not appear to substantially affect the number of patient failures.

Table 3.3 gives the power/error rates of the two sided chi-square test at the observed end of trial. The type I error of rejecting the null hypothesis is near nomial 0.05 level when there is no clinical difference for each method. The bivariate and univariate optimal methods perform well. Their power is as good as that achieved using balanced randomization. Our proposed bivariate optimal method decreases the number of patient deaths and preserves the power.

We are also interested in comparing the bivariate with the univariate optimal method with respect to allocation rate. In order to simplify the comparison, we present the median with interquartile range (IQR) after 25%, 50%, and 75% of patients have accrued.

For instance, the total sample size for the case 0.1 vs 0.2 is 526, then the 25 percentile of the visits is 132 (0.25*256). Table 3.4 shows the allocation rates assigned to treatment A with IQR at the 25th percentile visit. The allocation rates are similar at the 25th percentile visit. However, when the clinical difference increases to 0.2 (0.1 vs 0.3), the allocation rate is distinguishable between the two methods at the early stage. Due to a delay in the primary response, the univariate optimal method adopts the "preset" 0.5 allocation ratio. The median probability of assigning a patient to treatment A is 0.33-0.34 with a narrow IQR. When there is no clinical difference (0.3 vs 0.3), the allocation rate varies around 0.5. The bivariate optimal method has slightly narrower IQR than the univariate optimal method. As patients were accrued, table 3.5 displays the allocation rate at the 50th percentile visit. For the case 0.1 vs 0.3, we notice that the bivariate optimal method intends to put more patients on the beneficial treatment (trt B) than does the univariate optimal method. For the other two cases, the allocation rates are similar. The main reason is that we have more information on primary responses and weight less on auxiliary responses, as the adaptation proceeds longer (at large number of the visit). The results of using the bivariate method confirms with that of using the univariate optimal method at the end of the adaptation process, regardless of clinical difference and auxiliary outcome differences. Table 3.6 demonstrates the steadiness of the two methods at the end of adaptation process.

3.5 Discussion

Using auxiliary outcomes to facilitate classic response adaptive design is promising at early stages of a trial, particularly when primary outcomes are delayed. Unlike the uni-

variate optimal method (classic response design), the bivariate optimal method provides more information with which to allocate patients to beneficial treatment at the beginning of randomization. The bivariate optimal method results in fewer expected number of failures, while retains comparable power level as the univariate optimal method and balance allocation method. There is a substantial clinical reduction in treatment failures from balanced method to both univariate and bivariate optimal methods. Also, it seems that the level of auxiliary outcome does not directly impact the allocation ratio and direction during the adaptation process. This might be caused by the limitation of the correlation of bivariate beta distribution, which serves as prior for the correlated binary outcomes. To solve this problem, we will consider avoiding this restriction by using Copula methods in further study.

Table 3.1: Summary Number of Patients in Group B (receiving more effective treatment)

Primary		Auxilary		Sample Size	Method		
trtA	trtB	trtA	trtB		Bivariate [1]	Univariate [2]	Balance [3]
		0.4	0.7		104.0	95.4	80.9
		0.4	0.6		104.7	96.0	81.1
0.1	0.3	0.5	0.7	162	104.0	95.5	81.1
		0.5	0.6		104.8	95.2	80.8
		0.6	0.6		104.9	95.4	81.6
		0.4	0.7		313.4	304.8	263.5
		0.4	0.6		312.2	304.4	262.8
0.1	0.2	0.5	0.7	526	312.1	305.0	263.0
		0.5	0.6		313.6	305.4	263.2
		0.6	0.6		313.0	305.0	263.1
		0.4	0.6		99.7	99.9	100.3
		0.6	0.5		100.8	100.1	99.8
0.3	0.3	0.5	0.5	200	100.0	100.3	99.8
		0.7	0.5		97.1	99.9	100.0

[1] New proposed bivariate adaptive design.

[2] Optimal adaptive design.

[3] Balanced randomization design.

Table 3.2: Summary of Expected Number of Patient Failures (SD)

Primary		Auxilary		Method		
trtA	trtB	trtA	trtB	Bivariate	Univariate	Balance
		0.4	0.7	125.0 (5.64)	126.7 (5.58)	129.9 (4.98)
		0.4	0.6	124.6 (5.54)	126.4 (5.55)	129.5 (5.13)
0.1	0.3	0.5	0.7	125.1 (5.41)	126.8 (5.42)	129.5 (4.99)
		0.5	0.6	124.8 (5.38)	126.7 (5.34)	129.5 (5.20)
		0.6	0.6	124.8 (5.45)	126.6 (5.38)	129.7 (5.16)
		0.4	0.7	442.4 (8.59)	443.2 (8.48)	447.3 (7.94)
		0.4	0.6	442.1 (8.61)	443.0 (8.49)	447.1 (8.23)
0.1	0.2	0.5	0.7	441.8 (8.48)	442.5 (8.40)	447.5 (8.12)
		0.5	0.6	441.8 (8.89)	442.5 (8.77)	447.3 (8.11)
		0.6	0.6	442.2 (8.60)	443.1 (8.66)	447.4 (8.33)
		0.4	0.6	140.0 (6.84)	140.1 (6.76)	140.2 (6.79)
		0.6	0.5	140.0 (6.43)	140.0 (6.41)	140.0 (6.37)
0.3	0.3	0.5	0.5	139.8 (6.39)	139.9 (6.30)	139.9 (6.32)
		0.7	0.5	139.9 (6.46)	139.9 (6.50)	139.9 (6.38)

Table 3.3: Summary of Power/Error Rate

Primary		Auxiliary		Method		
trtA	trtB	trtA	trtB	Bivariate	Univariate	Balance
		0.4	0.7	0.91	0.91	0.91
		0.4	0.6	0.91	0.92	0.91
0.1	0.3	0.5	0.7	0.90	0.90	0.91
		0.5	0.6	0.92	0.92	0.91
		0.6	0.6	0.91	0.91	0.91
		0.4	0.7	0.92	0.91	0.91
		0.4	0.6	0.90	0.92	0.89
0.1	0.2	0.5	0.7	0.90	0.89	0.92
		0.5	0.6	0.91	0.91	0.90
		0.6	0.6	0.90	0.90	0.88
		0.4	0.6	0.06	0.06	0.06
		0.6	0.5	0.05	0.05	0.05
0.3	0.3	0.5	0.5	0.05	0.05	0.04
		0.7	0.5	0.04	0.04	0.05

Table 3.4: Summary of Allocation Rate (IQR) at the 25^{th} Percentile Visit for Trt A

Primary		Auxiliary		Method	
trtA	trtB	trtA	trtB	Bivariate	Univariate
0.1	0.3	0.4	0.7	0.34 (0.32, 0.37)	0.50 (0.50, 0.50)
		0.4	0.6	0.33 (0.32, 0.36)	0.50 (0.50, 0.50)
		0.5	0.7	0.33 (0.32, 0.37)	0.50 (0.50, 0.50)
		0.5	0.6	0.33 (0.31, 0.36)	0.50 (0.50, 0.50)
		0.6	0.6	0.33 (0.31, 0.35)	0.50 (0.50, 0.50)
0.1	0.2	0.4	0.7	0.41 (0.36, 0.45)	0.41 (0.37, 0.45)
		0.4	0.6	0.41 (0.36, 0.46)	0.41 (0.36, 0.46)
		0.5	0.7	0.41 (0.36, 0.46)	0.41 (0.37, 0.46)
		0.5	0.6	0.41 (0.37, 0.45)	0.41 (0.37, 0.45)
		0.6	0.6	0.41 (0.36, 0.45)	0.41 (0.37, 0.45)
0.3	0.3	0.4	0.6	0.50 (0.46, 0.54)	0.50 (0.45, 0.55)
		0.6	0.5	0.50 (0.46, 0.54)	0.50 (0.45, 0.55)
		0.5	0.5	0.50 (0.47, 0.53)	0.50 (0.45, 0.55)
		0.7	0.5	0.51 (0.46, 0.55)	0.50 (0.45, 0.55)

Table 3.5: Summary of Allocation Rate (IQR) at the 50^{th} Percentile Visit for Trt A

Primary		Auxiliary		Method	
trtA	trtB	trtA	trtB	Bivariate	Univariate
0.1	0.3	0.4	0.7	0.35 (0.32, 0.41)	0.38 (0.32, 0.44)
		0.4	0.6	0.35 (0.32, 0.41)	0.37 (0.32, 0.43)
		0.5	0.7	0.36 (0.32, 0.41)	0.38 (0.32, 0.43)
		0.5	0.6	0.36 (0.32, 0.41)	0.38 (0.32, 0.44)
		0.6	0.6	0.36 (0.32, 0.41)	0.38 (0.32, 0.43)
0.1	0.2	0.4	0.7	0.41 (0.38, 0.44)	0.41 (0.38, 0.44)
		0.4	0.6	0.41 (0.38, 0.44)	0.41 (0.38, 0.44)
		0.5	0.7	0.41 (0.38, 0.44)	0.41 (0.38, 0.44)
		0.5	0.6	0.41 (0.38, 0.44)	0.42 (0.38, 0.44)
		0.6	0.6	0.41 (0.38, 0.44)	0.41 (0.38, 0.44)
0.3	0.3	0.4	0.6	0.50 (0.47, 0.53)	0.50 (0.47, 0.53)
		0.6	0.5	0.50 (0.47, 0.53)	0.50 (0.47, 0.53)
		0.5	0.5	0.50 (0.47, 0.53)	0.50 (0.47, 0.53)
		0.7	0.5	0.50 (0.47, 0.53)	0.50 (0.47, 0.53)

Table 3.6: Summary of Allocation Rate (IQR) at the 75^{th} Percentile Visit for Trt A

Primary		Auxiliary		Method	
trtA	trtB	trtA	trtB	Bivariate	Univariate
0.1	0.3	0.4	0.7	0.36 (0.32, 0.40)	0.36 (0.31, 0.40)
		0.4	0.6	0.36 (0.32, 0.40)	0.36 (0.31, 0.40)
		0.5	0.7	0.36 (0.32, 0.40)	0.36 (0.32, 0.40)
		0.5	0.6	0.36 (0.31, 0.40)	0.36 (0.32, 0.40)
		0.6	0.6	0.36 (0.31, 0.40)	0.36 (0.32, 0.41)
0.1	0.2	0.4	0.7	0.41 (0.39, 0.43)	0.41 (0.39, 0.43)
		0.4	0.6	0.41 (0.39, 0.43)	0.41 (0.39, 0.43)
		0.5	0.7	0.41 (0.39, 0.43)	0.41 (0.39, 0.43)
		0.5	0.6	0.42 (0.39, 0.44)	0.42 (0.39, 0.44)
		0.6	0.6	0.41 (0.39, 0.44)	0.41 (0.39, 0.44)
0.3	0.3	0.4	0.6	0.50 (0.48, 0.52)	0.50 (0.48, 0.53)
		0.6	0.5	0.50 (0.48, 0.52)	0.50 (0.47, 0.52)
		0.5	0.5	0.50 (0.48, 0.52)	0.50 (0.48, 0.52)
		0.7	0.5	0.50 (0.48, 0.53)	0.50 (0.48, 0.53)

3.6 Appendix

3.6.1 Response adaptive allocation

$$\min[u(\theta)n_{A,i} + v(\theta)n_{B,i}] \text{ subject to } \frac{\sigma_A^2}{n_{A,i}} + \frac{\sigma_B^2}{n_{B,i}} = C \tag{3.11}$$

Define allocation ratio $R_i = \frac{n_{A,i}}{n_{B,i}}$, and $n_i = n_{A,i} + n_{B,i}$ is the total sample size at i^th stage. Using langrange multiplier method to find the local miminal of the function (3.11), subject to the constraint. The optimization problem (3.11) can be reformed as:

Minimize $h(R_i, n_i) = u(\theta)\frac{R_i}{1+R_i}n_i + v(\theta)\frac{1}{1+R_i}n_i$, subject to $g(R_i, n_i) = \frac{\sigma_A^2(1+R_i)}{R_i n_i} + \frac{\sigma_B^2(1+R_i)}{R_i n_i}$

We take derivatives of the function $\Lambda(R_i, n_i, \lambda) = h(R_i, n_i) + \lambda\big(g(R_i, n_i) - C\big)$ respect to R_i, n_i, λ and set derivatives to zero.

$$\frac{\partial\Lambda(R_i, n_i, \lambda)}{\partial R_i} = 0, \frac{\partial\Lambda(R_i, n_i, \lambda)}{\partial n_i} = 0, \frac{\partial\Lambda(R_i, n_i, \lambda)}{\partial\lambda} = 0 \tag{3.12}$$

Then, we have

$$\frac{u(\theta)n_i}{(1+R_i)^2} - \frac{v(\theta)n_i}{(1+R_i)^2} + \lambda\left[\frac{\sigma_B^2}{n_i} - \frac{\sigma_A^2}{n_i R_i^2}\right] = 0 \tag{3.13}$$

$$\frac{u(\theta)R_i}{1+R_i} + \frac{v(\theta)}{1+R_i} - \lambda\left[\frac{\sigma_A^2(1+R_i)}{R_i n_i} + \frac{\sigma_B^2(1+R_i)}{n_i^2}\right] = 0 \tag{3.14}$$

$$\frac{\sigma_A^2(1+R_i)}{R_i n_i} + \frac{\sigma_B^2(1+R_i)}{R_i n_i} = 0 \tag{3.15}$$

Through (3.14) and (3.15), we have

$$n_i = \frac{\big(1+R_i\big)\big(\sigma_A^2 + R_i\sigma_B^2\big)}{CR_i}, \qquad \lambda = \frac{\big(\sigma_A^2 + R_i\sigma_B^2\big)\big(u(\theta)R_i + v(\theta)\big)}{C^2R_i}$$

Substitute n_i and λ into (3.13), we have

$$\frac{\big(u(\theta) - v(\theta)\big)\big(\sigma_A^2 + R_i\sigma_B^2\big)}{(1+R_i)CR_i} + \frac{\big(u(\theta)R_i + v(\theta)\big)(\sigma_B^2R_i^2 - \sigma_A^2)}{(1+R_i)C^2R_i^2} = 0$$

$$\Rightarrow \big(u(\theta) - v(\theta)\big)\big(\sigma_A^2 + R_i\sigma_B^2\big)R_i + \big(u(\theta)R_i + v(\theta)\big)(\sigma_B^2R_i^2 - \sigma_A^2) = 0$$

$$\Rightarrow (R_i+1)(\sigma_B^2u(\theta)R_i^2 - \sigma_A^2v(\theta) = 0$$

$$\Rightarrow R_i = \frac{\sigma_A}{\sigma_B}\sqrt{\frac{v(\theta)}{u(\theta)}}$$

3.6.2 Conditonal expectation

Using Euler's formula

$$F(a,b;c;z) = \Gamma(c)[\Gamma(b)\Gamma(c-b)]^{-1} * \int_0^1 t^{b-1} * (1-t)^{c-b-1}(1-tz)^{-a}dt$$

we can easily calculate $E[P_2|P_1, X, Y]$.

$$\begin{aligned}
E[P_2|P_1,X,Y] &= \frac{\int_0^1 \frac{p_2^{y+\alpha_2}(1-p_2)^{\alpha_1+\beta-1+n_2-y}}{(1-p_1p_2)^{\alpha_1+\alpha_2+\beta}}dp_2}{\frac{\Gamma(y+\alpha_2)\Gamma(\alpha_1+\beta+n_2-y)}{\Gamma(\alpha_1+\alpha_2+\beta+n_2)} *_2 F_1(\alpha_1+\alpha_2+\beta, y+\alpha_2; \alpha_1+\alpha_2+\beta+n_2; p_1)} \\
&= \frac{\frac{\Gamma(y+\alpha_2+1)\Gamma(\alpha_1+\beta+n_2-y)}{\Gamma(\alpha_1+\alpha_2+\beta+n_2+1)} *_2 F_1(\alpha_1+\alpha_2+\beta, y+\alpha_2+1; \alpha_1+\alpha_2+\beta+n_2+1; p_1)}{\frac{\Gamma(y+\alpha_2)\Gamma(\alpha_1+\beta+n_2-y)}{\Gamma(\alpha_1+\alpha_2+\beta+n_2)} *_2 F_1(\alpha_1+\alpha_2+\beta, y+\alpha_2; \alpha_1+\alpha_2+\beta+n_2; p_1)} \\
&= \frac{y+\alpha_2}{\alpha_1+\alpha_2+\beta+n_2} * \frac{{}_2F_1(\alpha_1+\alpha_2+\beta, y+\alpha_2+1; \alpha_1+\alpha_2+\beta+n_2+1; p_1)}{{}_2F_1(\alpha_1+\alpha_2+\beta, y+\alpha_2; \alpha_1+\alpha_2+\beta+n_2; p_1)}
\end{aligned}$$

3.6.3 Correlation between X and Y

$$E[X|P_1] = n_1P_1, E[Y|P_2] = n_2P_2$$

$$E[XY] = E[E[XY|P_1,p_2] = E[E[X|P_1]E[Y|P2]] = n_1n_2E[P_1P_2]$$

$$Var(X) = Var(E(X|P_1]) + E(Var(X|P_1)] = n_1^2Var(P_1) + n_1\big(E(p_1) - E[p_1^2]\big)$$

$$= n_1^2Var(P_1) - n_1(\alpha_1+\beta+1)var(P_1) = n_1(\alpha_1+\beta+n_1)Var(P_1)$$

$$Var(Y) = n_2(\alpha_2+\beta+n_2)Var(P_2),\ \mathrm{corr}(X,Y) = \sqrt{\tfrac{n_1n_2}{(\alpha_1+\beta+n_1)(\alpha_2+\beta+n_2)}}\mathrm{corr}(P_1,P_2)$$

Therefore, the correlation of X and Y is linear function of the correlation of P_1 and P_2.

Chapter 4

Bivariate Response Adaptive Design Part II (written as manuscript)

4.1 Introduction

Randomized controlled trials (RCTs) are standard tools used to evaluate the effectiveness of interventions (treatments) on particular outcomes (e.g. death or survival rate). They help to ensure that patients are not consciously or unconsciously enrolled to receive treatments in a preferential manner. Therefore, RCTs help minimize selection and allocation bias (Schulz and Grimes, 2002). Response adaptive randomization has become popular in recent years in the study of Phase III trials, and utilizes accumulated outcomes from enrolled patients to assign an allocation ratio for the next patient or group of patients (Hu and Rosenberger, 2006).

For the purpose of ethicality, patients are favored to the beneficial treatment in the process of optimal allocation. However, outcome adaptive designs depend on the availability of outcomes. When the primary end point is a rare event or takes a long time to be

observed, adaptive designs do not offer much benefit. In order to solve this problem, we attempt to implement an ancillary outcome, which contains some information about the primary outcome, to the design. By means of joint modeling of the primary and ancillary outcomes, the allocation ratio is based on the inferences from the posterior distribution of primary efficacy conditional on the observed auxiliary efficacy.

In Chapter 3, we used a three-parameter bivariate beta distribution proposed by Olkin and Liu (2003) as prior distribution to describe the association of primary and auxiliary efficacies and then construct the model to estimate the allocation ratio for the adaptation. For a certain treatment with specified marginals, the three parameters of the bivariate distribution are reduced to 1 degree freedom. We have showed the limitation of the bivariate beta distribution, which imposes a small correlation range when the marginal means are fixed (see Chapter 2). This inflexibility stimulates us to find a bivariate prior distribution to reach a wider range of correlation given fixed marginal distributions.

Copula theory provides an approach to generate joint distributions in cases where a formal or intuitive joint distribution does not exist or is algebraically intractable. This theory also frees the constraints on association parameters due to their separate modeling from the fixed marginal distributions. The copula function can be viewed as a dependence function and links the multivariate distribution to several one-dimensional marginal distributions. Regardless of the form of the marginal distributions, we can modify the copula function to describe the dependence of the random variables. Due to its flexibility, the copula theory has been popularly applied to study quantitative risk management methodology within finance and insurance fields (Jaworski, 2010). Shih and Louis (1995) modeled the association of bivariate failure times (survival data) through copula functions, and Burzykowski et al. (2001) proposed two different copula models for

validating surrogate end points in multiple RCTs when the primary end point is treatment failure.

In this Chapter, we apply copula-based joint distribution to realize the adaptation of our proposed bivariate response adaptive design. Simulation studies are conducted to compare the results of the design using bivariate binomial-beta model (Chapter 3), in terms of number of patients assigned to treatments, expected number of patient failures, power/error rate, and allocation ratio at 25^{th}, 50^{th}, and 75^{th} percentile of enrollment.

4.2 Response adaptive design

The response adaptive design allows the allocation ratio (R) to be sequentially adapted based on the accumulated observed outcomes to hopefully assign patients to a "better" treatment, without violating the validation and integrity of the trial. Several adaptive procedures have been introduced in the clinical trial literature, such as randomized play the winner rule (Wei and Durham, 1978), optimal response adaptation (Rosenberger et al., 2001), and drop-the-loser (Ivanova, 2003). The response adaptive design focuses on the trade-off between the power (sample size) and number of patients receiving inferior treatment. In this paper, we develop a new response adaptive design based on the optimal response design. The optimal response adaptive design attains certain optimization criteria (e.g. number of patient failures, power) given that the information of each adaptation stage (the variance of efficacy end points) is constant.

Suppose two treatments (A and B) are studied, where $n_{A,i}$ is accumulated number of patients in treatment A and $n_{B,i}$ is accumulated number of patients in treatment B at i^{th} stage. The optimal problem for two treatments can be considered as (Jennison and

Turnbull):

$$R = \arg\min_{R}\{u(\theta)n_{A,i} + v(\theta)n_{B,i}\} \text{ , subject to } \eta(\theta, n_A, n_B) = C \tag{4.1}$$

where $u(\theta)$ and $v(\theta)$ are the positive weights on the treatment groups. θ represents the efficacy end points, and there are three commonly used end points: treatment difference $(p_A - p_B)$, relative risk $\left(\frac{p_A}{p_B}\right)$, and odds ratio $\left(\frac{p_A(1-p_B)}{p_B(1-P_A)}\right)$. $\eta(\theta, n_A, n_B)$ is the function of sample size and efficacy end point, which usually refers to the information known at each stage and retained equally. The consistency of the procedure information protects the power of the design, since the power of the analysis is a decreasing function of the procedure variability. In Chapter 3, we have discussed how the optimal allocation ratio can be solved using Lagrange multipliers when the treatment difference is the efficacy end point. The allocation ratio takes the form:

$$R_i = \frac{n_{A,i}}{n_{B,i}} = \frac{\sigma_A}{\sigma_B}\sqrt{\frac{v(\theta)}{u(\theta)}} \tag{4.2}$$

From equation 4.2, we notice that the optimization problem actually depends on how we establish $u(\theta)$ and $v(\theta)$ in the design assuming variances are known. For instance, they can be units to minimize the average sample size (Neyman allocation). Or, they can be weighted as failure rates to minimize the expected number of patients (so-called RSIHR or optimal allocation).

The efficacy end point associates with unknown parameter p_A and p_B, the primary efficacies, which are estimated with the observed primary outcomes. However, it is common for primary outcomes to not be available for up to weeks, months or years for any given subject. It is also common for the event to be rare, such as survival studies for cancer patients.

We incorporate the concept of auxiliary outcomes to $u(\theta)$ and $v(\theta)$ for both treatments.

An auxiliary outcome is not viewed as an replacement of the primary outcome, but as a positively associated companion of the primary outcome, which can be almost immediately collected from patients. Suppose the auxiliary outcome is collected at stage 1, and primary outcome becomes available at stage 2. Let p_1 be auxiliary outcome efficacy and p_2 be primary outcome efficacy. For the randomization with two treatments, the auxiliary outcome assists the primary outcome in assigning patients in the following way:

$$R_i = \frac{n_{A,i}}{n_{B,i}} = \frac{\sigma_A}{\sigma_B}\sqrt{\frac{v\left(\theta(p_{2,A}, p_{2,B})|p_{1,A}, p_{1,B}, \mathbf{X}, \mathbf{Y}\right)}{u\left(\theta(p_{2,A}, p_{2,B})|p_{1,A}, p_{1,B}, \mathbf{X}, \mathbf{Y}\right)}} \tag{4.3}$$

where $\theta(p_{2,A}, p_{2,B})$ is a function of primary outcome efficacies $p_{2,A}$, $p_{2,B}$, denoted as $\theta(\cdot)$. We assumed that (p_1, p_2) are joint distributed and follow a certain distribution. Therefore, we take the expectation of $v\left(\theta(\cdot)|p_{1,A}, p_{1,B}, \mathbf{X}, \mathbf{Y}\right)$ and $u\left(\theta(\cdot)|p_{1,A}, p_{1,B}, \mathbf{X}, \mathbf{Y}\right)$ to average all possible values over the range of $p_{2,A}$, $p_{2,B}$. In this paper, we are interested in minimizing the expected number of patients failures when $E[u\left(\theta|p_{1,A}, p_{1,B}, \mathbf{X}, \mathbf{Y}\right))] = E[p_{2,A}|p_{1,A}, X_A, Y_A]$ and $E[v\left(\theta|p_{1,A}, p_{1,B}, \mathbf{X}, \mathbf{Y}\right)] = E[p_{2,B}|p_{1,B}, X_B, Y_B]$. The allocation ratio is then updated as:

$$R_i = \frac{n_{A,i}}{n_{B,i}} = \frac{\sigma_A}{\sigma_B}\sqrt{\frac{E[p_{2,B}|p_{1,B}, X_B, Y_B]}{E[p_{2,A}|p_{1,A}, X_A, Y_A]}} \tag{4.4}$$

4.3 Method

Suppose $X|P_1$ and $Y|P_2$ are independent, and the joint distribution of X, Y, P_1 and P_2 is :

$$f(X, Y, P_1, P_2) = f(X|P_1)f(Y|P_2)f(P_1, P_2)$$

4.3.1 Prior distribution

Copula theory

The joint cumulative distribution function (c.d.f) of P_1 and P_2 is defined as $F(p_1, p_2) = Pr(P_1 \leq p_1, P_2 \leq p_2)$, and the marginal c.d.fs of P_1 and P_2 are defined as $u = F(x)$ and v=$F(y)$. Then, u and v are uniformly distributed. The quantile functions of the marginals are $p_1 = F^{-1}(u)$ and $p_2 = F^{-1}(v)$ (with one to one mapping). Then the joint c.d.f of X and Y can be reformed as:

$$\begin{aligned} F(p_1, p_2) =& Pr(P_1 \leq p_1, P_2 \leq p_2) \\ =& Pr(P_1 \leq F^{-1}(u), P_2 \leq F^{-1}(v)) \\ =& Pr(F(P_2) \leq u, F(P_2) \leq v) \\ =& Pr(U \leq u, V \leq v) \\ =& C(u, v) \end{aligned} \tag{4.5}$$

According to Sklar's theorem, C(u,v) is a uniquely defined copula function if the joint distribution is specified with marginals $F(x)$ and $F(y)$. Conversely, if $F(P_1)$, $F(P_2)$ and C(u, v) are distribution functions, then $F(x, y)$ is a unique joint distribution of P_1 and P_2 (Sklar, 1973). The joint density function takes the form:

$$\begin{aligned} f(p_1, p_2) =& \frac{\partial^2}{\partial p_1 \partial p_2} C(u, v) \\ =& \frac{\partial u}{\partial p_1} \frac{\partial v}{\partial p_2} \frac{\partial^2}{\partial u \partial v} C(u, v) \\ =& f(p_1) f(p_2) c(u, v) \end{aligned} \tag{4.6}$$

If $C(u, v)$ is the c.d.f, then $c(u, v)$ is the corresponding density function. From equation 4.6, we notice that the dependence between P_1 and P_2 is solely described by the function $c(u, v)$, which is independent from the marginals. This independence gives us the flexi-

bility to create a joint distribution by separately modeling the copula function and the marginal distributions.

The joint distribution $F(p_1, p_2)$ is naturally bounded by the Frèchet-Hoeffing bounds (Fréchet, 1951 and Hoeffding). The lower and upper bounds (for two variables) are defined as:

$$\begin{aligned} F_{Lower}(p_1, p_2) &= \max\left[F(p_1) + F(p_2) - 1, 0\right], \\ F_{Upper}(p_1, p_2) &= \min\left[F(p_1), F(p_2)\right] \end{aligned} \tag{4.7}$$

$F_{Lower}(p_1, p_2)$ and $F_{Upper}(p_1, p_2)$ are also copulas. Therefore, the graph of the copula C(u,v) is a continuous surface in between the graph of Frèchet-Hoeffing bounds within the unit cubic space. The bounds are corresponding to negative and positive dependence for the bivariate cases. In other words, the copula is equivalent to the upper bound when each of P_1 and P_2 is an almost surely increasing function of the other. The copula is equivalent to the lower bound when each of P_1 and P_2 is an almost surely decreasing function of the other.

Copula function

It is important to select an appropriate copula function for the data, because it describes the dependence structure between variables. The copula function can be categorized into two types: elliptical copulas and Archimedean copulas. The elliptical copulas are generated from the distribution that has elliptical contour. The Archimedean copulas are generated from a specified generator function (Nelsen, 2006), and they have simple forms and are able to capture wide ranges of dependence. For our purpose of study, we choose the Gaussian (Lee, 1983) and Clayton copula, the representatives of the two major copulas, to study the performance of the bivariate optimal design.

Gaussian Copula The Gaussian copula achieves the maximum lower and upper Frechet bounds. Also, a multivariate distribution based on a Gaussian copula is similar to the multivariate normal distribution (Xue-Kun Song, 2000). These properties are promising for the practical application on modeling dependence between random variables. The bivariate Gaussian copula takes the form:

$$C_G(u, v; \rho) = \Phi_G(\Phi(u)^{-1}, \Phi(v)^{-1}; \rho) \tag{4.8}$$

where $\Phi(\cdot)$ is the c.d.f of standard normal distribution, and $\Phi_G(u, v; \rho)$ is the c.d.f of standard bivariate normal distribution with correlation ρ.

Clayton Copula The Clayton (1978) copula is one type of the so-called Archimedean copula (Kimberling, 1974). And the following form in the bivariate case:

$$C_{Clayton}(u, v; \gamma) = (u^{-\gamma} + v^{-\gamma} - 1)^{-1/\gamma} \tag{4.9}$$

where γ indicates the level of dependence between u and v. The density function of the clayton copula is derived as:

$$\begin{aligned} c(u, v) =& \frac{\partial^2}{\partial u \partial v} C_{clyaton}(u, v; \gamma) \\ =& (\gamma + 1)(u^{-\gamma} + v^{-\gamma} - 1)^{-\frac{1}{\gamma}-2} u^{-\gamma-1} v^{-\gamma-1} \end{aligned} \tag{4.10}$$

The Kendall's *tao* correlation for this copula is $\frac{\gamma}{\gamma+2}$. In contrast to the Gaussian copula, the Clayton copula presents an asymmetric property on the dependence, where the asymptotic left tail dependence is $\lambda_L = 2^{-1/\gamma}$, and the asymptotic right tail dependence is close to zero (Fusai and Roncoroni, 2008), meaning it exhibits stronger dependence on left tail than on the right. Due to this property, the Clayton copula is suitable for the scenario when two outcomes exibit low event rates and high positive dependence. In

our scenario, we focus on low primary efficacy (0.1 to 0.3) and slightly higher auxiliary efficacy (0.4 to 0.6), both with strong positive dependence. Therefore, the Clayton copula is a good choice of comparison with Gaussian copula.

Marginal distribution

After choosing the copula function, we pick the marginal distribution based on our knowledge about the data. Since the parameter of interest is the probability of event occurrence, the beta distribution suits for the case and appropriately describes our initial beliefs about the auxiliary and primary efficacy (P_1 and P_2).

4.3.2 Metropolis-Hastings algorithm

The Metropolis-Hastings (MH) algorithm has been extensively used to generate a sequence of random samples from same target distribution, especially when the prior distribution and the likelihood function are not conjugated or where the full conditional posterior distribution is not in closed form. Unlike the Gibbs sampling method, the MH algorithm requires a function only be proportional to the target distribution. In our scenario, we are interested in sampling from $\pi(P_2|P_1, \boldsymbol{X}, \boldsymbol{Y})$, where P_2 is primary efficacy, P_1 is auxiliary efficacy, and $(\boldsymbol{X}, \boldsymbol{Y})$ are auxiliary and primary outcomes.

Selecting starting values p_1^0, p_2^0 from uniform(0,1), the MH algorithm proceeds as follow (for t in [1,T]):

1) Draw p_1^* and p_2^* from proposed transition kernel $q_1(\cdot|p_1^{t-1})$ amd $q_2(\cdot|p_2^{t-1})$.

2) Compute MH accept-reject ratio:

$$\alpha(p_1^*, p_1^{t-1}) = \min\left(\frac{\pi(p_1^*, p_2^{t-1}, \boldsymbol{X}, \boldsymbol{Y}) q_1(p_1^{t-1}|p_1^*)}{\pi(p_1^{t-1}, p_2^{t-1}, \boldsymbol{X}, \boldsymbol{Y}) q_1(p_1^*|p_1^{t-1})}, 1\right) \tag{4.11}$$

$$\alpha(p_2^*, p_2^{t-1}) = \min\left(\frac{\pi(p_2^*, p_1^{t-1}, \boldsymbol{X}, \boldsymbol{Y}) q_2(p_2^{t-1}|p_2^*)}{\pi(p_2^{t-1}, p_1^{t-1}, \boldsymbol{X}, \boldsymbol{Y}) q_2(p_2^*|p_2^{t-1})}, 1\right) \tag{4.12}$$

3) Accept p_1^* with probability $\alpha(p_1^*, p_1^{t-1})$ and p_2^* with probability $\alpha(p_2^*, p_2^{t-1})$, otherwise, p_1^t and p_2^t remain at the previous stage p_1^{t-1} and p_2^{t-1}.

4) Repeat (1), (2), (3) until the Markov chain has length of T.

4.3.3 Convergence diagnostic

The MH algorithm is based on the theory of Monte Carlo Markov Chains (MCMC), and we expect that the chain converges to the same stationary distribution (our target distribution) after a large number of iterations, regardless of the starting points selected for each parameter (i.e. P_1 and P_2). The simulation draws after t iterations are used as a sample from the stationary distribution. When the starting points are not in the high density region of the target distribution, the chain might converge slowly toward the target distribution. Therefore, it is important to have multiple starting points and diagnose the convergence of the Markov chains. Rubin and Gelman (1992) have constructed an approach to detect the convergence of the stochastic process using multiple sequences. Suppose we have M number of chains, and "burn in" (ie. remove) the first k draws. The Gelman and Rubin's diagnostics for each parameter are:

1) Calculate the average of the M within-chain variability, $W = \sum_{j=1}^{m} s_j^2/m$, where $s_j^2 = \sum_{i=1}^{n}(\tau_{ij} - \bar{\tau}_{\cdot j})^2)/(n-1)$ is the variance of j^{th} chain, $\bar{\tau}_{\cdot j}$ is the sample mean for j^{th} chain. n is number of draws in each chain (exluding the burn-in draws).

2) Calculate between-chain variability, $B = \sum_{j=1}^{m}(\bar{\tau}_{\cdot j} - \bar{\tau}_{\cdot\cdot})^2/(m-1)$, where $\bar{\tau}_{\cdot\cdot}$ is the overall sample mean for M chains.

3) The estimate of the target variance can be expressed as weighted average of within-

chain and between-chain variances, $Var(\tau) = \frac{n-1}{n}W + B$

4) The potential scale reduction ratio is the ratio of the estimated variance to within-variance, $\Omega = \sqrt{\frac{Var(\tau)df}{W(df-2)}}$. (Note that if $df \to \infty$ as $n \to \infty$, then $\frac{df}{df-2} \to 1$.) If the Markov chains converge to the target distribution, then the between-chain variance is diminished and the ratio is near 1. Otherwise, the ratio is greater than 1 and the Markov chains need more time to converge.

4.3.4 Simulation algorithm

Suppose the sample size that attained 90% power under balanced randomization is N, and the delayed primary outcome is D. Denote x and y as the auxiliary and primary outcomes.

The algorithm proceeds as follows:

1) Allocate patient i based on R_i; if i=1, R_i set to be 0.5.

2) Update R_{i+1} based on accumulated data from the 1st to the i^{th} patient. For treatment j, where j=A or B :

a. Obtain candidate $p^*_{1,j}$ and $p^*_{2,j}$ from kernel distributions $q(p^*_{1,j}, p^c_{1,j})$ and $q(p^*_{2,j}, p^c_{2,j})$.

- $p^c_{1,j}$ and $p^c_{2,j}$ are the current values of $p_{1,j}$ and $p_{2,j}$.
- kernel distributions $q(\cdot)$ for both $p_{1,j}$ and $p_{2,j}$ are chosen to be beta distributions; their location parameters α's are determined by $p^c_{1,j}$ and $p^c_{2,j}$ with specified scale βs. Scale βs are chosen to yield moderate variance. For simple reason, we fix βs to be 15.

b. Compute the densities $\pi\left(p^*_{2,j}, p^c_{1,j}, \boldsymbol{x_j}, \boldsymbol{y_j}\right)$, $\pi\left(p^*_{1,j}, p^c_{2,j}, \boldsymbol{x_j}, \boldsymbol{y_j}\right)$, and $\pi\left(p^c_{2,j}, p^c_{1,j}, \boldsymbol{x_j}, \boldsymbol{y_j}\right)$.

- $\boldsymbol{x^i_j}$, $\boldsymbol{y^i_j}$ are the corresponding vectors that contain auxiliary and primary outcomes

accumulated until the i^{th} stage in the j^{th} treatment group.

- $\pi(p_1, p_2, \boldsymbol{x}^i, \boldsymbol{y}^i) = \prod_{i=1}^{n_i} f(x_i, y_i|p_1, p_2) * f(p_1)f(p_2)c(p_1, p_2)$

c. Reject/accept $p^*_{1,j}$ and $p^*_{2,j}$ with probability α_1 and α_2.

- $\alpha_1 = \frac{\pi\left(p^*_{1,j}|p^c_{2,j},\boldsymbol{x_j},\boldsymbol{y_j}\right)}{\pi\left(p^c_{1,j}|p^c_{2,j},\boldsymbol{x_j},\boldsymbol{y_j}\right)} \frac{q(p^c_{1,j}|p^*_{1,j})}{q(p^*_{1,j}|p^c_{1,j})} = \frac{\pi\left(p^*_{1,j},p^c_{2,j},\boldsymbol{x_j},\boldsymbol{y_j}\right)}{\pi\left(p^c_{1,j},p^c_{2,j},\boldsymbol{x_j},\boldsymbol{y_j}\right)} \frac{q(p^c_{1,j}|p^*_{1,j})}{q(p^*_{1,j}|p^c_{1,j})}$

- $\alpha_2 = \frac{\pi\left(p^c_{1,j},p^*_{2,j},\boldsymbol{x_j},\boldsymbol{y_j}\right)}{\pi\left(p^c_{1,j},p^c_{2,j},\boldsymbol{x_j},\boldsymbol{y_j}\right)} \frac{q(p^c_{2,j}|p^*_{2,j})}{q(p^*_{2,j}|p^c_{2,j})}$

d. Repeat a, b, and c until convergence

e. Calculate $E(p_{2,j}|p_{1,j}, X_j, Y_j)$ by averaging overall MCMC sample of $p_{2,j}$ for M chains and $Var(p_2|p_1, X, Y)$ by averaging the mean square error for overall MCMC sample of $p_{2,j}$ or using sample variance.

3) Repeat (1) and (2) until i=N.

4.4 Result

In order to make results comparable with bivariate optimal design using bivariate beta distribution (see Chapter 3), we focus on the cases when primary efficacies are 0.1 vs 0.2, 0.1 vs 0.3 , and 0.3 vs 0.3 and auxiliary efficacies are 0.4 vs 0.6, 0.5 vs 0.6 with varying in associations, and there is a 30 patients lag between primary response and enrollment.

4.4.1 Gaussian copula simulation results

Recalling Chapter 3, we have used sample variances whenever it is available for the adaptation because it decreases the number of expected patient failures. In this copula

optimal design, we want to compare the performance of the adaptations when the estimates variances for both treatments are posterior variances and sample variances. Table 4.1 presents the number of patients allocated to the better treatment (trt B). It shows that cases using the sample variance assign more patients to better treatment than cases using the posterior variances. The difference in auxiliary efficacies does not affect the treatment assignments when the Pearson correlation of the copula is fixed ($\rho = 0.5$ or 0.8). When the difference of primary efficacies is smaller than that of auxiliary efficacies, it seems that moderate correlation ($\rho = 0.5$) case assigns more patients in treatment B than higher correlation ($\rho = 0.8$). It might be the impact of sample variances on assignment cancels out the correlation effect, so this trend is apparently showed in the scenario of posterior variance used.

Table 4.2 shows the expected number of patient failures. The Copula optimal method using the sample variance has a smaller number of expected patient failures at the cost of slightly larger standard deviation, compared with the method using posterior variance. Given paired primary efficacies, the numbers of patient failures are similar regardless of auxiliary efficacy difference and correlation for each method.

Table 4.3 depicts the power/error rate of the studies, which is estimated as the number of correct rejecting the null hypothesis out of all simulations. Error rate shows the chance of incorrectly rejecting the null hypothesis when there is no difference between treatments. Each case attains the level of 90% power no matter which estimated variances are utilized in the adaptation. It seems that the correlation associates with error rate level when auxiliary efficacy difference does not reflect the direction of treatment difference. When there is no treatment difference, the case with high correlation ($\rho = 0.8$) has slightly higher odds of making the wrong conclusion.

We are also interested in how the allocation rate adapts across the trials. Table 4.4 displays the allocation rate of treatment A at the 25^{th} percentile of visits (early stage). When treatment difference exists, the allocation rate starts to favor the better treatment (trt B) and assign less patient to treatment A. Otherwise, the allocation rate stays in balance (around 0.5). Compared with early stages, the allocation rate at the 50^{th} percentile visit (middle stage) gradually moves towards a smaller rate with a narrower IQR (see Table 4.5). When the study approaches the 75^{th} percentile visit, the allocation rate nearly remains at same rate as in middle stage of the trial (see Table 4.6).

4.4.2 Clayton copula simulation results

We have noticed that the design performs better using sample variance than using posterior variance in terms of expected number of patient failures. Therefore, the simulation for Clayton copula-based model will adopt sample variances to compare with the results of Gaussian copula-based model (i.e. we will not study the posterior variance). Table 4.7 presents the number of patients assigned to the better treatment group (group B). Compared with Gaussian copula-based model, the Clayton copula-based model assign less patients to group B by about 2 patients for the case when primary efficacy difference is small (0.1) or there is no difference. Regarding to expected number of patient failures (see Table 4.9), no difference exists between Gaussian and Clayton copula-based models. Table 4.8 shows the power/error rate of the studies using the Clayton copula-based model. When a treatment difference exists, the power remains around 90% as was observed in the Gaussian copula-based model. However, the type I error rates are smaller value using the Clayton copula-based model for the case when primary and auxiliary end points are highly correlated.

Table 4.10, 4.11, 4.12 summarize the shift of allocation rate of assigning patients to treatment A at the $25^{th}, 50^{th}, and 75^{th}$ percentiles of enrollment. At early stage of the enrollment, the allocation rates of assigning patient to treatment A for the cases (0.1 vs 0.2 and 0.3 vs 0.3) are slightly higher than that of Gaussian copula model. As the adaption continues, the allocation rates of Clayton copula-based model are similar to that of Guassian copula-based model.

4.5 Discussion

Unlike the bivariate binomial-beta model, copula-based bivariate response adaptive design gives us the flexibility to model the dependence between auxiliary and primary endpoints. We have compared the simulation results of applying Gaussian and Clayton copula in the model. We noticed that the Clayton copula-based model is slightly more conservative than Gaussian copula-based in terms of assigning patients to the better treatment when there is a small treatment difference. The Clayton copula-based model is better in preserving type I error for high dependent outcomes than the Gaussian copula based-model. We know that it is easy to control the power of the study by having sufficient sample size. However, the major challenge for an adaptive design is to control the inflated overall type I error when a shift of allocation rate exists (Feng et al., 2007). Due to the inherent unbalance allocation, the power of the study decreases. Therefore, the choice of the copula is critical to the study. As the primary efficacies are small, the number of expected patient failures would not be asymptotically different using either copula-based model or the bivariate binomial-beta model. The allocation rates depend on the unknown variance of the population. One thing we did notice is that variance estimates strongly

affect the allocation rates and indirectly change the expected number of patient failures. The adaptation performs better using empirical sample variances than using posterior variances. The problem with using sample variance is that it might not available at the early stage or with delayed responses. Hence, we suggest using the posterior variance at the early stage when sample variance is not estimable. Once the sample variance becomes available, we will apply sample variance instead of posterior variance.

Table 4.1: Summary Number of Patients in Group B (Gaussian Copula)

Primary		Auxiliary		Sample Size	Posterior variance [1]		Sample variance [2]	
trtA	trtB	trtA	trtB		ρ=0.5	ρ=0.8	ρ=0.5	ρ=0.8
0.1	0.3	0.40	0.60	162	89.50	90.00	99.00	98.30
		0.50	0.60		89.90	90.20	98.90	98.70
0.1	0.2	0.40	0.60	526	285.40	283.40	310.40	309.20
		0.50	0.60		285.70	283.80	309.60	308.20
0.3	0.3	0.40	0.60	200	99.70	97.80	100.30	99.30
		0.50	0.60		100.00	98.30	100.50	100.10

[1] using posterior variance to update the allocation rate.

[2] using sample variance to update the allocation rate.

ρ indicates the pearson correlation of the Gaussian copula function.

Table 4.2: Summary of Expected Number of Patient Failures (SD) (Gaussian Copula)

Primary		Auxiliary		Posterior variance		Sample variance	
trtA	trtB	trtA	trtB	ρ=0.5	ρ=0.8	ρ=0.5	ρ=0.8
0.1	0.3	0.40	0.60	127.7 (5.21)	127.7 (5.17)	125.9 (5.44)	126.4 (5.48)
		0.50	0.60	127.8 (5.18)	127.6 (5.20)	125.9 (5.30)	126.0 (5.33)
0.1	0.2	0.40	0.60	445.3 (8.28)	445.1 (8.03)	442.0 (8.47)	442.0 (8.77)
		0.50	0.60	444.8 (8.46)	445.1 (8.07)	442.1 (8.41)	442.2 (8.41)
0.3	0.3	0.40	0.60	139.8 (6.42)	139.8 (6.73)	139.9 (6.42)	140.3 (6.35)
		0.50	0.60	140.0 (6.42)	139.9 (6.23)	140.1 (6.53)	140.0 (6.45)

Table 4.3: Summary of Power/Error Rate (Gaussian Copula)

Primary		Auxiliary		Posterior variance		Sample variance	
trtA	trtB	trtA	trtB	ρ=0.5	ρ=0.8	ρ=0.5	ρ=0.8
0.1	0.3	0.40	0.60	0.92	0.92	0.90	0.90
		0.50	0.60	0.92	0.90	0.92	0.92
0.1	0.2	0.40	0.60	0.90	0.91	0.91	0.90
		0.50	0.60	0.90	0.91	0.91	0.90
0.3	0.3	0.40	0.60	0.06*	0.06*	0.05*	0.07*
		0.50	0.60	0.04*	0.06*	0.05*	0.06*

Note: $*$ indicates error rates.

Table 4.4: Summary of Allocation Rate (IQR) at the 25^{th} Percentile Visit for Trt A (Gaussian Copula)

Primary		Auxiliary		Posterior variance		Sample variance	
trtA	trtB	trtA	trtB	ρ=0.5	ρ=0.8	ρ=0.5	ρ=0.8
0.1	0.3	0.40	0.60	0.46 (0.45, 0.47)	0.46 (0.45, 0.47)	0.39 (0.30, 0.47)	0.39 (0.29, 0.47)
		0.50	0.60	0.46 (0.45, 0.47)	0.46 (0.44, 0.47)	0.36 (0.30, 0.47)	0.37 (0.29, 0.47)
0.1	0.2	0.40	0.60	0.46 (0.44, 0.47)	0.46 (0.45, 0.47)	0.44 (0.32, 0.48)	0.44 (0.32, 0.48)
		0.50	0.60	0.46 (0.45, 0.47)	0.46 (0.45, 0.47)	0.43 (0.32, 0.47)	0.44 (0.31, 0.48)
0.3	0.3	0.40	0.60	0.50 (0.49, 0.51)	0.50 (0.49, 0.51)	0.50 (0.43, 0.56)	0.50 (0.44, 0.56)
		0.50	0.60	0.50 (0.49, 0.50)	0.50 (0.49, 0.51)	0.49 (0.43, 0.53)	0.50 (0.43, 0.56)

Table 4.5: Summary of Allocation Rate (IQR) at the 50^{th} Percentile Visit for Trt A (Gaussian Copula)

Primary		Auxiliary		Posterior variance		Sample variance	
trtA	trtB	trtA	trtB	ρ=0.5	ρ=0.8	ρ=0.5	ρ=0.8
0.1	0.3	0.40	0.60	0.44 (0.43, 0.46)	0.44 (0.43, 0.46)	0.37 (0.30, 0.42)	0.38 (0.31, 0.43)
		0.50	0.60	0.44 (0.43, 0.46)	0.44 (0.42, 0.46)	0.37 (0.31, 0.42)	0.38 (0.31, 0.42)
0.1	0.2	0.40	0.60	0.46 (0.44, 0.47)	0.47 (0.45, 0.48)	0.41 (0.35, 0.46)	0.41 (0.35, 0.46)
		0.50	0.60	0.46 (0.44, 0.48)	0.46 (0.45, 0.48)	0.41 (0.34, 0.46)	0.41 (0.35, 0.46)
0.3	0.3	0.40	0.60	0.50 (0.49, 0.52)	0.52 (0.50, 0.53)	0.50 (0.47, 0.53)	0.50 (0.48, 0.53)
		0.50	0.60	0.50 (0.49, 0.52)	0.52 (0.50, 0.53)	0.50 (0.47, 0.53)	0.50 (0.48, 0.53)

Table 4.6: Summary of Allocation Rate (IQR) at the 75^{th} Percentile Visit for Trt A (Gaussian Copula)

Primary		Auxiliary		Posterior variance		Sample variance	
trtA	trtB	trtA	trtB	ρ=0.5	ρ=0.8	ρ=0.5	ρ=0.8
0.1	0.2	0.40	0.60	0.46 (0.44, 0.47)	0.47 (0.45, 0.48)	0.41 (0.35, 0.45)	0.41 (0.36, 0.45)
		0.50	0.60	0.46 (0.44, 0.47)	0.46 (0.44, 0.48)	0.41 (0.36, 0.45)	0.41 (0.37, 0.46)
0.1	0.3	0.40	0.60	0.44 (0.42, 0.45)	0.44 (0.42, 0.45)	0.37 (0.32, 0.41)	0.38 (0.32, 0.41)
		0.50	0.60	0.44 (0.42, 0.45)	0.43 (0.42, 0.45)	0.37 (0.32, 0.41)	0.37 (0.33, 0.41)
0.3	0.3	0.40	0.60	0.51 (0.49, 0.52)	0.52 (0.5, 0.53)	0.50 (0.48, 0.52)	0.51 (0.48, 0.53)
		0.50	0.60	0.50 (0.49, 0.52)	0.52 (0.50, 0.53)	0.50 (0.48, 0.52)	0.50 (0.48, 0.53)

Table 4.7: Summary Number of Patients in Group B (Clayton Copula)

Sample Size	Primary		Auxiliary		Sample variance	
	trtA	trtB	trtA	trtB	γ=2	γ=8
162	0.1	0.3	0.40	0.60	99.00	98.70
			0.50	0.60	99.30	99.30
526	0.1	0.2	0.40	0.60	307.40	304.4
			0.50	0.60	308.40	– –[1]
200	0.3	0.3	0.40	0.60	98.80	98.00
			0.50	0.60	99.00	98.60

[1] Based on results of the cases where $\gamma = 8$, I terminate the simulation of this case.

Table 4.8: Summary of Power/Error rate (Clayton Copula)

Primary		Auxiliary		Sample Variance	
trtA	trtB	trtA	trtB	γ=2	γ=8
0.1	0.3	0.40	0.60	0.92	0.91
		0.50	0.60	0.91	0.91
0.1	0.2	0.40	0.60	0.91	0.91
		0.50	0.60	0.90	– –
0.3	0.3	0.40	0.60	0.06	0.06
		0.50	0.60	0.05	0.04

Table 4.9: Summary of Expected Number of Patient Failures (SD) (Clayton Copula)

Primary		Auxiliary		Sample Variance	
trtA	trtB	trtA	trtB	γ=2	γ=8
0.1	0.3	0.40	0.60	126.1 (5.64)	126.3 (5.38)
		0.50	0.60	126.1 (5.65)	126.1 (5.66)
0.1	0.2	0.40	0.60	442.6 (8.92)	443.3 (8.59)
		0.50	0.60	442.3 (8.3)	– –
0.3	0.3	0.40	0.60	140.3 (6.64)	139.8 (6.46)
		0.50	0.60	139.8 (6.49)	140.2 (6.28)

Table 4.10: Summary of Allocation Rate (IQR) at the 25^{th} Percentile Visit for Trt A (Clayton Copula)

Primary		Auxiliary		Sample Variance	
trtA	trtB	trtA	trtB	γ=2	γ=8
0.1	0.3	0.40	0.60	0.41 (0.33, 0.48)	0.4 (0.32, 0.48)
		0.50	0.60	0.39 (0.32, 0.48)	0.38 (0.32, 0.47)
0.1	0.2	0.40	0.60	0.44 (0.36, 0.51)	0.44 (0.35, 0.52)
		0.50	0.60	0.43 (0.35, 0.50)	– –
0.3	0.3	0.40	0.60	0.5 (0.43, 0.57)	0.5 (0.44, 0.56)
		0.50	0.60	0.51 (0.45, 0.57)	0.51 (0.44, 0.57)

Table 4.11: Summary of Allocation Rate (IQR) at the 50^{th} Percentile Visit for Trt A (Clayton Copula)

Primary		Auxiliary		Sample Variance	
trtA	trtB	trtA	trtB	γ=2	γ=8
0.1	0.3	0.40	0.60	0.37 (0.30, 0.42)	0.38 (0.31, 0.42)
		0.50	0.60	0.37 (0.31, 0.42)	0.37 (0.30, 0.42)
0.1	0.2	0.40	0.60	0.41 (0.35, 0.47)	0.42 (0.36,0.47)
		0.50	0.60	0.41 (0.34, 0.47)	– –
0.3	0.3	0.40	0.60	0.51 (0.48, 0.54)	0.52 (0.48, 0.55)
		0.50	0.60	0.51 (0.48, 0.54)	0.51 (0.48, 0.54)

Table 4.12: Summary of Allocation Rate (IQR) at the 75^{th} Percentile Visit for Trt A (Clayton Copula)

Primary		Auxiliary		Sample Variance	
trtA	trtB	trtA	trtB	γ=2	γ=8
0.1	0.3	0.40	0.60	0.37 (0.32, 0.40)	0.38 (0.33, 0.41)
		0.50	0.60	0.37 (0.32, 0.41)	0.37 (0.32, 0.41)
0.1	0.2	0.40	0.60	0.41 (0.36, 0.46)	0.42 (0.37, 0.46)
		0.50	0.60	0.41 (0.36, 0.46)	– –
0.3	0.3	0.40	0.60	0.51 (0.48, 0.54)	0.52 (0.49, 0.54)
		0.50	0.60	0.51 (0.48, 0.54)	0.51 (0.49, 0.54)

Chapter 5

Discussion and Future Work

Discussion

Using both primary and auxiliary outcomes provides us an alternative adaptive allocation method, where we attempt to shift the probability of allocation toward assigning more patients to the "better" treatment when a primary outcome can not be rapidly collected. The adaptive design method is derived from the minimization of a specified loss function with constraints on the variability of each adaptive stage, where the loss function is a linear combination of weights and number of patients in each treatment group. The traditional adaptive design has the weight related to primary efficacy. Therefore, we proposed a statistical model to connect auxiliary and primary outcomes and calculate the posterior conditional expectation of primary efficacy (P_2, which is parameter of interest), given auxiliary efficacy (P_1). We then substitute conditional expectations as the weights on the treatment groups. The most promising advantage of this modeling is that we allow the adaptive allocation without neglecting information of primary outcome, which allows quicker adaptation through the use of the auxiliary outcome.

The two-dimensional Bayesian model is complicated, in the terms of prior selection and model fitting. The first statistical model presented is based on the two dimensional beta-binomial distribution to describe the correlated auxiliary and primary outcome data. We studied the performance of a three-parameter bivariate beta distribution (Olkin and Liu, 2003) and a Morgenstern type five-parameter bivariate beta distribution with fixed marginal distributions. We noted that the three-parameter bivariate beta distribution does not allow correlation to range freely between $[-1, 1]$ when the marginals are fixed. If both marginal means are small, the correlation of the marginal is negligible or weak. As both marginal probabilities increase, the correlation falls in range from moderate to strong. Compared with the three-parameter bivariate beta distribution, the five-parameter bivariate beta distribution is even more restrictive on the correlation, where the maximum absolute correlation is $\frac{1}{3}$. In general, the three-parameter bivariate beta distribution is a better choice as prior for its simplicity and flexibility.

We ran a simulation study using the proposed statistical model with the three-parameter bivariate beta prior distribution. The purpose of the simulation is to see how the bivariate adaptive method performs when the prior correlation information is different from that which the simulated data is based on. In this study, we focused on particular cases where the primary outcome is a rare event and in delay, where the probability varies from 0.1 to 0.3. The auxiliary outcome is a short-term response and can be collected relatively quicker (at least quicker than patient enrollment), where its probability ranges from 0.4 to 0.7. Compared with the univariate optimal adaptive design, the bivariate optimal adaptive design assigns more patients to the better treatment and reduces patient deaths by 1 or 2 when a treatment difference exists, while maintaining similar power and error rates. As we have seen, the advantage of the bivariate optimal method lies in the practi-

cal clinical scenarios, especially for survival studies (e.g. cancer studies) where complete remission is the primary endpoint. It seems that the auxiliary efficacy difference does not affect the simulation results as long as the auxiliary efficacy difference has the same direction of the treatment difference. The similarity might be a result of weak correlation between auxiliary and primary efficacies when primary efficacies are relatively small.

We therefore introduced copula-based bivariate distributions to our statistical model. The copula method allows us to separately model marginal distributions and the dependence structure of the joint distribution. And we are also able to compare the results of bivariate optimal adaptive method when the association of auxiliary and primary outcomes is moderate and strong. The choice of the copula is important and difficult. On one hand, all information about the dependence structure is contained in the copula function. On the other hand, no systematic method guarantees the selected copula to approximate the real dependence structure. We first studied the Gaussian copula, due to its simplicity, to construct the joint bivariate distribution. We then also applied the Clayton copula to the proposed model, based on the characteristics of clinical scenarios which require high dependence between correlated small events data. Comparing both results, we have noticed that the allocation rates using the Clayton copula favors the "better" treatment more consistently than that using the Gaussian copula. In other words, the IQR of the allocation rate using the Clayton copula is much wider than that of using Gaussian copula, even though the medians are equivalent. This leads to the phenomena that the expected number of patient failures is comparable using both copulas but the standard deviation using the clayton copula is slightly larger than that of using Gaussian copula.

The copula-based prior gives us the flexibility to model the dependence of correlated

data free of specified marginal distributions, even though the simulated adaptation results using the copula-based prior and the three-parameter bivariate beta distribution are similar. The copula-based prior also makes the bivariate optimal adaptive method more applicable to a clinical trial study in practice, where we take advantage of historical information of the endpoints to modify marginal distributions of priors as well as the dependence structure. Our proposed method is suitable for the clinical scenario when the delay in response in not stochastic and the endpoints take a long time to observe.

Future work

In this study, we fixed the number of delayed patient to be 30, meaning the primary response becomes available when there are 30 patients enrolled in the trial, which does not reflect real clinical trial setting, though it does accurately reflect a delayed response. The delayed response depends on response time and treatment assignment. For instance, the delay in responses will be longer for the inferior treatment group which has smaller efficacy. Therefore, to model the delayed response for each treatment following parametric distribution (e.g. exponential distribution) would be more appropriate and promising. Meanwhile, it is common in long-term clinical trials to encounter the presence of censoring, such as patient death or loss follow-up. The consideration of censoring in the simulation of delayed primary responses is important to study the performance of the bivariate optimal design.

It is important and challenging to make a right choice of copula for the bivariate joint distribution that suits for describing the dependence of auxiliary and primary outcomes. In this study, we only chose the Gaussian and Clayton copulas to be the representatives for empirical copulas and Archimedean copulas. In the future, we may explore other copulas

such as student t copula, frank copula, etc. to construct the joint bivariate distribution with marginal beta distributions. To make our results comparable with the model using the bivariate beta distribution, we choose to apply a joint copula-based prior instead of using copula method to model joint distribution of primary and auxiliary outcomes. Other future work may involve jointly modeling the likelihood using copula method and have independent priors for the bivariate optimal adaptive design. In the earlier chapters, we have mentioned that the concurrent bivariate binomial distributions are not an idea choice to be the likelihood for the positively correlated binomial random variables. Using copula methods, we are able to overcome the limitation of the existing bivariate binomial distribution on the correlation.

References

M. Abramowitz and I.A. Stegun. Handbook of mathematical functions. 55, 1964.

AC Aitken and HT Gonin. On fourfold sampling with and without replacement. *Proceedings of the Royal Society of Edinburgh*, 55:114–125, 1935.

ZD Bai, Feifang Hu, and William F Rosenberger. Asymptotic properties of adaptive designs for clinical trials with delayed response. *The Annals of Statistics*, 30(1):122–139, 2002.

Wilfrid Norman Bailey. Generalized hypergeometric series. 1935.

Bo Martin Bibby and Michael Væth. The two-dimensional beta binomial distribution, 2011.

Atanu Biswas and Jing-Shiang Hwang. A new bivariate binomial distribution. *Statistics & probability letters*, 60(2):231–240, 2002.

Tomasz Burzykowski, Geert Molenberghs, Marc Buyse, Helena Geys, and Didier Renard. Validation of surrogate end points in multiple randomized clinical trials with failure time end points. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(4):405–422, 2001.

Shein-Chung Chow, Mark Chang, and Annpey Pong. Statistical consideration of adaptive

methods in clinical development. *Journal of biopharmaceutical statistics*, 15(4):575–591, 2005.

Persi Diaconis and Donald Ylvisaker. Conjugate priors for exponential families. *The Annals of statistics*, 7(2):269–281, 1979.

Huaibao Feng, Jun Shao, and Shein-Chung Chow. Adaptive group sequential test for clinical trials with changing patient population. *Journal of biopharmaceutical statistics*, 17(6):1227–1238, 2007.

Thomas R Fleming, Ross L Prentice, Margaret S Pepe, and David Glidden. Surrogate and auxiliary endpoints in clinical trials, with potential applications in cancer and aids research. *Statistics in medicine*, 13(9):955–968, 1994.

Maurice Fréchet. Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon Sect. A*, 9:53–77, 1951.

Gianluca Fusai and Andrea Roncoroni. Implementing models in quantitative finance: methods and cases, 2008.

Paul Gallo, Christy Chuang-Stein, Vladimir Dragalin, Brenda Gaydos, Michael Krams, and José Pinheiro. Adaptive designs in clinical drug developmentan executive summary of the phrma working group. *Journal of biopharmaceutical statistics*, 16(3):275–283, 2006.

AK Gupta and CF Wong. On three and five parameter bivariate beta distributions. *Metrika*, 32(1):85–91, 1985.

MA Hamdan. Canonical expansion of the bivariate binomial distribution with unequal

marginal indices. *International Statistical Review/Revue Internationale de Statistique*, 40(3):277–280, 1972.

MA Hamdan and DR Jensen. A bivariate binomial distribution and some applications 1. *Australian Journal of Statistics*, 18(3):163–169, 1976.

W Hoeffding. Masstabinvariante korrelationstheorie für diskontinuiereliche verteilungen. *Arkiv fr matematischen Wirtschaften und Sozialforschung*.

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

Feifang Hu and William F Rosenberger. *The theory of response-adaptive randomization in clinical trials*, volume 525. Wiley, 2006.

Anastasia Ivanova. A play-the-winner-type urn design with reduced variability. *Metrika*, 58(1):1–13, 2003.

Piotr Jaworski. *Copula theory and its applications: proceedings of the workshop held in Warsaw, 25-26 September 2009*, volume 198. Springer, 2010.

Christopher Jennison and Bruce W Turnbull. *Group sequential methods with applications to clinical trials*. CRC Press.

MC Jones. Multivariate t and beta distributions associated with the multivariate f distribution. *Metrika*, 54(3):215–231, 2002.

W David Kelton and Averill M Law. *Simulation modeling and analysis*. McGraw Hill Boston, MA, 2000.

Lung-Fei Lee. Generalized econometric models with selectivity. *Econometrica: Journal of the Econometric Society*, 51(2):507–512, 1983.

Rajiv Mahajan and Kapil Gupta. Adaptive design clinical trials: Methodology, challenges and prospect. *Indian journal of pharmacology*, 42(4):201, 2010.

Albert W Marshall and Ingram Olkin. A family of bivariate distributions generated by the bivariate bernoulli distribution. *Journal of the American Statistical Association*, 80(390):332–338, 1985.

Vincent Melfi and Connie Page. Variability in adaptive designs for estimation of success probabilities. *Lecture Notes-Monograph Series*, 34:106–114, 1998.

Dietrich Morgenstern. Einfache beispiele zweidimensionaler verteilungen. *Mitt. Math. Statist*, 8(1):234–235, 1956.

Roger B Nelsen. *An introduction to copulas*. Springer, 2006.

Ingram Olkin and Ruixue Liu. A bivariate beta distribution. *Statistics & Probability Letters*, 62(4):407–412, 2003.

TG Pham and N Turkkan. Bayes binomial sampling by attributes with a general-beta prior distribution. *Reliability, IEEE Transactions on*, 41(2):310–316, 1992.

Ross L Prentice. Surrogate endpoints in clinical trials: definition and operational criteria. *Statistics in medicine*, 8(4):431–440, 1989.

William F Rosenberger and John M Lachin. The use of response-adaptive designs in clinical trials. *Controlled clinical trials*, 14(6):471–484, 1993.

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

William R Schucany, William C Parr, and John E Boyer. Correlation structure in farlie-gumbel-morgenstern distributions. *Biometrika*, 65(3):650–653, 1978.

Kenneth F Schulz and David A Grimes. Generation of allocation sequences in randomized trials: chance, not choice. *The Lancet*, 359(9305):515–519, 2002.

Joanna H Shih and Thomas A Louis. Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, pages 1384–1399, 1995.

Abe Sklar. Random variables, joint distribution functions, and copulas. *Kybernetika*, 9 (6):449–460, 1973.

Henry Teicher. On the multivariate poisson distribution. *Scandinavian Actuarial Journal*, 1954(1):1–9, 1954.

William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933.

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

Peter Xue-Kun Song. Multivariate dispersion models generated from gaussian copula. *Scandinavian Journal of Statistics*, 27(2):305–320, 2000.

Vita

Shuxian Z. Sinks was born on Oct 2, 1987 in Zunyi, Guizhou and grew up in Changsha, Hunan, China. In 2009, she received her Bachelor of Science Degree in Pharmacy and Marketing from Shenyang Pharmaceutical University. She came to the United States and began her graduate education in the Department of Biostatistics at Virginia Commonwealth University in Richmond, Virginia in the fall of 2009.