Practical and Optimal Crossover Designs for Clinical Trials

by

Su Hwan Kim

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Statistics

Department of Mathematical and Statistical Sciences

University of Alberta

Abstract

This thesis investigates various issues arising from crossover designs, which received great attention for their efficiency. Crossover designs gain advantage over parallel designs for efficient estimation of treatment effects and smaller required sample size as the within-subject variability is in general smaller than between-subject variability. We investigate optimal crossover designs under various assumptions and objectives.

The within-subject comparisons allow subjects to serve as their own controls. Such within-subject comparisons remove nuisance subject effects. The direct treatment effects and carryover effects, which are portion of the direct treatment effects being carried from one period to the next, are defined through various assumptions. Often, washout periods between treatments are applied between periods to minimize the carryover effects. However, it is often difficult to precisely determine or practically implement the sufficiently long washout period. For this reason, optimal designs were constructed with carryover effects.

First, we investigate the effect of unequal treatment variances. Traditionally, the optimal designs were constructed with an assumption that all treatments being tested have equal variances. However, this assumption may be too naive to describe designs and experiments that are increasingly becoming more complex. Therefore, we investigate how the unequal treatment variability affects the efficiency of the existing optimal designs and construct appropriate optimal designs with unequal treatment variability.

Second, we investigate the assumption that the carryover effects are proportional to the direct treatment effects. When a constant washout period is applied and failed to remove the carryover effects completely, the existing carryover effects may be described as a proportion of the direct treatment effect with an assumption that the proportion may be similar for all treatments being compared. Under this model, we investigate the benefits of adding baseline measurements where a portion of the direct treatment effects remain and can be described by another proportion.

Lastly, we investigate response adaptive crossover designs with two different objective functions. Adaptive designs were designed to allow clinical trials to respond to the information acquired during the trials to achieve specific objectives, which could include but not limited to allocating more subjects to superior treatments, improving statistical efficiency, reducing the sample size for cost consideration, increasing the sample size to maintain pre-specified statistical power, or including covariates. In this chapter, we apply a multiple objective function to find balance between treatment effects and statistical efficiency and propose a new allocation method that achieves balance between the two objectives.

Acknowledgements

First of all, I would like to thank my supervisor Dr. Keumhee Chough for all her supports, encouragements, advice, caring. Without her continuous care and support, I would not have finished this thesis. Through the course of her supervision, I was able to learn the knowledge, skills, professionalism, and work ethics required for statistician and researcher. I am truly grateful to her for introducing me to the study of crossover designs, a wonderful and fulfilling field of research that I plan to continue my research on.

I would like to thank my supervisory committee members as well as the exam committee members, Dr. Bei Jiang, Dr. Adam Kashlak, Dr. Peter Minev and Dr. Vernon Chinchilli, for all the valuable and insightful comments they provided from the Candidacy Exam to the end. Also, I would like to thank Dr. Weng Kee Wong for his valuable suggestions and comments.

I would like to thank the Department of Mathematical and Statistical Sciences, Dr. Jochen Kuttler, Tara Schuetz, and all the other supporting staff members who provided excellent support during the course of my program.

Lastly I would like to thank my family. Without the support of my wife, Xi Chen, I could never have started journey nor finished it.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

This thesis investigates three interesting issues arising from crossover designs, which is a special case of repeated measure designs. Crossover designs allocate subjects to sequences of treatments in two or more periods. In crossover trials, subjects serve as its own control, thereby effectively remove subject specific nuisance effects. With a general notion that within-subject variability is smaller than between subject variability, crossover designs have advantage over parallel designs with respect to statistical efficiency. As with the nature of repeated measure designs, within-subject measurements are most often correlated. In addition, crossover designs must deal with compound treatment effects and may use washout periods to prevent the effects of treatments in a period from affecting the subsequent periods. Such lasting effects beyond the period of administration are called carryover effects and washout periods are hoped to minimize them. However, it is difficult to determine and implement an adequate length of washout periods. For this reason, studies on optimal crossover designs proposed various assumptions on the carryover effects. Moreover, optimality of crossover designs is model dependent on the model, the assumptions, the parameters included in the model, and the type of responses, which could be either continuous or binary.

Grizzle (1965) showed that cross-over designs for two treatments and two periods can be constructed so that direct treatment effects and carryover effects can be estimated separately. The carryover effects were mostly assumed

to last one period and depend on the treatment administered in the prior period. Many investigators conducted crossover designs and the analysis under this model and thus we call this model as the traditional model(Carriere (1994), Carriere and Reinsel (1992), Carriere and Reinsel (1993), Cheng and Wu (1980), Freeman (1989) , Hedayat and Afsarinejad (1975), Hedayat and Afsarinejad (1978), Kershner and Federer (1981), Kifer (1973), Kiefer (1975), Kunert (1983), Kunert (1985)). Bishop and Jones (1984), Bose and Mukherjee (2000), and Divecha and Gondaliya (2015) considered higher order carryover effects where the carryover effects were assumed to last more than one period. Moreover, Hedayat and Afsarinejad (2002) and Kunert and Stufken (2002, 2008) considered a self and mixed carryover effects where the carryover effects depend on a pair of treatments administered on the current and previous periods.

Kiefer (1975) proposed a strategy of constructing $A/D/E/T$ optimal designs by using the properties of the Information matrix and the covariance matrix of parameters. A design d is said to be universally optimal, i.e. $A/D/E$ optimal, if the Information matrix I is completely symmetric and its trace is maximized (T optimal). The result is a T optimal design. Completely symmetric matrices can be written in the form of $aI_p + b1_p1_p'$ and have zero row sums. This has been the basis for many researches investigating optimal designs with various other model assumptions, which include Kunert (1983) and Kushner (1997).

Baseline measurements were introduced to improve the statistical efficiency of crossover designs and have been utilized in various ways. A natural approach for the use of the baseline measurements was to take a difference from baseline measurements as the responses (Wallenstein (1979), Kershner and Federer (1981), Laska, Meisner, and Kushner (1983)) whereas other researchers pointed out that it may not be appropriate as this approach often results in bias (Hills and Armitage (1979), Willan and Pater (1986), Wallensteins and Fisher (1977)). Another approach was developed under an assumption that baseline measurements are positively correlated with treatment responses (Wallensteins and Fleiss (1988), Kenward and Roger (2009), Jemielita (2017)). This method

treats baseline measurements or combinations of baseline measurements as covariates, eliminating the problem of biased estimation but optimal designs can not be constructed under this method.

Most optimal designs are fixed in the planning stage of experiments in a sense that a predetermined sample of subjects are recruited as planned. Adaptive designs are gaining popularity as they utilize the information gathered during the course of experiments. When one treatment happens to be superior to another, ethical issues may arise as some subjects receive more of better treatments and vice versa. This led to the development of numerous adaptive allocation schemes based on the treatment effects (Zelen (2003), Wei and Durham 1978, Bandyopadhyay, Biswas, and Mukherjee (2009a), Bandyopadhyay, Biswas, and Mukherjee (2009b)). Furthermore, acquired responses were used to verify the determined sample size, which was estimated in the planning stage of the experiments. In these adaptive designs, the sample size and power may be updated (Armitage (1975), Wang (2014)).

1.2 Thesis Overview

In Chapter 2, we deal with the first of the three issues and construct optimal designs when the treatment variances are unequal. The notion of unequal treatment variance has been widely studied and discussed in the case of parallel designs. Levene (1960) developed Levene's test to measure the difference in the variances of treatments and Welch (1947) developed Welch's t-test as an alternative to parametric and non-parametric t tests for treatments with unequal variances (Gosset (1908), Wilcoxon (1945)). However, it did not receive much attention in the construction of optimal crossover designs. For this reason, we discuss the effects of unequal treatment variances on the construction of optimal two or three-period and two-treatment crossover designs. The optimal designs are constructed using the orthogonality method proposed by Kunert (1983).

In Chapter 3, we discuss the second issue, namely the use of baseline measurements, and construct optimal designs when the carryover effects are

proportional to the direct treatment effects. The use of baseline measurements were discussed and debated over the years. The baselines were first considered as a reference to the treatment responses and the difference from baselines were used as responses. Wallensteins and Fleiss (1988) showed that this method is appropriate only when the correlation matrix is determined by a correlation parameter and a length of washout period. Subsequent studies considered baselines or functions of baselines as covariates (Kenward and Roger (2009),Jemielita (2017)). Liang and Carriere (2010) applied a modeling approach to baseline measurements and investigated the effect of including baseline measurements in the design efficiency. The modeling approach indeed utilizes the most information we can obtain from the baseline measurements and thus is preferred over the use of changes from baselines and covariate modelling approaches. In this chapter, the carryover effects in the treatment responses and baselines are assumed to be the traditional carryover effects and are proportional to the direct treatment effects. We update the universal optimality theorem of Kushner (1997) for baseline models and link the universal optimal designs under the traditional model to the E optimal designs under the proportional model. Then, we construct optimal designs for various proportionality parameters.

Lastly, chapter 4 addresses the third issue, which is the thical issue arising from clinical trials in the use of optimal designs. Response adaptive designs were developed to allow the utilization of the information gathered in the course of experiments to fulfill various objectives. Such adaptive designs are constructed from a single objective and often shown to be not practical. Liang and Carriere (2009) proposed a multiple objective response adaptive design where multiple objectives are compromised using weight parameters. Also, Yi and Wang (2009) and Li and Wang (2012) proposed a variance penalized criterion where a scalar multiple of the variance of the total number successes is subtracted from the mean total number of successes to maintain balance between two goals: to allocate more subjects to better treatments and to reduce the variance. In this chapter we construct response adaptive designs based on the two criteria above using the Generalized Estimating Equations approach

and compare them to find the better of the two methods with respect to the treatment effects and efficiencies. Then, we compare the method to other response adaptive designs suggested by Li (2017) and Bandyopadhyay, Biswas, and Mukherjee (2007) with respect to the allocations, treatment effects, and efficiencies.

Chapter 2

Optimal Crossover Design Under Unequal Treatment Variance Assumption

Abstract

Crossover designs allow within-subject differences to estimate the effect of treatments being studied. These designs gain advantages over parallel designs in terms of estimation efficiency and reduced sample size as the within-subject variability is in general smaller than between-subject variability in repeated measures data. Optimal crossover designs have been generally constructed based on an equal variance and covariance assumptions. Although it has served its purpose well, such an assumption may be too naive to describe designs and experiments that are increasingly becoming more complex. As these designs often involve small samples, and some can actually be inappropriate or inferior in practice, we construct optimal crossover designs under realistic considerations of variance assumptions. We also develop and extend a test for variance equality to repeated measures data in the presence of treatment effects. We find that the efficiency of existing optimal designs can be as bad as less than 60% of the constrained T optimal designs under the unequal variance assumptions and this is even more so with the popular two-treatment two-period designs.

2.1 Introduction

A crossover design is a special case of repeated measurement designs where experimental units receive multiple treatments and one measurement is taken from each treatment measurements are taken from each. In crossover designs, within-subject measurements are compared for the efficacy of the test treatment. Such within-subject comparisons remove nuisance subject effects and are shown to have less variability compared to between-subject comparisons. This property alone makes crossover designs popular with increased efficiency, requiring fewer samples than parallel designs. In every subsequent period of crossover designs, some effects are carried over from previous periods.

Optimal crossover designs for continuous responses have been constructed, based on the Fisher's Information, obtainable from considering a plausible model. However, they are highly model dependent, which means changing the model assumptions can have critical impact on the optimality of the constructed designs. Several carryover effect assumptions are available. Traditionally the carryover effects were assumed to last only one period and remain constant for any combination of treatments that are administered in the current and previous periods. It is also assumed that carryover effects are not present in the first period. Saha (1983), Laska, Meisner, and Kushner (1983), Jones, Kunert, and Wynn (1992) and Carriere and Reinsel (1992) considered the traditional model with random subject effects, where the repeated responses of subjects were considered to have an equicorrelated correlation structure. Bose and Mukherjee (2000) and Carriere (1994) considered higher order carryover effects. Kunert and Stufken (2002) proposed a self and mixed carryover effects model where carryover effects exist for one period and depend on the pair of treatments. Cheng and Wu (1980) assumed that subject effects are fixed whereas Carriere and Reinsel (1993) assumed that subject effects are random. Hence, the literature on optimal crossover designs have heavily focused mostly on how we assume these carryover effects and the general strategy has been to assume non-ignorable carryover effects. This chapter also considers finding optimal crossover designs with unequal carryover effects,

while focusing on another practical issue inherent with these designs.

When involving multiple groups, it is natural to observe that populations have varying degrees of magnitude in variances. In response to this issue, Welch (1947) proposed a generalization of the Student's t-test for populations with unequal variances. The Welch's t-test was proven to be robust against unequal variances and almost as efficient as the Student's t test and Mann-Whitney U test even when the population variances are equal and sample sizes are balanced (Derrick and White (2016)). Bartlett (1937) proposed a test of equality of variances based on the normality assumption and Levene (1960) proposed a robust test against the normality assumption, emphasizing its importance in ensuring tests. However, the studies on optimal crossover designs have not given considerations on the effect of unequal treatment variances. Therefore, it is unclear whether the previously studied optimal designs can still remain optimal or robust when treatment variances are unequal.

A number of crossover trials reported results, indicating that variances of responses for multiple treatments are markedly different. For example, Kovalchuk et al. (2018) reported that TEd90-100 group and the control group had a variance ratio of 5.22. Romano et al. (2019) found that heart rates for the fentanyl group had a variance of 289 beats per minutes whereas the saline group had a variance of 16 beats per min, exhibiting a variance ratio of 18 times over the saline group. Clearly, these exhibit a variance inequality problem. Therefore, there is a need to ensure that the design being adopted in practice is defensible, optimal or robust against unequal variances. Further, a test of unequal treatment variances for repeated measure data is needed to readily detect the problem, if existed, along with the effect of unequal treatment variances on the efficiency of existing popular crossover designs.

Crossover designs with $p = 2$ and $t = 2$ are of particular importance as these designs are widely popular in practice. Carriere and Reinsel (1993) proved that design that equally allocates on all four sequences (AA/BB/ AB/BA) is universally optimal in the presence of carryover effects. However, the crossover design allocating equally only on sequences AB and BA is naturally popular in practice. Grizzle (1965), Freeman (1989), and other

researchers discussed the limitations of the design with sequences AB and BA in the presence of carryover effects and proposed to use baselines, more treatment sequences, or more periods as possible solutions. As a result, optimal three-period crossover designs gained some popularity. However, all assumes the typical equal variance assumption in measurement errors.

This chapter aims to construct optimal crossover designs under the assumption of unequal treatment variances, unequal carryover effects and random subject effects. We consider efficiencies of the new optimal designs over the existing designs and discuss some practical alternatives. Section 2 will first review crossover design models and optimality criteria. Section 3 constructs new optimal designs and explores the comparative efficiencies. Finally, Section 4 will develop tests for equal treatment variances. followed by the conclusion.

2.2 Model

2.2.1 Crossover Design Models

Denote the class of crossover designs with t treatments, p periods, and n subjects by $\Omega_{t,p,n}$. Let Y_{ij} denote the continuous response variable in the period i from subject j, and we consider the following model:

$$y_{ij} = \mu + \pi_i + \tau_{d(i,j)} + \gamma_{d(i-1,j)} + \xi_j + \epsilon_{ij} \tag{2.1}$$

where μ, π_i, $\tau_{d(i,j)}$, $\gamma_{d(i-1,j)}$, ξ_j and ϵ_{ij} refer to the overall mean effect, periodic effects, direct treatment effects, carryover effects, random subject, and measurement errors for $i = 1, 2, \cdots, p$, $j = 1, 2, \cdots, n$, and $k = 1, 2, \cdots, K$. The ξ_j and ϵ_{ij} are identically and independently distributed with a mean of 0 and variances of σ_s^2 and σ_ϵ, respectively. Then, $cor(Y_{ij}, Y_{i'j}) = \rho = \sigma_s^2/(\sigma_s^2 + \sigma_\epsilon^2)$.

A general form of correlation structure can be written as

$$\begin{pmatrix} 1 & \rho & \cdots & \rho^{1+(p-2)Q} \\ \rho & 1 & \cdots & \rho^{1+(p-3)Q} \\ \vdots & & \ddots & \vdots \\ \rho^{(1+(p-2)Q)} & \rho^{(1+(p-3)Q)} & \cdots & 1 \end{pmatrix}.$$

If $\rho = 0$, then responses from a subject are uncorrelated. If $\rho \neq 0$ and $Q = 1$, then responses have an auto-regressive covariance structure. If $\rho \neq 0$ and $Q = 0$ then responses have an equicorrelated covariance structure. Often responses from the same subject are assumed to have an equicorrelated structure. For example, the covariance matrix for $\Omega(t, 2, n)$ can be written as

$$\begin{pmatrix} \sigma_s^2 + \sigma_\epsilon^2 & \sigma_s^2 \\ \sigma_s^2 & \sigma_s^2 + \sigma_\epsilon^2 \end{pmatrix} = \begin{pmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{pmatrix}$$

Furthermore, the covariance matrices of errors vary for treatment sequences when the treatment variances are unequal. We define $\xi'_{ij} = \epsilon_{ij} + \xi_j$ and $\boldsymbol{\xi}'_j = (\xi'_{1j}, \cdots, \xi'_{pj})$. Then for any subject j in treatment sequence k, the covariance matrix of $\boldsymbol{\xi}'_j$ for $\Omega(2, 2, n)$ denoted by $COV(\boldsymbol{\xi}'_j) = V_k$ is defined as,

$$V_k = \begin{pmatrix} \tilde{\sigma}^2_{d_{1k}} & \rho\tilde{\sigma}_{d_{1k}}\tilde{\sigma}_{d_{2k}} \\ \rho\tilde{\sigma}_{d_{1k}}\tilde{\sigma}_{d_{2k}} & \tilde{\sigma}^2_{d_{2k}} \end{pmatrix}, \tag{2.2}$$

where $\epsilon'_{ijA}s$ and $\epsilon'_{ijB}s$ are mutually independent.

For $\Omega(2, 3, n)$, the covariance structure under the unequal treatment variance assumption can be defined as

$$V_k = \begin{pmatrix} \tilde{\sigma}^2_{d_{1k}} & \rho\tilde{\sigma}_{d_{1k}}\tilde{\sigma}_{d_{2k}} & \rho\tilde{\sigma}_{d_{1k}}\tilde{\sigma}_{d_{3k}} \\ \rho\tilde{\sigma}_{d_{1k}}\tilde{\sigma}_{d_{2k}} & \tilde{\sigma}^2_{d_{2k}} & \rho\tilde{\sigma}_{d_{2k}}\tilde{\sigma}_{d_{3k}} \\ \rho\tilde{\sigma}_{d_{1k}}\tilde{\sigma}_{d_{3k}} & \rho\tilde{\sigma}_{d_{2k}}\tilde{\sigma}_{d_{3k}} & \tilde{\sigma}^2_{d_{3k}} \end{pmatrix}.$$

2.2.2 Optimality Criteria

The Fisher information matrix measures the amount of information of the parameters that the observed random variable Y contains. Given any vector of parameters $\boldsymbol{\theta}$ and a probability density function $f(\boldsymbol{y}; \boldsymbol{\theta})$, the Fisher Information matrix can be written as

$$I(\boldsymbol{\theta}) = E\left[(\frac{\delta}{\delta\boldsymbol{\theta}} log f(\boldsymbol{y}; \boldsymbol{\theta})^2 | \boldsymbol{\theta}\right] \tag{2.3}$$

$$= E\left[-\frac{\delta^2}{\delta\boldsymbol{\theta}^2} log f(\boldsymbol{y}; \boldsymbol{\theta}) | \boldsymbol{\theta}\right]. \tag{2.4}$$

Linear models, including the traditional crossover design model, can be written in the following general form.

$$\boldsymbol{y} = 1_{np}\mu + P\boldsymbol{\pi} + T_d\boldsymbol{\tau} + F_d\boldsymbol{\gamma} + \boldsymbol{\epsilon} = X\boldsymbol{\theta} + \boldsymbol{\epsilon} \tag{2.5}$$

where $\boldsymbol{y}$ denotes the $np\times 1$ vector of responses and X denotes the $np\times(1+p+2t)$ design matrix for the $(1+p+2t)\times 1$ vector of parameters $\boldsymbol{\theta} = \begin{pmatrix}\mu & \boldsymbol{\pi} & \boldsymbol{\tau} & \boldsymbol{\gamma}\end{pmatrix}'$, and $\boldsymbol{\epsilon}$ denotes the $np \times 1$ vector of errors with mean $\mathbf{0}$ and covariance matrix of $\sigma^2\Sigma$. When treatments have unequal variances, the covariance matrix Σ takes a specific form, which we will describe next. The Kronecker product noted by $*$ is used to define the covariance matrix. The covariance matrix is defined as $\Sigma = I_n * \boldsymbol{V}$ where $\boldsymbol{V}$ is a $n \times n$ block diagonal matrix with the l^{th} diagonal block element for the covariance matrix, V_l, corresponds to the treatment sequence that patient l receives, for $l = 1, 2, \cdots, n$. We can partition the design matrix X from (2.5) into 3 blocks in the following way. $X_{1d} = (1_{np}|P)$, $X_{2d} = (F_d)$, and $X_{3d} = T_d$ with a corresponding scalar or vector of parameters $\boldsymbol{\theta}_1 = \begin{pmatrix}\mu & \boldsymbol{\pi}'\end{pmatrix}$, $\boldsymbol{\theta}_2 = (\boldsymbol{\gamma})'$ and $\boldsymbol{\theta}_3 = \begin{pmatrix}\boldsymbol{\tau}\end{pmatrix}'$. The covariance matrix Σ is positive definite so that $\Sigma^{-1/2}$ exists. Multiplying $\Sigma^{-1/2}$ on the left of all terms in the above model gives

$$\Sigma^{-1/2}\boldsymbol{y} = \Sigma^{-1/2}X\boldsymbol{\theta} + \Sigma^{-1/2}\boldsymbol{\epsilon}. \tag{2.6}$$

The mean and variance of the new error term are $\mathrm{E}(\Sigma^{-1/2}\boldsymbol{\epsilon}) = \mathbf{0}$ and $Var(\Sigma^{-1/2}\boldsymbol{\epsilon}) = \sigma^2 I_{np}$. Now the responses, $(\Sigma^{-1/2}\boldsymbol{y})$, are independent and identically distributed with mean 0 and variance of σ^2. Define $W = \Sigma^{-1/2}\boldsymbol{y}$ and $Z = \Sigma^{-1/2}X$ then

$$I(\boldsymbol{\theta}) = E\left[-\frac{d^2}{d\boldsymbol{\theta}^2}(-\frac{1}{2}log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(W - Z\boldsymbol{\theta})^T(W - Z\boldsymbol{\theta})|\boldsymbol{\theta}\right] \tag{2.7}$$

$$= E\left[-\frac{d^2}{d\boldsymbol{\theta}^2}(-\frac{1}{2\sigma^2}(W^TW - \boldsymbol{\theta}^TZ^TW - WZ\boldsymbol{\theta}^T + \boldsymbol{\theta}^TZ^TZ\boldsymbol{\theta}))|\boldsymbol{\theta}\right] \tag{2.8}$$

$$= E\left[-\frac{d^2}{d\boldsymbol{\theta}^2}(-\frac{1}{2\sigma^2}\boldsymbol{\theta}^TZ^TZ\boldsymbol{\theta})|\boldsymbol{\theta}\right] \tag{2.9}$$

$$= \frac{1}{\sigma^2}Z^TZ = \frac{1}{\sigma^2}X^T\Sigma^{-1}X. \tag{2.10}$$

The Information Matrix can be obtained as follows

$$I_d = \begin{pmatrix} X'_{1d}\Sigma^{-1}X_{1d} & X'_{1d}\Sigma^{-1}X_{2d} & X'_{1d}\Sigma^{-1}X_{3d} \\ X'_{2d}\Sigma^{-1}X_{1d} & X'_{2d}\Sigma^{-1}X_{2d} & X'_{2d}\Sigma^{-1}X_{3d} \\ X'_{3d}\Sigma^{-1}X_{1d} & X'_{3d}\Sigma^{-1}X_{2d} & X'_{3d}\Sigma^{-1}X_{3d} \end{pmatrix} \tag{2.11}$$

and the Information Matrix for estimating the direct treatment effects is

$$I_d(\boldsymbol{\theta}_3) = X'^*_{3d} pr^{\perp}(X^*_{1d}|X^*_{2d})X^*_{3d} \tag{2.12}$$

The inverse of the Fisher information is a lower bound of the variance of any unbiased estimators of a parameter and the asymptotic variance of maximum likelihood estimators. For this reason, the approaches to find optimal designs have utilized the Information matrix. Define T-optimality criterion as a way to measure the trace of the Information Matrix, $I(\boldsymbol{\theta})_d$, of competing designs to find the design with a maximum trace.

Kiefer (1975) showed that the universally optimal design, which is D(determinant), A(average), and E(Minimum Eigenvalue) optimal, maximizes $\Phi(I_d)$ for any Φ satisfying the following conditions.

1. Φ is concave.
2. $\Phi(S'IS) = \Phi(I)$ for any permutation matrix S.
3. $\Phi(bI)$ is non-decreasing in the scalar $b > 0$.

Proposition 1. *Kiefer (1975). Any design, d^*, satisfying the following conditions is universally optimal.*

1. *A_{d^*} is completely symmetric that $A_{d^*} = aI_t + b1_t1_t'$ for some constant a and b.*
2. *$tr(A_{d^*}) \geq tr(A_d)$ for all $d \in D$, that is d^* is T optimal.*

Define the following notations $X^*_{sd} = \Sigma^{-1/2}X_{sd}$, a projection matrix $pr(X_{sd})$ $= X_{sd}(X'_{sd}X_{sd})^{-1}X'_{sd}$, and $pr^{\perp}(X_{sd}) = I - pr(X_{sd})$ for s =1, 2, 3. Then the matrix X information matrix of crossover design d for the estimation of treatment effects $\boldsymbol{\theta}_3$ is given as follows

$$M_d(\boldsymbol{\theta}_3) = X^{*'}_{3d} pr^{\perp}(X^*_{1d}|X^*_{2d})X^*_{3d}. \tag{2.13}$$

Define a simpler model using the partitioned design matrices as $\boldsymbol{y}_d = X_{1d}\boldsymbol{\theta}_1 + X_{3d}\boldsymbol{\theta}_3 + \boldsymbol{\epsilon}$.

Proposition 2. *Kunert (1983). The trace of the information matrix for estimating the treatment effects $\boldsymbol{\theta}_3$, $tr(M_d(\boldsymbol{\theta}_3)$, has an upper bound of $tr(X'_{3d}$ $pr^{\perp}(X_{1d})X_{3d})$ and this upper bound is achieved if any of the following conditions are satisfied.*

1. $X^{*'}_{3d}pr^{\perp}(X^{*}_{1d})X^{*}_{2d} = 0_{t\times p}$
2. $X^{*'}_{3d}pr^{\perp}(X^{*}_{1d})X^{*}_{2d}(X^{*'}_{2d}pr^{\perp}(X^{*}_{1d})X^{*}_{2d})^{-} = 0_{t\times p}$

Proof. Using the properties of a block matrix we obtain,

$$\begin{aligned} &pr(X^{*}_{1d}|X^{*}_{2d}) = pr(X^{*}_{1d}) + pr\{pr^{\perp}(X^{*}_{1d})X^{*}_{2d}\} \\ &\Rightarrow pr^{\perp}\{(X^{*}_{1d}|X^{*}_{2d})\} \\ &= pr^{\perp}(X^{*}_{1d}) - pr^{\perp}(X^{*}_{1d})X^{*}_{2d}(X^{*'}_{2d}pr^{\perp}(X^{*}_{1d})X^{*}_{2d})^{-}X^{*'}_{2d}pr^{\perp}(X^{*}_{1d}) \end{aligned}$$

Then multiply $X^{*'}_{3d}$ and X^{*}_{3d} on the left and right side of the terms to get the Information matrix. Using the property of traces of sums of matrices that $tr(A+B) = tr(A) + tr(B)$ and the property of projection matrices that all eigenvalues are 1 or 0, the upper bound of the trace of the information matrix is attained if any of the above three conditions are satisfied. □

2.2.3 Strategies for Finding Optimal Designs

Kunert (1983) proposed the following strategies to construct the universally optimal design. Let Δ be a class of crossover designs and define Δ^* to be the set of all $d \in \Delta$ such that $X'_{3d}pr^{\perp}(X_{1d})X_{3d}$ is completely symmetric and has maximum trace in the class Δ, then all designs in Δ^* are universally optimal in estimating treatment effects under the simpler model.

Proposition 3. *Suppose that there is a design $d^* \in \Delta^*$ such that $X^{*}_{3d*}pr^{\perp}(X^{*}_{1d*})X_{2d^{*}*} = \mathbf{0}$. Then d^* is the universally optimal design for estimation of treatment effects under the simpler model by proposition 1 and 2. Also, the set of all D optimal designs and A optimal designs for the estimation of treatment effects is equal to the subset of Δ^*, which fulfills the orthogonality condition.*

Proof. The proof for this proposition is straightforward from Proposition 1 and 2. □

This proposition provides a simple tool to find universally optimal designs but it is possible that no design $d \in \Delta^*$ satisfies the orthogonality condition. Then another strategy can be applied.

Proposition 4. *First find $d^* \in \Delta^*$ such that I_{d^*} is completely symmetric and has maximum trace over Δ^* and $tr(I_{d^*}) \geq tr(X_{3d}pr^{\perp}(X_{1d}^*)X_{3d})$ for all d in $\tilde{\Delta}$, which is a subset of Δ. Then d^* is universally optimal for estimating treatment effects over $\Delta^* \cup \tilde{\Delta}$.*

2.3 Optimal Designs

In this chapter, we investigate crossover designs with $t = 2$. For any treatment sequence k, there is a dual sequence k', which is obtainable by replacing treatment A with B and vice versa. For example, AA is a dual sequence of BB.

2.3.1 Uncorrelated Responses $\sigma_A^2 \neq \sigma_B^2$

First, we consider the case where responses from the same subject are independent. From this point, we denote treatment A and B by 1 and 2 respectively and allocations of treatment sequence k is also denoted with numbers.

Lemma 1. *For any design $d \in \Omega_{p,2,n}$, the upper bound of the Information matrix for estimating treatment effects, $X_{3d}^{*\prime}pr^{\perp}(X_{1d}^*)X_{3d}^*$, is completely symmetric where $X_{1d}^* = 1_{np}$.*

Proof. The upper bound of the Information matrix for estimating treatment effects is defined as

$$\begin{aligned}
&X_{3d}^{*\prime}pr^{\perp}(X_{1d}^*)X_{3d}^* \\
&= X_{3d}'\Sigma^{-1}X_{3d} - X_{3d}'\Sigma^{-1}X_{1d}(X_{1d}'\Sigma^{-1}X_{1d})^{-}X_{1d}'\Sigma^{-1}X_{3d} \\
&= \begin{pmatrix} \frac{q_{d1}}{\sigma_1^2} & \frac{q_{d2}}{\sigma_2^2} \end{pmatrix} I_2 - \frac{1}{q_{d1}/\sigma_1^2 + q_{d2}/\sigma_2^2} \begin{pmatrix} \frac{q_{d1}}{\sigma_1^2} \\ \frac{q_{d2}}{\sigma_2^2} \end{pmatrix} \begin{pmatrix} \frac{q_{d1}}{\sigma_1^2} & \frac{q_{d2}}{\sigma_2^2} \end{pmatrix} \\
&= 2\frac{q_{d1}q_{d2}}{q_{d1}\sigma_2^2 + q_{d2}\sigma_1^2} I_2 - \frac{q_{d1}q_{d2}}{q_{d1}\sigma_2^2 + q_{d2}\sigma_1^2} 1_2 1_2'
\end{aligned}$$

where $q_{d1} = n(2p_{11} + p_{12} + p_{21})$ and $q_{d2} = n(2p_{22} + p_{12} + p_{21})$. □

Lemma 2. *For any design $d \in \Omega_{p,2,n}$, the class of designs Δ^* that maximizes the trace of the upper bound of the Information matrix for estimating direct treatment effects is defined as $q_{1d} = np/(1 + r)$.*

Proof. The trace of the upper bound is defined as

$$\begin{aligned}
T(q_{d1}) &= trace(X_{3d}'pr^{\perp}(X_{1d})X_{3d}) \\
&= \frac{2q_{d1}q_{d2}}{q_{d1}\sigma_2^2 + q_{d2}\sigma_1^2} = \frac{2q_{d1}(pn - q_{d1})}{q_{d1}r^2\sigma_1^2 + (pn - q_{d1})\sigma_1^2}
\end{aligned}$$

where $q_{d1} + q_{d2} = pn$ and $r = \sigma_B/\sigma_A$.

Then, we solve for design d^*, which satisfies the optimizing conditions below.

$$\frac{\partial}{\partial q_{d1}} T(q_{d^*1}) = -\frac{2q_{d^*1}^2(r^2-1) + 4npq_{d^*1} - 2n^2p^2}{\sigma_1^2(np + (r^2-1)q_{d^*1})} = 0, \tag{2.14}$$

$$\frac{\partial^2}{\partial q_{d1}^2} T(q_{d^*1}) = -\frac{4n^2p^2r^2}{\sigma_1^2(np + (r^2-1)q_{d^*1})^3} < 0. \tag{2.15}$$

The second derivative condition is satisfied as r, σ_1^2, n, p, $q_{d1} > 0$ for any design d. The first condition is satisfied for any design d^* that satisfies $q_{d^*1} = np/(1+r)$. □

Kunert (1983) proposed that universally optimal designs can be obtained using the universal optimality criteria of Kiefer (1975) and block matrix decomposition of the Information matrix. Therefore, design d^* is universally optimal in estimating the direct treatment effects if and only if d^* maximizes the trace of the Information matrix, has completely symmetric Information matrix under the simpler model, and satisfies one of the orthogonality conditions in Proposition 2.

In the case of $p = 2$, the first orthogonality conditions, (1) of Proposition 2, are defined as

$$X_{3d}'\Sigma^{-1}P_2 - X_{3d}'\Sigma^{-1}1_{np}1_{np}'\Sigma^{-1}P_2/1_{np}'\Sigma^{-1}1_{np} = \begin{pmatrix}0\\0\end{pmatrix} \tag{2.16}$$

$$X_{3d}'\Sigma^{-1}F_d - X_{3d}'\Sigma^{-1}1_{np}1_{np}'\Sigma^{-1}F_d/1_{np}'\Sigma^{-1}1_{np} = \begin{pmatrix}0 & 0\\0 & 0\end{pmatrix} \tag{2.17}$$

where $P_2 = \begin{pmatrix}0\\1\end{pmatrix} \otimes 1_n$

Lemma 3. *For crossover designs in $\Omega_{2,2,n}$, the designs that satisfy the equation* (2.16) *are balanced on dual* (AB/BA).

Proof.

$$\begin{aligned} &X_{3d}'\Sigma^{-1}P_2 - X_{3d}'\Sigma^{-1}1_{np}1_{np}'\Sigma^{-1}P_2/1_{np}'\Sigma^{-1}1_{np} \\ &= \begin{pmatrix}l_{d12}/\sigma_1^2\\ l_{d22}/\sigma_2^2\end{pmatrix} - \frac{l_{d12}/\sigma_1^2 + l_{d22}/\sigma_2^2}{q_{d1}/\sigma_1^2 + q_{d2}/\sigma_2^2}\begin{pmatrix}q_{d1}/\sigma_1^2\\ q_{d2}/\sigma_2^2\end{pmatrix}. \end{aligned}$$

The above equation reduces to $q_{d1} = l_{d12}q_{d2}/l_{d22}$. With $l_{d12} = n(p_{11} + p_{21})$, $l_{d22} = n(p_{22} + p_{12})$, $q_{d1} = n(2p_{11} + p_{12} + p_{21})$, $q_{d2} = n(2p_{22} + p_{12} + p_{21})$, and $p_{11} + p_{22} + p_{12} + p_{21} = 1$, the equality holds iff $p_{12} = p_{21}$. □

Lemma 4. *Among the set of crossover designs in $\Omega_{2,2,n}$ with $p_{12} = p_{21}$, satisfying the equation in* (2.16)*, designs with $p_{11} = p_{12}^2/p_{22}$ satisfy the second condition in* (2.16)*.*

Proof.

$$
\begin{aligned}
&X_{3d}'\Sigma^{-1}F_d - X_{3d}'\Sigma^{-1}1_{np}1_{np}'\Sigma^{-1}F_d/1_{np}'\Sigma^{-1}1_{np} \\
&= \begin{pmatrix} p_{11}/\sigma_A^2 & p_{21}/\sigma_A^2 \\ p_{12}/\sigma_B^2 & p_{22}/\sigma_B^2 \end{pmatrix} - \frac{1}{q_{d1}/\sigma_A^2 + q_{d2}/\sigma_B^2} \begin{pmatrix} q_{d1}/\sigma_A^2 \\ q_{d2}/\sigma_B^2 \end{pmatrix} \left(\tfrac{p_{11}}{\sigma_A^2} + \tfrac{p_{12}p_{21}}{\sigma_B^2\sigma_A^2} + \tfrac{p_{22}}{\sigma_B^2} \right) \\
=&\frac{1}{q_{d1}/\sigma_A^2 + q_{d2}/\sigma_B^2} \begin{pmatrix} p_{11}p_{22} - p_{12}^2 & -(p_{11}p_{22} - p_{12}^2) \\ -(p_{11}p_{22} - p_{12}^2) & p_{11}p_{22} - p_{12}^2 \end{pmatrix}.
\end{aligned}
$$

□

Theorem 1. *For crossover designs in $\Omega_{2,2,n}$, the universally optimal design for estimating the direct treatment effects is given by $p_{11} = 1/(1+r)^2$, $p_{22} = r^2/(1+r)^2$, and $p_{12} = p_{21} = r/(1+r)^2$*

Proof. The proof for Theorem 5 is straightforward by using Proposition 1 (Kiefer (1975)), Proposition 2 (Kunert (1983)), lemma 1-4, and some arithmetic. □

Naturally, the condition of unequal variance in treatment effects alters the optimal designs that we have become used to. Allocation to sequences $AA/BB/AB/BA$ is no longer equal and requires more subjects to treatments with higher variances. For $p = 2$, the pattern is rather simple.

Three-Period Two-Treatment Designs

In $\Omega_{2,3,n}$ there are 2^3 treatment sequences namely (AAA, AAB, ABB, ABA, BBB, BBA, BAA, BAB). As the Propositions 1 and 2 still hold for p periods, the orthogonality conditions are defined as

Table 2.1: Universally optimal two-period designs with optimal allocations for $p = 2$ and $t = 2$ with $\sigma_A^2 \neq \sigma_B^2$.

$r = \sigma_B/\sigma_A$	p_{AA}	p_{BB}	p_{AB}	p_{BA}.
0.1	0.83	0.01	0.08	0.08
0.25	0.64	0.04	0.16	0.16
0.5	0.44	0.12	0.22	0.22
1	0.25	0.25	0.25	0.25
2	0.12	0.44	0.22	0.22
4	0.04	0.64	0.16	0.16
10	0.01	0.83	0.08	0.08

$$X'_{3d}\Sigma^{-1}P_3 - X'_{3d}\Sigma^{-1}1_{np}1'_{np}\Sigma^{-1}P_3/1'_{np}\Sigma^{-1}1_{np} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \tag{2.18}$$

$$X'_{3d}\Sigma^{-1}F_d - X'_{3d}\Sigma^{-1}1_{np}1'_{np}\Sigma^{-1}F_d/1'_{np}\Sigma^{-1}1_{np} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \tag{2.19}$$

where $P_3 = \begin{pmatrix} 00 \\ 10 \\ 01 \end{pmatrix} \otimes 1_n$.

Lemma 5. *For crossover designs in $\Omega_{2,3,n}$, the designs that satisfy equation* (2.18) *are the ones that satisfy the following equations,*

$$l_{d12} = \frac{q_{d1}l_{d22}}{q_{d2}}, \tag{2.20}$$

$$l_{d13} = \frac{q_{d1}l_{d23}}{q_{d2}}, \tag{2.21}$$

where $l_{d12} = p_{111} + p_{112} + p_{211} + p_{212}$ and $l_{d13} = p_{111} + p_{121} + p_{211} + p_{221}$.

Lemma 6. *For crossover designs in $\Omega_{2,3,n}$, the designs that satisfy equation* (2.19) *are the ones that satisfy the following equations,*

$$m_{d11} = \frac{m_{d12}r_{d1}}{q_{d2}} \tag{2.22}$$

$$m_{d22} = \frac{m_{d21}q_{d2}}{q_{d1}} \tag{2.23}$$

where $m_{d11} = 2p_{111} + p_{112} + p_{211}$, $m_{d22} = 2p_{222} + p_{221} + p_{122}$, $m_{d12} = p_{211} + p_{212} + p_{221} + p_{121}$, and $m_{d21} = p_{121} + p_{122} + p_{112} + p_{212}$.

Proof. The proofs for lemma 5 and 6 are analogous to that of lemma 3 and 4. □

Using Equations (20)-(23), we come up with a sufficiency condition for orthogonality criteria.

$$q_{d1} = \frac{l_{d13}q_{d2}}{l_{d23}} = \frac{l_{d12}q_{d2}}{l_{d22}} = \frac{m_{d11}q_{d2}}{m_{d12}} = \frac{m_{d21}q_{d2}}{m_{d22}}. \tag{2.24}$$

Using the identity $q_{d1} + q_{d2} = 2np$, we get

$$q_{d1} = \frac{npl_{d13}}{l_{d23} + l_{d13}} = \frac{npl_{d12}}{l_{d22} + l_{d12}} = \frac{npm_{d11}}{m_{d12} + m_{d11}} = \frac{npm_{d21}}{m_{d22} + m_{d21}}. \tag{2.25}$$

Theorem 2. *Among the crossover designs in $\Omega_{2,3,n}$, if design d satisfies*

$$q_{d1} = \frac{npl_{d13}}{l_{d23} + l_{d13}} = \frac{npl_{d12}}{l_{d22} + l_{d12}} = \frac{npm_{d11}}{m_{d12} + m_{d11}} = \frac{npm_{d21}}{m_{d22} + m_{d21}} = \frac{np}{1+r}, \quad (2.26)$$

then d is the universally optimal design for estimating the treatment effects.

Proof. The proof is straightforward from Proposition 1, Proposition 2, Lemma 1, Lemma 2, and equation (2.25). □

The set of designs that satisfy (2.26) can be classified into 4 designs, namely Design I, II, III, IV with 4, 8, 6, and 6 treatment sequences. Design I with 4 treatment sequences is a direct extension from $t = 2$ and $p = 2$. Whereas Design II consists of all 8 treatment sequences. From Design II, we can derive Design $III(IV)$ by allocating $p_{ABB}(p_{AAB}) = p^*$ as shown in Table 2.3 and some examples of these designs are shown in Table 2.4. Table 2.4 only shows the case of $r > 1$ as the case of $r < 1$ is analogous to $r > 1$ and can be obtained by interchanging the allocations of dual sequences.

In Designs $II - IV$, we can observe the dual balanced property on the sequences ABA/BAB, ABB/BAA and AAB/BBA while the sequences AAA/BBB depends on the ratio of treatment variances. It can be noted that optimal designs depend on the ratio r and need to allocate more subjects to AAA/BBB as r increases/decreases from 1. Design I is a strictly extended version of the existing optimal two-period design and Design II-IV are slight variations of it. Design III and IV use the same allocation on dual treatment sequences AAB/BBA and ABB/BAA. It can be shown that under the equal treatment variability assumption of $r = 1$, Design III reduces to the universally optimal design suggested by Laska, Meisner, and Kushner (1983) and Design IV reduces to the nearly optimal design suggested by Carriere (1994).

Table 2.2: Proposed universally optimal designs under unequal treatment variances for $p = 3$ and $t = 2$.

Design	K	Treatment Sequences	Allocations	Conditions
I	4	ABA	$\frac{r}{(1+r)^2}$	any $r > 0$
		BAB	$\frac{r}{(1+r)^2}$	
		AAA	$\frac{1}{(1+r)^2}$	
		BBB	$\frac{r^2}{(1+r)^2}$	
II	8	AAB	p_{AAB}	$p_{AAB} + p_{ABB} = p^*$ $p^* \leq \frac{2}{(1+r)^2}$ for $r > 1$ $p^* \leq \frac{2r^2}{(1+r)^2}$ for $0 < r < 1$
		BBA	p_{AAB}	
		ABB	p_{ABB}	
		BAA	p_{ABB}	
		ABA	$(2r - p^*(1+r)^2)/2(1+r)^2$	
		BAB	$(2r - p^*(1+r)^2)/2(1+r)^2$	
		AAA	$(2 - p^*(1+r)^2)/2(1+r)^2$	
		BBB	$(2r^2 - p^*(1+r)^2)/2(1+r)^2$	

Table 2.3: Special cases of Design II from Table 2.2.

Design	K	Treatment Sequences	Allocations	Conditions
III	6	AAB	p^*	$p^* \leq \frac{2}{(1+r)^2}$ for $r > 1$ $p^*, \leq \frac{2r^2}{(1+r)^2}$ for $0 < r < 1$
		BBA	p^*	
		ABA	$(2r - p^*(1+r)^2)/2(1+r)^2$	
		BAB	$(2r - p^*(1+r)^2)/2(1+r)^2$	
		AAA	$(2 - p^*(1+r)^2)/2(1+r)^2$	
		BBB	$(2r^2 - p^*(1+r)^2)/2(1+r)^2$	
IV	6	ABB	p^*	$p^* \leq \frac{2}{(1+r)^2}$ for $r > 1$ $p^*, \leq \frac{2r^2}{(1+r)^2}$ for $0 < r < 1$
		BAA	p^*	
		ABA	$(2r - p^*(1+r)^2)/2(1+r)^2$	
		BAB	$(2r - p^*(1+r)^2)/2(1+r)^2$	
		AAA	$(2 - p^*(1+r)^2)/2(1+r)^2$	
		BBB	$(2r^2 - p^*(1+r)^2)/2(1+r)^2$	

Table 2.4: Proposed universally optimal designs under unequal treatment variances with 6 sequences.

r	Allocations					
	ABB(AAB)	BAA(BBA)	ABA	BAB	AAA	BBB
$r > 1$	$\frac{2}{(r+1)^2}$	$\frac{2}{(r+1)^2}$	$\frac{r-1}{(r+1)^2}$	$\frac{r-1}{(r+1)^2}$	0	$\frac{r-1}{r+1}$
1.5	0.32	0.32	0.08	0.08	0	0.2
2	0.22	0.22	0.11	0.11	0	0.34
3	0.125	0.125	0.125	0.125	0	0.5
4	0.08	0.08	0.12	0.12	0	0.60

From the table 2.3, we can choose the design with the largest value for $p = p_{ABB} = p_{BAA}$. Then, the proportion of subjects assigned to AA/BB in their first two periods, $p_{AAA} + p_{AAB} + p_{BBB} + p_{BBA}$, is minimized and one of the two sequences AAA/BBB is removed from the design depending on the value of r. As a result, $2r/(1+r)$ or $2/(1+r)$ of all subjects would be assigned to the sequences AB/BA in the first two periods, while only $(1-r)/(1+r)$ or $(r-1)/(1+r)$ would be assigned to AAA/BBB depending on the value of r. For example, 1/3 of subjects would be allocated to the sequences AAA/BBB for $p = 3$ designs when $r = 0.5$ whereas 5/9 of subjects would be allocated to the sequences AA/BB for $p = 2$ designs. This means that more subjects are assigned to the treatment sequences with treatments crossing over.

2.3.2 Correlated Responses

Responses from the same subjects are more likely correlated. We now consider correlated responses. As a first step to accommodating unequal variances, we assume a constant correlation of ρ. We have three choices for simpler models to acquire the upper bound for the trace of the Information matrix. That is to choose one of the followings as X_1^*: overall mean effect, period effects, or carryover effects. The orthogonality conditions for the upper bounds constructed from the first two effects cannot be satisfied. However, the orthogonality conditions for the upper bounds created from having carryover effects as X_1^* are satisfied for any designs in the $\Omega_{2,2,n}$. That is, the Information matrix for estimating the direct treatment effects after accounting for the other effects

can simply be written as $T_d'^* pr^{\perp}(F_d'^*)T_d^*$.

Lemma 7. *For any crossover design $d \in \Omega_{2,2,n}$, the Fisher's Information matrix for estimating direct treatment effects accounting for other effects can be reduced to $I_d = T_d'^* pr^{\perp}(F_d'^*)T_d^*$. Thus, any design that satisfies the optimality criteria on the Information matrix is optimal.*

Proof. Refer to Appendix. □

It can be shown that there is no universally optimal design, a design that has a completely symmetric Information matrix and is T optimal. Therefore, we examine designs that are optimal in some subset of $\Omega_{2,2,n}$. Two subsets are being investigated in this chapter: 1) the set of designs that are uniform on the first period, with the same number of subjects assigned. 2) the set of designs that follow the ratio adjustment for independent case. The first set of designs are the ones that satisfy the constraints $p_{11}+p_{12} = p_{22}+p_{21} = 0.5$ and the second set of designs are the ones that satisfy the constraints $p_{11} + p_{12} = 1/(1+r)$ and $p_{22} + p_{21} = r/(1+r)$. These sets include the universally optimal designs constructed under the independent assumption of the responses.

Theorem 3. *The following designs are T optimal in the subset of $\Omega_{2,2,n}$ with allocations to $AA/BB/AB/BA$.*

1. $d_1 : \frac{1}{2(1+r)}, \frac{r}{2(1+r)}, \frac{r}{2(1+r)}, \frac{1}{2(1+r)}$ *for* $p_{11} + p_{12} = p_{22} + p_{21}$.
2. $d_2 : \frac{1}{(1+r)^2}, \frac{r^2}{(1+r)^2}, \frac{r}{(1+r)^2}, \frac{r}{(1+r)^2}$ *for* $p_{11}+p_{12} = \frac{1}{(1+r)}$ *and* $p_{22}+p_{21} = \frac{r}{(1+r)}$.

Proof. Refer to the Appendix □

In the case of $p = 3$, the Fisher's information matrix for estimating direct treatment effects after accounting for other effects has no explicit form. Therefore, numerical "Nelder-Mead" method was used to compute the T-optimal designs.

2.3.3 Efficiency

The efficiencies of T-optimal designs under the unequal variance with constraints are compared to the existing universally optimal design ($AA/BB/AB/BA$ with equal allocation) and also to the design (AB/BA with

equal allocation) and shown in Table the 2.5. The efficiencies are computed as the ratio of the variances of estimated treatment effects $\hat{\tau}$.

Table 2.5: Entries are ratio of variances of estimated treatment effects for designs T-optimal designs under the unequal variance assumption over the designs $AA/BB/AB/BA$ and AB/BA with an equal allocations. The constraints are d_1, $p_{11} + p_{12} = p_{21} + p_{22} = 0.5$, for the first half and d_2, $p_{11} + p_{12} = 1/(1+r)$ and $p_{22} + p_{21} = r/(1+r)$, for the second half.

Constraints	$r \setminus \rho$	0	0.1	0.3	0.5	0.7	0.9
d_1	1/10	0.75	0.76	0.77	0.79	0.83	0.91
$AA/BB/AB/BA$	1/4	0.85	0.85	0.86	0.86	0.87	0.91
	1/3	0.89	0.89	0.89	0.89	0.9	0.92
	1/2	0.95	0.95	0.95	0.94	0.94	0.95
	1	1	1	1	1	1	1
d_1	1/10	0.37	0.39	0.42	0.43	0.43	0.42
AB/BA	1/4	0.42	0.43	0.44	0.43	0.4	0.32
	1/3	0.44	0.45	0.45	0.43	0.38	0.28
	1/2	0.47	0.48	0.47	0.43	0.36	0.22
	1	0.5	0.5	0.48	0.43	0.34	0.16
d_2	1/10	0.6	0.62	0.65	0.7	0.76	0.88
$AA/BB/AB/BA$	1/4	0.74	0.74	0.76	0.78	0.81	0.89
	1/3	0.8	0.8	0.81	0.82	0.84	0.9
	1/2	0.9	0.9	0.9	0.91	0.91	0.93
	1	1	1	1	1	1	1
d_2	1/10	0.3	0.32	0.35	0.38	0.4	0.4
AB/BA	1/4	0.37	0.38	0.39	0.39	0.37	0.31
	1/3	0.4	0.41	0.41	0.4	0.36	0.27
	1/2	0.45	0.45	0.44	0.41	0.35	0.22
	1	0.5	0.5	0.48	0.43	0.34	0.16

The efficiency tables show that the T-optimal designs in the restricted class of $\omega(2, 2, n)$ are more efficient than AB/BA or $AA/BB/AB/BA$ with equal allocations when treatment effects have unequal variances. The T optimal designs under the constraint 2, d_2, demonstrates better efficiency than d_1. However, under the constraint d_2, the design allocates more subjects to the treatment sequences AA/BB than d_1. The larger variance inequality, the fewer subjects available for within subject contrasts.

The Design d_1 is balanced on the first period, which means that the proportions of subjects allocated to treatment A and B in the first period are

equal (50%). The sum of allocations to the sequences AA/BB is equal to that of AB/BA. The universally optimal design under equal variance assumption has the same property. That is to say, the new constrained T-optimal design can replace the equal allocation design without much changes as these two designs assign the same proportion of subjects to the crossover sequences AB/BA. This indicates that a researcher using the equal allocation design $AA/BB/AB/BA$ may test the equality of the treatment variances during the experiment/trial and adjust the allocations based on the estimated ratio r in the case of sequential designs.

Table 2.6: Ratio of variances of estimated treatment effects for T optimal designs under the unequal variance assumption for $p = 3$ over the design ABB/BAA with an equal allocation.

$r \setminus \rho$	0	0.1	0.3	0.5	0.7	0.9
1/10	0.599	0.651	0.740	0.819	0.893	0.965
1/4	0.735	0.774	0.837	0.889	0.933	0.977
1/3	0.800	0.834	0.884	0.921	0.952	0.982
1/2	0.900	0.925	0.951	0.959	0.967	0.984
1	1	1	1	1	1	1

The T optimal designs for $\omega_{2,3,n}$ have allocations similar to that of Table 2.3 when ρ is small. As r increases, the proportion of allocation to the treatment sequence BBB increases. As ρ increases with a fixed r, the proportion of sequences ABB/BAA increases while the proportion to the treatment sequence BBB decreases with some allocation ($20\% \sim 40\%$ combined) to the sequence BBA and BAB. For example, the T-optimal design is $p_{ABB} = 0.3$, $p_{BAA} = 0.15$, $p_{BBA} = 0.14$, $p_{BAB} = 0.17$, and $p_{BBB} = 0.24$ for $r = 3$ and $\rho = 0.5$. The overall relative efficiency of the existing universally optimal design under the equal treatment variance, ABB/BAA, is consistently high at 80% or higher compared to the T-optimal design under unequal variance assumption when $\rho \geq 0.5$ or $r \geq 1/3$.

2.4 Test of Equality of Treatment Variances

The results from the previous subsections are relevant if and only if the treatment variances are unequal and we can estimate the difference. We review the methods for testing equality of treatment variances and extend it to 2-period 2-treatment 4-sequence designs. Suppose that errors η_{jk}'s and ϵ_{ijk}'s have the same means and variances as in (2.1). Further, assume that these errors follow the normal distribution. Then,

$$\boldsymbol{y}_{jk} = X_{jk}\boldsymbol{\theta} + \boldsymbol{\epsilon}_{jk}, \text{ for } j = 1, \cdots, n_k \tag{2.27}$$

where $\boldsymbol{\theta} = (\mu, \pi_1, \cdots, \pi_p, \tau_1, \cdots, \tau_t, \gamma_1, \cdots, \gamma_t)^T$ and $\boldsymbol{\epsilon}_{jk} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_k)$.

Define the within sequence sample means $\hat{\boldsymbol{\mu}}_k = \Sigma_{j=1}^{n_k}(\boldsymbol{y}_{jk})/n_k$ for $k = 1$, 2, ..., K. Then, the covariance matrix for responses from the same subject can be estimated by $p \times p$ matrix $\hat{V}_k = \Sigma_{j=1}^{n_k}(\boldsymbol{y}_{jk} - \hat{\boldsymbol{\mu}}_k)(\boldsymbol{y}_{jk} - \hat{\boldsymbol{\mu}}_k)'$ for each k. $\hat{\boldsymbol{\mu}}_k$'s and $\hat{V}_k$'s are unbiased estimates for $\boldsymbol{\mu}_k$'s and V_k's respectively and $(n_k - 1)\hat{V}_k \sim W_p(n_k, V_k)$.

Chinchilli et al. (2005) proposed a maximum likelihood and restricted maximum likelihood methods for parameter estimation, where variance components are estimated in closed forms. Also, Chinchilli (1996) proposed a test of unequal variance in which the likelihood-based estimates of the variances of the treatment sequences were used. Then, the ratio of the variances of treatment sequences were tested using a F test statistic based on the two estimated variances, which are chi square random variables. Shanga (2003) and Jung (2009) presented a likelihood ratio test for equality of treatment variances for some designs (AB/BA, $ABC/BCA/CAB$). For-two period two-treatment two-sequence design, Jung (2009) showed that,

$$V_{AB} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} V_{BA} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$$\hat{V}_{AB} + \hat{V}^*_{BA} = \begin{pmatrix} \hat{V}_{11} & \hat{V}_{12} \\ \hat{V}_{12} & \hat{V}_{22} \end{pmatrix} \sim W_2(n_{AB} + n_{BA}, V_{AB}),$$

with the idea that V_{AB} can be transformed to V_{BA} by $\hat{V}^*_{BA} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \hat{V}_{BA} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$,

to provide maximum likelihood estimators and the likelihood ratio test statistics. However, this result is limited on the equal allocation on sequence AB/BA.

We introduce two tests of unequal variances, one using AA/BB, which complements the Jung's test, and the other incorporating all four sequences $AA/BB/AB/BA$. The test, t_1, based on Design with sequences AA/BB can be considered a test of equal variance from a repeated measures design with two distinct treatments. Test, t_2, based on Design with sequences $AA/BB/AB/BA$ uses estimates from t_1, which we give details below for testing $H_0 : \sigma_A^2 = \sigma_B^2$.

First, we construct the first test, t_1. Assume that

$$\hat{V}_{AA} = \hat{V}_1 = \begin{pmatrix} \hat{v}_{11,1} & \hat{v}_{12,1} \\ \hat{v}_{12,1} & \hat{v}_{22,1} \end{pmatrix} \sim W_2(n_{AA}, V_{AA}),$$
$$\hat{V}_{BB} = \hat{V}_2 = \begin{pmatrix} \hat{v}_{11,2} & \hat{v}_{12,2} \\ \hat{v}_{12,2} & \hat{v}_{22,2} \end{pmatrix} \sim W_2(n_{BB}, V_{BB}).$$

The likelihood function and the log-likelihood function can be derived as

$$L = \Pi_{k=1}^{2} \frac{c_k}{det(V_k)^{(n_k-1)/2}} exp(-tr(\hat{V}_k V_k^{-1})/2), \tag{2.28}$$
$$logL = \Sigma_{k=1}^{2} \left(c_k' - \frac{n_k - 1}{2} log(det(V_k)) - tr(\hat{V}_k V_k^{-1})/2 \right). \tag{2.29}$$

Under H_0, the partial derivatives with respect to $\sigma = \sigma_A = \sigma_B$ and ρ are given as, with an assumption $n_{AA} = n_{BB} = n$,

$$\frac{\partial logL}{\partial \sigma} = \frac{4\sigma^2(n-1)(\rho^2-1) + tr(\hat{V}_1) + tr(\hat{V}_2) - 2\rho\hat{v}_{12,1} - 2\rho\hat{v}_{12,2}}{\sigma^3(1-\rho^2)}, \tag{2.30}$$
$$\frac{\partial logL}{\partial \rho} = \frac{1}{\sigma^2(1-\rho^2)^2}((1+\rho^2)(\hat{v}_{12,1} + \hat{v}_{12,2}) \tag{2.31}$$
$$-\rho(tr(\hat{V}_1) + tr(\hat{V}_2)) + 2\sigma^2(n-1)\rho(1-\rho^2)). \tag{2.32}$$

The MLEs for σ and ρ are given as

$$\hat{\sigma} = \sqrt{\frac{tr(\hat{V}_1) + tr(\hat{V}_2)}{4(n-1)}}, \tag{2.33}$$

$$\hat{\rho} = \frac{2(v_{12,AA} + v_{12,BB})}{tr(\hat{V}_1) + tr(\hat{V}_2)}. \tag{2.34}$$

Under Ha, the partial derivatives with respect to σ_A, σ_B, and ρ are given as,

$$\frac{\partial logL}{\partial \sigma_A} = \frac{2\sigma_A^2(n-1)(\rho^2-1) + tr(\hat{V}_1)}{\sigma_A^3(1-\rho^2)} - \frac{2\rho\hat{v}_{12,1}}{\sigma_A^3(1-\rho^2)}, \tag{2.35}$$

$$\frac{\partial logL}{\partial \sigma_B} = \frac{2\sigma_B^2(n-1)(\rho^2-1) + tr(\hat{V}_2)}{\sigma_B^3(1-\rho^2)} - \frac{2\rho\hat{v}_{12,2}}{\sigma_B^3(1-\rho^2)}, \tag{2.36}$$

$$\frac{\partial logL}{\partial \rho} = \frac{\hat{v}_{12,1}(1+\rho^2) - \rho tr(\hat{V}_1)}{\sigma_A^2(1-\rho)^2(1+\rho)^2} + \frac{\hat{v}_{12,2}(1+\rho^2) - \rho tr(\hat{V}_2)}{\sigma_B^2(1-\rho)^2(1+\rho)^2} + \frac{2\rho(n-1)}{(1-\rho^2)}. \tag{2.37}$$

The MLEs for σ_A, σ_B, and ρ are given as

$$\hat{\sigma}_A = \sqrt{\frac{tr(\hat{V}_1)tr(\hat{V}_2)^2 - 4tr(\hat{V}_1)\hat{v}_{12,2} + tr(\hat{V}_2)\sqrt{tr(\hat{V}_1)^2 - 4\hat{v}_{12,1}^2}\sqrt{tr(\hat{V}_2)^2 - 4\hat{v}_{12,2}^2}}{4(n-1)(tr(\hat{V}_2) - 4\hat{v}_{12,2})}}, \tag{2.38}$$

$$\hat{\sigma}_B = \sqrt{\frac{tr(\hat{V}_2)tr(\hat{V}_1)^2 - 4tr(\hat{V}_2)\hat{v}_{12,1} + tr(\hat{V}_1)\sqrt{tr(\hat{V}_2)^2 - 4\hat{v}_{12,2}^2}\sqrt{tr(\hat{V}_1)^2 - 4\hat{v}_{12,1}^2}}{4(n-1)(tr(\hat{V}_2) - 4\hat{v}_{12,2})}}, \tag{2.39}$$

$$\hat{\rho} = \frac{tr(\hat{V}_1)tr(\hat{V}_2) + 4\hat{v}_{12,1}\hat{v}_{12,2} - \sqrt{tr(\hat{V}_1)^2 - 4\hat{v}_{12,1}^2}\sqrt{tr(\hat{V}_2)^2 - 4\hat{v}_{12,2}^2}}{2tr(\hat{V}_1)\hat{v}_{12,2} + 2tr(\hat{V}_2)\hat{v}_{12,1}}. \tag{2.40}$$

Then, the test statistic t_1 becomes

$$
\begin{aligned}
t_1 &= -2\log\left(\frac{L\left(\hat{\boldsymbol{\theta}}|H_0\right)}{L\left(\hat{\boldsymbol{\theta}}|Ha\right)}\right) \\
&= -2\Sigma_{k=1}^{2}\left(\frac{(n_k-1)}{2}\log\left(\frac{det(V_k|_{\hat{\boldsymbol{\theta}}_{Ha}})}{det(V_k|_{\hat{\boldsymbol{\theta}}_{H_0}})}\right) - \frac{1}{2}tr(\hat{V}_k(V_k|_{\hat{\boldsymbol{\theta}}_{H_0}} - V_k|_{\hat{\boldsymbol{\theta}}_{Ha}}))\right) \\
&= 2(n-1)\log\left(\frac{\hat{\sigma}^4(1-\hat{\rho}^2)}{\hat{\hat{\sigma}}_A^2\hat{\hat{\sigma}}_B^2(1-\hat{\hat{\rho}}^2)}\right) + \frac{tr(\hat{V}_1)+tr(\hat{V}_2)-2\hat{\rho}(\hat{v}_{12,1}+\hat{v}_{12,2})}{(1-\hat{\rho}^2)\hat{\sigma}^2} \\
&- \frac{\hat{\hat{\sigma}}_A^2(tr(\hat{V}_2)-2\hat{\hat{\rho}}\hat{v}_{12,2})+\hat{\hat{\sigma}}_B^2(tr(\hat{V}_1)-2\hat{\hat{\rho}}\hat{v}_{12,1})}{\hat{\hat{\sigma}}_A^2\hat{\hat{\sigma}}_B^2(1-\hat{\hat{\rho}}^2)},
\end{aligned}
$$

where we used a $\hat{\theta}L = \Pi_{k=1}^{4}\frac{c_k}{det(V_k)^{(n_k-1)/2}}exp(-tr(\hat{V}_k V_k^{-1})/2), logL = \Sigma_{k=1}^{4}\left(c_k' - \frac{n_k-1}{2}log(det(V_k)) - t$

Under H_0, the partial derivatives with respect to $\sigma = \sigma_A = \sigma_B$ and ρ are given as,

$$
\frac{\partial logL}{\partial \sigma} = -\frac{8(\sigma^2(n-1)(\rho^2-1)+\hat{v}_A+\hat{v}_B-2\rho\hat{v}_{12,1}-2\rho\hat{v}_{12,2}-2\rho\hat{v}_{12,3}+\hat{v}_{12,4})}{\sigma^3(\rho^2-1)} \tag{2.41}
$$

$$
\frac{\partial logL}{\partial \rho} = \frac{-4(n-1)\left(\rho^2-1\right)\rho\sigma^2-\rho(\hat{v}_A+\hat{v}_B)+(\rho^2+1)(\hat{v}_{12,1}+\hat{v}_{12,2}+\hat{v}_{12,3}+\hat{v}_{12,4})}{\left(\rho^2-1\right)^2\sigma^2}. \tag{2.42}
$$

where $\hat{v}_A = \hat{v}_{11,1}+\hat{v}_{22,1}+\hat{v}_{11,3}+\hat{v}_{22,4}$ and $\hat{v}_B = \hat{v}_{11,2}+\hat{v}_{22,2}+\hat{v}_{22,3}+\hat{v}_{11,4}$.

The MLEs for σ and ρ are given as

$$
\hat{\sigma} = \sqrt{\frac{\hat{v}_A+\hat{v}_B}{8(n-1)}}, \tag{2.43}
$$

$$
\hat{\rho} = \frac{2(\hat{v}_{12,1}+\hat{v}_{12,2}+\hat{v}_{12,3}+\hat{v}_{12,4})}{\hat{v}_A+\hat{v}_B}. \tag{2.44}
$$

Under Ha, the partial derivatives with respect to σ_A, σ_B, and ρ are given as,

$$\frac{\partial logL}{\partial \sigma_A} = \frac{-4\sigma_A^2\sigma_B(n-1)\left(\rho^2-1\right)+\sigma_A\rho(\hat{v}_{12,3}+\hat{v}_{12,4})-\sigma_B(\hat{v}_A-2\rho\hat{v}_{12,1})}{\sigma_A^3\sigma_B\left(\rho^2-1\right)} = 0, \tag{2.45}$$

$$\frac{\partial logL}{\partial \sigma_B} = \frac{-4\sigma_B^2\sigma_A(n-1)\left(\rho^2-1\right)+\sigma_B\rho(\hat{v}_{12,3}+\hat{v}_{12,4})-\sigma_A(\hat{v}_B-2\rho\hat{v}_{12,2})}{\sigma_B^3\sigma_A\left(\rho^2-1\right)} = 0, \tag{2.46}$$

$$\begin{aligned}\frac{\partial logL}{\partial \rho} = & \frac{-\rho\hat{v}_A+(1+\rho^2)\hat{v}_{12,1}}{\sigma_A^2(\rho-1)^2(\rho+1)^2}+\frac{-\rho\hat{v}_B+(1+\rho^2)\hat{v}_{12,2}}{\sigma_B^2(\rho-1)^2(\rho+1)^2} \\ & +\frac{(1+\rho)^2(\hat{v}_{12,3}+\hat{v}_{12,4})}{\sigma_A\sigma_B(\rho-1)^2(\rho+1)^2}-\frac{4(n-1)\rho}{\rho^2-1} = 0.\end{aligned} \tag{2.47}$$

The above equations are satisfied for

$$\begin{aligned}\hat{\sigma}_B &= \frac{\hat{\sigma}_A\sqrt{\hat{v}_B}}{\sqrt{8\hat{\sigma}_A^2(n-1)-\hat{v}_A}}, \\ \hat{\rho} &= \frac{\hat{v}_B\left(\hat{v}_A-4\hat{\sigma}_A^2(n-1)\right)}{\hat{v}_{12,2}\left(\hat{v}_A-8\hat{\sigma}_A^2(n-1)\right)+\hat{v}_B\hat{v}_{12,1}}.\end{aligned}$$

The test statistic for LRT for the second test, t_2, AA/BB/AB/BA is given as follows

$$\begin{aligned}t_2 &= -2\log\left(\frac{L\left(\hat{\boldsymbol{\theta}}|H_0\right)}{L\left(\hat{\boldsymbol{\theta}}|Ha\right)}\right) \\ &= -2\Sigma_{k=1}^4\left(\frac{(n_k-1)}{2}\log\left(\frac{det(V_k|_{\hat{\boldsymbol{\theta}}_{Ha}})}{det(V_k|_{\hat{\boldsymbol{\theta}}_{H_0}})}\right)-\frac{1}{2}tr(\hat{V_k}(V_k|_{\hat{\boldsymbol{\theta}}_{H_0}}-V_k|_{\hat{\boldsymbol{\theta}}_{Ha}}))\right) \\ &= 4(n-1)\log\left(\frac{\hat{\sigma}^4(1-\hat{\rho}^2)}{\hat{\sigma}_A^2\hat{\sigma}_B^2(1-\hat{\rho}^2)}\right)+\frac{\hat{v}_A+\hat{v}_B-2\hat{\rho}(\hat{v}_{12,1}+\hat{v}_{12,2}+\hat{v}_{12,3}+\hat{v}_{12,4})}{(1-\hat{\rho}^2)\hat{\sigma}^2} \\ &- \frac{(\hat{\hat{\sigma}}_B^2(\hat{v}_A-2\hat{\hat{\rho}}\hat{v}_{12,2})+(\hat{\hat{\sigma}}_A^2(\hat{v}_B-2\hat{\hat{\rho}}\hat{v}_{12,1})-2\hat{\hat{\sigma}}_A\hat{\hat{\sigma}}_B(\hat{v}_{12,3}+\hat{v}_{12,4})}{\hat{\hat{\sigma}}_A^2\hat{\hat{\sigma}}_B^2(1-\hat{\hat{\rho}}^2)}.\end{aligned}$$

We used an optimization algorithm in the family of quasi-Newton methods that approximates the Broyden-Fletcher-Goldfarb-Shanno algorithm (L-BFGS) method, which uses a limited amount of computer memory to obtain numerical solutions. Then, the two LRTs, t_1 and t_2, are determined and assumed to follow χ_1^2, which will be confirmed via simulations.

The following steps were used to conduct simulations. The covariance matrix for sequence k, V_k, can be factored into LU decomposition with the matrix U defined as

$$U_k = \begin{pmatrix} \sigma_{d_{(1,k)}} & \rho\sigma_{d_{(2,k)}} \\ 0 & \sigma_{d_{(2,k)}}\sqrt{(1-\rho^2)} \end{pmatrix}.$$

Now suppose that $\boldsymbol{x} = [x_1, x_2]' \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right]$, then

$$\boldsymbol{y} = U_k'\boldsymbol{x} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, V_k\right].$$

We generate $\boldsymbol{x}$ for each subject in all sequences for different values of parameters and allocate an equal number of subjects to each of the sequences $AA/BB/AB/BA$ and denote this sample size for each sequence as n_k. For this simulation, $n_k = \{3, 6, 12, 18, 24\}$ and $\rho = \{0, 0.1, 0.3, 0.5, 0.7, 0.9\}$ were considered with $\alpha = 0.05$ and 5000 runs.

Table 7 summarizes the empirical sizes and powers for testing the variance inequality. The rows for $r = 1$ exhibit the empirical type I errors of the two tests obtained from 5000 simulations. These empirical type I errors converges to the true $\alpha = 0.05$ as n_k increases. It seems that $n_k = 12$ or 18 is sufficiently large enough sample sizes for each sequence k in the respective designs to approximate the likelihood ratio test statistic to χ^2 random variables with degrees of freedom of 1. The power for testing the equality of variances increases as n_k increases and r^2 deviates from 1. The statistical powers for t_1 appear to be independent of ρ, whereas the statistical powers for t_2 appear to be dependent on ρ similar to the case of Jung's AB/BA test. For example, the statistical power for t_2 achieves over 80% when $n_k \geq 6$ or $\rho = 0.9$ for $r^2 = 4$ whereas the power for t_1 is higher than 80% for $n_k \geq 12$. The more r^2 deviates from 1, the more powerful the test t_1 and t_2 become. In summary, our new test, t_2, detects the unequal variances of two treatments successfully for various values of σ_A^2, σ_B^2, ρ, and n_k, for the design with sequences $AA/BB/AB/BA$. We have used the MLEs for t_1 as the starting values in obtaining numerical solutions for t_2, as t_1 is based on sequences that are subsets for t_2.

Table 2.7: Entries are powers for tests t_1 (left panel) and t_2 (right panel), for testing $H_0 : \sigma_A^2 = \sigma_B^2$, when $\mu_A = 1$, $\mu_B = 0$, and $\gamma_A = 0.5$ and $\gamma_B = 0$.

$r^2 = \sigma_B^2/\sigma_A^2$	$\rho \backslash n_k$	3	6	12	18	24	3	6	12	18	24
1	0	0.056	0.052	0.052	0.050	0.051	0.068	0.056	0.053	0.055	0.050
	0.1	0.061	0.057	0.053	0.052	0.048	0.065	0.058	0.051	0.048	0.050
	0.3	0.060	0.056	0.052	0.053	0.052	0.066	0.057	0.054	0.054	0.054
	0.5	0.060	0.052	0.052	0.053	0.053	0.067	0.057	0.051	0.053	0.054
	0.7	0.056	0.055	0.053	0.054	0.053	0.070	0.060	0.054	0.054	0.053
	0.9	0.056	0.053	0.053	0.055	0.054	0.076	0.056	0.057	0.053	0.049
2	0	0.093	0.178	0.348	0.516	0.636	0.161	0.330	0.633	0.812	0.911
	0.1	0.097	0.176	0.350	0.507	0.643	0.173	0.330	0.633	0.813	0.914
	0.3	0.090	0.169	0.345	0.509	0.636	0.179	0.374	0.646	0.830	0.926
	0.5	0.093	0.168	0.338	0.502	0.639	0.197	0.376	0.686	0.860	0.949
	0.7	0.090	0.175	0.354	0.504	0.638	0.233	0.468	0.790	0.928	0.977
	0.9	0.093	0.177	0.342	0.510	0.634	0.400	0.739	0.970	0.997	1.000
4	0	0.212	0.527	0.877	0.977	0.997	0.477	0.861	0.995	1.000	1
	0.1	0.208	0.517	0.882	0.976	0.995	0.478	0.852	0.995	1.000	1
	0.3	0.212	0.521	0.874	0.974	0.996	0.492	0.874	0.995	1.000	1
	0.5	0.216	0.524	0.878	0.973	0.995	0.523	0.910	0.999	1.000	1
	0.7	0.216	0.529	0.883	0.973	0.997	0.621	0.946	1.000	1.000	1
	0.9	0.219	0.533	0.880	0.975	0.996	0.833	0.995	1.000	1.000	1
8	0	0.403	0.858	0.996	1	1	0.804	0.996	1	1	1
	0.1	0.407	0.860	0.996	1	1	0.805	0.994	1	1	1
	0.3	0.415	0.851	0.997	1	1	0.812	0.993	1	1	1
	0.5	0.419	0.857	0.996	1	1	0.850	0.998	1	1	1
	0.7	0.402	0.856	0.997	1	1	0.901	0.999	1	1	1
	0.9	0.406	0.854	0.996	1	1	0.973	1.000	1	1	1

We also observed mean square errors and biases of the estimates of variances and correlations. Figures 1 and 2 are depicted for t_2 to show similar results as we observed in power calculations such that mean square errors and bias approaches to zero as the sample sizes increases. The tests appear to be valid and estimation of variances consistent.

2.5 Conclusion

In this chapter, we have investigated how the unequal treatment variances affect the optimality of crossover designs. The ratio of treatment standard deviations, r, correlation coefficient for within subject measurements, ρ, and sample size for each sequence, n_k, were assumed to be fixed and classes of crossover designs $\Omega_{2,2,n}$ and $\Omega_{3,2,n}$ were investigated.

For $p = 2$, the T optimal designs for estimating direct treatment effects under two constraints were independent of $\rho > 0$ and showed that optimal

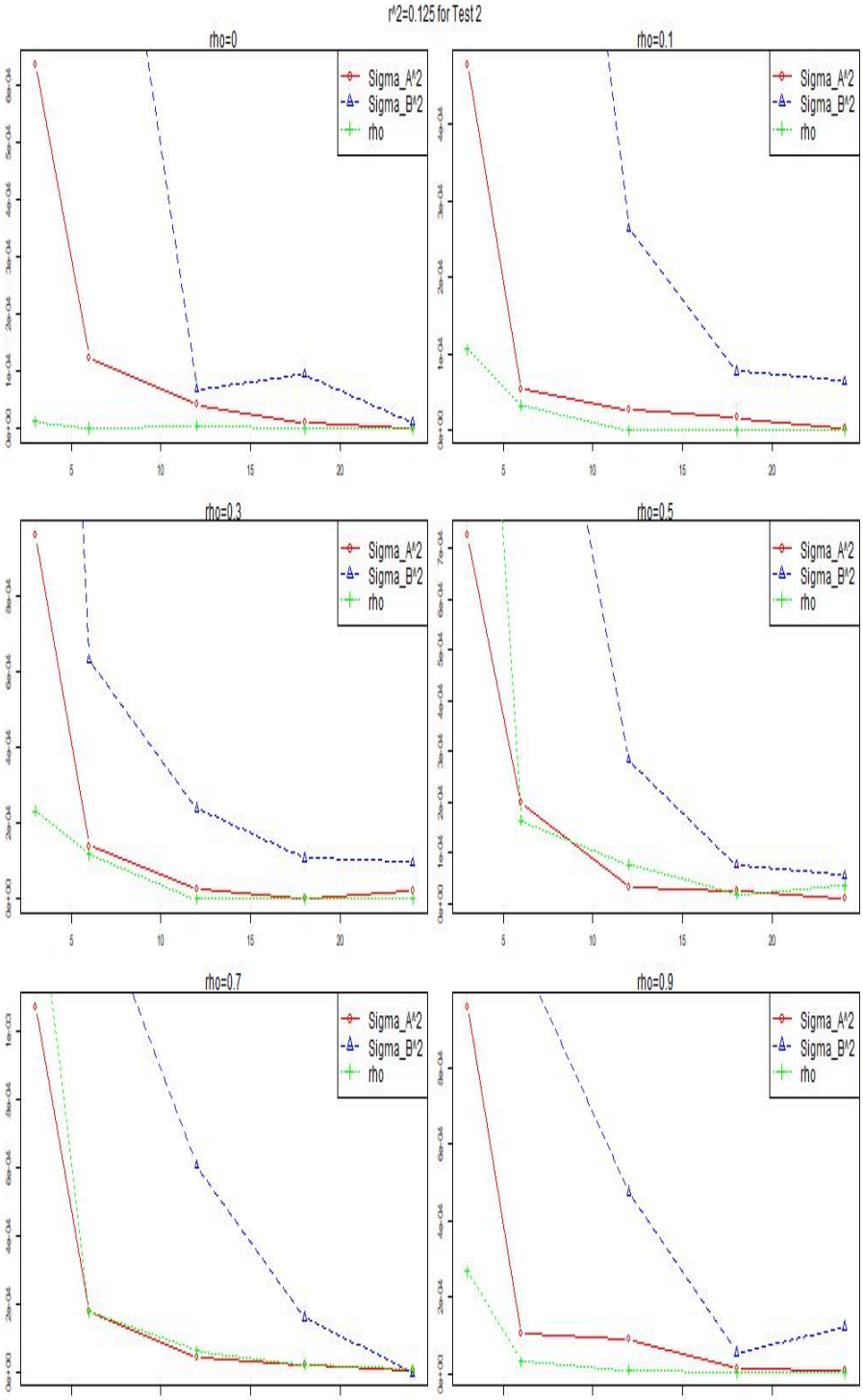

Figure 2.1: Plot of squared biases in estimates of σ_A^2, σ_B^2, and ρ as a function of sample size n_k for t_2: $r^2 = 1$ for $\sigma_A^2 = \sigma_B^2$ and $r^2 = 8$ for $\sigma_B^2 = 8 \times \sigma_A^2$

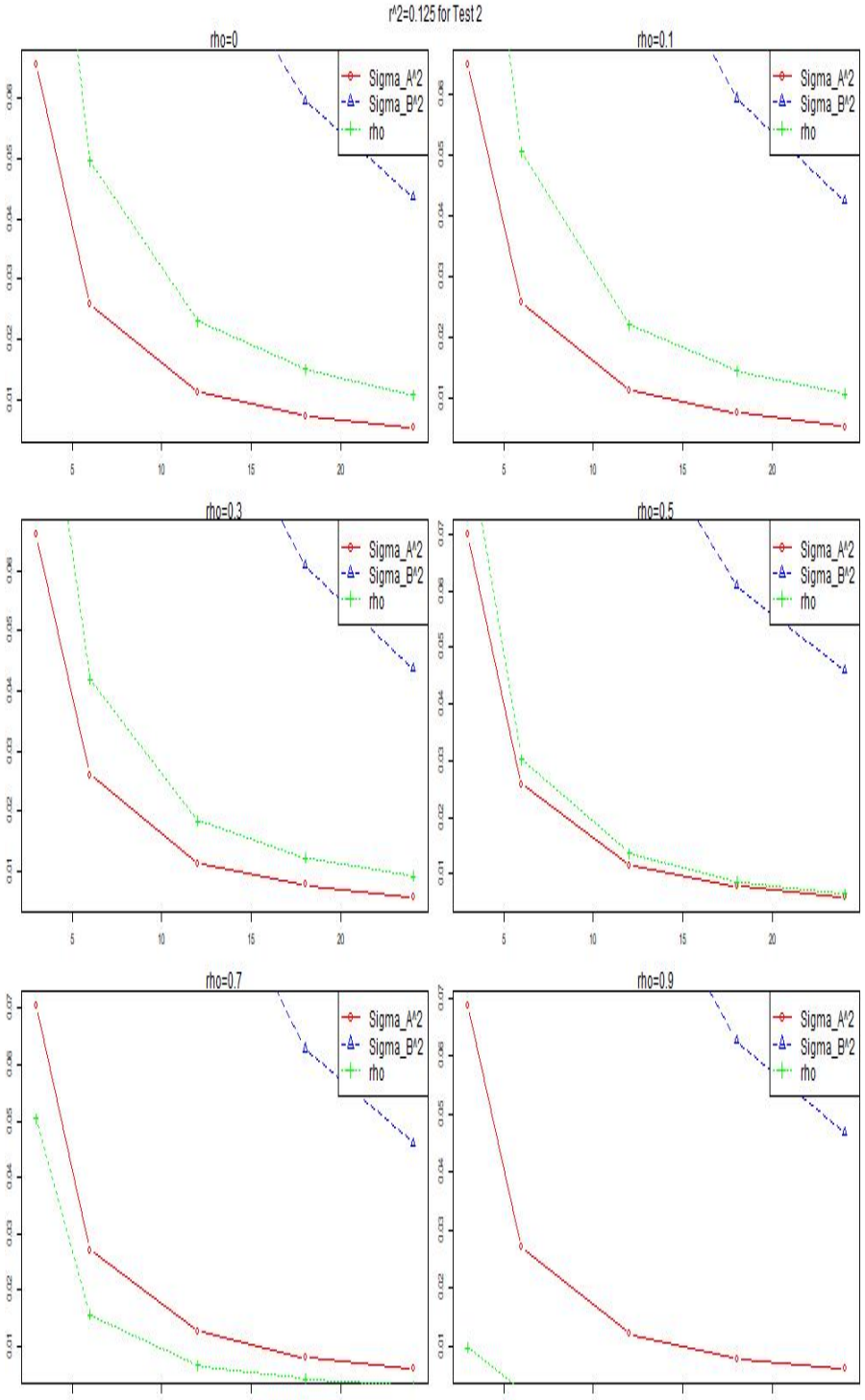

Figure 2.2: Plot of mean squared errors in estimates of σ_A^2, σ_B^2, and ρ as a function of sample size n_k for t_2: $r^2 = 1$ for $\sigma_A^2 = \sigma_B^2$ and $r^2 = 8$ for $\sigma_B^2 = 8 \times \sigma_A^2$

designs allocate subjects based on the ratio of the treatment standard deviations, r. For $p = 3$, the T optimal designs for estimating direct treatment effects depend on both r and ρ. As the ratio r deviates from 1, the T-optimal designs require the use of sequences AAA/BBB. But a strong positive within-subject correlation alleviates the problem and the existing universally optimal design under the equal variance assumption, ABB/BAA, remains competitive against inequality of treatment variances with high efficiency of over 80% when $\rho \geq 0.5$ and/or $1/3 \leq r \leq 3$. We also introduced a test of unequal treatment variances based on 4 treatment sequences among $\Omega_{2,2,n}$, which appear to work well for moderate sample sizes, especially when the ratio of variances are large.

It is clear that unequal treatment variances affect the optimality of crossover designs and must be accounted for in the designing stage through prior information or adaptively implemented during the trial, in which the test of equal treatment variance may be used. For example, a researcher may begin the experiment with 50/50 allocation on AB/BA sequences. Once sufficient information about variability is acquired, the researcher may decide to add AA/BB based on the obtained estimate of r or to extend the trial to $p = 3$, ABB/BAA, if ρ is sufficiently large. In the presence of carryover effects, researchers may choose to plan their trials with $AA/BB/AB/BA$. If unequal variances are found in the course of the trials, researchers may adapt their design to the T optimal design under the constraint (2) and benefit from the enhanced efficiency of the new constrained T optimal design. Suppose that we are conducting a two-period two-treatment four-sequence crossover trial allocating $N = 50$ subjects. Suppose that we have allocated $n_k = 6$ to each of the sequences. Moreover, we observed that the ratio of variances $\hat{r}$ is 2 and the test of unequal variances reject the null hypothesis. Then, the design may be updated to $n = (6, 22, 11, 11)$. This kind of adaptation can reduce the effect arising from existing carryover effects as well as unequal treatment variability and emphasizes the importance and usefulness of the universal optimal designs $AA/BB/AB/BA$ and ABB/BAA built under the equal variance assumption.

Chapter 3

Optimal Crossover Design with Proportional Carryover Effects and Baselines

Abstract

A model where carryover effects are proportional to direct treatment effects are considered to obtain efficient and practical designs. This chapter investigates the effect of introducing baseline measurements where carryover effects are also proportional to direct treatment effects. $A/D/E/T$ optimal designs, which are based on the average of inverse, products, minimum, and average of the eigenvalues of the Information matrix, are constructed. In this chapter, we find that designs assigning all distinct treatments or repeating last two periods are optimal or highly efficient for $t \geq p - 1$ for the negative and positive proportional carryover effects. We then compare these designs with optimal proportional designs without baseline measurements.

3.1 Introduction

Define the class of all crossover designs with p periods, t treatments, and n subjects as $\Omega_{p,t,n}$ as in Chapter 2. In this crossover design, the response of each subject $j \in \{1, 2, \cdot, n\}$ in period $i \in \{1, 2, \cdots, p\}$ is modeled with fixed and random effects. Traditionally we have considered response y_{ij}, from subject j in period i, with the following model:

$$y_{ij} = \mu + \alpha_i + \beta_j + \tau_{d(i,j)} + \gamma_{d(i-1,j)} + \epsilon_{ij} \tag{3.1}$$

where μ, α, β, τ, and γ refer to overall mean, period, subject, direct treatment, and carryover effects and ϵ is random error term. Also $d(i, j)$ refers to the treatment being administered on ith period of jth subject. Optimal designs were constructed from the above model, originally suggested by Hedayat and Afsarinejad (1975, 1978) for various p, t, and covariance structures of responses. The sufficiency condition for the universally optimal design was proposed by Kifer (1973) and Kushner (1997) proposed sufficiency and necessity equations for the universally optimal design.

Baseline measurements were considered important and useful measures to improve efficiencies. Some suggested the use of change from baseline measurements to analyze repeated measure data (Wallenstein (1979), Kershner and Federer (1981), Laska, Meisner, and Kushner (1983)), whereas others pointed out that the change from baseline approach may not always be appropriate as it may results in biased estimate of the treatment effects (Hills and Armitage (1979) and Willan and Pater (1986)). Wallensteins and Fisher (1977) considered the use of baseline measurements under various covariance structures and washout periods, where they assumed that the traditional carryover effects exist, and these carryover effects were assumed to be distinct from those of treatment responses. Jemielita (2017) considered using a linear contrast of baseline measurements as covariates with an multivariate normal assumption and various covariance structures among and between the responses and baseline measurements. Chen and Chinchilli (2010) expanded the general correlation coefficient (Chinchilli et al. (2005)) to multivariate random variables,

which can be applied to the case of treatment responses and baseline measurements to precisely identify the type of correlation within responses and baseline measurements as well as between responses and baseline measurements. Liang and Carriere (2010) further considered various carryover effects model for treatment responses and baseline measurements, and compared the efficiencies of different designs in $\Omega_{(2,3,4),2,n}$. Baseline measurements $x_{i,j}$'s for jth subject and ith period with traditional carryover effects can be modeled as the followings:

$$x_{ij} = \mu + \alpha'_i + \beta_j + \eta_{d(i-1,j)} + \epsilon_{xij} \tag{3.2}$$

where α'_i refers to the periodic effects in baseline measurements in ith baseline period and this is distinct from α_i in (3.1) and the η refers to the traditional carryover effects in baseline measurements, which are assumed to be distinct from the γ.

In crossover designs, carryover effects are inevitably related to direct treatment effects given that sequential dependencies exist (Cross (1973), and Cross and Decarlo (1990)). Cross (1973) and Schifferstein and Oudejans (1996) reported that carryover effects may depend on the magnitude of the direct treatment effects positively or negatively. Ferris (1999), in his PhD dissertation, proposed proportional carryover effects. Kempton, Ferris, and David (2001) investigated optimal crossover designs under proportional carryover effects, and Zheng (2013) applied the optimal equations methods to the proportional model.

This chapter extends the optimality equations of Kushner (1997) to incorporate baseline measurements, construct optimal designs under a proportional traditional carryover effects model with baseline measurements, and compare its efficiencies to the optimal proportional designs without baseline measurements.

3.2 Information Matrix and Optimality Equations

We can write the models in (3.1) and (3.2) in a matrix form as,

$$z_{i,j} = \begin{pmatrix} x_{i,j} \\ y_{i,j} \end{pmatrix} = \begin{pmatrix} \mu + \alpha_i' + \beta_j + \eta_{d(i,j)} + \epsilon_{xij} \\ \mu + \alpha_i + \beta_j + \tau_{d(i,j)} + \gamma_{d(i,j)} + \epsilon_{yij} \end{pmatrix},$$

$$\boldsymbol{Z_d} = \begin{pmatrix} z_{1,1} \\ \vdots \\ z_{n,p} \end{pmatrix},$$

and we can rearrange the $\boldsymbol{Z_d}$ by separating baseline measurements and outcome responses as below,

$$\boldsymbol{Z_d'} = \begin{pmatrix} \boldsymbol{X_d} \\ \boldsymbol{Y_d} \end{pmatrix} = \begin{pmatrix} 1_{np}\mu + W_1\boldsymbol{\alpha'} + U_1\boldsymbol{\beta} + F_d\boldsymbol{\eta} + \boldsymbol{\epsilon} \\ 1_{np}\mu + W_1\boldsymbol{\alpha} + U_1\boldsymbol{\beta} + T_d\boldsymbol{\tau} + F_d\boldsymbol{\gamma} + \boldsymbol{\epsilon} \end{pmatrix},$$

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{pmatrix} = A_1\boldsymbol{\theta_1} + A_2\boldsymbol{\theta_2} + \boldsymbol{\epsilon}, \tag{3.3}$$

$$= \begin{pmatrix} 1_{np} & 0_{np\times p} & W_1 & U_1 \\ 1_{np} & W_1 & 0_{np\times p} & U_1 \end{pmatrix} \begin{pmatrix} \mu \\ \boldsymbol{\alpha} \\ \boldsymbol{\alpha'} \\ \boldsymbol{\beta} \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{np\times t} & \mathbf{0}_{np\times t} & F_d \\ T_d & F_d & \mathbf{0}_{np\times t} \end{pmatrix} \begin{pmatrix} \boldsymbol{\tau} \\ \boldsymbol{\gamma} \\ \boldsymbol{\eta} \end{pmatrix} + \boldsymbol{\epsilon}. \tag{3.4}$$

Let $\mathrm{Cov}(\boldsymbol{Y}, \boldsymbol{X}) = \Sigma = I_n \otimes V$, where V is $2p \times 2p$ matrix with diagonal elements of $\sigma_\epsilon^2 + \sigma_s^2$ and off-diagonal element σ_s^2. The covariance matrix V may depend on various assumptions. For this chapter, we assume that it has an equicorrelated structure. Then, the Fisher Information matrix for direct treatment effects and two types of carryover effects can be derived as the following.

$$\begin{aligned} C_d(\boldsymbol{\tau}, \boldsymbol{\gamma}, \boldsymbol{\eta}) &= A_2'\Sigma^{-1}A_2 - A_2'\Sigma^{-1}A_1(A_1'\Sigma^{-1}A_1)^-A_1'\Sigma^{-1}A_2 \\ &= A_2'(\Sigma^{-1} - \Sigma^{-1}A_1(A_1'\Sigma^{-1}A_1)^-A_1'\Sigma^{-1})A_2, \end{aligned} \tag{3.5}$$

where the vector 1_{2np} is in the column space of design matrices for period effects and fixed subject effects.

We can simplify the above Information matrix by the following steps. We start by defining the inverse of V, a $2p \times 2p$ covariance matrix, as V^{-1} and

another notation for a matrix $H_w = I_w - 1_w 1_w'/w$ for any integer w. Then, the inner most term in the equation (3.5) can be seen as,

$$A_1'\Sigma^{-1}A_1 = \begin{bmatrix} nV^{-1} & 1_n' \otimes V^{-1}1_{2p} \\ 1_n \otimes 1_{2p}'V^{-1} & \delta I_n \end{bmatrix}$$

and

$$A_1'\Sigma^{-1} = A_1'\Sigma^{-1}A_1 \begin{pmatrix} (1_n' \otimes I_{2p})/(n) \\ (H_n \otimes 1_{2p}'V^{-1})/\delta \end{pmatrix} = \begin{pmatrix} 1_n' \otimes V^{-1} \\ I_n \otimes 1_{2p}'V^{-1} \end{pmatrix}.$$

If V is completely symmetric, then

$$\Sigma^{-1}A_1(A_1'\Sigma^{-1}A_1)^- A_1'\Sigma^{-1} = \begin{pmatrix} 1_n \otimes I_{2p}/n & H_n \otimes V^{-1}1_{2p}/\delta \end{pmatrix} \begin{pmatrix} 1_n' \otimes V^{-1} \\ I_n \otimes 1_{2p}'V^{-1} \end{pmatrix}$$
$$= \frac{1}{n}1_n 1_n' \otimes V^{-1} + \frac{1}{\delta}H_n \otimes V^{-1}1_{2p}1_{2p}'V^{-1}.$$

Then it follows,

$$\Sigma^{-1} - \Sigma^{-1}A_1(A_1'\Sigma^{-1}A_1)^- A_1'\Sigma^{-1} = H_n \otimes V^{-1} - \frac{1}{\delta}H_n \otimes V^{-1}1_{2p}1_{2p}'V^{-1}$$
$$= H_n \otimes (V^{-1} - \frac{1}{\delta}V^{-1}1_{2p}1_{2p}'V^{-1}) = H_n \otimes \tilde{V}$$

so that

$$C_d(\boldsymbol{\tau}, \boldsymbol{\gamma}, \boldsymbol{\eta}) = A_2'(H_n \otimes \tilde{V})A_2 \tag{3.6}$$

note that if covariance matrix is compound symmetric then $\tilde{V}$ is compound symmetric as well.

Now define $T_d^1 = \begin{pmatrix} T_d & \mathbf{0}_{np\times t} \end{pmatrix}'$, $T_d^2 = \begin{pmatrix} F_d & \mathbf{0}_{np\times t} \end{pmatrix}'$, $T_d^3 = \begin{pmatrix} \mathbf{0}_{np\times t} & F_d \end{pmatrix}'$, $T_{d,j}^1 = \begin{pmatrix} T_{d,j} & \mathbf{0}_{p\times t} \end{pmatrix}'$, $T_{d,j}^2 = \begin{pmatrix} F_{d,j} & \mathbf{0}_{p\times t} \end{pmatrix}'$, $T_{d,j}^3 = \begin{pmatrix} \mathbf{0}_{p\times t} & F_{d,j} \end{pmatrix}'$.

With the above notations, define,

$$M^u = n^{-1}\Sigma_{j=1}^n T_j^u.$$

The information matrix of direct treatment effects and carryover effects is derived as,

$$C_d(\tau, \gamma, \eta) = \begin{pmatrix} C_{d11} & C_{d12} & C_{d13} \\ C_{d21} & C_{d22} & C_{d23} \\ C_{d31} & C_{d32} & C_{d33} \end{pmatrix}, \tag{3.7}$$

where $C_{duv} = (T^u)' H_n \otimes \tilde{V}(T^u)$ for $1 \leq u, v \leq 3$.

These sub-matrices can be defined as

$$C_{duv} = \Sigma_{j=1}^{n} (T_j^u - M^u)' \tilde{V} (T_j^v - M^v) \text{ for } 1 \leq u, v \leq 3. \tag{3.8}$$

We denote p_d as the proportion of allocation to the treatment sequence d. Also define,

$$T = \begin{pmatrix} J_{pt}/t & J_{pt}/t & 0_{pt} \\ 0_{pt} & 0_{pt} & J_{pt}/t \end{pmatrix}.$$

Then the information matrix becomes,

$$\begin{aligned} C_d(\tau, \gamma, \eta) =& n\Sigma_{k=1}^{K} p_k (T_k - M_d)' \tilde{V} (T_k - M_d) && (3.9) \\ =& n\Sigma_{k=1}^{K} p_k (T_k - T)' \tilde{V} (T_k - T) - N(T - M_d)' \tilde{V} (T - M_d) && (3.10) \\ =& n\Sigma_{k=1}^{K} p_k (\hat{T}_k^u)' \tilde{V} \hat{T}_k^v - n\hat{M}_d' \tilde{V} \hat{M}_d && (3.11) \\ =& n\hat{L}_d - n\hat{M}_d' \tilde{V} \hat{M}_d, && (3.12) \end{aligned}$$

where the second term in equation (3.11) is $\mathbf{0}$ for symmetric designs. Symmetric designs refer to the design that allocates the same number of patients to the elements of the set of treatment sequences, which can be obtained by relabeling treatments in the sequence. For example, $ABC/ACB/BCA/BAC/CAB/CBA$ are in the symmetric set for $\Omega(3, 3, n)$ and we will refer this set as symmetric block of $< ABC >$.

Now, define a quadratic function, $q_k(x, y)$, associated to a treatment sequence k as

$$\begin{aligned} q_k(x, y) &= tr\left[(\hat{T}_d^1 + \hat{T}_d^2 x + \hat{T}_d^3 y)' \tilde{V} (\hat{T}_d^1 + \hat{T}_d^2 x + \hat{T}_d^3 y) \right] \\ &= q_{k11} + 2q_{k12}x + 2q_{k13}y + q_{k23}xy + q_{k22}x^2 + q_{k33}y^2, \text{ for } x, y \in (-\infty, \infty). \end{aligned} \tag{3.13}$$

Note: The quadratic functions for treatment sequences in a symmetric block are identical.

Define a symmetric group by S_t and a new permuted design $P_\sigma \leftrightarrow d_\sigma$ for each $\sigma \in S_t$. Then, there is a $t \times t$ permutation matrix H_σ representing the permutation σ. We have,

$$T^u_{d_\sigma,j} = T^u_{d,j}H_\sigma,\ \ M^u_{d_\sigma,j} = M^u_{d,j}H_\sigma, \text{ for } \sigma \in S_t, 1 \le u, v \le 3 \tag{3.14}$$

$$C_{d_\sigma uv} = H'_\sigma C_{duv} H_\sigma \tag{3.15}$$

$$C_{d_\sigma}(\tau,\gamma,\eta) = (I_3 \otimes H_\sigma)' C_d(\tau,\gamma,\eta)(I_3 \otimes H_\sigma) \tag{3.16}$$

$$T^u_{\sigma k} = T^u_k H_\sigma. \tag{3.17}$$

Let P be the array of proportions representing a design d. For any $P \leftrightarrow d$ the quadratic equation of a design d is defined as,

$$\begin{aligned} Q(x,y,P) =& \Sigma_{k=1}^K p_k q_k(x,y) \\ =& q_{11}(P) + 2q_{12}(P)x + 2q_{13}(P)y + 2q_{23}(P)xy \\ &+ q_{22}(P)x^2 + q_{33}(P)y^2. \end{aligned} \tag{3.18}$$

Any symmetric design $P \leftrightarrow d$ satisfies the following.

$$C_d(\tau,\gamma,\eta) = N \begin{pmatrix} q_{d11}(P) & q_{d12}(P) & q_{d13}(P) \\ q_{d21}(P) & q_{d22}(P) & q_{d23}(P) \\ q_{d31}(P) & q_{d32}(P) & q_{d33}(P) \end{pmatrix} \otimes H_t/(t-1), \tag{3.19}$$

$$\begin{aligned} C_d(\tau) =& N(q_{d11} + \frac{q^2_{d13}q_{d22} - 2q_{d12}q_{d13}q_{d23} + q^2_{d12}q_{d33}}{q^2_{d23} - q_{d22}q_{d33}}) \otimes (tI_t - J_t)/(t(t-1)) \\ =& N \min_{-\infty<x,y<\infty} Q(x,y,P)(tI_t - J_t)/(t(t-1)), \end{aligned} \tag{3.20}$$

$$\hat{M^u_d} = 0 \text{ for } 1 \le u \le 3. \tag{3.21}$$

Proof. If d is symmetric, then $M_d = X \otimes 1'_t$ with X being $p \times 3$ matrix. Then $M_d - T_d = 0$, leading to equation (3.21). Further, we use the permutation invariant properties of the block matrices of the Information matrix. C_{duv} must be of form $a * I + b * 1_t 1'_t$. Now using the property that row/column sums are 0, we can prove equation (3.19). Equation (3.20) is straight forward from Schur's complement of equation (3.19).

Theorem 4. *Kushner (1997) Thm.4.3*

Let

$$b = \max_{P} \min_{-\infty<x,y<\infty} Q(x, y, P) \tag{3.22}$$

.

Then

(i)

$$\begin{aligned} \max_{d} \Phi(C_d(\tau)) &= \max_{dsymmetric} \Phi(C_d(\tau)) && (3.23)\\ &= N\Phi(tI_t - J_t) \max_{P} \min_{-\infty<x,y<\infty} Q(x, y, P)/(t(t-1)) \\ &= Nb\Phi(tI_t - J_t)/(t(t-1)) && (3.24) \end{aligned}$$

for any Φ*,*

(ii) d *is* Φ*-optimal for* Φ *strictly concave (resp. universally optimal) if and only if its treatment effects information matrix is*

$$C_d(\tau) = Nb(tI_t - J_t)/(t(t-1)). \tag{3.25}$$

For the set of quadratics $\{q_k(x, y)\}$, define a_1, a_2, b, and $q(x, y)$ as the followings,

$$\begin{aligned} q(x, y) &= \max_{k}\{q_k(x, y)\}, && (3.26)\\ b &= \min_{-\infty<x,y<\infty} q(x, y), && (3.27)\\ q(a_1, a_2) &= \min_{-\infty<x,y<\infty} \max_{q_k(x,y)} = b. && (3.28) \end{aligned}$$

Suppose R is $3t \times 3t$ matrix with R_{11} a $t \times t$ matrix, R_{22} a $2t \times 2t$ diagonal sub-matrices, and $R_{12} = R_{21}^t$ a $t \times 2t$ off diagonal sub-matrices. Let R_{uv}^+ denotes Moor-Penrose inverse of R_{uv} for square matrix R_{uv}. For a non-zero scalar c, c^+ would mean reciprocal.

Lemma 8. *If* $C \geq 0$ *is a* $3t \times 3t$ *matrix, then*

$$trC^s \leq tr(C_{d11}) - \frac{(tr(C_{d13})^2 tr(C_{d22}) - 2tr(C_{d12})tr(C_{d13})tr(C_{d23}) + tr(C_{d12})^2 tr(C_{d33}))}{(tr(C_{d23}^2 - tr(C_{d22}tr(C_{d33}))} \tag{3.29}$$

Proof. The proof follows from Pukelsheim (2006) and Kushner (1997).

Let R be the $3t \times 3t$ Information matrix in (3.7) and $T = (I_t, X_t)$ where X_t is a $t \times 2t$ matrices. Then,

$$\begin{aligned} T'RT &= R_{11} + X_t'R_{21} + R_{12}X_t + X_t'R_{22}X_t \\ &= R^s + X_t'R_{22}X_t + X_t'R_{21} + R_{12}X_t + R_{12}R_{22}^+R_{21} \\ &= R^s + (X_t + R_{22}^+R_{21})'R_{22}(X_t + R_{22}^+R_{21}). \end{aligned} \tag{3.30}$$

.

Moreover, $C^+ \leq T'CT = C_{d11}+2xC_{d12}+2yC_{d13}+2xyC_{d23}+x^2C_{d22}+y^2C_{d33}$, which leads to

$$tr(C^s) \leq tr(C_{d11} + 2xC_{d12} + 2yC_{d13} + 2xyC_{d23} + x^2C_{d22} + y^2C_{d33}).$$

Now choose,

$$\begin{aligned} x &= (trC_{d13}trC_{d23} - trC_{d12}trC_{d33})/(trC_{d22}trC_{d33} - trC_{d23}^2) = a_1 \\ y &= (trC_{d13}trC_{d22} - trC_{d12}trC_{d23})/(trC_{d22}trC_{d33} - trC_{d23}^2) = a_2 \end{aligned}$$

to get (3.29)

□

Theorem 5. *A crossover design d with a_1, a_2, b, and $\mathcal{F}$ from* (3.26), (3.27), *and* (3.28) *is a universally optimal design if the following equations are satisfied.*

$$\Sigma_k p_k((\hat{T}_k^1)'\tilde{V}\hat{T}_k^1 + a_1(\hat{T}_k^1)'\tilde{V}\hat{T}_k^2 + a_2(\hat{T}_k^1)'\tilde{V}\hat{T}_k^3) = \frac{b(tI_t - J_t)}{t(t-1)}, \tag{3.31}$$

$$\Sigma_k p_k \begin{pmatrix} (\hat{T}_k^2)'\tilde{V}\hat{T}_k^1 + a_1((\hat{T}_k^2)'\tilde{V}\hat{T}_k^2 + (\hat{T}_k^2)'\tilde{V}\hat{T}_k^3), \\ (\hat{T}_k^3)'\tilde{V}\hat{T}_k^1 + a_2((\hat{T}_k^3)'\tilde{V}\hat{T}_k^2 + (\hat{T}_k^3)'\tilde{V}\hat{T}_k^3) \end{pmatrix} = 0_{2t\times t}, \tag{3.32}$$

$$\tilde{V}(\Sigma_k p_k(\hat{T}_k^1 + a_1\hat{T}_k^2 + a_2\hat{T}_k^3) = 0_{t\times t}, \tag{3.33}$$

$$p_k = 0, \quad if \quad k \notin \mathcal{F}. \tag{3.34}$$

Proof. As in Kushner (1997), we prove this theorem in the following equivalent

form.

$$\hat{L}_{d11} + a_1\hat{L}_{d12} + a_2\hat{L}_{d13} = \frac{b(tI_t - J_t)}{t(t-1)}, \tag{3.35}$$

$$\begin{pmatrix} \hat{L}_{d21} + a_1(\hat{L}_{d22} + \hat{L}_{d23}) \\ \hat{L}_{d31} + a_2(\hat{L}_{d23} + \hat{L}_{d33}) \end{pmatrix} = 0_{2t\times t}, \tag{3.36}$$

$$\tilde{V}(\hat{M}_d^1 + a_1\hat{M}_d^2 + a_2\hat{M}_d^3) = 0. \tag{3.37}$$

Let d be an universally optimal design and f be a symmetric and universally optimal design. Define $g = d/2 + f/2$. Let $A = \hat{L}_f$, $B = \hat{L}_d$, and $D = \hat{L}_g = A/2 + B/2$. A and B are divided into block matrices where A_{11} and B_{11} are $t \times t$ diagonal block matrices and A_{22} and B_{22} are $2t \times 2t$ diagonal block matrices.

Then,

$$\begin{aligned}&((X_t + A_{22}^+A_{21})'A_{22}(X_t + A_{22}^+A_{21}) + (X_t + B_{22}^+B_{21})'B_{22}(X_t + B_{22}^+B_{21}))/2\\ &= (X_t + D_{22}^+D_{21})'D_{22}(X_t + D_{22}^+D_{21}).\end{aligned}$$

If we choose $X_t = -D_{22}^+D_{21}$, then

$$\begin{aligned}&(A_{22}^+A_{21} - D_{22}^+D_{21})'A_{22}(A_{22}^+A_{21} - D_{22}^+D_{21})\\ &+ (B_{22}^+B_{21} - D_{22}^+D_{21})'B_{22}(B_{22}^+B_{21} - D_{22}^+D_{21})\\ &= Y'A_{22}Y + W'B_{22}W = 0.\end{aligned}$$

This means that $Y'A_{22}Y = W'B_{22}W = 0$ and $A_{22}Y = B_{22}W = 0$. As f is symmetric design, $A_{22} \neq 0$ and $A_{22}^+A_{12} = \begin{pmatrix} a_1 & a_2 \end{pmatrix} \otimes (I_t - J_t/t)$ with $A_{21} = A_{12}'$. This leads to,

$$\begin{aligned}Y &= -D_{22}^+D_{21} - \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \otimes (I_t - J_t/t) = 0_{2t\times t}\\ \implies W &= \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \otimes (I_t - J_t/t) + B_{22}^+B_{21},\end{aligned}$$

which proves (3.36). Now apply this result to (3.20). We now get,

$$b(tI_t - J_t)/(t(t-1)) = n^{-1}C_d(\tau) = \hat{L}_{d11} - D_{d12}D_{d22}^{+}D_{d21}$$
$$= \hat{L}_{d11} + \begin{pmatrix} a_1 & a_2 \end{pmatrix} \otimes (I_t)\hat{D}_{d21}$$
$$= \hat{L}_{d11} + a_1\hat{L}_{d12} + a_3\hat{L}_{d13} = (3.19).$$

From equations (3.10), (3.23), and (3.29) we have

$$\begin{aligned} Nb =& tr(C_d(\tau)) \leq tr(C_{d11} + 2xC_{d12} + 2yC_{d13} + 2xyC_{d23} + x^2C_{d22} + y^2C_{d33}) \\ &= NQ(x, y, P) - N(tr(\hat{M}_d^1 + x\hat{M}_d^2 + y\hat{M}_d^3)'\tilde{V}(\hat{M}_d^1 + x\hat{M}_d^2 + y\hat{M}_d^3)). \end{aligned}$$

Setting $x = a_1$ and $y = a_2$,

$$\begin{aligned} b \leq\ & b - tr((\hat{M}_d^1 + a_1\hat{M}_d^2 + a_2\hat{M}_d^3)'\tilde{V}(\hat{M}_d^1 + a_1\hat{M}_d^2 + a_2\hat{M}_d^3)) \\ \implies\ & tr((\hat{M}_d^1 + a_1\hat{M}_d^2 + a_2\hat{M}_d^3)'\tilde{V}(\hat{M}_d^1 + a_1\hat{M}_d^2 + a_2\hat{M}_d^3)) = 0 \\ \implies\ & \tilde{V}(\hat{M}_d^1 + a_1\hat{M}_d^2 + a_2\hat{M}_d^3) = 0. \end{aligned}$$

□

3.3 Proportional Model

Consider the assumption that carryover effects in responses and baselines are proportional to the direct treatment effects in the responses. That is, there exist λ_{1o}, $\lambda_{2o} \in (-1, 1)$ such that $\boldsymbol{\gamma}_o = \lambda_{1o}\boldsymbol{\tau}_o$ and $\boldsymbol{\eta}_o = \lambda_{2o}\boldsymbol{\tau}_o$. These proportionality parameters can be estimated based on previous experiences or adaptively during the trials. Moreover, the proportional modeling approach can improve the estimation of the direct treatment effects by reducing the number of parameters to be estimated (Kempton, Ferris, and David (2001), Bailey and Kunert (2006), Bose and Stufken (2007)).

Then, equation (3.3) becomes,

$$\boldsymbol{Z}_d = \begin{pmatrix} \boldsymbol{Y} \\ \boldsymbol{X} \end{pmatrix} = A_1\boldsymbol{\theta_1} + A_2\boldsymbol{\theta_2} + \boldsymbol{\epsilon} \tag{3.38}$$

$$= \begin{pmatrix} 1_{np} & W_1 & 0_{np\times p} & U_1 \\ 1_{np} & 0_{np\times p} & W_1 & U_1 \end{pmatrix} \begin{pmatrix} \mu \\ \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} + \begin{pmatrix} T_d & F_d & \mathbf{0}_{np\times t} \\ \mathbf{0}_{np\times t} & \mathbf{0}_{np\times t} & F_d \end{pmatrix} \begin{pmatrix} \boldsymbol{\tau} \\ \lambda_1\boldsymbol{\tau} \\ \lambda_2\boldsymbol{\tau} \end{pmatrix} + \boldsymbol{\epsilon}. \tag{3.39}$$

This proportional model depends on values of the parameters λ_1, λ_2, and $\boldsymbol{\tau}$. The Fisher Information matrix for estimating direct treatment effects under the proportional carryover effects with baseline measurements can be defined as the following. First, we apply the Taylor's expansion to (3.38) and obtain the following linearized model.

$$\tilde{Z}_d \approx W\boldsymbol{\alpha} + U\boldsymbol{\beta} + (T_d + F_d\lambda_{1o} + \tilde{F}_d\lambda_{2o})\boldsymbol{\tau} + F_d\boldsymbol{\tau}_o\lambda_1 + \tilde{F}_d\boldsymbol{\tau}_o\lambda_2 + \boldsymbol{\epsilon}. \quad (3.40)$$

We use the projection on the design matrices of the fixed period effects and subject effects to derive the information matrix for estimating direct treatment effects with unknown true values of λ_{1o}, λ_{2o}, and $\boldsymbol{\tau}_o$. Recall $pr(B1|B2) = pr(B1) + pr(pr^{\perp}(B1)B2)$. Then,

$$\begin{aligned} C_d(\tau) &= (T_d + F_d\lambda_{1o} + \tilde{F}_d\lambda_{2o})' pr^{\perp}(W|U|F_d\tau_o|\tilde{F}_d\tau_o)(T_d + F_d\lambda_{1o} + \tilde{F}_d\lambda_{2o}) \\ &= (T_d + F_d\lambda_{1o} + \tilde{F}_d\lambda_{2o})'(pr^{\perp}(W|U) \\ &\quad - pr(pr^{\perp}(W|U)(F_d\tau_o|\tilde{F}_d\tau_o))(T_d + F_d\lambda_{1o} + \tilde{F}_d\lambda_{2o}), \end{aligned}$$

and

$$\begin{aligned} C_{d,\tau_o,\lambda_{1o},\lambda_{2o}}(\tau) &= C_{d11} + \lambda_{1o}(C_{d12} + C_{d21}) + \lambda_{2o}(C_{d13} + C_{d31}) \\ &\quad + \lambda_{1o}\lambda_{2o}(C_{d23} + C_{d32}) + \lambda_{1o}^2 C_{d22} + \lambda_{2o}^2 C_{d33} \\ &\quad - A'\boldsymbol{\tau}_o \begin{pmatrix} \boldsymbol{\tau}_o' C_{d22}\boldsymbol{\tau}_o & \boldsymbol{\tau}_o' C_{d23}\boldsymbol{\tau}_o \\ \boldsymbol{\tau}_o' C_{d32}\boldsymbol{\tau}_o & \boldsymbol{\tau}_o' C_{d33}\boldsymbol{\tau}_o \end{pmatrix}^{-1} \boldsymbol{\tau}_o' A, \end{aligned} \quad (3.41)$$

where $A' = \begin{pmatrix} C_{d12} + \lambda_{1o}C_{d22} + \lambda_{2o}C_{d32} & C_{d13} + \lambda_{1o}C_{d23} + \lambda_{2o}C_{d33} \end{pmatrix}$. The above Information matrix depends on λ_{1o}, λ_{2o}, and $\boldsymbol{\tau}_o$. Bose and Stufken (2007) suggested that the unknown proportionality parameter may be fixed(known) at the stage of the experiment and showed that the information matrix no longer depends on $\boldsymbol{\tau}_o$. But this approach results in a biased estimation. Kempton, Ferris, and David (2001) used a computer algorithm to find A optimal designs for the proportional model without baselines. Bailey and Kunert (2006) used the A optimality and proved that there exists totally balanced design, which is A optimal over all $\Omega_{p,t,n}$ for limited sets of p, t, and λ_{1o}. Zheng (2013) devised $A/D/T/E$ optimality criteria. These criteria are based on the Fisher's

Information matrix. The A optimality criterion aims to maximize the geometric mean of the eigenvalues, whereas the D optimality criterion aims to maximize the determinant. Also, the E optimality criterion aims to maximize the smallest eigenvalue and the T optimality criterion aims to maximize the trace of the Information matrix. Zheng (2013) showed that universally optimal designs for a non-proportional model is E optimal in proportional model and that is near-optimal with respect to $A/D/T$ optimality criteria for all λ_{1o} (Equivalence Theorem).

Define a probability measure assigning an equal allocation to all elements $\{\sigma\tau_o|\sigma \in \mathcal{P}\}$ as δ_{τ_o}, and define a Bayesian type of optimality criteria ϕ as

$$\begin{aligned}\phi_{(g,\lambda_{1o},\lambda_{2o})}(d) &= \int \Phi(C_{d,\tau_o,\lambda_{1o},\lambda_{2o}}(\tau)g(\tau_o)d(\tau_o) \\ &= E_g(\Phi(C_{d,\tau_o,\lambda_{1o},\lambda_{2o}}(\tau)).\end{aligned} \tag{3.42}$$

Theorem 6. *(Kushner (1997), Zheng (2013)) In an approximate design theory, there exists a symmetric design d^* for any real values of λ_{1o}, λ_{2o}, and $\boldsymbol{\tau}_o$ that satisfies the following,*

$$\phi_{\delta_{\tau_o},\lambda_{1o},\lambda_{2o}}(d) \leq \phi_{\delta_{\tau_o},\lambda_{1o},\lambda_{2o}}(d^*) \text{ for all } d \tag{3.43}$$

Proof. The proof uses the properties of symmetric designs and is straightforward from the proof in Zheng (2013) Theorem 3. □

Theorem 7. *Regardless of the value of the treatment effects, $\boldsymbol{\tau}_o$, the Fisher Information matrix for estimating the direct treatment effects for a symmetric design d under the proportional model with baselines, $C_{d,\tau_o,\lambda_{1o},\lambda_{2o}}(\tau)$, has the following eigenvalues with multiplicities of 1, 1, and $t-2$.*

- κ_1=*0,*
- $\kappa_2 = c_{d11} - (c_{d13}^2 c_{d22} - 2c_{d12}c_{d13}c_{d23} + c_{d12}^2 c_{d33})/(c_{d22}c_{d33} - c_{d23}^2)$,
- $\kappa_3 = c_{d11} + 2\lambda_{1o}c_{d12} + 2\lambda_{2o}c_{d13} + 2\lambda_{1o}\lambda_{2o}c_{d23} + \lambda_{1o}^2 c_{d22} + \lambda_{2o}^2 c_{d33}$.

Proof. Recall that for a symmetric design d, the block matrices of the Information matrix C_{uv} have 0 row/column sums and take the form of $C_{duv} =$

$c_{duv}H_t/(t-1)$ with $c_{duv} = tr(C_{duv})$. Then (3.41) becomes,

$$
\begin{aligned}
&(t-1)C_{d,\tau_o,\lambda_{1o},\lambda_{2o}}(\tau) \\
&= (c_{d11} + 2\lambda_{1o}c_{d12} + 2\lambda_{2o}c_{d13} + 2\lambda_{1o}\lambda_{2o}c_{d23} + \lambda_{1o}^2 c_{d22} + \lambda_{2o}^2 c_{d33})H_t \\
&- \frac{(c_{d12} + \lambda_{1o}c_{d22} + \lambda_{2o}c_{d32})^2 c_{d33} + (c_{d13} + \lambda_{1o}c_{d23} + \lambda_{2o}c_{d33})^2 c_{d22}}{c_{d22}c_{d33} - c_{d23}^2} \frac{\boldsymbol{\tau}_o\boldsymbol{\tau}_o'}{\boldsymbol{\tau}_o'\boldsymbol{\tau}_o} \\
&+ \frac{(c_{d12} + \lambda_{1o}c_{d22} + \lambda_{2o}c_{d32})(c_{d13} + \lambda_{1o}c_{d23} + \lambda_{2o}c_{d33})c_{d23}}{c_{d22}c_{d33} - c_{d23}^2} \frac{\boldsymbol{\tau}_o\boldsymbol{\tau}_o'}{\boldsymbol{\tau}_o'\boldsymbol{\tau}_o},
\end{aligned} \tag{3.44}
$$

where $H_t 1_t = 0$ and $H_t\boldsymbol{\tau}_o = \boldsymbol{\tau}_o$. Suppose $l_1, \cdots, l_{t-2}$ are orthonormal vectors orthogonal to 1_t and $\boldsymbol{\tau}_o$ forming the t eigenvectors. The 3 corresponding eigenvalues have multiplicities of 1, 1, and $t-2$.

□

Zheng (2013) proposed a new optimality criteria $\mathcal{E}_{(g,\lambda_{1o},\lambda_{2o})}(d)$ where the second smallest eigenvalue is used to evaluate the optimality criteria $\phi_{(g,\lambda_{1o},\lambda_{2o})}(d)$.

Proposition 5. *In the approximate design theory, regardless of λ_{1o}, λ_{2o}, and the prior distribution g, as long as g is exchangeable, a design d is $\mathcal{E}_{(g,\lambda_{1o},\lambda_{2o})}$ optimal if and only if $\mathcal{E}_{(g,\lambda_{1o},\lambda_{2o})}(d) = nb/(t-1)$ with b defined in Theorem 4.*

Proof. The proof takes same steps as in Proposition 1 in Zheng (2013). First, it is easy to show $\kappa_1 < \kappa_2$ as κ_1 is the minimum of quadratic function $q(\lambda_{1o}, \lambda_{2o})$ for any design d. This proposition is proved by Theorem 4 (ii). □

Theorem 8. *A crossover design d with a_1, a_2, b, and $\mathcal{F}$ from* (3.26), (3.27), *and* (3.28) *is $\mathcal{E}_{(g,\lambda_{1o},\lambda_{2o})}$ optimal for any values of λ_{1o} and λ_{2o} and any exchangeable prior g if following equations are satisfied.*

$$\Sigma_k p_k((\hat{T}_k^1)'\tilde{V}\hat{T}_k^1 + a_1(\hat{T}_k^{\,1})'\tilde{V}\hat{T}_k^2 + a_2(\hat{T}_k^1)'\tilde{V}\hat{T}_k^3) = \frac{b(tI_t - J_t)}{t(t-1)}, \tag{3.45}$$

$$\Sigma_k p_k \begin{pmatrix} (\hat{T}_k^2)'\tilde{V}\hat{T}_k^1 + a_1((\hat{T}_k^2)'\tilde{V}\hat{T}_k^2 + (\hat{T}_k)'\tilde{V}\hat{T}_k^3), \\ (\hat{T}_k^3)'\tilde{V}\hat{T}_k^1 + a_2((\hat{T}_k^3)'\tilde{V}\hat{T}_k^2 + (\hat{T}_k^3)'\tilde{V}\hat{T}_k^3) \end{pmatrix} = 0_{2t\times t}, \tag{3.46}$$

$$\tilde{V}(\Sigma_k p_k(\hat{T}_k^1 + a_1\hat{T}_k^2 + a_2\hat{T}_k^3) = 0_{t\times t}, \tag{3.47}$$

$$p_k = 0, \quad if \quad k \in \mathcal{F}. \tag{3.48}$$

Proof. By the properties of the Information matrix,

$$\mathcal{E}_{(\tau_o,\lambda_{1o},\lambda_{2o})}(\tau) = E_g\left[\min_{l'1_t=0, l'l=1} l' C_{(\tau_o,\lambda_{1o},\lambda_{2o})}(\tau) l\right].$$

We know that $C_{d22} = C_{d33}$, and off-diagonal submatrices C_{d12}, C_{d13}, and C_{d23} are symmetric. Applying (3.45) - (3.48) to (3.41) we have,

$$\begin{aligned} C_{(\tau_o,\lambda_{1o},\lambda_{2o})}(\tau) = \frac{b}{t-1} H_t + (a_1-\lambda_{1o})^2 C_{d22} + (a_2-\lambda_{2o})^2 C_{d22} \\ + (a_1^2 + a_2^2 - 2\lambda_{1o}a_1 - 2\lambda_{2o}a_2 + 2\lambda_{1o}\lambda_{2o}) C_{d23} \\ - A' \boldsymbol{\tau}_o \begin{pmatrix} \boldsymbol{\tau}_o' C_{d22} \boldsymbol{\tau}_o & \boldsymbol{\tau}_o' C_{d23} \boldsymbol{\tau}_o \\ \boldsymbol{\tau}_o' C_{d32} \boldsymbol{\tau}_o & \boldsymbol{\tau}_o' C_{d33} \boldsymbol{\tau}_o \end{pmatrix}^{-1} \boldsymbol{\tau}_o' A, \end{aligned} \tag{3.49}$$

where $A' = \begin{pmatrix} (\lambda_{1o}-a_1)C_{d22} + (\lambda_{2o}-a_1)C_{d23} & (\lambda_{2o}-a_1)C_{d23} + (\lambda_{2o}-a_2)C_{d22} \end{pmatrix}$.

Now let $\{0, e_{1,1}, \cdots, e_{1,t-1}\}$ be eigenvalues of C_{d22} with corresponding eigenvectors of $\{1_t, l_{1,1}, \cdots, l_{1,t-1}\}$ and $\{0, e_{2,1}, \cdots, e_{2,t-1}\}$ be eigenvalues of C_{d23} with corresponding eigenvectors of $\{1_t, l_{2,1}, \cdots, l_{2,t-1}\}$. This means that $C_{d22} = \Sigma_{i=1}^{t-1} e_{1,i} l_{1,i} l_{1,i}'$ and $C_{d23} = \Sigma_{i=1}^{t-1} e_{2,i} l_{2,i} l_{2,i}'$. By definition, $\boldsymbol{\tau}_o' 1_t = 0$, which implies that there exist $\{w_{1,1}, \cdots, w_{1,t-1}\}$ and $\{w_{2,1}, \cdots, w_{2,t-1}\}$ such that $\boldsymbol{\tau}_o = \Sigma_{i=1}^{t-1} w_{1,i} l_{1,i} = \Sigma_{i=1}^{t-1} w_{2,i} l_{2,i}$. For any vector l such that $l'1_t = 0$, we can write $l = \Sigma_{i=1}^{t-1} y_{1,i} l_{1,i} = \Sigma_{i=1}^{t-1} y_{1,i} l_{2,i}$ and $l' H_t l = 1$. Then (3.49) becomes,

$$l' C_{(\tau_o,\lambda_{1o},\lambda_{2o})}(\tau) l \tag{3.50}$$

$$\begin{aligned} &= \frac{nb}{t-1} + (a_1-\lambda_{1o})^2 \Sigma_{i=1}^{t-1} e_{1,i} y_{1,i}^2 + (a_2-\lambda_{2o})^2 \Sigma_{i=1}^{t-1} e_{1,i} y_{1,i}^2 \\ &+ (a_1^2 + a_2^2 - 2\lambda_{1o}a_1 - 2\lambda_{2o}a_2 + 2\lambda_{1o}\lambda_{2o}) \Sigma_{i=1}^{t-1} e_{2,i} y_{2,i}^2 \\ &- l' A' \boldsymbol{\tau}_o \begin{pmatrix} \Sigma_{i=1}^{t-1} e_{1,i} w_{1,i}^2 & \Sigma_{i=1}^{t-1} e_{1,i} w_{2,i}^2 \\ \Sigma_{i=1}^{t-1} e_{1,i} w_{2,i}^2 & \Sigma_{i=1}^{t-1} e_{1,i} w_{1,i}^2 \end{pmatrix}^{-1} \boldsymbol{\tau}_o' A l \end{aligned} \tag{3.51}$$

$$\geq \frac{nb}{t-1} \text{ equality holds iff } l = \frac{\boldsymbol{\tau}_o}{\|\boldsymbol{\tau}_o\|}, \text{ with} \tag{3.52}$$

$$\boldsymbol{\tau}_o' A l = \begin{pmatrix} (\lambda_{1o}-a_1)\Sigma_{i=1}^{t-1} w_{1,i} y_{1,i} e_{1,i} + (\lambda_{2o}-a_1)\Sigma_{i=1}^{t-1} w_{2,i} y_{2,i} e_{2,i}. \\ (\lambda_{2o}-a_1)\Sigma_{i=1}^{t-1} w_{2,i} y_{2,i} e_{2,i} + (\lambda_{2o}-a_2)\Sigma_{i=1}^{t-1} w_{1,i} y_{1,i} e_{1,i}. \end{pmatrix}$$

The equality in (3.52) holds as the equation (3.46) implies that $a_1 = a_2 = 0$ as $c_{d12} = c_{d13} = 0$. □

This theorem implies that the universally optimal design for (3.3) is $\mathcal{E}$ optimal for (3.38). Zheng (2013) proposed a new set of $A/D/T/E$ optimality

criteria for $t \times t$ Information matrices, C, with eigenvalues of $\{0 \leq a_1 \leq \cdots \leq a_{t-1}\}$.

$$\Phi_A(C) = (t-1)\left(\Sigma_{i=1}^{t-1} a_i^{-1}\right)^{-1},$$

$$\Phi_D(C) = \left(\Pi_{i=1}^{t-1} a_i\right)^{\frac{1}{t-1}},$$

$$\Phi_T(C) = (t-1)^{-1}\Sigma_{i=1}^{t-1} a_i,$$

$$\Phi_E(C) = \min_{i>0} a_i.$$

3.4 Optimal Designs Utilizing Baselines

From Theorem 5 and Theorem 6, we know that universally optimal designs under model (3.3) is E optimal in (3.38). Also, Zheng (2013) investigated the proportional model without baselines and showed that E optimal designs are equivalent to $A/D/T$ optimal designs. In this section, we will investigate $A/D/E/T$ optimal designs for $\{\rho, \lambda_{1o}, \lambda_{2o}\} \in \{-0.5, 0, 0.5\}$. After identifying the optimal designs, we further investigate the efficiencies with respect to $A/D/E/T$ optimality criteria for various designs to give an idea on how designs perform in practice. Then, the E optimal designs constructed from Zheng's E criteria and algorithm are compared to the new proposed D optimal designs. The D optimal designs were chosen as these demonstrate the variation of efficiency most vividly. The four columns under A, D, E, and T show the efficiencies based on $A/D/E/T$ optimality criteria (the closer to 1 the better the row design is), and the column $E_{nobaseline}$ shows the D optimality efficiencies against the E optimal design without baseline measurements. The efficiencies in this column with > 1 indicate that optimal designs with baselines are better than ones without baselines, indicating efficiency improved by the use of baselines. The Equivalence theorem provided by Zheng (2013) suggests that the E optimal design is nearly optimal in other criteria. For this reason,

we chose E optimal design and D-optimality criterion in our comparison.

The proportionality parameters λ_{1o} and λ_{2o} would naturally take the same sign but not necessarily the same magnitude. For this reason, we performed a preliminary investigation on how unequal magnitude of parameters affect the optimal designs. It was found that the differences in these parameters did not affect the optimal designs much, as long as they had same sign. Therefore, $\lambda_{1o} = \lambda_{2o}$ was assumed for the construction of optimal designs. The optimal designs would consist of treatment sequences using distinct treatments for the case where carryover effects are succesfully removed by the washout periods or when $t \geq p$. In this chapter, we investigate the cases where washout periods did not remove carryover effects and that carryover effects are negatively or positively proportional as well as the case when carryover effects are successfully removed, $(\lambda_{1o} = \lambda_{2o} \in \{-0.5, 0, 0.5\})$. In addition to this, we will investigate how optimal designs change with a correlation coefficient $\rho \in \{-0.5, 0, 0.5\}$.

We first adopt the notations from Zheng (2013) and define $d_{<di>}$ and $d_{<re>}$ as design of treatment sequences with p and $(p-1)$ distinct treatments where the latter one repeats its $(p-1)$th treatment on the last period. Designs $d_{<di>}$ and $d_{<re>}$ allocate an equal proportion of subjects to all treatment sequences, $p_i = 1/|d|$. Moreover, we will use a symmetric Design $d_{<abc>}$, which is defined as a set of treatment sequences that can be obtained by relabeling the given treatment sequence abc for $1 \leq a, b, c \leq t$. In the case of $t = 2$, these design blocks are called dual designs. For example, $d_{<di>} = \{AB, BA\}$ and $d_{<re>} = \{AA, BB\}$ for $p = 2$ and $d_{<re>} = \{ABB, BAA\}$ for $p = 3$. For $t = 3$, $d_{<di>} = \{ABC, ACB, BAC, BCA, CAB, CBA\}$ and $d_{<re>} = \{ABB, ACC, BAA, BCC, CAA, CBB\}$. Some patterns appeared during the construction of $A/D/E/T$ optimal designs under the proportional carryover effects model with baselines. The two designs, $d_{<di>}$ and $d_{<re>}$, appear repeatedly as optimal designs for many p and t combination settings. The Design $d_{<di>}$ is feasible only for $t \geq p$ and the Design $d_{<re>}$ is feasible only for $t + 1 \geq p$. When $t < p-1$, several symmetric blocks of treatment sequences allocating the same treatment on the last two periods play important role in the optimal designs.

3.4.1 Two Period Designs

The A optimal designs of Bailey and Kunert (2006) did not include two period designs as their optimality was restricted to $2 < p \leq t$. The case for $p = 2$ and $t = 2$ pose no practical advantage as the proportional model approach requires 6 parameters fo, the overall mean effect, and the period effects, the direct treatment effects, and the two proportionality parameters, whereas the traditional model requires only 5 parameters with $\tau = (\tau_A - \tau_B)/2$, $\gamma = (\gamma_A - \gamma_B)/2$, and $\eta = (\eta_A - \eta_B)/2$. Therefore, we will look at $p = 2$ and $t \geq 3$. Some interesting results are found for two period designs. When $t \geq 3$, the optimal designs depend on t and $(\lambda_{1o}, \lambda_{2o})$ in the presence of carryover effects. As in the traditional models, the optimal two period crossover designs are independent of ρ. The optimal designs for the negative proportionality parameters differ by criteria. However, these designs allocate a small portion of subjects to $d_{<re>}$ and the rest on $d_{<di>}$. for the positive proportionality parameters, the $A/D/E/T$ optimal designs resemble that of Carriere and Reinsel (1993), where the universally optimal design for the traditional model with random effects and a compound symmetric structure were shown to be $p_{d<re>} = 1/t$ and $p_{d<di>} = (t-1)/t$. The optimal designs and their efficiencies are shown in Table 3.1. For the tables below, the numbers under $A/D/E/T$ columns refer to the efficiency of the design presented on that row against $A/D/E/T$ optimal designs in the corresponding criterion. In these columns, efficiencies are ≤ 1 indicating how efficient the proposed design is when compared with the corresponding $A/D/E/T$ optimal designs. Moreover, the last column presents the efficiency against E optimal design without baselines computed from the criteria and algorithm from Zheng (2013) with respect to D optimality. The main reason for choosing E optimal designs from Zheng's approach was based on the equivalence theorem provided by Zheng where E optimal designs are equivalent to $A/D/T$ optimal designs for the proportional model without baselines. When the value on the last column is greater than 1, it indicates that the use of baselines in the proportional model is beneficial. Table 3.1 shows that all 3 suggested the designs are highly efficient when the proportional carryover

effects assumption holds.

Table 3.1: Efficiencies of designs, d_{re} and d_{di}, against $A/D/E/T$ optimal designs when carryover effects are proportional to the direct treatment effects for $p = 2$ and $t = 3$.

Designs	$\lambda_{1o}, \lambda_{2o}$	A	D	E	T	$E_{\text{no baseline}}$
$d_{[1]}$	-0.5	1.0000	0.9877	0.9956	0.9478	3.2800
$d_{[2]}$	-0.5	0.9432	0.9813	0.8543	1.0000	3.2186
$d_{[1]}$	0.5	0.9953	0.9966	0.9956	1.0000	4.6196
$d_{[2]}$	0.5	0.9189	0.8551	0.8543	0.9442	3.9638
$d_{[3]}$	0.5	1	1	1	1	4.6633

[1] Nearly $A/D/E$ optimal design for the negative $(\lambda_{1o}, \lambda_{2o})$. $p_{<re>} = 1/4$ and $p_{<di>} = 3/4$.
[2] T optimal design for the negative $(\lambda_{1o}, \lambda_{2o})$. $p_{<di>} = 1$.
[3] E optimal design for the positive $(\lambda_{1o}, \lambda_{2o})$. $p_{<re>} = 1/t = 1/3$ and $p_{<di>} = (t-1)/t = 2/3$
$< re > = \{AA, BB, CC\}$ and $< di > = \{AB, AC, BA, BC, CA, CB\}$.

3.4.2 Three Period Designs

For $p > 2$, the $A/D/E/T$ optimal designs are constructed for two separate cases, $t \geq p$ and $t < p$. When the number of treatments is less than the number of periods, that is $t < p$, symmetric blocks of sequences that repeat the last two periods often form $A/D/E/T$ optimal designs for $-1 \leq \rho \leq 1$ and $(\lambda_{1o}, \lambda_{2o}) \in \{-0.5, 0.5\}$. One particular example is when $t = 2$ and $p = 3$, the optimal design is ABB/BAA with an equal allocation, which Laska, Meisner, and Kushner (1983) proved to be the universally optimal design under the traditional crossover model with random subject effects. Hence, we proved the usefulness of this design with respect to the proportional carryover effects assumption with baseline measurements as well.

For $t \geq p = 3$, the $A/D/E/T$ optimal designs change based on the values of $(\lambda_{1o}, \lambda_{2o})$. When the carryover effects are adversely related with the direct treatment effects, $(\lambda_{1o} = \lambda_{2o}) \in (-1, 0]$, $d_{<di>}$ is $A/D/E/T$ optimal for all ρ. On the other hand, $d_{<re>}$ is $A/D/T$ optimal for all ρ when $(\lambda_{1o}, \lambda_{2o}) \in (0, 1)$ and $d_{<di>}$ is E optimal.

Table 3.2: Efficiencies of d_{re} and d_{di} against $A/D/E/T$ optimal designs when carryover effects are negatively proportional to the direct treatment effects for $p = 3$, $t = 3$.

Designs	$\lambda_{1o}, \lambda_{2o}$	A	D	E	T	$E_{\text{no baseline}}$
d_{re}	-0.5	0.8660	0.6862	0.9874	0.9250	0.7924
$d_{di}^{[1]}$	-0.5	1	1	1	1	1.1172
$d_{re}^{[2]}$	0.5	1	1	0.9874	1	1.4009
$d_{di}^{[3]}$	0.5	0.9332	0.8544	1	0.9156	1.1968

[1] $A/D/E/T$ optimal design for the negative $(\lambda_{1o}, \lambda_{2o})$.
[2] $A/D/T$ optimal design for the positive $(\lambda_{1o}, \lambda_{2o})$.
[3] E optimal design for the positive $(\lambda_{1o}, \lambda_{2o})$.
$< re > = \{ABB, ACC, BAA, BCC, CAA, CBB\}$ and $< di > = \{ABC, ACB, BAC, BCA, CAB, CBA\}$.

3.4.3 Four Period Designs

For $t = 2$, the Design $d_{<AABB>}$ is $A/D/E/T$ optimal for all $(\lambda_{1o}, \lambda_{2o})$ and ρ. This symmetric block is also part of optimal design proposed by Laska, Meisner, and Kushner (1983) ($AABB$, $BBAA$, $ABAB$, $BABA$). As discussed earlier, we need to estimate more parameters under the proportional model for $t = 2$. Moreover, the $A/D/E/T$ optimal design under the proportional model assumption uses less favorable treatment sequences compared to the universally optimal design by Laska, Meisner, and Kushner (1983) under the traditional model. Therefore use of the proportional model approach may not be recommendable for this case.

For $t = 3$, the following optimal designs were found with negatively proportional carryover effects. Symmetric blocks such as $< ABCC >$ (which is $< re >$), $< ABCB >$, $< ABBC >$, $< ABCA >$, and $< AABC >$ form various optimal designs. $< ABCC >$ is E optimal, whereas $< AAAA >$, $< ABCC >$, $< ABCB >$, and $< ABCA >$ form A/D optimal designs. $< ABCB >$ and $< ABCA >$ form a T optimal design. Exact proportions of the optimal designs are given in the Table 3.3. We can see that $A/D/E/T$ optimal designs for the negatively proportional carryover effects are nearly equivalent except that the T optimal design has slightly lower efficiency with respect to the E optimality criterion. The Design $d_{<re>}$ performs well not only

Table 3.3: Efficiencies of d_{re} and d_{di} against $A/D/E/T$ optimal designs for $p = 4$ and $t = 3$.

Designs	$\lambda_{1o}, \lambda_{2o}$	A	D	E	T	$E_{\text{no baseline}}$
$d_{re}^{[1]}$	-0.5	0.9973	0.9540	1	0.9250	1.2343
$d_{A}^{[2]}$	-0.5	1	0.9645	0.9890	0.9328	1.2480
$d_{D}^{[3]}$	-0.5	0.9983	1	0.9210	0.9687	1.2939
$d_{T}^{[4]}$	-0.5	0.9516	0.9840	0.8065	1	1.2121
$d_{re}^{[1]}$	0.5	1	1	1	1	1.2343

[1] E optimal for the negative $(\lambda_{1o}, \lambda_{2o})$ and $A/D/E/T$ optimal for the positive $(\lambda_{1o}, \lambda_{2o})$.
[2] A optimal for the negative $(\lambda_{1o}, \lambda_{2o})$ with $p_{<AAAA>} = 0.02$, $p_{<ABCC>} = 0.86$, $p_{<ABCB>} = 0.07$, and $p_{<ABCA>} = 0.05$.
[3] D optimal for the negative $(\lambda_{1o}, \lambda_{2o})$ with $p_{<AAAA>} = 0.04$, $p_{<ABCC>} = 0.37$, $p_{<ABCB>} = 0.56$, and $p_{<ABCA>} = 0.03$.
[4] T optimal for the negative $(\lambda_{1o}, \lambda_{2o})$ with $p_{<ABCB>} = 0.5$ and $p_{<ABCA>} = 0.5$

for the negatively proportional carryover effects but it is also $A/D/E/T$ optimal for the positively proportional carryover effects. Therefore, use of $d_{<re>}$ is recommended for $p = 4$ and $t = 3$ crossover designs. The Design $d_{<ABCC>}$ is $A/D/E/T$ optimal when the carryover effects are positively proportional to the direct treatment effects as well as when there are no carryover effects, $\lambda_{1o} = \lambda_{2o} = 0$.

We further investigated $t = 4$ and found that the Design $d_{<di>}$ is $A/D/E/T$ optimal and the Design $d_{<re>}$ is highly efficient in the E optimality criterion for the negative proportionality parameters. When there are no carryover effects, the Design $d_{<di>}$ is $A/D/E/T$ optimal as expected. The Design $d_{<re>}$ is $A/D/T$ optimal and nearly E optimal, with efficiency ≈ 0.99 whereas the Design $d_{<di>}$ is E optimal for the positive proportionality parameters. In addition to these two designs we compare a Design d_{mix}, which consists of $p_{<re>} = p_{<di>} = 0.5$.

3.4.4 Five Period Designs

For $t = 2$, design with equal allocations on $< ABBAA >$ and $< AABBA >$ is the $A/D/E/T$ optimal design for all $\lambda_{1o} = \lambda_{2o} \in (-1, 1)$. However, the case for

Table 3.4: Efficiencies of d_{re} and d_{di} against $A/D/E/T$ optimal designs for $p = 4$ and $t = 4$.

Designs	$\lambda_{1o}, \lambda_{2o}$	A	D	E	T	$E_{\text{no baseline}}$
$d_{re}^{[1]}$	-0.5	0.8760	0.6250	0.9991	0.8372	0.7884
$d_{di}^{[2]}$	-0.5	1	1	1	1	1.2615
$d_{mix}^{[3]}$	-0.5	0.9544	0.8278	1	0.9259	1.0443
$d_{re}^{[1]}$	0.5	1	1	0.9991	1	1.8756
$d_{di}^{[2]}$	0.5	0.9176	0.7535	1	0.9031	1.4136
$d_{mix}^{[3]}$	0.5	0.9721	0.9012	0.9996	0.9603	1.6901

[1] $A/D/T$ optimal for the positive $(\lambda_{1o}, \lambda_{2o})$ where $p_{<ABCC>} = 1$
[2] $A/D/E/T$ optimal for the negative and E optimal for the positive $(\lambda_{1o}, \lambda_{2o})$ where $p_{<ABCD>} = 1$
[3] $p_{<ABCC>} = p_{<ABCD>}$=0.5

$t = 3$ involves more symmetric blocks. $< ACABB >$ and $< AABCB >$ form $A/D/E$ optimal designs and $< ABCBC >$, $< ABCAC >$, $< ABCAB >$, and $< ACABC >$ form a T optimal designs for the negative proportional carryover effects. On the other hand, $< ABBCC >$ and $< AABCC >$ form $A/D/T$ optimal designs whereas $< ACABB >$ and $< AABCB >$ form an E optimal design when the carryover effects are positively proportional. $< ABACC >$ and $< AABCB >$ form $A/D/E/T$ optimal designs when there are no carryover effects.

When $t = 4 < p = 5$, $< ABCDC >$, $< ABCDB >$, and $< ABCDA >$ form $A/D/T$ optimal designs, and $< ABCDD >$ is an E optimal design for the negative proportional carryover effects. Also, $d_{<ABCDD>}$ is $A/D/E/T$ optimal design for no carryover effects or positively proportional carryover effects.

For $t = 5$, the Design $d_{<di>}$ is $A/D/E/T$ optimal for the negatively proportional carryover effects or no carryover effects. It is E optimal for the positively proportional carryover effects and $A/D/T$ optimal design as well for weak positive proportional carryover effects ($\lambda_{1o} = \lambda_{2o} \leq 0.12$). However, the Design $d_{<re>}$, is A/D optimal and $< ABBCC >$ and $< AABCC >$ form a T optimal design for $\lambda_{1o} = \lambda_{2o} > 0.12$.

Table 3.5: Efficiencies of d_{re} and d_{di} against $A/D/E/T$ optimal designs for $p = 5$ and $t = 4$.

Designs	$\lambda_{1o}, \lambda_{2o}$	A	D	E	T	$E_{\text{no baseline}}$
$d_{re}^{[1]}$	-0.5	0.9462	0.7725	1	0.8938	1.4450
$d_{ADT}^{[2]}$	-0.5	1	1	0.9093	0.8372	1.8708
$d_{re}^{[1]}$	0.5	1	1	1	1	1.4450
$d_{ADT}^{[2]}$	0.5	0.8462	0.5919	0.9093	0.8337	0.8552

[1] E optimal for the negative and $A/D/E/T$ optimal for the positive $(\lambda_{1o}, \lambda_{2o})$
where $p_{<ABCDD>}$=1
[2] $A/D/T$ optimal for the negative $(\lambda_{1o}, \lambda_{2o})$ where $p_{<ABCDC>} = p_{<ABCDB>} = p_{<ABCDA>} = 1/3$

3.5 Conclusion

In the last section, we constructed optimal designs for the proportional model with baselines and observed the following. The $A/D/E/T$ optimal designs take various forms depending on t and p as well as $(\lambda_{1o}, \lambda_{2o}) \in (-1, 0]$ or $(0, 1)$. For two-period crossover designs, the universally optimal design under the traditional model proposed by Carriere and Reinsel (1993) is still $A/D/E/T$ optimal for the positive proportionality parameters. In this case, the use of baseline measurements significantly improved efficiency under the proportional model. When $2 \leq t < p$ and $(\lambda_{1o}, \lambda_{2o}) \in (-1, 0]$, symmetric blocks without identical treatments in two consecutive periods are optimal, i.e. $< ABCAB >$ for $p = 5$ and $t = 3$. On the other hand, designs that assign $p - 1$ distinct treatments and then repeating on the last period is optimal for $2 < p \leq t$ and $(\lambda_{1o}, \lambda_{2o}) \in (0, 1)$. Especially when $t = p - 1$, the optimal designs for the positive proportionality parameters is the Design $d_{<re>}$. For example, $< ABCDD >$ is the optimal design for $t = 4$ and $p = 5$. With $t < p$, the Design $d_{<re>}$ is often the $A/D/E/T$ optimal design for the positive proportionality parameters whereas the Design $d_{<di>}$ is $A/D/E/T$ optimal for the negative proportionality parameters. With $t \geq p$, the Design $d_{<di>}$ is $A/D/E/T$ optimal or nearly $A/D/E/T$ optimal for any proportionality parameters $\in (-1, 1)$. Even when these two designs are not necessarily

Table 3.6: Efficiencies of existing optimal designs and other practical designs against $A/D/E/T$ optimal designs and E optimal design under no baseline measurements for $p = 5$ and $t = 5$.

Designs	$\lambda_{1o}, \lambda_{2o}$	A	D	E	T	$E_{\text{no baseline}}$
$d_{re}^{[1]}$	-0.5	0.8921	0.5964	0.9998	0.8682	0.8771
$d_{di}^{[2]}$	-0.5	1	1	0.9998	1	1.4709
$d_{re}^{[1]}$	0.5	1	1	0.9998	0.9944	2.4278
$d_{di}^{[2]}$	0.5	0.9204	0.6987	1	0.9038	1.6960
$d_{T}^{[3]}$	0.5	0.9783	0.9739	0.8715	1	2.3639

[1] $A/D/E/T$ optimal for the positive $(\lambda_{1o}, \lambda_{2o})$ where $p_{<ABCDD>}$=1.
[2] $A/D/E/T$ optimal for the negative and E optimal for the positive $(\lambda_{1o}, \lambda_{2o})$
where $p_{<ABCDE>} = 1$.
[3] T optimal for the positive $(\lambda_{1o}, \lambda_{2o})$ where $p_{<ABBCC>} = p_{<AABCC>} = 1/2$.

$A/D/E/T$ optimal designs, they maintain a high level of efficiencies compared to the $A/D/E/T$ optimal designs for various p, t, and $(\lambda_{1o}, \lambda_{2o})$.

Some interesting phenomena were observed when the new optimal designs under the proportional models in the presence of baseline measurements are compared to the optimal designs without baseline measurements. For any p, the efficiency of the optimal designs with baselines against the optimal designs without baselines increases as t increases. When t is fixed, the efficiencies decreases as p increases. When $t = p = 3$ and proportionality parameters are positive, the $A/D/T$ optimal Design $d_{<ABB>}$ is 1.4 times efficient than the E optimal design without baseline measurements when compared with respect to D optimality criterion. For $t = p = 6$, $d_{<re>}$ design with baselines is 3 times more efficient than the optimal design without baselines when proportionality parameters are positive. For $t \geq p$ and $(\lambda_{1o}, \lambda_{2o}) \geq 0$, the Design $d_{<di>}$ is nearly A/T optimal and often E optimal. The efficiency of the Design $d_{<di>}$ against the optimal design without baselines is still relatively high. Given that the Design $d_{<di>}$ is preferred to the Design $d_{<re>}$ or any other designs by the experimenters, this result suggests that researchers may incorporate baseline measurements when they have information that the carryover effects are

proportional or negligible. The case for no carryover effects were included as the information about carryover effects are, often, not known in the planning stage of the trials thus it is worth to verify whether the Design $< di >$ (the optimal design for traditional model with no carryover effects) is still optimal design under the proportional model with baseline measurements. This is a new discovery from the earlier guidelines where $d_{<di>}$ was recommended only when carryover effects are negligible. The information about the proportionality of carryover effects may be known before the trials or in the course of trials, which would allow some option for response adaptive scheme. Upon successfully detecting the proportionality parameters, researchers may utilize that information by allocating more subjects to a particular block of treatment sequences that may be more efficient under the proportional model.

Chapter 4

Multiple Objective Response Adaptive Crossover Designs for Binary Responses Using the Generalized Estimating Equations

Abstract

In clinical trials, adaptive allocation schemes are considered to utilize the information acquired during the trials to fulfill various objectives. A multiple objective response adaptive scheme for binary responses has been proposed previously. We introduce new approaches incorporating multiple objectives, while advocating a regression type estimation approach via the Generalized Estimating Equations method. We then compare the new approaches to identify appropriate weight parameters and to propose a new adaptive allocation scheme that maximizes the benefits from the superior treatments while maintaining a sufficiently high level of design efficiency. We find that the multiple objective criterion successfully constructs spectrum of designs ranging over various efficiencies and success ratios. Moreover, the new adaptive allocation scheme successfully construct designs that fall right in the recommended range of the weight parameter.

4.1 Introduction

Crossover designs have advantages over parallel designs, such as completely randomized design in terms of statistical efficiencies. Equal or balanced allocations play an important role in the construction of optimal designs under various model assumptions. However, equal allocations may pose ethical dilemma when researchers start to suspect that one treatment may be superior to the other. All trials start with the null hypothesis that the effects of

the treatments being tested are the same. At some point in the trial, we may find an evidence indicating that the effects of treatments are notably different. Then, we may wonder whether to equally allocate remaining subjects to the treatments as per the protocol or to adapt to the findings and alter the allocation scheme to reflect the trial phenomena. Connor et al. (1994) studied a mother to infant HIV transmission drug named AZT. Among 477 pregnant mothers with HIV, 239 were assigned to a placebo and 238 were assigned to the AZT. The trial resulted in 60 infants diagnosed with HIV from the placebo group and 20 infants diagnosed with HIV from the AZT group. A decade later, Tymofyeyev, Rosenberger, and Hu (2007) suggested that use of 50-50 allocation was ethically improper given the seriousness of the outcome of the study and he recommended to use response-adaptive allocation. Tymofyeyev, Rosenberger, and Hu (2007) utilized the Play the Winner Rule(PWR) allocation (Zelen (2003) and Wei and Durham (1978)) and showed that 360 and 117 pregnant mothers are adaptively allocated to the AZT and placebo, respectively. The results of Hu's simulation showed that 60 infants were expected to be diagnosed with HIV in two groups combined, which revealed some of the benefits of the adaptive allocations.

Response adaptive designs may have several other goals. Zelen (2003), Wei and Durham (1978), Bandyopadhyay, Biswas, and Mukherjee (2009a, 2009b) aimed at allocating more subjects to a better treatment. Armitage (1975) aimed at reducing the sample size and Wang (2014) aimed at increasing the sample size based on the pre-specified statistical power and the data acquired. Furthermore, Bandyopadhyay and Biswas (2001) investigated the method of introducing covariates in response adaptive designs. Sorkness et al. (2019) proposed designs that were adaptive to the prevalence of events, in which sample size re-calculation was done to remedy the loss of statistical power arising from the imbalance in the prevalence. However, these studies utilized

the acquired information for a single objective. Liang and Carriere (2009), Liang et al. (2014), and Li (2017) proposed a multiple objective response adaptive design where they defined an objective function with two components, controlled by a weight parameter. On the other hand, Yi and Wang (2009) and Li and Wang (2012) proposed a response adaptive randomization based on a variance-penalized mean criterion, which takes account of both the mean and variance of the total number of successes. This chapter incorporates these two multiple objective functions for two-treatment designs with two or three periods and binary responses.

The binary responses are modelled differently from continuous responses in a way that the information matrix is a function of the responses, and the standard logistic regression assumes that the responses are independent. Therefore, the Generalized Estimating Equations (GEE) method, which can incorporate a desired covariance structure of responses, is used for the analysis of crossover designs in this chapter. Liang and Zeger (1986) proposed the GEE, which takes into account for the time dependencies of the data by allowing different working correlation matrices. The GEE method estimates parameters by solving the system of equations based on the Quasi-Likelihood function. The advantage of Quasi-Likelihood method is that it does not need to provide joint distribution of the data. It only requires the marginal distribution and its mean and variance. GEE estimates are consistent under a mild regularity conditions(Liang and Zeger (1986)). Valois (1997) utilized GEE in the analysis of crossover designs.

This chapter aims to construct multiple objective response adaptive designs for two treatments with binary outcomes using the GEE. Adaptive designs are first constructed using simulations and some examples will be provided for various weights of multiple objective functions. We first review the theoretical grounds for crossover designs with binary outcome and the GEE method.

Then, we compare the two suggested response adaptive allocations for the construction of adaptive two-treatment two/three-period crossover designs. The two adaptive allocation schemes will be compared in terms of the success ratios and relative efficiencies. We also compare the GEE methods to the previous approaches done by Li (2017). Lastly, we develop a new strategy for maximizing the success ratio while maintaining certain level of statistical efficiency, which does not involve selecting the weight parameter for the two objectives.

4.2 Multiple objective response adaptive designs with GEE

4.2.1 Model and Information Matrix

Agresti (2014) studied the Generalized Linear Model (GLM) for an exponential family of distributions. Suppose Y follows a distribution in an exponential family with parameters (β, ϕ). Then the pdf of Y can be written as,

$$f(y|\beta,\phi) = exp((y\beta - b(\beta))/a(\phi) + c(y,\phi)). \tag{4.1}$$

Taking a log on the likelihood function we get,

$$l(\beta,\phi|y) = logf(y;\beta,\phi) = \frac{y\beta - b(\beta)}{a(\phi)} + c(y,\phi), \tag{4.2}$$

$$\begin{aligned}
&E\left(\frac{\partial l}{\partial \beta}\right) = E\left(\frac{y - b'(\beta)}{a(\phi)}\right) = 0,\\
&E\left(\frac{\partial^2 l}{\partial \beta^2}\right) - E\left(\frac{\partial l}{\partial \beta}\right)^2 = \frac{b''(\beta)}{a(\phi)} - \frac{Var(Y)}{a^2(\phi)} = 0,\\
&\Rightarrow E(Y) = b'(\beta) \text{ and } Var(Y) = b''(\beta)a(\phi).
\end{aligned}$$

Suppose that Y_{ijk} denotes the binary response of ith period of jth subject in kth treatment sequence and $Y_{ijk} \sim$ Bernoulli(p_{ijk}). Now suppose X is a design matrix for an overall mean effect, period effects, direct treatment effects, and

carryover effects with the corresponding vector of parameters θ. With a logit link function $g()$, we can define the following model,

$$\eta_{ijk} = g(E(Y_{ijk})) = g(P(Y_{ijk} = 1) = logit(P(Y_{ijk} = 1)) \tag{4.3}$$

$$= log\left(\frac{P(Y_{ijk} = 1)}{1 - P(Y_{ijk} = 1)}\right) = \mu + \alpha_i + \tau_{d(i,j,k)} + \gamma_{d(i-1,j,k)} = X'\beta. \tag{4.4}$$

It is easy to see that the mean and variance of Y_{ijk} are defined as

$$E(Y_{ijk}) = \mu_{ijk} = b'(\beta_{ijk}) = \frac{exp(X'_{ijk}\theta)}{1 - exp(X'_{ijk}\theta)}, \tag{4.5}$$

$$Var(Y_{ijk}) = b''(\beta_{ijk}) = \frac{exp(X'_{ijk}\theta)}{(1 - exp(X'_{ijk}\theta))^2}. \tag{4.6}$$

Now define $\partial\mu_j/\partial\beta_j = B_j$ and the quasi likelihood function as in McCullagh and Nelder (1999) as the following,

$$Q(\mu_j|y_j) = \int_{y_j}^{\mu_j} \frac{y_j - t}{var(Y_j)}\partial t = \int_{y_j}^{\mu_j} \frac{y_j - t}{V(\mu_j)}\partial t. \tag{4.7}$$

We can derive the Information matrix by taking the derivative of $Q(\mu_j, y_j)$ with respect to θ and θ',

$$\begin{aligned}
I_j &= -E\left(\frac{\partial^2 Q(\mu_j|y_j)}{\partial\theta\partial\theta'}\right) \\
&= -E\left[\frac{\partial}{\partial\theta'}\left(\left(\frac{\partial\mu_j}{\partial\theta}\right)' V^{-1}(\mu_j)(y_j - \mu_j)\right)\right] \\
&= -E\left[\frac{\partial}{\partial\theta'}\left(\left(\frac{\partial\mu_j}{\partial\beta_j}\frac{\partial\beta_j}{\partial\theta}\right)' V^{-1}(\mu_j)(y_j - \mu_j)\right)\right] \\
&= -E\left[\frac{\partial}{\partial\theta'}\left(X'_j B_j V^{-1}(\mu_j)(y_j - \mu_j)\right)\right] \\
&= X'_j B_j V^{-1}(\mu_j) B_j X_j.
\end{aligned}$$

Then, Bose and Dey (2009) showed that the covariance matrix for parameters θ is defined as follows

$$Var(\hat{\theta}) = \left(\Sigma_{k\in\Omega} n_k I_k\right)^{-1}, \tag{4.8}$$

with the design matrices being identical for subjects in the same treatment sequence when a covariance matrix is correctly specified. Otherwise, a sandwich variance estimator is suggested.

$$Var(\hat{\theta}) = A\left(\Sigma_{k\in\Omega} n_k X_k' B_k V^{-1}(\mu_k) Cov(Y_k) V^{-1}(\mu_k) B_k X_k\right) A, \tag{4.9}$$

where $A = (\Sigma_{k\in\Omega} n_k X_k' B_k V^{-1}(\mu_k) B_k X_k)^{-1}$. This sandwich variance estimator is known to be consistent(Liang and Zeger (1986)).

4.2.2 Multiple Objective Function

Liang and Carriere (2009) proposed the following multiple objective function for the continuous responses,

$$\Phi_{j,k} = \lambda \frac{\Delta(\hat{I}^k_{j+1}(\theta))}{\Delta(\hat{I}^{k'}_{j+1}(\theta))} + (1-\lambda)\frac{f_{j,k}}{f_{j,k''}}, \tag{4.10}$$

where $\hat{I}^k_{j+1}(\theta)$ is the Fisher's Information matrix for subject $j+1$ allocated to treatment sequence k with Δ being an optimality criterion of choice and $f_{j,k}$ is an evaluation function for treatment sequence k based on the first j subjects in the trial. In this function, treatment sequence k' refers to the sequence with maximum $\hat{I}^k_{j+1}(\theta)$ and k'' refers to the sequence with maximum $f_{j,k}$, which may not necessarily be identical. Among the two terms in the objective function, the first term of the function investigates the efficiency of design with respect to the Fisher's Information matrix given that subject $(j+1)$ is allocated to treatment sequence k. This is represented as a ratio over the sequence with maximum information so that the component may take value in $[0, 1]$. The second term of the function is called the evaluation function that evaluates the total efficacy of treatment sequences based on the estimated treatment effects. When $\lambda = 0$, the objective function considers only the efficiency of the design and ignores any superiority/inferiority of the treatments being tested. On

the other hand, the objective function with $\lambda = 1$ would construct adaptive designs based solely on the positive effects of treatments being tested.

Liang et al. (2014) and Li (2017) extended their multiple objective function to binary responses and derived the Information Matrix for estimated success probabilities for binary responses. The observed number of successes for each treatment sequence was used for the evaluation function f. The sandwich estimate of the covariance matrix is known to be robust, and consistent for heteroscedasticity and mis-specification of covariance matrices for within-subject responses. As the analysis of crossover design mainly focuses on direct treatment effects, we choose the inverse of the variance of estimated treatment effects, $1/var(\hat{\tau})$, as the criterion for comparing the efficiency of various treatment sequences. McCullagh (2005) showed that Quasi-likelihood estimates are invariant under a linear transformation. That is, $\hat{\mu_k}$ maximizes the quasi-likelihood function.

Throughout this chapter, we will refer to the equation (4.10) as the multiple objective function and choose the first term $\Delta(\hat{I}^k_{j+1}(\theta))$ as the variance of the estimated treatment effects, $var(\hat{\tau}_{j+1,k})$. The data acquired from the first j subjects are modelled using the GEE approach and predictions for subject $j+1$ are made for all of the K treatment sequences. Then, we include the predicted responses of subject $j+1$ into the model and obtain the variance of an estimated treatment effects of each treatment sequence. Then, we evaluate the efficacy of each treatment sequence by using $\Sigma^p_{i=1}\hat{\eta}_{i,j,k}$. The $\eta's$ take any values in $\mathbb{R}$ where large values correspond to a better treatment sequence. We transform these values to positive numbers so that a larger value indicates a better sequence and the ratios could be easily implemented. For this reason, we choose $f_{j,k} = logit(\Sigma^p_{i=1}\hat{\eta}_{i,j,k})$, which falls in $(0, 1)$.

Yi and Wang (2009) proposed a variance-penalized mean criterion for response adaptive designs for parallel design where the goal was to maximize the

function $E(\Sigma_{j=1}^{n} Y_j) - \psi Var(\Sigma_{j=1}^{n} Y_j)$ for $\psi \geq 0$. When $\psi = 0$, this is equivalent to (4.10) with $\lambda = 0$. We can see that increasing ψ moves the focus of the maximization on reducing the variance. We adopt this idea of penalizing treatment effects by its variance and define a new variance penalized criterion Φ' as follows,

$$\Phi'(j,k) = \Sigma_{i=1}^{p} \hat{\eta}_{i,j,k} - \psi Var(\Sigma_{i=1}^{p} \hat{\eta}_{i,j,k}), \tag{4.11}$$

where the total effect of each treatment sequence is penalized by its own variance with a weight parameter $\psi \geq 0$.

4.3 Two Allocation Methods

We compare the performance of the two allocation functions, namely multiple objective function in (4.10) and variance penalized mean function in (4.11). For the comparison, we construct two-treatment two-period designs and two-treatment three-period designs based on the parameter settings from Li (2017), which are shown in Table 4.1 with a slight modification on the values to incorporate the GEE modeling approach. Initially, four subjects are assigned to each treatment sequence. Afterwards, new subjects are introduced sequentially and are assigned to the treatment sequence with the highest (4.10) or (4.11). When all subjects are assigned, the variance of the estimated treatment effects, $var(\hat{\tau}_N)$, is computed and compared to the variance obtained from the optimal fixed designs suggested by Mukhopadhyay, Singh, and Dey (2015). Mukhopadhyay, Singh, and Dey (2015) conducted simulation study for the optimal fixed crossover design with binary outcomes using the GEE method and showed that $AA/AB/BB/BA$ is optimal for p=2 and $ABB/AAB/BAA/BBA$ is optimal for p=3 under the compound symmetric covariance structure with binary outcomes.

Table 4.1: Parameter values for simulation in construction of multiple-objective response adaptive crossover design with binary outcomes.

p	Parameters	Treatment Sequences	Success Probabilities	Expected Success per Period
2	μ=-0.22	AA	$(0.60, 0.70)$	0.65
	α_2=0.018	AB	$(0.60, 0.40)$	0.50
	τ=0.63	BA	$(0.30, 0.50)$	0.40
	γ=0.42	BB	$(0.30, 0.22)$	0.26
3	μ=-0.22	AAA	$(0.60, 0.70, 0.65)$	0.65
	α_2=0.018	AAB	$(0.60, 0.70, 0.35)$	0.55
	α_3=-0.21	ABA	$(0.60, 0.40, 0.44)$	0.48
	τ=0.63	ABB	$(0.60, 0.40, 0.19)$	0.40
	γ=0.4	BAA	$(0.30, 0.50, 0.65)$	0.48
		BAB	$(0.30, 0.50, 0.35)$	0.38
		BBA	$(0.30, 0.22, 0.44)$	0.32
		BBB	$(0.30, 0.22, 0.19)$	0.23

4.3.1 Two Period Design

There are 4 treatment sequences for two-treatment two-period crossover trials. Carriere and Reinsel (1993) showed that an equal allocation on all sequences $AA/BB/AB/BA$, denoted as $d_{opt,p2}$, is universally optimal for a continuous response and Mukhopadhyay, Singh, and Dey (2015) confirmed that the same design is numerically optimal even when responses are binary. We assign 4 subjects to each of the four sequences and allocate the rest based on the two objective functions in (4.10) and (4.11). The following tables show the allocations of the adaptive designs, their efficiency compared with the optimal design, and their success ratio for different values of λ and ψ and N.

When $\lambda = 0$ or $\psi = 0$, the resulting allocation focuses on the treatment sequence AA with very few assigned to the rest of the sequences due to randomness during the initial stage of the trial. We can see that the allocation to the sequence AA decreases as λ or ψ increases. However, the multiple objective response adaptive designs behave differently from the penalized mean response adaptive designs. For the multiple objective response adaptive designs, the

allocations move toward a dual balanced design $d_{opt,p2}$, which assigns equal allocations to all 4 sequences. The relative efficiency of the proposed multiple objective adaptive design against $d_{opt,p2}$ is low for $\lambda = 0$ and approaches 1 as λ increases to 1. The success ratio is close to the expected success shown in Table 4.1 when $\lambda = 0$ and decreases as λ increases. Therefore, we must find a reasonable compromise between efficiency and a success ratio. For $N = 40$, $\lambda \in (0.85, 0.9)$ would construct an efficient design (efficiency> 0.8) with a sufficiently higher success ratio ($5\% \sim 8\%$ increased) than $\lambda = 1$. For $N = 80$, $\lambda \in (0.9, 0.95)$ would construct a similar design (efficiency > 0.8 and success ratio improved by $5\% \sim 8\%$). For $N = 100$, something happens drastically around $\lambda \in (0.9, 0.95)$ that efficiency drops from 0.8957 to 0.7096 while the success ratio increases from 0.5168 to 0.5638 showing that the choice of suitable λ may vary significantly by the sample size N.

The adaptive designs in Tables 4.2 and 4.3 behave similarly when $\lambda = 0$ and $\psi = 0$, as expected. We can observe that the allocations to treatment sequences AB/BA increase, and the allocations to treatment sequence AA/BB decrease as ψ increases. This indicates that the sequences with different treatments, AB/BA, have a smaller variance of the estimated sum of η's. As ψ increases further, the allocation to the sequence AA decreases significantly and most of the allocations are assigned to the two sequences AB/BA. However, the efficiency of the resulting adaptive design does not improve, and the success ratios drop significantly as ψ increases. These losses in success ratios did not accompany any gains in efficiencies.

For $p = 2$ and $t = 2$, the expected value of the first term in the equation (4.11) is defined as $\eta_{1,AA(BB)} + \eta_{2,AA(BB)} = 2\mu + \alpha_2 + 2\tau(-2\tau) + \gamma(-\gamma)$ for

Table 4.2: Allocation, Efficiency, and Success Ratio for two-period designs using the multiple objective criteria in (4.10).

N	λ	**AA**	**AB**	**BA**	**BB**	**Efficiency**	**Success Ratio**
40	0	26.97	4.37	4.51	4.15	0.5679	0.5635
	0.3	26.46	4.40	4.94	4.20	0.5696	0.5596
	0.7	25.42	5.46	4.89	4.23	0.6378	0.5576
	0.8	22.22	7.60	5.63	4.55	0.7615	0.5420
	0.9	16.63	9.49	7.28	6.60	0.9152	0.5042
	1	10.20	10.01	9.66	10.13	1.0141	0.4534
80	0	65.81	4.53	5.46	4.20	0.2998	0.6046
	0.3	66.05	4.65	5.13	4.17	0.3012	0.6055
	0.7	64.37	5.95	5.43	4.26	0.3554	0.6020
	0.8	59.16	9.95	6.28	4.60	0.4844	0.5896
	0.9	45.08	16.64	10.56	7.72	0.7582	0.5507
	0.95	33.68	18.98	13.95	13.39	0.9368	0.5048
	1	20.35	19.81	18.96	20.88	1.0076	0.4532
100	0	85.14	4.71	5.85	4.31	0.2859	0.6126
	0.3	85.84	4.56	5.45	4.15	0.2773	0.6136
	0.7	84.57	6.03	5.25	4.16	0.3184	0.6114
	0.9	61.75	19.36	11.01	7.89	0.7096	0.5638
	0.95	45.77	23.49	16.30	14.44	0.8957	0.5168
	1	25.41	24.40	23.66	26.54	1.0258	0.4525

sequences $AA(BB)$ and $2\mu + \alpha_2 + \gamma(-\gamma)$ for treatment sequences $AB(BA)$. Consider the followings,

$$V_1 = Var(2\hat{\mu} + \hat{\alpha_2} + 2\hat{\tau} + \hat{\gamma}) > Var(2\hat{\mu} + \hat{\alpha_2} + \hat{\gamma}) = V_3,$$

$$\Rightarrow 4Var(\hat{\tau}) + 4Cov(\hat{\mu}, \hat{\tau}) + 2Cov(\hat{\alpha_2}, \hat{\tau}) + 2Cov(\hat{\tau}, \hat{\gamma}) > 0.$$

$$V_2 = Var(2\hat{\mu} + \hat{\alpha_2} - 2\hat{\tau} - \hat{\gamma}) > Var(2\hat{\mu} + \hat{\alpha_2} - \hat{\gamma}) = V_4,$$

$$\Rightarrow 4Var(\hat{\tau}) - 4Cov(\hat{\mu}, \hat{\tau}) - 2Cov(\hat{\alpha_2}, \hat{\tau}) + 2Cov(\hat{\tau}, \hat{\gamma}) > 0.$$

$$V_1 = Var(2\hat{\mu} + \hat{\alpha_2} + 2\hat{\tau} + \hat{\gamma}) > Var(2\hat{\mu} + \hat{\alpha_2} - \hat{\gamma}) = V_4,$$

$$\Rightarrow 4Var(\hat{\tau}) + 4Cov(\hat{\mu}, \hat{\tau}) + 2Cov(\hat{\alpha_2}, \hat{\tau}) + 2Cov(\hat{\tau}, \hat{\gamma}) + 4Cov(\hat{\mu}, \hat{\gamma})$$

$$+ 2Cov(\hat{\alpha_2}, \hat{\gamma}) > 0.$$

$$V_2 = Var(2\hat{\mu} + \hat{\alpha_2} - 2\hat{\tau} - \hat{\gamma}) > Var(2\hat{\mu} + \hat{\alpha_2} + \hat{\gamma}) = V_3,$$

$$\Rightarrow 4Var(\hat{\tau}) - 4Cov(\hat{\mu}, \hat{\tau}) - 2Cov(\hat{\alpha_2}, \hat{\tau}) + 2Cov(\hat{\tau}, \hat{\gamma}) - 4Cov(\hat{\mu}, \hat{\gamma})$$

$$- 2Cov(\hat{\alpha_2}, \hat{\gamma}) > 0.$$

The consistent estimates for the above terms can be obtained by replacing the parameters with their GEE estimates. Also, the variance of the estimated

Table 4.3: Allocation, Efficiency, and Success Ratio for two-period designs using the penalized mean criteria in (4.11).

N	ψ	**AA**	**AB**	**BA**	**BB**	**Efficiency**	**Success Ratio**
40	0	26.66	4.45	4.72	4.17	0.5587	0.5637
	1	24.40	5.48	6.02	4.10	0.5719	0.5513
	4	10.24	15.19	10.52	4.05	0.5431	0.4853
	200	4.71	17.31	13.77	4.21	0.5355	0.4593
80	0	65.65	4.42	5.74	4.19	0.2999	0.6042
	1	59.26	7.30	9.29	4.15	0.3002	0.5877
	4	23.15	32.51	20.27	4.07	0.2961	0.5057
	200	5.70	38.62	31.10	4.58	0.2893	0.4572
100	0	85.89	4.67	5.29	4.15	0.2790	0.6139
	1	78.36	7.08	10.29	4.27	0.2908	0.5988
	4	30.12	41.29	24.35	4.24	0.2962	0.5106
	200	6.44	48.59	40.12	4.85	0.2658	0.4570

η's can easily be computed using the sandwich covariance matrix from GEE. The treatment sequences with a smaller variance do not necessarily improve efficiency in this case and the efficiency depends on the covariance matrix of the estimates of parameters. This covariance matrix, in turn, does not have a closed form as in the continuous response case. The simulation results indicate that the penalized mean criterion function behaves in a lot more complex manner than the multiple objective function in (4.10). Thus, it seems that the variance penalized criterion may not be suitable for crossover design.

4.3.2 Three Period Design

Three-period two-treatment crossover designs constructed from the multiple objective response adaptive approach behave similarly as the two-period two-treatment designs. When $\lambda = 0$, the majority of the subjects are allocated to the treatment sequence AAA, which has the highest success ratio per period. For small sample size, $n = 40$, the efficiencies remain high and the success ratios are improved (-5% efficiency and 5% success ratio) for any values of $\lambda < 1$. This is largely due to the conditions of the design, where $3\times8=24$ subjects out of 40 are assigned evenly to all 8 sequences and thus only 16 subjects are allocated based on the multiple objective response adaptive schemes.

Therefore, the relative efficiency, which is computed based on the optimal design suggested by Mukhopadhyay, Singh, and Dey (2015), remains high and the success ratio is improved only to a degree. However, in the case of $N = 80$, the success ratio increases from 0.4323 to 0.5647 and the efficiency decreases from 1.0370 to 0.5793 as λ decreases from 1 to 0. It is notable that the relative efficiencies of multiple objective response adaptive designs for $\lambda = 1$ and $n = \{40, 80, 100\}$ are greater than 1, indicating that these designs are slightly better than the optimal design suggested by Mukhopadhyay, Singh, and Dey (2015) for the given set of parameters. For $\lambda = 1$ and $N = 80$, the expected success per period is 0.4375, which is close to the observed success ratios of 0.4323. The design with $\lambda = 0.95$ is as efficient as the optimal design, relative efficiency of 1.0055, and yet shows a higher success ratio, $0.4708 > 0.4323$, with an expected success ratio of $0.4696 > 0.4375$. In the case of $\lambda = 0.9$, the relative efficiency decreases to 0.9220 while the success ratio increases to 0.5050 from 0.4323. Looking at the design with $\lambda = 0.85$, we see that the relative efficiency decreases to 0.8133 while the success ratio increases to 0.5290. These two designs with $\lambda = 0.9$ and $\lambda = 0.85$ indicate that we could improve the success ratio of the design by $7\% \sim 10\%$ at the cost of relative efficiency between 0.1 and 0.2. When $N = 100$, the designs show a similar performance to the case of $N = 80$ with respect to efficiency and the success ratio except that efficiencies drop faster as λ decreases. We can observe a rapid decrease in the efficiency for $\lambda \in (0.85 \sim 0.9)$.

In summary, the above tables show that adaptive schemes could benefit more subjects without much loss of efficiency for the given set of parameters. But it is important to find an appropriate λ to improve the success ratios while maintaining a sufficient level of statistical efficiency. In this case, $\lambda \in (0.85, 0.9)$ is recommended for both $N = 80$ and $N = 100$. However, we can see that the decrease in efficiency is more rapid for $N = 100$ than that of $N = 80$,

indicating that sample size N is another player determining the recommended λ. The resulting designs would have success ratios increased by $9\% \sim 12\%$ when compared to the optimal fixed design($\lambda = 1$). Taking a smaller value of λ can benefit further but the gain in success ratio decreases marginally as the λ decreases. This shows that the choice of λ depends not only on the parameters but also on the sample size.

Table 4.4: Allocation, Efficiency, and Success Ratio for three-period design using the multiple objective criteria in (4.10).

N	λ	**AAA**	**AAB**	**ABA**	**ABB**	**BAA**	**BAB**	**BBA**	**BBB**	**Efficiency**	**Success Ratio**
40	0	11.98	4.00	4.01	4.00	4.00	4.00	4.01	4.00	0.9603	0.4797
	0.3	11.95	4.02	4.00	4.00	4.01	4.00	4.02	4.00	0.9631	0.4785
	0.7	10.97	4.77	4.12	4.12	4.11	4.01	4.00	4.00	0.9891	0.4767
	0.8	9.24	5.62	4.73	4.56	5.01	4.55	4.36	4.23	0.9931	0.4678
	0.9	6.94	5.62	4.73	4.56	5.01	4.55	4.36	4.23	1.0075	0.4566
	1	5.04	5.03	4.73	4.96	4.98	4.93	4.97	5.36	1.0302	0.4377
80	0	51.98	4.01	4.00	4.00	4.00	4.00	4.01	4.00	0.5793	0.5647
	0.3	51.95	4.00	4.01	4.00	4.01	4.00	4.00	4.00	0.5848	0.5656
	0.7	49.58	5.91	4.24	4.01	4.23	4.01	4.02	4.00	0.6043	0.5619
	0.8	41.09	10.41	6.07	4.24	5.89	4.24	4.06	4.00	0.7223	0.5458
	0.85	33.85	12.40	7.48	5.14	7.47	5.20	4.39	4.07	0.8133	0.5290
	0.9	25.20	13.16	8.67	6.74	9.10	6.94	5.63	4.56	0.9220	0.5050
	0.95	16.76	12.06	9.02	8.48	10.27	8.61	7.94	6.86	1.0055	0.4708
	1	10.09	9.97	8.58	9.95	9.83	9.75	9.91	11.92	1.0370	0.4323
100	0	71.97	4.01	4.00	4.00	4.00	4.00	4.02	4.00	0.4972	0.5817
	0.3	71.98	4.01	4.00	4.00	4.01	4.00	4.00	4.01	0.5081	0.5805
	0.7	69.41	6.03	4.28	4.02	4.25	4.01	4.00	4.00	0.5379	0.5812
	0.8	57.12	6.09	5.56	6.20	6.01	5.95	6.18	6.90	0.6345	0.5445
	0.85	54.13	12.65	7.646	4.94	7.32	5.00	4.28	4.05	0.6951	0.5553
	0.9	37.37	16.65	10.20	7.39	10.46	7.62	5.80	4.50	0.8696	0.5231
	0.95	23.93	15.91	10.85	10.17	12.35	10.44	9.17	7.19	0.9858	0.4794
	1	12.45	12.50	10.38	12.57	12.10	12.18	12.49	15.32	1.0435	0.4315

When $\psi = 0$, the penalized mean criterion allocates most of the subjects to treatment sequence AAA. The variance component of the penalized mean criterion favors treatment sequences $ABA/ABB/BAB$ and their allocations increase as ψ increases. For $N = 40$, the efficiency and success ratios move in an unpredictable direction as ψ increases. The allocations change similarly in the case of $N = 80$. However, it is hard to find a good value of ψ that

finds a balance between the success ratio and efficiency, where the efficiencies ranged from 0.5 to 0.6. For instance, the penalized mean adaptive design at $\psi = 1$ and $n = 80$, with relative efficiency of 0.6327 and success ratio of 0.5120, shows moderate efficiency and success ratio. While the penalization method may have a value in scrutinizing varying variances in treatments, there seems to be uncontrolled and unexplainable factors and we need to understand fully before we can use it adaptively.

Table 4.5: Allocation, Efficiency, and Success Ratio for three-period design using penalized mean criteria in (4.11).

N	ψ	**AAA**	**AAB**	**ABA**	**ABB**	**BAA**	**BAB**	**BBA**	**BBB**	**Efficiency**	**Success Ratio**
40	0	12.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	0.9523	0.4844
	1	5.91	5.62	8.43	4.01	4.03	4.00	4.00	4.00	0.9751	0.4548
	4	4.00	4.30	10.63	4.89	4.00	4.18	4.00	4.00	0.8427	0.4371
	200	4.00	4.02	9.30	6.11	4.00	4.57	4.00	4.00	0.9143	0.4444
80	0	52.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	0.5890	0.5643
	1	25.90	10.26	23.58	4.13	4.13	4.00	4.00	4.00	0.6327	0.5120
	4	4.00	5.75	45.79	7.81	4.27	4.38	4.00	4.00	0.5427	0.4605
	200	4.00	4.49	37.11	16.01	4.00	6.39	4.00	4.00	0.5655	0.4448

4.4 Comparison of Two Modeling Approaches

Bandyopadhyay, Biswas, and Mukherjee (2007) utilized an example of a three-period crossover trial of two-treatments for hypertension. In this trial, 68 subjects were equally assigned to the treatment sequences $ABB/BAA/ABA/BAB$. Li (2017) used the last two periods of this trial to obtain a crossover design with $AA/BB/AB/BA$. The response variable was continuous measurements of systolic blood pressure. Thus two binary response variables were computed by dichotomizing the blood pressures at "135 or more" and "140 or more" and denoting the responses as failures. Two sets of success probabilities were estimated from this data. $(\hat{v}_{A1}, \hat{v}_{A2}, \hat{v}_{B1}, \hat{v}_{B2}) = (0.24, 0.24, 0.24, 0.35)$ and $(\hat{v}_{A1}, \hat{v}_{A2}, \hat{v}_{B1}, \hat{v}_{B2}) = (0.35, 0.5, 0.35, 0.53)$ where v is the probability of success

with the letters denoting treatments and numbers denoting periods. These estimated probabilities were considered as actual success probabilities, and the multiple objective response adaptive technique was applied with $\lambda = 1$ and $\lambda = 0.9$. A comparison of allocations, efficiencies, and success ratios of three methods are provided below. We introduce fixed group effects, β_k, to the model in (4.3) to incorporate those success probabilities. The parameters and other settings are provided in the Table 4.6, and the results of simulations are provided in 4.7.

$$\eta_{ijk} = \mu + \alpha_i + \beta_k + \tau_{d(i,j,k)} + \gamma_{d(i-1,j,k)} \tag{4.12}$$

Table 4.6: Parameter values and expected success probabilities based on the crossover trial of Bandyopadhyay, Biswas, and Mukherjee (2007).

Probabilities	**Parameters**	**Treatment Sequences**	**Success Probabilities**	**Expected Success per Period**
$\hat{v}_{A1} = 0.24$	$\mu = -1.89$, $\beta_1 = 1$	AA	$(0.24, 0.24)$	0.240
$\hat{v}_{A2} = 0.24$	$\alpha_2 = 0.27$, $\beta_2 = 1$	AB	$(0.24, 0.35)$	0.295
$\hat{v}_{B1} = 0.24$	$\tau = -0.27$, $\beta_3 = 0.47$	BA	$(0.24, 0.24)$	0.240
$\hat{v}_{B2} = 0.35$	$\gamma = -0.27$, $\beta_4 = 0.47$	BB	$(0.24, 0.35)$	0.295
$\hat{v}_{A1} = 0.35$	$\mu = -1.56$, $\beta_1 = 1$	AA	$(0.35, 0.50)$	0.425
$\hat{v}_{A2} = 0.5$	$\alpha_2 = 0.68$, $\beta_2 = 1$	AB	$(0.35, 0.53)$	0.440
$\hat{v}_{B1} = 0.35$	$\tau = -0.06$, $\beta_3 = 0.88$	BA	$(0.35, 0.50)$	0.425
$\hat{v}_{B2} = 0.53$	$\gamma = -0.06$, $\beta_4 = 0.88$	BB	$(0.35, 0.53)$	0.440

The efficiencies in Table 4.7 were computed against the equal allocation design, the fixed optimal for two-period and two-treatment designs. First, we compare multiple objective response adaptive designs with $\lambda = 1$. We see that when the difference of the expected success probabilities between the sequences are small (0.425 vs 0.44, second example in Table 6), Li's strategy allocates an extensive number of the subjects to the treatment sequences AB/BB and results in a substantial loss of efficiency. Moreover, the gain in the expected success over an equal allocation design is minimal (0.4352 vs

Table 4.7: Allocation, Efficiency, and Success Ratio for two-period designs.

Parameters	**Design**	λ	**AA**	**AB**	**BA**	**BB**	**Efficiency**	**Expected Success**
$(\hat{v}_{A1}, \hat{v}_{A2}, \hat{v}_{B1}, \hat{v}_{B2})$	d_1^B		15.75	16.92	17.01	18.32	0.9912	0.2685
$(0.24, 0.24, 0.24, 0.35)$	d_2^L	1	13.13	21.03	13.03	20.80	0.9143	0.2738
	d_3^L	0.1	14.69	19.06	13.64	20.62	0.9522	0.2729
	d_4^P	1	12.81	20.85	12.49	21.85	0.8913	0.2745
	d_5^P	0.1	15.22	19.58	14.19	19.01	0.9829	0.2713
	d_6^E		17.00	17.00	17.00	17.00	1.0000	0.2675
$(0.35, 0.50, 0.35, 0.53)$	d_7^B		13.00	16.42	16.46	22.12	0.9769	0.4335
	d_8^L	1	7.32	16.35	14.88	29.46	0.8376	0.4352
	d_9^L	0.1	12.38	16.71	15.80	23.11	0.9627	0.4338
	d_{10}^P	1	16.22	17.89	15.40	18.49	0.9970	0.4330
	d_{11}^P	0.1	16.76	17.53	16.80	16.91	0.9983	0.4326
	d_6^E		17.00	17.00	17.00	17.00	1.0000	0.4325

$[B]$ Bandyopadhyay, Biswas, and Mukherjee (2007)
$[L]$ Li (2017)
$[P]$ The proposed multiple objective response adaptive design
$[E]$ Equal allocation design

0.4325). The simulations confirm this observation, and d_8 has relative efficiency of 0.8376 without much gain as a result. On the other hand, d_{10} adapts to the small differences in the sequences in a careful manner, and it assigns about 3 more subjects to better treatment sequences AB/BB without losing efficiency (0.9970). d_{10} allocates fewer subjects to AB/BB compared to d_7, d_8, and d_9.

It is noticeable that the pattern is not the same when there is a moderate difference in the expected success probabilities between the treatment sequences (0.24 vs 0.295). Design d_2 allocates 41.83 subjects to better sequences AB/BB whereas d_4 allocates 42.7 subjects. The designs allocate more subjects to better treatment sequences than d_1 while maintaining a high level of efficiency.

The designs constructed using the multiple objective response adaptive method with GEE are more responsive to the differences in treatments better than Bandyopadhyay, Biswas, and Mukherjee (2007) and Li (2017), while maintaining a high level of efficiency when there is a moderately large difference

in the treatment effects. Our revised method assigns more subjects to the better treatment sequence when the treatment differences are large. Moreover, the resulting designs are close to the optimal design with an equal allocations on all 4 sequences, when the treatment differences are negligible. This assures that even if the treatment difference is not as large as expected, the multiple objective response adaptive method is robust and creates an efficient design.

4.5 A Revised Adaptive Allocation Strategy

In Tables 4.2 and 4.4, we observed that the decrease in efficiency following the decrease in λ is not consistent for differing sample sizes. That is, if we wish to maintain some level of relative efficiency with respect to the known fixed optimal design while applying the multiple objective adaptive allocation scheme, we must fully understand the behaviors of this adaptive allocation scheme and find the suitable λ, which is determined by the true parameters as well as the sample size. The simulations on this scheme may help suggest some λ's but that is limited to the specific scenarios being studied. Therefore, we propose a sensible new approach to the multiple objective based allocation scheme without having to precisely know which λ to use.

Instead of defining the multiple-objective function as in (4.10), we define two objective functions separately.

$$H_{1,j,k} = \frac{\Delta(\hat{I}^k_{j+1}(\theta))}{\Delta(\hat{I}^{k'}_{j+1}(\theta))}, \tag{4.13}$$

$$H_{2,j,k} = \frac{f_{j,k}}{f_{j,k''}}, \tag{4.14}$$

which are the first and second terms of the equation (4.10).

The allocation takes the following steps.

1. Select an acceptable relative efficiency r^*.

2. Acquire the first j subjects.

3. Construct a logistic regression model and obtain the quasi-likelihood estimates of the parameters, μ, π_i's, τ, γ, etc.

4. Generate another set of data with the same number of total subjects as the current dataset with allocations according to the optimal design $d_{opt,p2}$.

5. Obtain estimates of the parameters and sandwich covariance matrices of the estimated parameters from the new data.

6. Compare the efficiencies of two designs, one being the adaptive design and the other being the fixed optimal design (Mukhopadhyay, Singh, and Dey (2015)). In this study, we compare the efficiencies of the two designs by the variance of estimated treatment effects, i.e. $r = var(\hat{\tau}_{opt})/var(\hat{\tau}_{Adaptive})$.

7. If $r < r^*$ then use $H_{1,j,k}$ as the allocation function for subject $j+1$, otherwise use $H_{2,j,k}$ as the allocation function for subject $j+1$.

8. Return to step 3 until all subjects are allocated.

Now we apply the above method to the parameters in Table 4.1 with the aim of constructing a response adaptive design with a relative efficiency around $r^* > 0.8$. First, we construct 2-period 2-treatment response adaptive designs with $N = 40$, 80 and 100. We present the results for 3-period 2-treatment designs with $N = 80$ and 100. The case for $N = 40$ was excluded as all adaptive designs constructed using (4.10) with any λ have relative efficiencies > 0.9.

From Table 4.8, we can see that the designs constructed using our adaptive allocation method, denoted as $d_{Adaptive}$, have relative efficiencies close to 0.8 or

Table 4.8: Comparison of new revised response adaptive two-period design with the results from Table 4.2.

N	Designs	AA	AB	BA	BB	Efficiency	Success Ratio
40	$d_{(0.8)}$	22.22	7.60	5.63	4.55	0.7615	0.5420
	$d_{(0.9)}$	16.63	9.49	7.28	6.60	0.9152	0.5042
	$d_{(1)}$	10.20	10.01	9.66	10.13	1.0141	0.4534
	$d_{Adaptive}$	21.75	6.11	6.21	5.94	0.8465	0.5319
80	$d_{(0.9)}$	45.08	16.64	10.56	7.72	0.7582	0.5507
	$d_{(0.95)}$	33.68	18.98	13.95	13.39	0.9368	0.5048
	$d_{(1)}$	20.35	19.81	18.96	20.88	1.0076	0.4532
	$d_{Adaptive}$	43.58	12.35	12.24	11.83	0.8430	0.5309
100	$d_{(0.9)}$	61.75	19.36	11.01	7.89	0.7096	0.5638
	$d_{(0.95)}$	45.77	23.49	16.30	14.44	0.8957	0.5168
	$d_{(1)}$	25.41	24.40	23.66	26.54	1.0258	0.4525
	$d_{Adaptive}$	57.09	14.42	14.12	14.37	0.7972	0.5391

slightly larger than that while the success ratios are increased by 9% compared to the designs for $\lambda = 1$. For $N = 40$, our adaptive design follows the pattern of changes in the allocations, efficiency, and success ratio so that we can find one between $d_{(0.8)}$ and $d_{(0.9)}$. For example, the allocation to the treatment sequence AA is $21.75(d_{Adaptive})$, which is between $16.63(d_{(0.8)})$ and $22.22(d_{(0.9)})$. This pattern is also the case for all other columns in the table for $N = 80$ and 100. Our adaptive designs appear to be constructed in a similar manner as the multiple objective response adaptive designs as if they were constructed with the λ in the suggested range of $(0.8, 0.9)$. Similarly, the $d_{Adaptive}$ designs for $N = 80$ and $N = 100$ fall right in between $d_{(0.9)}$ and $d_{(0.95)}$.

From Table 4.9, the relative efficiencies of our adaptive designs are 0.7999 and 0.7854 for $N = 80$ and $N = 100$ respectively. These efficiencies are very close to our target $r^* = 0.8$ while the success ratios are improved by approximately 9%. We can see that the allocation for treatment sequence AAA, relative efficiency, and the success ratio for the new adaptive designs $d_{Adaptive}$ follow the same pattern as the multiple objective response adaptive designs. The allocations to the other sequences are relatively small and do

Table 4.9: Comparison of our new revised response adaptive three-period design with the results from Table 4.4.

N	Designs	AAA	AAB	ABA	ABB	BAA	BAB	BBA	BBB	Efficiency	Success Ratio
80	$d_{(0.8)}$	41.09	10.41	6.07	4.24	5.89	4.24	4.06	4.00	0.7223	0.5458
	$d_{(0.9)}$	25.20	13.16	8.67	6.74	9.10	6.94	5.63	4.56	0.9220	0.5050
	$d_{(0.95)}$	16.76	12.06	9.02	8.48	10.27	8.61	7.94	6.86	1.0055	0.4708
	$d_{(1)}$	10.09	9.97	8.58	9.95	9.83	9.75	9.91	11.92	1.0370	0.4323
	$d_{Adaptive}$	39.49	5.25	7.35	5.42	4.98	5.02	6.88	5.61	0.7999	0.5267
100	$d_{(0.7)}$	69.41	6.03	4.28	4.02	4.25	4.01	4.00	4.00	0.5379	0.5812
	$d_{(0.8)}$	57.12	6.09	5.56	6.20	6.01	5.95	6.18	6.90	0.6345	0.5445
	$d_{(0.9)}$	37.37	16.65	10.20	7.39	10.46	7.62	5.80	4.50	0.8696	0.5231
	$d_{(0.95)}$	23.93	15.91	10.85	10.17	12.35	10.44	9.17	7.19	0.9858	0.4794
	$d_{(1)}$	12.45	12.50	10.38	12.57	12.10	12.18	12.49	15.32	1.0435	0.4315
	$d_{Adaptive}$	50.48	5.93	9.52	6.45	5.65	5.78	9.11	7.08	0.7854	0.5278

not seem to affect the efficiency that much as long as the allocation to AAA is well controlled. The above strategy successfully leads us to obtain desired success ratios and maintain efficiency to a pre-specified level without having to determine what the necessary λ is.

4.6 Conclusion

Binary responses have distinct properties that are different from continuous responses in that their means and variances are functions of the responses. As a result, binary response designs are response dependent. Due to this characteristic, the construction of optimal designs for binary responses requires special attention. There are limited studies on response adaptive designs and optimal designs in the literature. In this chapter, we compared two approaches of constructing response adaptive designs. Also, we conducted a simulation study based on an actual data example to investigate the performance of the multiple objective response adaptive designs using the GEE over the other two methods.

In section 4.3, we constructed response adaptive designs using the two objective functions, namely the multiple objective function and the variance-penalized mean function. The designs constructed using the multiple objective

function were highly efficient, successful with respect to treatment outcomes, and more importantly sensible by varying the values of weight parameter λ. The penalized mean criterion and its functionality, on the other hand, depend highly on the covariance structure of the estimated parameters. Compared with predicting the variance of estimated treatment effects, $var(\hat{\tau})$, in the multiple objective method, the required covariance structure is unknown before any trial and even harder to predict in advance.

In Tables 4.2 and 4.4, we observe that the choice of λ for an efficient and successful design would depend on the sample size and the true values of μ, $\pi_i's$, τ, and γ, which may not be known before the trial. However, the efficiencies drop significantly when N increases or λ decreases. Those designs may have significantly higher success ratios but may also have significantly low efficiency, which is undesirable.

In section 4.4, we compared our approach to the multiple objective adaptive designs using the GEE to the response adaptive design by Mukhopadhyay, Singh, and Dey (2015) and multiple objective adaptive designs using binary probability modeling approach by Li (2017) for 2 period 2 treatment crossover designs. The proposed designs responded to the differences in the treatment effects in a rather robust manner. When the treatment difference is very small, the proposed designs were very close to the optimal design with an equal allocation on 4 treatment sequences, $AA/AB/BA/BB$. On the other hand, the other two methods assign too large a proportion of subjects to treatment sequences BB and lose efficiencies for very small gain in success ratios. When the treatment difference is moderately large, the proposed design with $\lambda = 1$ assigns more subjects to a better treatment sequences compared to the other 2 designs proposed by Bandyopadhyay, Biswas, and S. Mukherjee (2009a) and Li (2017).

Previously in section 4.3 we observed that the choice of λ is very important

in finding a balance between the relative efficiency and a success ratio. One may suggest some appropriate range of λ but that is valid for only a certain set of parameters and sample size, and the true parameters are usually unknown. To overcome this challenge, we devised a multiple objective response adaptive scheme, which utilizes all of the two components of (4.10), not simultaneously but in a sequential manner. The simulation results show that this adaptive scheme can construct designs with desired relative ratios without having to select the weight parameter λ. Our scheme allows researchers to run an adaptive trials knowing that their design would find the balance of the two important components of the trial, being efficient and being ethically defensible.

4.7 Discussion

Response adaptive designs have so much potential to complement the traditional experimental designs. The use of the data acquired during the trial may benefit the trial in numerous ways such as improving the statistical power or reducing the cost of the trial by re-calculating the required sample size, assigning more subjects to a better treatment or treatment sequences, or utilizing the information acquired from the covariates to improve efficiency. The multiple objective criteria may incorporate more components or select various other sets of components such as cost efficiency versus statistical efficiency and many others.

The results in section 4.5 were limited to some sets of parameters, which are common scenarios selected to show how the proposed designs and our adaptive scheme work. Further studies may be needed to verify whether our adaptive scheme is valid on larger sets of parameters and various model assumptions such as the self and mixed carryover effects model (Hedayat and Afsarinejad (2002), Kunert and Stufken (2002)), higher order carryover effects model (Bishop and Jones (1984), Bose and Mukherjee (2000)), proportional

carryover effects model (Bailey and Kunert (2006), Bose and Stufken (2007), Kempton, Ferris, and David (2001), Zheng (2013)), and under other allocation conditions such as group sequential allocations.

The two response adaptive allocation methods investigated in this chapter are not pre-planned. Therefore, introducing penalty or significance level adjustment for adaptive designs may further be investigated.

Chapter 5

Conclusion

This thesis focused on some of many practical and ethical questions arising in the crossover trials. We have deliberated various impractical model assumptions used in constructing optimal designs. We investigated the optimality of crossover designs for the cases when treatment variances are unequal, how proportional carryover effects affect the optimal designs. We also advocated the use of the GEE in adaptively planning the multiple objective designs and developed a new allocation method to achieve the maximum success ratio while maintaining a sufficiently high level of statistical efficiency.

5.1 Main Contribution

5.1.1 Optimal Crossover Design Under an Unequal Treatment Variance Assumption

This chapter addressed the problem of unequal variances in the design of crossover trials. Under the unequal variance assumption, treatment sequences can have distinct and unequal covariance structures of the responses, although the correlation coefficients may be assumed to be the same for all treatment sequences. We proposed the universally optimal designs for independent responses and constrained T optimal designs for dependent responses when $t = 2$ and $p = \{2, 3\}$. For two-period two-treatment crossover designs, some may favor the use of the design AB/BA as all subjects can be allocated to the treatment sequences that allow within-subject comparisons, although the design $AA/BB/AB/BA$ is optimal. We constructed new optimal designs and

find that

1. The ratio of treatment standard deviations, $r = \sigma_B/\sigma_A$, plays an important role. Moreover, when $\rho = 0$, the optimal 2-period and 3-period crossover designs can be derived analytically. For $p = 2$, the optimal design allocates $(1/(1+r)^2, r^2/(1+r)^2, r/(1+r)^2, r/(1+r)^2)$ to treatment sequences $AA/BB/AB/BA$. For $p = 3$, there are sets of optimal designs and one of them is an extension of two-period optimal design and others are slight variations of it.

2. For $\rho \neq 0$ and $p = 2$, we applied two constraints and derived optimal designs under each constraint. These optimal designs can be represented as $(1/(2(1+r)), r/(2(1+r)), r/(2(1+r)), 1/(2(1+r)))$ and $(1/(1+r)^2, r^2/(1+r)^2, r/(1+r)^2, r/(1+r)^2)$. The more unequal treatment variances get, the more efficient these designs become when compared with the existing universally optimal design with an equal allocation on four sequences. Also, the weaker within-subject correlation is, the more efficient these designs are. We found the range of ρ and r where the traditional optimal design may be robust against the unequal treatment variances.

3. For example, when $p = 3$, the optimal design under the equal variance assumption, ABB/BAA, is robust against unequal variances for $\rho \geq 1/4$ or $1/3 \geq r \geq 3$.

4. The above findings can be integrated with the test of unequal treatment variances in a way that researchers conducting crossover trials with $p = 2$, $t = 2$, and treatment sequences AB/BA or $AA/BB/AB/BA$ may obtain the MLE for variances and correlation and make decision on whether to extend one period or adjust their allocations to adapt to the information acquired.

5.1.2 Optimal Crossover Design with Proportional Carryover Effects and Baselines

We extended the scope of the universal optimality criteria of Kiefer (1975) to the traditional carryover effects model with baseline measurements with an equicorrelated covariance matrix. We further investigated the effect of the proportional assumption of carryover effects to the traditional model with baseline measurements. We find that

1. Optimal designs differ based on the values of proportionality parameters λ_{1o} and λ_{2o} as well as the number of treatments t and number of periods p.

2. The universally optimal design under the traditional model with baseline is E optimal under the proportional model assumption and baseline measurements.

3. The E optimal designs under the proportional carryover effects model with baseline measurements perform relatively well with respect to other optimality criteria ($A/D/T$).

4. When $p = 2$ and $t = 3$, the universally optimal designs suggested by Carriere and Reinsel (1993) with $p_{<re>} = 1/t = 1/3$ and $p_{<di>} = (t-1)/t = 2/3$ is $A/D/E/T$ optimal even in the case of proportional carryover effects model. This remains to be true for larger t. In this case, the $< re >$ consists of treatment sequences $\{AA,\ BB,\ CC\}$ and $< di >$ consists of $\{AB,\ AC,\ BA,\ BC,\ CA,\ CB\}$. The design $d = d_{<di>} * (t-1)/t + d_{<re>} * 1/t$ is $A/D/E/T$ optimal for $t = 4$, 5, and 6.

5. When $p = 3$, the optimal designs depend on t and $\{\lambda_{1o}, \lambda_{2o}\}$. For $p = 3$ and $t = 2$, the optimal design suggested by Laska, Meisner, and Kushner (1983), ABB/BAA, is $A/D/E/T$ optimal under the proportional carryover effects model with baseline measurements. For $t \geq 3$ and $p = 3$, the $A/D/E/T$ optimal designs differ now for $\{\lambda_{1o}, \lambda_{2o}\}$. Design with $< di >$ is $A/D/E/T$ optimal for negatively proportional carryover effects as well

as no carryover effects and this design is E optimal for positively proportional carryover effects. Also, $< re >$ is $A/D/T$ optimal for positively proportional carryover effects.

6. For $p = 4$ and $t = 3$, the $A/D/E/T$ optimal designs depend on negatively proportional carryover effects. $< re >$ is E optimal for negatively proportional carryover effects and nearly optimal with respect to $A/D/T$ criteria as well. Also it is $A/D/E/T$ optimal for positively proportional carryover effects. Therefore, $< re >$ is recommended for $p = 4$ and $t = 3$ crossover designs.

7. For $p = 4$ and $t \geq 4$, $< di >$ is $A/D/E/T$ optimal for negatively proportional carryover effects as well as no carryover effects and this design is E optimal for positively proportional carryover effects. Also, $< re >$ is $A/D/T$ optimal for positively proportional carryover effects.

8. For $p = 5$, numerous symmetric blocks of treatment sequences form $A/D/E/T$ optimal designs. However, $< di >$ and $< re >$ are shown to be efficient again for $t \geq p - 1$.

9. In general, we do see that the benefit of baselines reduces when p increases while t is fixed. But the use of baseline measurements improves the design efficiency when $t \geq p - 1$. Also, the use of baseline improves efficiency, if and only if, right designs are selected for the proportionality parameters. For example, use of $< re >$ design on negative proportional carryover effects for trials with $p = 5$ and $t = 5$ would result in relative efficiency of 0.8771 with respect to the E optimal design without baseline measurements. On the other hand, correctly specified designs may improve the efficiencies significantly. However, the proportionality between direct treatment effects and carryover effects is unknown. Such information may be obtainable from previous studies or we may need to apply an adaptive allocation scheme.

5.1.3 Multiple Objective Response Adaptive Crossover Designs for Binary Responses Using the Generalized Estimating Equations

We extended the strategy of building crossover designs with binary outcomes to utilize the generalized estimating equations method, as it is known to be robust against mis-specification of covariance matrix of the responses. Then we compared two types of response adaptive allocation methods that consider treatment effects and efficiency (variance) of the treatment sequences, namely the multiple objective response adaptive allocation and the penalized mean response adaptive allocation. These two methods use weight parameters λ and ψ that assigns weights to two objectives, treatment effect and efficiency (variance). We also compared the design building approach using the GEE against the probability modeling approach in terms of sensitivity, efficiency, and success ratios. Lastly, we proposed a new allocation method that achieves a higher success ratio while maintaining a sufficiently large level of statistical efficiency. This method does not involve selecting appropriate λ, which can vary for different parameters and sample size. In this chapter, we find that

1. The multiple objective response adaptive allocation function constructs designs that are easier to predict and successfully fulfilling the multiple objectives. The penalized mean objective function becomes highly complex in repeated measurement designs.

2. The new revised multiple objective response adaptive allocation method constructs designs that are equally responsive to large differences in treatment effects and are robust against small difference in treatment effects. The latter is important because researchers do not know the true treatment difference prior to the trial. These findings can provide researchers with confidence in using the multiple objective response adaptive design.

3. Our new adaptive allocation method successfully constructed designs that maximize the success ratios while achieving the target efficiency,

(relative efficiency in the range from 0.7854 to 0.8465 with the target efficiency set to be 0.8). This method does not involve selecting an appropriate λ before the allocations.

5.2 Future Research

This thesis addressed three practical issues, unequal treatment variances, proportional carryover effects, and binary response designs. When treatment variances are unequal or carryover effects are proportional to direct treatment effects, the optimality of designs change as they are model dependent. These new optimal designs can largely improve the efficiency when the assumptions are met. However, information about the assumptions are not known at the planning stage of trials. For this reason, correctly checking whether various assumptions are met, and applying adaptive allocation schemes based on the identified assumptions may further improve the design efficiency.

There are other issues to remain to be further explored.

1. The optimal designs under unequal variances must be extended to other model assumptions such as the mixed and self carryover effects model suggested by Hedayat and Afsarinejad (2002) and Kunert and Stufken (2002).

2. As an natural extension, the optimal designs and response adaptive designs with binary outcome and baseline measurements must be investigated.

3. The optimal designs under the proportional carryover effects must be extended to other model assumptions and efficient ways of estimating the proportionality parameters must be developed. One possible issue is to devise a way of grouping multiple carryover effects by their proportionality and constructing optimal designs for proportional model with multiple proportionality parameters.

4. The multiple objective response adaptive designs must be verified for a

larger set of parameter values and the new adaptive allocation scheme must be tested as well.

5. In this thesis, we proposed various ways to keep the design robust by exploring when the traditional assumptions are violated. Future studies may incorporate them into response adaptive allocation scheme.

References

Agresti, A. (2014). *Categorical Data Analysis*. Hoboken: Wiley. 65

Armitage, P. (1975). *Sequential Medical trials*. Oxford: Blackwell. 3, 63

Bailey, R. A. and J. Kunert (2006). "On optimal crossover designs when carryover effects are proportional to direct effects." In: *Biometrika* 93, pp. 613–625. 46, 47, 53, 85

Bandyopadhyay, U and A. Biswas (2001). "Adaptive designs for normal responses with prognostic factors." In: *Biometrika* 88, pp. 409–419. 63

Bandyopadhyay, U, A Biswas, and S. Mukherjee (2009a). "Randomized play-the-winner rule in two-period two-treatment repeated measurement design." In: *Austrian J Statistics* 38, pp. 151–169. 3

— (2009b). "Adaptive two-treatment two-period crossover design for binary treatment responses incorporating carry-over effects." In: *Statistical Methods and Applications* 18, pp. 13–33. 3

— (2007). "Adaptive two-treatment two-period crossover design for binary treatment responses." In: *Statistica Neerlandica* 61.3, pp. 329–334. 5, 76–78

Bartlett, M. S. (1937). "Properties of sufficiency and statistical tests." In: *Proceedings of the Royal Statistical Society, Series A* 160, pp. 268–282. 9

Bishop, S.H. and B. Jones (1984). "A review of higher order crossover designs." In: *Journal of Applied Statistics* 11, pp. 29–50. 2, 84

Bose, M. and A. Dey (2009). *Optimal crossover designs*. New Jersey: World Scientic. 66

Bose, M. and B. Mukherjee (2000). "Crossover designs in the presence of higher order carryovers." In: *Australian New Zealand Journal of Statistics* 42(2), pp. 235–244. 2, 84

— (2003). "Optimal crossover designs under a general model." In: *Statistics Probability Letters* 62, pp. 413–418.

Bose, M. and J. Stufken (2007). "Optimal crossover designs when carryover effects are proportional to direct effects." In: *J. Statist. Plann. Inference.* 137, pp. 3291–3302. 46, 47, 85

Carriere, K.C. (1994). "Crossover designs for clinical trials." In: *Statistics in Medicine* 13(10), pp. 1063–1069. 8, 19

Carriere, K.C. and G.C. Reinsel (1992). "Investigation of Dual-Balanced Crossover Designs for Two Treatments." In: *Biometrics* 48(4), pp. 1157–1164. 8

— (1993). "Optimal Two-Period Repeated Measurement Designs with Two or More Treatments." In: *Biometrika* 80(4), pp. 924–929. 8, 9, 53, 58, 70, 88

Chen, V. and V.M. Chinchilli (2010). "A Generalized Correlation Coefficient for the General Two-Treatment Crossover Design." In: *Communications in Statistics - Theory and Methods* 39(11), pp. 2063–2080. 37

Cheng, C.S. and C.F. Wu (1980). "Balanced Repeated Measurements Designs." In: *The Annals of Statistics* 8(6), pp. 1272–1283. 2, 8

Chinchilli, V.M. (1996). "Design And Analysis Of Intra-Subject Variability In Cross-Over Experiments." In: *Statistics in Medicine* 15.15, pp. 1619–1634. 25

Chinchilli, V.M. et al. (2005). "A General Class of Correlation Coefficients for the 2 2 Crossover Design." In: *Biometrical Journal* 47(5), pp. 644–653. 25, 37

Choi, S. C. and V.M. Chinchilli (1987). "Analysis of Within- and Across-Subject Correlations." In: *Multivariate Statistical Modeling and Data Analysis*, pp. 77–93.

Connor, EM et al. (1994). "Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with zidovudine treatment." In: *New England Journal of Medicine* 331, pp. 1173–1180. 63

Cross, D.V. (1973). "Sequential dependencies and regression in psychophysical judgements." In: *Percept. Psychophys* 14, pp. 547–552. 38

Cross, D.V. and L.T. Decarlo (1990). "Sequential effects in magnitude scaling: models and theory." In: *Journal of Experimental Psychology* 119, pp. 375–396. 38

Derrick, B. and P. White (2016). "Why Welchs test is Type I error robust." In: *The Quantitative Methods for Psychology* 12(1), pp. 30–38. 9

Divecha, J. and J. Gondaliya (2015). "Estimation of treatment and carryover effects in optimal cross-over designs for clinical trials." In: *Statistics in Biopharmaceutical Research* 7(2), pp. 95–105. 2

Ferris, J.F. (1999). "Response Models and Efficient Designs for Change-Over Experiments with Treatment Carryover." PhD thesis. University of Edinburgh. 38

Freeman, P.R. (1989). "The performance of the two-stage analysis of two-treatment, two-period crossover trials." In: *Statistics in Medicine* 8(12), pp. 1421–1432. 2, 9

Gelfand, J.M. et al. (2019). "A Phase IV, Randomized, Double-Blind, Placebo-Controlled Crossover Study of the Effects of Ustekinumab on Vascular Inflammation in Psoriasis (the VIP-U trial)." In: *Journal of Investigative Dermatology*.

Gosset, W.S. (1908). "The Probable Error of a Mean." In: *Biometrika* 6.1, pp. 1–25. 3

Grizzle, J.E. (1965). "The two-period change-over design and its use in clinical trials." In: *Biometrics* 21, pp. 467–480. 1, 9

Hedayat, A. and K. Afsarinejad (1975). "Repeated measurements designs." In: *In A Survey of Statistical Design and Linear Models (N. Srivastava, ed.)* Pp. 229–240. 2, 37

— (1978). "Repeated measurements designs. II." In: *Annals of Statistics* 6, pp. 619–628. 2

Hedayat, A. and K. Afsarinejad (2002). "Repeated Measurements Designs for a Model With Self and Mixed Carryover Effects." In: *Journal of Statistical Planning and Inference* 106, pp. 449–459. 2, 84, 91

Hills, M. and P. Armitage (1979). "The two-period crossover clinical trial." In: *B. J. of Clinical Pharmacology* 8, pp. 7–20. 2, 37

Jemielita, T. (2017). "Efficient baseline utilization in crossover clinical trials through linear combinations of baselines: parametric, nonparametric, and model selection approaches." PhD thesis. Philadelphia, Pennsylvania, USA: University of Pennsylvania. 2, 4, 37

Jones, B., J. Kunert, and H. P. Wynn (1992). "Information matrices for mixed effects models with applications to the optimality of repeated measurements designs." In: *Journal of Statistical Planning and Inference* 78, pp. 307–316. 8

Jung, Y. (2009). "Tests for Unequal Treatment Variances in Crossover Designs." PhD thesis. Kansas State University. 25

Kempton, R.A., S. J. Ferris, and O. David (2001). "Optimal Change-Over Designs When Carry-Over Effects Are Proportional to Direct Effectsof Treatments." In: *Biometrika* 88(2), pp. 391–399. 38, 46, 47, 85

Kenward, M. G. and J. H. Roger (2009). "The use of baseline covariates in crossover studies." In: *Biostatistics* 11(1), pp. 1–17. 2, 4

Kershner, R. P. and W. T. Federer (1981). "Two-treatment crossover designs for estimating a variety of effects." In: *Journal of American Statistical Association* 76, pp. 612–619. 2, 37

Kiefer, J. (1975). "Construction and optimality of generalized Youden designs." In: *A Survey of Statistical Designs and Linear Models* 3(1), pp. 109–118. 2, 13, 16, 17, 88

Kifer, J. (1973). "Construction and optimality of generalized Youden designs." In: *In A Survey of Statistical Design and Linear Models (Proc. Internat. Sympos., Colorado State Univ., Ft. Collins, Colo., 1973) (J. N. Srivastava, ed.)*. Pp. 333–353. 2, 37

Knuth, Donald E. (1968). *The Art of Computer Programming, Volume I: Fundamental Algorithms*. Addison-Wesley.

Kovalchuk, M.O. et al. (2018). "Acute effects of riluzole and retigabine on axonal excitability in patients with ALS: a randomized, double-blind, placebo-controlled, cross-over trial." In: *Clinical Pharmacology Therapeutics* 104(6), pp. 1136–1145. 9

Kunert, J. (1983). "Optimal Design and Refinement of the Linear Model with Applications to Repeated Measurements Design." In: *Annals of Statistics* 11(1), pp. 247–257. 2, 3, 14, 16, 17

— (1984). "Optimality of balanced uniform repeated measurements designs." In: *Annals of Statistics* 12, pp. 1006–1017.

— (1985). "Optimal repeated measurements designs for correlated observations and analysis by weighted least squares." In: *Biometrika* 72(3), pp. 375–389. 2

Kunert, J. and J. Stufken (2002). "Optimal Crossover Designs in a Model With Self and Mixed Carryover Effects." In: *Journal of American Statistical Association* 97(459), pp. 898–906. 2, 8, 84, 91

— (2008). "Crossover Designs for Two Treatments in the Presence of Mixed and Self-Carryover Effects." In: *Journal of the American Statistical Association* 103(484), pp. 1641–1647.

Kushner, H.B. (1997). "Optimal repeated measurements designs: the linear optimality equations." In: *Annals of Statistics* 25(6), pp. 2328–2344. 2, 4, 37, 38, 43, 44, 48

— (1998). "Optimal and efficient repeated-measurements designs for uncorrelated observations." In: *Journal of American Statistical Association* 93, pp. 1176–1187.

Laska, E., M. Meisner, and H.B. Kushner (1983). "Optimal crossover designs in the presence of carryover effects." In: *Biometrics* 39, pp. 1087–1091. 2, 8, 19, 37, 54, 55, 88

Levene, H. (1960). "Robust tests for equality of variances." In: pp. 278–292. 3, 9

Li, X and X. Wang (2012). "Variance-penalized response-adaptive randomization with mismeasurement." In: *J of Statistical Planning and Inference* 142.7, pp. 2128–2135. 4, 64

Li, Y (2017). "Optimal Crossover Designs in Clinical Trials." PhD thesis. Edmonton AB, Canada: University of Alberta. 5, 64, 65, 68, 69, 76, 78,

Liang, KY and S. Zeger (1986). "Longitudinal data analysis using generalized linear models." In: *Biometrika* 73.1, pp. 13–22. 64, 67

Liang, Y. and K.C. Carriere (2009). "Multiple-objective response-adaptive repeated measurement designs for clinical trials." In: *Journal of Statistical Planning and Inference* 139, pp. 1134–1145. 4, 64, 67

— (2010). "On the role of baseline measurements for crossover designs under the self and mixed carryover effects model." In: *Biometrics* 66, pp. 140–148. 4, 38

Liang, Y. et al. (2014). "Multiple-objective response-adaptive repeated measurement designs in clinical trials for binary responses." In: *Statistics in Medicine* 33.4, pp. 607–617. 64, 68

Lin, H. M. and V.M. Chinchilli (2001). "A Measure Of Association Between Two Sets Of Multivariate Responses." In: *Journal of Biopharmaceutical Statistics* 11(3), pp. 139–154.

McCullagh, P (2005). "Quasi-Likelihood Functions." In: *The Annals of Statistics* 11.1, pp. 59–67. 68

McCullagh, P and JA. Nelder (1999). *Generalized Linear Models*. Boca Raton: Chapman and Hall/CRC. 66

Mukhopadhyay, S, S.P Singh, and A. Dey (2015). *Optimal two-treatment crossover designs for binary response models*. eprint: `arXiv:1505.02488`. 69, 70, 74, 80, 83

Pukelsheim, F. (2006). *Optimal Design of Experiments*. Society for Industrial and Applied Mathematics. 44

Romano, M. et al. (2019). "Changes in serum cortisol and blood glucose concentrations in anesthetized, pain-free dogs administered fentanyl." In: *Veterinary Anaesthesia and Analgesia* 46, pp. 488–491. 9

Saha, R. (1983). "Repeated measurements designs." In: *Calcutta Statistical Association Bulletin* 32, pp. 153–168. 8

Schifferstein, H.N.J. and I.M. Oudejans (1996). "Determinants of cumulative successive contrast in saltiness intensity judgement." In: *Percept. Psychophys.* 58, pp. 713–724. 38

Shanga, G.M (2003). "Analysis of crossover designs when treatments have unequal variances." PhD thesis. Kansas State University. 25

Sorkness, C.A. et al. (2019). "Adapting clinical trial design to maintain meaningful outcomes during a multicenter asthma trial in the precision medicine era." In: *Contemporary Clinical Trials* 77, pp. 98–103. 63

Tymofyeyev, Y., W. F. Rosenberger, and F. Hu (2007). "Implementing Optimal Allocation in Sequential Binary Response Experiments." In: *Journal of the American Statistical Association* 102.477, pp. 224–234. 63

Valois, MF. (1997). "Evaluation of the performance of the generalized estimating equations method for the analysis of crossover design." MA thesis. Montreal QC, Canada: McGill University. 64

Wallenstein, S. (1979). "Inclusion of baseline values in the analysis of crossover designs." In: *Biometrics* 35, p. 894. 2, 37

Wallensteins, S. and A.C. Fisher (1977). "The analysis of the two period repeated measurements crossover design with applications to clinical trials." In: *Biometrics* 33, pp. 261–269. 2, 37

Wallensteins, S. and J.L. Fleiss (1988). "The two-period crossover design with baseline measurements." In: *Communications in Statistics-Theory and Methods* 17(10), pp. 3333–3343. 2, 4

Wang, J. (2014a). "Adaptive Optimal Two Treatment Crossover Designs with Binary Endpoint." In: *University of Illinois, Chicago.* 3, 63

Wang, M. (2014b). "Generalized Estimating Equations in Longitudinal Data Analysis: A Review and Recent Developments." In: *Advances in Statistics* 2014, pp. 1–11.

Wei, LJ and S Durham (1978). "The randomized play-the-winner rule in medical trials." In: *J American Statistical Association* 73, pp. 840–843. 3, 63

Welch, B. L. (1947). "The generalization of Student's problem when several different population variances are involved." In: *Biometrika* 34.(1-2), pp. 28–35. 3, 9

Wilcoxon, F. (1945). "Individual comparisons by ranking methods." In: *Biometrics Bulletin* 1.6, pp. 80–83. 3

Willan, A. R. and J. L. Pater (1986). "sing baseline measurements in the two period cross-over trial." In: *Controlled Clinical Trials* 7, pp. 282–289. 2, 37

Yi, Y and X. Wang (2009). "Response Adaptive Designs with a Variance-penalized Criterion." In: *Biometrical Journal* 51.5, pp. 763–773. 4, 64, 68

Zelen, M (2003). "Play the winner rule and the controlled clinical trial." In: *J American Statistical Association* 64, pp. 131–146. 3, 63

Zheng, W. (2013). "Optimal crossover designs for the proportional model." In: *Annals of Statistics* 41(4), pp. 2218–2235. 38, 47–53, 85

Appendix A

Some Proofs for Chapter 2

A.1 Proof for Lemma 7

The information matrix can be written as,

$$X_{3d}' pr^{\perp}(X_{1d}^*) X_{2d}$$

$$= \begin{pmatrix} \frac{p_4(r^2p_4+p_2-\rho(rp_4+p_2))}{\sigma_A^2(\rho^2-1)(r^2p_4+p_2)} + \frac{(rp_1(\rho-1)+\rho p_3)((\rho-1)r^2p_1-p_3+r\rho p_3)}{\sigma_A^2 r(\rho^2-1)(r^2p_1+p_3)} + \frac{\rho(p_4+2rp_1+p_3)-r(p_4+2p_1+p_3)}{\sigma_A^2 r(\rho^2-1)} & 0 \\ \frac{(r^2p_4+p_2-\rho(rp_4+p_2))(\rho(rp_4+p_2)-p_2)}{\sigma_B^2(\rho^2-1)(r^2p_4+p_2)} + \frac{p_3(r^2p_1(\rho-1)-p_3(1-r\rho))}{\sigma_B^2(\rho^2-1)(r^2p_1+p_3)} + \frac{p_4(r\rho-1)+2p_2(\rho-1)+p_3(r\rho-1)}{\sigma_B^2(\rho^2-1)} & 0 \end{pmatrix}$$

$$X_{2d}' pr^{\perp}(X_{1d}^*) X_{2d}|_{1,1}$$

$$= -\frac{p_1\left(\sigma_A^2\rho^2+2\sigma_B^2\rho-2\sigma_B^2\right)}{\sigma_B^2(\rho-1)(\rho+1)} + \frac{\left(\sigma_A^2p_4-\sigma_A\sigma_B\rho p_4-\sigma_B^2\rho p_2+\sigma_B^2p_2\right)^2}{\sigma_A^2(\rho-1)(\rho+1)\left(\sigma_A^2p_4+\sigma_B^2p_2\right)} - \frac{p_4\left(\sigma_A^2-2\sigma_A\sigma_B\rho+\sigma_B^2\right)}{\sigma_A^2(\rho-1)(\rho+1)} + \frac{2\sigma_B^2\rho p_2}{\sigma_A^2(\rho-1)(\rho+1)}$$

$$-\frac{2\sigma_B^2p_2}{\sigma_A^2(\rho-1)(\rho+1)} + \frac{\sigma_A^4\rho^2p_1^2-2\sigma_A^3\sigma_B\rho^2p_1^2+\sigma_A^2\sigma_B^2\rho^2p_1^2}{\sigma_B^2(\rho-1)(\rho+1)\left(\sigma_A^2p_1+\sigma_B^2p_3\right)} + \frac{2\sigma_A\rho p_1}{\sigma_B(\rho+1)} + \frac{2\sigma_A\rho p_1}{\sigma_B(\rho-1)(\rho+1)} + \frac{2\rho p_1}{(\rho-1)(\rho+1)} - \frac{3p_1}{(\rho-1)(\rho+1)} + p_3$$

$$X_{2d}' pr^{\perp}(X_{1d}^*) X_{2d}|_{2,2} = 0$$

$$X_{2d}' pr^{\perp}(X_{1d}^*) X_{2d}|_{1,2} = X_{2d}' pr^{\perp}(X_{1d}^*) X_{2d}|_{2,1} = \frac{\rho\left(\sigma_A^2p_1+\sigma_A(p_3+p_4)\sigma_B+y\right)}{\sigma_A^2\left(\rho^2-1\right)}$$

$$\Rightarrow X_{3d}^{*'} pr^{\perp}(X_{1d}^*) X_{2d}^* (X_{2d}^{*'} pr^{\perp}(X_{1d}^*) X_{2d}^*)^- = 0_{2\times 2}.$$

With these results, we can easily see that condition (2) from Proposition 2 is satisfied.

A.2 Proof for Theorem 3

The trace of the Information matrix for estimating direct treatment effects under the simpler model using carryover effects is given as below,

$$
\begin{aligned}
&f(p_1,p_2,p_3,p_4) = tr({X^*}'_{3d} pr^{\perp}(F^*{}_d) X^*{}_{3d}) \\
&= \frac{2\rho r p_1}{\sigma_A^2(1+\rho)} + \frac{2\rho p_2 p_4}{\sigma_A^2(1+\rho) r(r^2 p_2 p_4)} + \frac{p_2(\rho^2 p_2 - p_4 - p_2)}{\sigma_A^2(\rho^2-1) r^2 (r^2 p_4 + p_2)} \\
&+ \frac{p_3^2(\rho^2-1) - p_1 p_2 (3 - 2\rho + \rho^2 r^2) - p_3((3-2\rho) r^2 p_1 + \rho^2 r^4 p_1 + (3-2\rho) p_2)}{\sigma_A^2(\rho^2-1)(r^2 p_4 + p_2)} \\
&+ \frac{p_3}{\sigma_A^2} + \frac{r^2 p_1^2 (2 + 2\rho(r-1) + \rho^2 (r-1)^2)}{\sigma_A^2(\rho^2-1)(r^2 p_1 + p_3)}.
\end{aligned}
$$

The derivatives can be obtained as below

$$
\begin{aligned}
\frac{\partial f}{\partial p_1} &= \frac{r^4 p_1^2(\rho^2-1) + p_3^2((2\rho-3) + 2(\rho-1)\rho r - r^2\rho^2) + 2r^2 p_1 p_3(\rho^2-1)}{\sigma_A^2(\rho^2-1)(r^2 p_1 + p_3)^2} \\
\frac{\partial f}{\partial p_2} &= \frac{p_2^2(\rho^2-1) + p_4^2 r^2 (2r\rho(r-1) - 3r^2 + 2r\rho^2 - r^2\rho^2) + 2p_2 p_4 r^2(\rho^2-1)}{\sigma_B^2(\rho^2-1)(r^2 p_1 + p_3)^2} \\
\frac{\partial f}{\partial p_3} &= \frac{p_3^2(\rho^2-1) + p_1^2 r^2((2 - 2\rho + \rho^2) + 2(\rho-1)\rho r - r^2) + 2r^2 p_1 p_3(\rho^2-1)}{\sigma_A^2(\rho^2-1)(r^2 p_1 + p_3)^2} \\
\frac{\partial f}{\partial p_4} &= \frac{p_2^2(2(\rho-1)\rho r - 1 - r^2(2 - 2\rho + \rho^2) + r^2 p_4^2(r^2(\rho^2-1)) + 2p_2 p_4(\rho-1)\rho r}{\sigma_B^2(\rho^2-1)(r^2 p_1 + p_3)^2}.
\end{aligned}
$$

Using the Lagrange optimization, we can find the solution that satisfies $\partial f/\partial p_1 = \partial f/\partial p_2$ is $p_3 = r * p_1$ and the solution that satisfies $\partial f/\partial p_2 = \partial f/\partial p_4$ is $p_2 = r * p_4$. However, $\partial f/\partial p_1 = \cdots = \partial f/\partial p_4$ only if $r = 1$ or $r = -1 - \rho + 2\rho^2 \pm \sqrt{\rho(2 - 3\rho - 4\rho^2 + 4\rho^3}$.

The rest of the derivation is long but straight forward. Also, the Hessian matrix of $f(p_1,p_2,p_3,p_4) - h(p_1,p_2,p_3,p_4)$ can be shown to be negative semidefinite at the solutions. This means that for any non-zero vector z $z'Hz \leq 0$ which satisfies the sufficiency condition in KKT optimization. So it is a local maximum.