Optimal Crossover Designs in Clinical Trials

by

Yin Li

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 specializes in statistical issues involving crossover designs, a very popular design in clinical trials for comparing non-curative treatments for their efficacy. The popularity stems from the fact that each experimental subject receives a sequence of trial treatments rather than one single treatment as in parallel designs, and thereby requires fewer experimental subjects. Further, it reduces variability in treatment comparisons because subjects serve as their own controls and between-subject variations are eliminated.

One distinct feature in crossover designs is that the treatment assigned to subjects may have lasting effects, called carryover effects, on their responses to treatments in subsequent applications. Crossover designs are well-deliberated for its controversy involving the non-orthogonal key parameters of direct and carryover treatments, which leads to completely different experimental designs depending on which is the primary interest. There are several issues that we address in this thesis.

First, when building optimal designs, there are often competing objectives that the investigator desires to optimize. These multiple objectives can include two or more parameters or some functionals, ultimately requiring simultaneous considerations. We revisit the controversy from the point of view of constrained and compound designs for better understanding.

Second, we focus on the construction of optimal designs to that of individual-based designs. Typically, designs were constructed to optimize the average subjects and not ideal in clinical and medical applications. N-of-1 trials are randomized multi-crossover experiments using two or more treatments on a

single patient. They provide evidence and information on an individual patient, thus optimizing the management of the individual's chronic illness. We build one sequence N-of-1 universally optimal designs. We also construct optimal N-of-1 designs for two treatments. Then, we discuss the extension to optimal aggregated N-of-1 designs, which will be optimal for an overall treatment effect.

Third, we extend the response adaptive allocation strategy for continuous responses to construct those for binary responses with the goal of allocating more patients to better treatment sequences without sacrificing much estimation precision. Although design efficiency in terms of mean squared error may drop sharply, increase in allocated patients to the treatment with beneficial effect is evident. We show a balance can be achieved between various competing multiple objectives.

Fourth, we advocate the convex optimization techniques to construct optimal crossover designs where analytic solutions are not feasible. Upon identifying the unique problems and conditions for constructing optimal designs to that of convex optimization problem, we apply them to find optimal designs relatively simply.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Keumhee Chough Carriere, for her continuous support of my Ph.D. study, and for her many helpful and crucial comments and advice on my research. She guided me with her professionalism in teaching, researching and consulting. Without her, I would never be able to become a qualified Ph.D. candidate and complete my dissertation.

I would also like to thank my supervisory and examining committee members Dr. Sunita Vohra, Dr. Bei Jiang, Dr. Peter Minev and Dr. Weng Kee Wong for reading my thesis and providing inspiring and insightful comments.

My sincere thanks also goes to Department of Mathematical and Statistical Sciences and Dr. Giseon Heo, for offering me the teaching assistantship and research assistantship, which gave me financial support during my Ph.D. study.

I wish to thank my parents, Xinyuan Li and Hua Yan. Their constant encouragement and believing in me greatly supported me during my pursuit of the Ph.D. degree. Finally, I thank my wife, Xiaoqing Niu, who is also a great colleague and friend. She is always there standing by me through the good times and bad.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Crossover designs

Designs in which each experimental subject receives a sequence of different exposures or treatments are called crossover designs (CODs) or change-over designs. The use of crossover designs is popular in clinical trials for comparing non-curative treatments for their efficacy.

CODs are composed of several treatment sequences and each subject receives one treatment in each period. For example, in a two-treatment (A and B) two-period crossover design, a subject who is assigned a treatment sequence of AB receives treatment A in the first period and crosses over to treatment B in the second period.

For the designs discussed in this thesis, it is assumed that each subject is treated in the same number of periods, generally denoted by p. The class of all CODs that compares t treatment over p period with a total of n experiment subjects is denoted by COD(t,n,p).

In CODs, the impact of a treatment assigned in a period for a subject's

response is called a treatment effect or a direct treatment effect. The typical objective of crossover studies in clinical trials is estimating the treatment effects. The optimality of the designs has been studied by many researchers in constructing designs which optimize the statistical precision of the estimations (Kiefer 1975; Cheng and Wu 1980; Kershner and Federer 1981; Kunert 1984; Laska and Meisner 1985; Carriere and Reinsel 1992; Kunert and Stufken 2002, 2008).

One advantage of CODs is that as each subject is measured on p (> 1) periods, it requires fewer experiment subjects than designs which only use each experiment unit for one observation in a parallel design. For this reason, CODs have been popular when experiment units are scarce or expensive. The major statistical reason for its popularity is that such designs reduce the variability in treatment comparisons, because subjects serve as their own controls and between-subject variations are eliminated.

There are drawbacks in using CODs as well. The treatments assigned to subjects may have lasting effects on their responses to treatments in subsequent periods. These effects which are carried over from the treatments in previous periods are called carryover or residual treatment effects. It is often assumed that the carryover effects last for only one period. In this case, they are called the first-order carryover effects.

In many cases, one can attempt to remove the carryover effects by including 'washout' periods between the experimental periods. However, including washout periods usually require a longer study duration for the experiments and may raise the risk of subject loss to follow-up. Also, there is no guarantee that washout periods will completely remove the carryover effect, and in certain medical applications there may be ethical issues in the use of washout

periods.

In this thesis, we assume that carryover effects exist and consider statistical models to adjust and remove their effects and to obtain unbiased direct treatment effects.

1.2 Statistical Models

Suppose that d is a design in COD(t,n,p), and the treatment assigned to subject j in the period i is denoted by d(i,j). The response variable is assumed to be continuous. Let Y_{ij} denote the response variable in the ith period from the jth subject. A linear model that is most widely used for these observations is:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \tau_{d(i,j)} + \gamma_{d(i-1,j)} + \epsilon_{ij}. \tag{1.1}$$

In the model, μ denotes the general mean, α_i denotes the ith period effect, β_j denotes the jth fixed subject effect, $\tau_{d(i,j)}$ and $\gamma_{d(i-1,j)}$ denote, respectively, the direct treatment effect due to the treatment on the ith period and the first-order carryover effect due to the treatment on the $(i-1)$th period. To account for the absence of carryover effects in the first period, $\gamma_{d(0,j)}$ is assigned as 0, for all j. The $\epsilon_{ij}^{'}$s are assumed to be independent, have zero mean and constant variance σ^2. This model has traditionally been considered in the literature and is commonly referred to as the *traditional* model.

The traditional model assumes that the carryover effects last for only one period. In the literature, more complex models have also been considered, incorporating higher-order carryover effects (Bose and Mukherjee, 2003). The

subject effects are considered fixed in the traditional model. Models with random subject effects are considered in Saha (1983), Laska and Meisner (1985), Jones, Kunert and Wynn (1992) and Carriere and Reinsel (1993). Also, the traditional model assumes no carryover effects for the observations in the first period. An alternative model which has carryover effects in the first period as well is built by giving subjects a pre-period or baseline period (Kunert, 1984; Afsarinejad, 1988).

It should also be noted that the traditional model assumes that the carryover effects only depend on the treatment assigned for the previous period. This is unrealistic when carryover effects depend on the treatment contributing the direct effect. Taking this into account, Kunert and Stufken (2002) presented a model with self and mixed carryover effects. The self carryover effect occurs when the treatment administered in the current and the previous period are the same; otherwise, we have a mixed carryover effect.

The model with the self and mixed carryover effects is:

$$Y_{ij} = \begin{cases} \mu + \alpha_i + \beta_j + \tau_{d(i,j)} + \gamma_{s,d(i-1,j)} + \epsilon_{ij}, \text{ if d(i,j)=d(i-1,j)} \\ \mu + \alpha_i + \beta_j + \tau_{d(i,j)} + \gamma_{m,d(i-1,j)} + \epsilon_{ij}, \text{ if d(i,j)}\neq\text{d(i-1,j)}, \end{cases} \tag{1.2}$$

where α_i, β_j, $d(i,j)$ and $\tau_{d(i,j)}$, are defined as in model (1.1); $\gamma_{s,d(i-1,j)}$ and $\gamma_{m,d(i-1,j)}$ represent the self and mixed carryover effects of the treatment assigned to the $(i-1)$th period, respectively. In our further discussion, this model is referred to as the *self and mixed carryover* model.

There are other modifications of the traditional model. One among them considers if carryover effects are proportional to the direct treatment effects (Kempton et al., 2001). Another model includes the interaction effects between

the treatment contributing the direct treatment effect and the treatment contributing the carryover effect (Sen and Mukerjee, 1987).

In the traditional model, the assumption on the variances of error terms is that they are independent identically distributed with mean 0 and variance σ^2. This assumption is however not always reasonable. It is sometimes more appropriate to introduce dependence among the error terms. In our discussion, three possible covariance structures are considered for the error terms.

A general form of the covariance structure is:

$$\Sigma = \begin{bmatrix} 1 & \rho & \rho^{1+K} & \rho^{1+2K} & \cdots & \rho^{1+(p-2)K} \\ \rho & 1 & \rho & \rho^{1+K} & \cdots & \rho^{1+(p-3)K} \\ \rho^{1+K} & \rho & 1 & \rho & \cdots & \rho^{1+(p-4)K} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ \rho^{1+(p-2)K} & \rho^{1+(p-3)K} & \rho^{1+(p-4)K} & \rho^{1+(p-5)K} & \cdots & 1 \end{bmatrix}.$$

For $0 \leq \rho < 1$ and $K = 0$, 1. Three types of dependency are:

- Uncorrelated covariance: $\rho = 0$;
- Equal-correlated covariance: $\rho \neq 0$ and $K = 0$;
- Auto-regressive covariance: $\rho \neq 0$ and $K = 1$.

1.3 Information matrix

Fisher information measures the amount of information carried by an observable random variable Y for the parameters, which the distribution of the Y is dependent on. Let θ denote the vector of the parameters and $f(Y;\theta)$ denote

the probability density function. Then, the fisher information is defined as,

$$I(\theta) = E\left[(\frac{\partial}{\partial\theta} log f(Y;\theta))^2 \middle| \theta \right], \tag{1.3}$$

or equivalently, when the second derivative of $log f(Y;\theta)$ exists, as

$$I(\theta) = E\left[-\frac{\partial^2}{\partial\theta^2} log f(Y;\theta) \middle| \theta \right]. \tag{1.4}$$

The Fisher information is widely used in optimal experiment designs. Since it is reciprocal to estimator-variance, maximizing the information corresponds to minimizing the variance.

Both of the traditional model and the self and mixed carryover effect model belong to a class of models which can be written in a general form:

$$y = X\theta + \epsilon, \tag{1.5}$$

where y denotes the vector of random variables that are assumed with normal distribution, X denotes the design matrix, and ϵ denotes the vector of errors that have zero means and a general variance matrix $\sigma^2\Sigma$. The variance matrix Σ is positive definite, so that $\Sigma^{-\frac{1}{2}}$ exists. Applying $\Sigma^{-\frac{1}{2}}$ to the both side of the model (1.5), we have

$$\Sigma^{-\frac{1}{2}} y = \Sigma^{-\frac{1}{2}} X\theta + \Sigma^{-\frac{1}{2}}\epsilon. \tag{1.6}$$

Since $E(\Sigma^{-\frac{1}{2}}\epsilon) = 0$ and $Var(\Sigma^{-\frac{1}{2}}\epsilon) = \Sigma^{-\frac{1}{2}}\Sigma\Sigma^{-\frac{1}{2}} = \sigma^2 I$, the elements of $\Sigma^{-\frac{1}{2}} y$ are independently, identically distributed with mean zero and a constant variance, σ^2.

Let $w = \Sigma^{-\frac{1}{2}} y$ and $Z = \Sigma^{-\frac{1}{2}} X$, then the log likelihood function can be written as,

$$l(\theta|y) = -\frac{1}{2} log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(w - Z\theta)^T(w - Z\theta) \tag{1.7}$$

According to the second definition of the Fisher information (1.3), we have

$$\begin{aligned}
I(\theta) &= E\left[-\frac{\partial^2}{\partial\theta^2}[-\frac{1}{2} log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(w - Z\theta)^T(w - Z\theta)] \middle| \theta \right] \\
&= E\left[-\frac{\partial^2}{\partial\theta^2}[-\frac{1}{2\sigma^2}(w^T w - \theta^T Z^T w - w Z\theta^T + \theta^T Z^T Z\theta)] \middle| \theta \right] \\
&= E\left[-\frac{\partial^2}{\partial\theta^2}[-\frac{1}{2\sigma^2}\theta^T Z^T Z\theta] \middle| \theta \right] \\
&= \frac{1}{\sigma^2} Z^T Z \\
&= \frac{1}{\sigma^2} X^T \Sigma^{-1} X.
\end{aligned}$$

When independent errors are assumed, i.e. $\Sigma = I$, the information matrix for estimating the parameter θ is reduced to $X^T X$.

In order to minimize the estimator-variance, one can maximize the information matrix $I(\theta)$ by applying optimality criteria. Taking some commonly applied optimality criteria for examples, A-optimality seeks to minimize the trace of $I^{-1}(\theta)$; D-optimality seeks to maximize the determinant: $|I(\theta)|$; and E-optimality maximizes the minimum eigenvalue of $I(\theta)$. In the next section, we show some optimality results in crossover designs.

1.4 Some optimality results in crossover designs

Kiefer (1975) conducted a systematic study of various optimality criteria (including A-, D- and E-optimality criteria) and related optimal designs. Kiefer introduced the notion of universal optimality. If a design is universally optimal, then it also satisfies the three criteria of A-, D- and E-optimality. Then sufficient conditions were provided on the information matrix, C, of the parameters of interest, which ensure the design to be universally optimal. The conditions are: (i) C is completely symmetric, which requires that all diagonal elements are equal, all off-diagonal elements are equal and all row sums are zero; (ii) trace of C which is the sum of diagonal elements (often denoted by $trace(C)$) is maximal on the class of designs over which the universal optimality is claimed. The conditions have been used by many researchers in determining the universal optimality of their crossover designs (Carriere and Reinsel, 1993).

Hedayat and Afsarinejad (1978) used the conditions to derive the universal optimality over uniform designs (also see Kunert 1984).

A **uniform design** is defined as a design in which each treatment is assigned equally often (uniform on periods) and each treatment appears in the same number of periods (uniform on subjects). Therefore, the uniform design requires that $n = \lambda_1 t$ and $p = \lambda_2 t$ for some integers λ_1,λ_2.

In the study of Hedayat and Afsarinejad (1978), λ_2 was set to 1 and the subclass of the designs could be denoted by COD(t,λt,t), for some integer λ. Their main result was that balanced uniform designs are universally optimal for estimation of direct treatment effects, and also for the estimation of first-

order carryover effects, over the uniform designs in COD(t,λt,t).

A **balanced design** is defined as a design in which each treatment is preceded equally often by every other treatment, but never by itself. Balanced uniform designs were found to exist in COD(t,t,t) when t is even, and in COD(t,$2t$,t) when t is odd. Cheng and Wu (1980) produced some extensions of Hedayat and Afsarinejad's work. The findings include:

1. If a design is constructed from a balanced uniform design in COD(t,λt,t) by repeating the last period to obtain an extra period, then it is universally optimal for the estimation of direct and carryover effects, over COD(t,λt,$t+1$);

2. Strongly balanced designs are universally optimal for the estimation of direct and carryover effects, over COD(t,n,p).

A **strongly balanced design** is defined as a design in which each treatment is preceded equally often by every treatment, including itself.

The existence of strongly balanced designs requires $n = \lambda_1 t^2$ for integer λ_1 and $p = \lambda_2 t$ for integer $\lambda_2 > 1$. If λ_2 is even, then such designs must exist.

Other authors continued the study of balanced uniform designs and more optimality results can be found in Hedayat and Yang (2003, 2004), Bate and Jones (2008).

There also is literature which generates the optimality results for COD with just two treatments (say A and B). A universally optimal design for the estimation of direct treatment effects over COD(2,n,p) is the design that minimizes the variance of the direct treatment effect contrast (often defined as $(\tau_A - \tau_B)/2$). Similarly, if carryover effects are of interest, a universally

optimal design would minimize the variance of the carryover effect contrast (often defined as $(\gamma_A - \gamma_B)/2$).

Laska et al. (1983) pointed out that optimal designs for the estimation of the direct or carryover effects can be found immediately. If p is even, a strongly balanced uniform design is optimal; if p is odd, an extra-period design is optimal. Under model (1.1), the optimal designs for $p = 2$, 3 and 4 are respectively, (i) AB, BA, AA, BB; (ii) ABB, BAA; and (iii) $AABB$, $BBAA$, $ABBA$, $BAAB$.

The optimality and efficiency of two-treatment CODs have also been studied under various modifications of the traditional model. Kershner and Federer (1981) studied a model with sequence effects instead of subject effects. The authors calculated the variance of estimators for direct and carryover effect contrasts in a number of two-treatment CODs. Laska and Meisner (1985) assumed random subject effects and generated optimal designs. The model in their study was with equal-correlated covariance structure for general p and with auto-regressive covariance structure for $p = 2$,3 and 4. Matthews (1987) studied optimal and efficient designs in models with auto-regressive covariance structure. Results for general p were given by Matthews (1990) under the uncorrelated errors, and by Kunert (1991) under the autoregressive errors.

Carriere and Reinsel (1992) assumed a model with sequence effects and random subject effects to obtain optimal and efficient two-treatment CODs. The generalized least square estimator for both direct treatment effect and carryover effect was calculated and design efficiency was discussed in two-period four-sequence designs with unequal allocation. The authors remarked that with or without baseline measurements, the use of the two-period four-sequence design with 80% of the subjects assigned to AB and BA treatment

sequence and the rest of subjects to AA and BB treatment sequences was recommended.

Other optimality results can be found in Laska, Meisner and Kushner (1983), Ebbutt (1984) and Carriere and Huang (2000).

1.5 Thesis Overview

In chapter 2, we discuss the use of two equivalent approaches in constructing multiple objectives crossover designs. We apply the approaches to obtain optimal and efficient designs in two-treatment two-period crossover designs, in which both the efficiency of estimating the direct treatment effect and the carryover effect are the objectives of the design construction. Assuming the traditional model with an equal-correlated covariance, the four-sequence design with equal allocation is optimal for estimating the treatment effects, yet is not optimal for estimating carryover effects. In this scenario, there would be no single design that is optimal for both types of effects. Two common options in constructing an optimal design for this two-objective crossover experiment are constrained optimal designs and compound optimal designs. A constrained optimal design is a design, which optimizes the secondary objective while satisfying a constraint on the primary objective. In this scenario, the constraint on the primary objective can be a minimum precision requirement in estimating the treatment effects and the secondary objective can be the precision in estimating the carryover effects. A compound optimal design is a design optimizing an objective, which is compound from the two original objectives. One such example of this kind of objectives can be found in attempting to improve the estimation efficiency of a linear combination of

treatment and carryover effects. Optimal designs constructed from each of the two approaches are presented in this chapter and the equivalence of the two approaches is also discussed.

Chapter 3 is the first of two chapters that focus on building optimal designs for N-of-1 trials. In patient based or evidence based clinical medicine, N-of-1 trials, also known as multi-crossover single-patient trials, are often employed when there are concerns about making the best possible treatment decision for an individual patient. Most randomized controlled trials are focused on optimizing the treatment effect for an average patient. However, individuals enrolled in a trial may be better or worse than the average patient, and the available optimal designs are not capable of offering such individual-based treatment decisions. The simplest two-treatment N-of-1 trial uses the AB (or BA) sequence as a within-patient comparison. With the rising cost of patient care, N-of-1 trials have the potential to be extremely useful, as it can minimize clinic visits and time spent on suboptimal treatments. In this chapter, we present universal optimality results in N-of-1 trials with $t \geq 2$ treatments. A sufficient condition of universally optimal designs is given for N-of-1 trials with general t and the design results are presented.

Continued discussion in optimal 2-treatment N-of-1 trials is in chapter 4. Straight application of the two-treatment optimal designs in literature with A to AB and B to BA suggests that optimal N-of-1 trials would need to use $ABBA$, $ABAB$ and their duals for two within-patient comparisons, $ABBABA$ and its dual for three within-patient comparisons, and the sequences $ABBABAAB$, $ABABBABA$ and their duals for four within-patient comparisons. It is not yet known whether all of these sequences are indeed optimal and required for 4, 6 and 8-period N-of-1 trials. Further, it would

require at least 2 patients to utilize these existing designs. In this section, we prove that these designs are not optimal for N-of-1 trials for estimating individual-based treatment effects. Optimal N-of-1 designs are constructed for two treatments under a variety of conditions about the carryover effects, the covariance structure, and the number of planned periods. Extension to optimal aggregated N-of-1 designs is also discussed.

In chapter 5, we investigate the applicability of the adaptive allocation strategy for two-treatment RMDs with binary responses. The utility of the proposed multiple-objective response-adaptive repeated measurement designs will be demonstrated on several practically useful designs with two, three, and four periods. There are various types of response-adaptive designs, depending on the goals of a particular study. In a typical case, the adaptive treatment allocation is used to fulfill a single objective such as increasing the number of patients assigned to the better treatment group, reducing the sample size in a trial, or increasing the estimation precision of a treatment effect.

Recently, Liang and Carriere (2009) developed a new treatment allocation scheme to construct multiple-objective response-adaptive repeated measurement designs (RMD) for continuous responses, where study subjects can receive two or more treatments (not necessarily the same treatments) over a period of time. Their adaptive allocation strategy can simultaneously achieve two objectives: potentially preventing patients from being exposed to inferior treatments and enhancing the precision of the estimates of parameters. In this chapter, we investigate the applicability of the same adaptive allocation strategy for two-treatment RMDs with binary responses. Through simulations, the utility of the new proposed multiple-objective response-adaptive RMDs will be demonstrated on several practically useful designs with two, three, and four

periods.

In Chapters 2-5, we experienced that in many situations, a closed form of optimal designs was often not possible. In Chapter 6, we investigate an alternate strategy advocating the convex optimization solutions for possible resolution in such cases. We show that upon specifying certain conditions, we could represent the problem of constructing optimal designs to that of optimizing convex functions. We give examples of selected problems where we could not analytically determine the optimal designs in earlier chapters, and show how easily the convex optimization strategy we outlined could give solutions relatively quickly and simply.

In Chapter 7, we give the summary of the main contribution of the thesis and suggestions for future research.

Chapter 2

Constrained and compound optimal designs in two-treatment two-period crossover trials

Abstract

When building optimal designs, there are often competing objectives that the investigator desires to optimize. These multiple objectives can include two or more parameters or some functionals, ultimately involving simultaneous considerations. Crossover designs are well known for their controversy involving non-orthogonal key parameters of direct and residual or crossover treatment effects, which can lead to completely different experimental designs depending on the primary parameter. We revisit this controversy from the point of view of constrained and compound designs. Cook and Wong (1994) showed that constrained and compound optimal designs are basically equivalent. In this chapter, we provide an alternate simpler proof for crossover designs.

2.1 Introduction

In crossover trials, two-treatment two-period designs are popular by clinicians. With A and B denoting the two treatments (or a treatment and a placebo), the possible treatment sequences in a two-period design are AB, BA, AA and BB.

When carryover effects are absent, the optimal design for estimating direct treatment effects is the two-sequence design AB and BA with equal allocation (Laska, Meisner and Kushner 1983). As considered to be optimal, this design provides unbiased estimators for direct treatment effects with the minimum variance. When carryover effects are present, the two-sequence design cannot

provide an unbiased estimation of direct treatment effects. Instead, the four-sequence design AB, BA, AA and BB with equal allocation is the optimal design for direct treatment effects (Laska and Meisner 1985; Laska, Meisner, and Kushner 1983; Kershner and Federer 1981).

The traditional model assumes that the variances of error terms are independent identically distributed. Under the traditional model, the four-sequence design is also optimal for carryover effects (Laska and Meisner 1985). However, in many situations it is realistic to introduce dependence among the error terms. We shall assume the error terms are equal-correlated.

It can be shown that the optimal design for the direct treatment effects and that for the carryover treatment effects are different. In other words, the efficiencies of estimating these two type of effects cannot always be optimized using the same design.

Assuming the precision of estimation of carryover effects is considered equally important as the direct treatment effects, we aim to obtain optimal designs to improve both the estimation of direct treatment effects and carryover effects. Accordingly, we investigate the issues of constrained and compound designs.

2.2 Optimal two-treatment two-period crossover design

In this chapter, the traditional model with random subject effects is used in constructing the optimal two-treatment crossover designs. We define the contrast of treatment effects as $\tau = (\tau_A - \tau_B)/2$ and the contrast of carryover

effects as $\gamma = (\gamma_A - \gamma_B)/2$. Then the traditional model can be written as

$$Y_{ij} = \mu + \alpha_i + \Phi_{d(i,j)} \cdot \tau + \Phi_{d(i-1,j)} \cdot \gamma + \xi_j + \epsilon_{ij}, \tag{2.1}$$

where

$$\Phi_{d(i,j)} = \begin{cases} 1, \text{ if } d(i,j) = A \\ -1, \text{ if } d(i,j) = B \end{cases}$$

and

$$\Phi_{d(i-1,j)} = \begin{cases} 0, \text{ if } i = 1 \\ 1, \text{ if } i \neq 1 \text{ and } d(i,j) = A \\ -1, \text{ if } i \neq 1 \text{ and } d(i,j) = B. \end{cases}$$

The term ξ_j denotes the random subject effect with zero mean and constant variance σ_ξ^2 and ϵ_{ij} denotes the random error with zero mean and constant variance σ_ϵ^2. The covariance matrix of the vector $Y_j = (Y_{1j}, Y_{2j}, \cdots Y_{pj})'$ is

$$C = \sigma_\epsilon^2 \mathbf{I}_p + \sigma_\xi^2 \mathbf{1}_p \mathbf{1}_p^T$$

and the correlation between $y_{i,j}$ and $y_{i',j}$ with $i \neq i'$, called the within-subject correlation, is

$$\rho = \frac{\sigma_\xi^2}{\sigma_\epsilon^2 + \sigma_\xi^2}.$$

For two sequences assigned to j and j', if $\Phi_{d(i,j)} = -\Phi_{d(i,j')}$ holds for all $i = 1, \cdots, p$, then the two sequences are called dual sequences. The designs are defined as **dual-balanced** if the designs allocate equal numbers of subjects to the sequences and their dual sequences. In finding the optimal designs in

two-treatment crossover trials, duality is a useful concept. Laska and Meisner (1985) applied the Lagrange Multiplier solution to the two-treatment optimal design problem. Optimal designs under various situations were presented and the variances of the Best Linear Unbiased Estimators (BLUE) were calculated for the direct treatment effect contrast and carryover effect contrast. The authors also proved that if an optimal design exists then a dual-balanced design can be constructed upon it and will not be worse. This finding halved the number of sequences to determine optimality upon the BLUE, as any two dual sequences have the same weights and only one of them is needed to be considered.

Consider a dual-balanced design d in COD(2,n,2), which assigns m subjects to the treatment sequence AB. The number of subjects assigned to BA, AA and BB are m, $\frac{n}{2}-m$, and $\frac{n}{2}-m$, respectively. Using the Lagrange multiplier approach, Laska and Meisner (1985) obtained the variance of the BLUE of τ and γ in a two-treatment two-period design.

$$var(\hat{\tau}_d) = \frac{n(\rho+1)}{n^2(2-\rho^2)-(n-4m)^2},$$
$$var(\hat{\gamma}_d) = \frac{2(4m\rho+n-n\rho)(\rho+1)}{n^2(2-\rho^2)-(n-4m)^2}$$

It can be shown that the minimum of $var(\hat{\tau}_d)$ equals to $(1+\rho)/[n(2-\rho^2)]$ and the optimal design for τ is the design AB, BA, AA and BB with equal allocation $(n/4)$ to each sequence. The minimum of $var(\hat{\gamma}_d)$ equals to $(1+\rho)/n$ and is achieved at $m = (1-\rho)n/4$. It can be seen that the optimal design for γ depends on the value of ρ. When $\rho = 0$, which implies independent errors, the optimal design is the sequences AB, BA, AA and BB with equal allocation $(n/4)$ to each sequence. When $\rho \to 1$, the optimal design for estimation of γ is AA, BB with $n/2$ subjects allocated to each sequence (also see Kershner

and Federer 1981).

The efficiency of a design d can be defined, in terms of the variance of estimation of a parameter of interest compared to the optimal design, as

$$Eff_\theta(d) = var(\hat{\theta}_{d^*_\theta})/var(\hat{\theta}_d),$$

where d^* denotes the optimal design for estimation of θ. The value of the efficiency is between 0 and 1. For an optimal design d, the efficiency equals to 1.

For a design d, the efficiency of the estimation of τ and the efficiency of the estimation of γ are given as

$$Eff_\tau(d) = \frac{n^2(2-\rho^2) - (n-4m)^2}{n^2(2-\rho^2)} \tag{2.2}$$

$$Eff_\gamma(d) = \frac{n^2(2-\rho^2) - (n-4m)^2}{2n(4m\rho + n - n\rho)(\rho+1)} \tag{2.3}$$

One can directly show that the optimal design for estimating τ has an efficiency of $(2-\rho^2)/2$ for estimation of γ, and conversely the optimal design for estimating γ has an efficiency of $(2-2\rho^2)/(2-\rho^2)$ for estimation of τ. Therefore, when ρ is large, the optimal designs for estimating τ and γ can be different. When ρ gets close to 1, the optimal design for τ can only estimate γ with 50% efficiency.

In situations when estimating both direct treatment effects and carryover treatment effects is of interest, a design balancing the efficiencies of estimation of τ and γ is appropriate. In the next two sections, we will use two approaches to resolve the design issues, which aim to optimize for 2 parameters simulta-

neously.

2.3 Constrained optimal designs in COD(2,n,2)

Consider the primary interest or objective is to find a design optimizing the estimation of direct treatment effect contrast τ and the secondary interest is to optimize the estimation of the carryover effect contrast γ. A constrained optimal design is a design, which optimizes the secondary objective while satisfying a constraint on the primary objective (Clyde and Chaloner, 1996; Lee, 1988; Mandal, Torsney and Carriere, 2005).

The Constrained Optimal Design Problem:

Maximize $Eff_\gamma(d)$ *, while satisfying the constraint* $Eff_\tau(d) \geq c$, $c \in [0,1]$.

Theorem 2.3.1: *Under model (2.1), the optimal design for the constrained optimal design problem is d with*

$$\begin{cases} m = \frac{1}{4}n(1-\rho), & \text{if } c \leq 1 - \frac{\rho^2}{2-\rho^2}; \\ m = \frac{n}{4} - \frac{n}{4}\sqrt{(1-c)(2-\rho^2)}, & \text{if } c > 1 - \frac{\rho^2}{2-\rho^2}. \end{cases}$$

Proof:

Let $x = 1 - 4m/n$ for the convenience of the proof. As m can take values in $[0, n/2]$, the range of x is $[-1, 1]$. The efficiencies of the estimation of τ and γ are functions of x. The constraint on the efficiency of the estimation of τ can be written as,

$$Eff_\tau(d) = \frac{n^2(2-\rho^2) - n^2x^2}{n^2(2-\rho^2)} \geq c \tag{2.4}$$

By solving the inequality (2.4), the condition on x can be obtained to satisfy the about constraint,

$$-\sqrt{(1-c)(2-\rho^2)} \leq x \leq \sqrt{(1-c)(2-\rho^2)}.$$

Then we optimize $Eff_\gamma(d)$ with the above condition. The efficiency of the estimation of γ can be written as

$$\begin{aligned} Eff_\gamma(d) &= \frac{(1+\rho)/n \cdot [n^2(2-\rho^2) - n^2x^2]}{2(n-n\rho x))(\rho+1)} \\ &= \frac{2-\rho^2-x^2}{2(1-\rho x)} \end{aligned}$$

To find the maximum of $Eff_\gamma(d)$, we note its derivatives with respect to x are

$$(Eff_\gamma(d))'_x = \frac{(x-\rho)(\rho x - 2 + \rho^2)}{(1-\rho x)^2} \tag{2.5}$$

$$(Eff_\gamma(d))''_x = -\frac{2(\rho^2-1)^2}{(1-\rho x)^3} \tag{2.6}$$

Letting the first derivative in (2.5) equal to zero, we can find that the maximum of $Eff_\gamma(d)$ is obtained when $x = \rho$ or $x = (2-\rho^2)/\rho$. Recall that the value of x is within the range (-1, 1) and that of ρ is within (0, 1). Therefore, $x = (2-\rho^2)/\rho$ holds and be valid only when $x = \rho$.

It is easy to verify that the second derivative in (2.6) is negative for any $\rho \in (0, 1)$, indicating the $Eff_\gamma(d)$ has a maximum value and it is monotonically increasing on $x \in (0, \rho)$. Therefore, when the value of ρ is within the interval

$$(-\sqrt{(1-c)(2-\rho^2)}, \sqrt{(1-c)(2-\rho^2)}),$$

the optimal design for γ, with the constraint on the estimation of τ, is obtained at $x = \rho$. When the value of ρ is outside the interval, the optimal constrained design for γ is obtained at the positive boundary of the interval $x = \sqrt{(1-c)(2-\rho^2)}$. That is,

$$x = \rho, \text{ if } c \leq 1 - \frac{\rho^2}{2-\rho^2}; \tag{2.7}$$

$$x = \sqrt{(1-c)(2-\rho^2)}, \text{ if } c > 1 - \frac{\rho^2}{2-\rho^2}. \tag{2.8}$$

By replacing x by $1 - 4m/n$, we obtain the results as given in Theorem 2.3.1. □

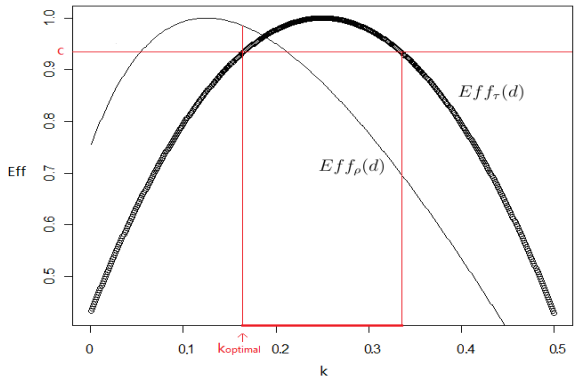

Figure 2.1: Efficiency of estimation of direct and carryover effects for the constrained optimal design problem. Note that $k = m/n$, the within-subject correlation is set to $\rho = 0.5$ and the constraint is set to $c = 0.93$.

Figure (2.1) shows the efficiencies in estimating of the direct and carryover effects with respect to $k = m/n$. The horizontal line $Eff = c$ joins the curve of $Eff_\tau(d)$ and the interval bounded by the two intersections contains all the values for k satisfying the constraint. By maximizing the $Eff_\gamma(d)$ within this interval, the optimal constrained design is determined.

Table (2.1) shows the results of the constrained optimal designs under

Table 2.1: Efficiencies of the estimation of the direct treatment effects and the carryover effects for the constrained optimal design.

ρ	c	$\frac{1-\rho^2}{2-\rho^2}$	x	$Eff_\tau(d)$	$Eff_\gamma(d)$	AB (%)	AA(%)
0.3	0.7	0.953	0.300	95.3%	100%	17.5%	32.5%
0.3	0.8	0.953	0.300	95.3%	100%	17.5%	32.5%
0.3	0.9	0.953	0.300	95.3%	100%	17.5%	32.5%
0.3	0.95	0.953	0.300	95.3%	100%	17.5%	32.5%
0.3	0.99	0.953	0.138	99%	0.986%	17.5%	32.5%
0.5	0.7	0.857	0.500	85.7%	100%	12.5%	37.5%
0.5	0.8	0.857	0.500	85.7%	100%	12.5%	37.5%
0.5	0.9	0.857	0.418	90.0%	99.6%	14.5%	35.5%
0.5	0.95	0.857	0.296	95.0%	97.6%	17.6%	32.4%
0.5	0.99	0.857	0.132	99.0%	92.8%	14.5%	35.5%
0.7	0.7	0.675	0.673	70.0%	99.9%	8.2%	41.8%
0.7	0.8	0.675	0.550	80.0%	98.2%	11.3%	38.7%
0.7	0.9	0.675	0.389	90.0%	93.3%	15.3%	34.7%
0.7	0.95	0.675	0.275	95.0%	88.9%	18.1%	31.9%
0.7	0.99	0.675	0.123	99.0%	81.8%	21.9%	28.1%
0.9	0.7	0.319	0.597	70.0%	90.1%	10.1%	39.9%
0.9	0.8	0.319	0.488	80.0%	84.9%	12.8%	37.2%
0.9	0.9	0.319	0.345	90.0%	77.7%	16.4%	33.6%
0.9	0.95	0.319	0.244	95.0%	72.4%	18.9%	31.1%
0.9	0.99	0.319	0.109	99.0%	65.3%	22.3%	27.7%

various settings of ρ and c. For each setting, the value of x is calculated and so are the efficiencies of the estimation of both of the effects. Proportions of the sequences AB and AA are provided according to the value of x. It can be observed that when $\rho = 0.3$ or 0.5, the designs are efficient on the estimation of both of the effects. When $\rho = 0.7$ or 0.9, the efficiency of the estimation of the carryover effects would decline dramatically when the constraint requires higher efficiency of the estimation of the direct treatment effects. This result is particularly useful, when there is evidence of large within-subject correlation, to prevent potentially poor estimation of the carryover effects.

2.4 Compound optimal designs in COD(2,n,2)

A compound optimal design is a design optimizing an objective, which is compound from the two original objectives. The compound objective can be

a linear combination of the efficiency in estimating treatment effects and the efficiency in estimating carryover effects.

The Compound Optimal Design Problem:

Maximize $\eta Eff_\tau(d) + (1-\eta)Eff_\gamma(d)$ *, where* $\eta \in [0,1]$.

Theorem 2.4.1: *Under model 2.1, the optimal design for the compound optimal problem is d with* $x = 1 - 4m/n$ *and* x *is the real root of the following equation,*

$$x^3 - (\frac{4}{\rho} + \frac{2-\rho^2}{2\rho}\frac{1-\eta}{\eta})x^2 + (\frac{4}{\rho^2} + \frac{4-2\rho^2}{\rho^2}\frac{1-\eta}{\eta})x - \frac{(2-\rho^2)^2}{2\rho}\frac{1-\eta}{\eta} = 0 \quad (2.9)$$

Proof:

Replacing $(1-4m/n)$ by x in (2.2) and (2.3), the linear combination of the efficiencies in the compound optimal design problem can be written as

$$\eta Eff_\tau(d) + (1-\eta)Eff_\gamma(d) = \eta(1 - \frac{x^2}{2-\rho^2}) + (1-\eta)\frac{2-\rho^2-x^2}{2(1-\rho x)}. \quad (2.10)$$

To find the maximum of the linear combination of efficiencies, we set the first derivative of the right-hand side of (2.10) with respect to x to zero. The result is equation (2.9) and solving the equation generates the design maximizing the combination of the efficiencies in the compound optimal design problem. □

As an example, suppose the within-subject correlation is set to $\rho = 0.5$ and the weight in the linear combination of the efficiencies is set to $\eta = 0.5$. The optimal design can be obtained by solving (2.9), and the optimal design is $x = 0.19$, which implies $m \approx 0.20n$.

2.5 Equivalence of the two approaches

To discuss equivalence of the constrained and compound optimal crossover designs in our example, we can compare the optimal designs discussed in previous sections. It can be seen from Figures 1 and 2 that all optimal designs for either of the problems have k in $[\frac{1-\rho}{4}, \frac{1}{4}]$; and each k within the interval is an optimal design for a certain case of the constrained optimal design problem and a certain case of the compound optimal design problem.

In fact, for any $k \in (\frac{1-\rho}{4}, \frac{1}{4}]$, we observe that there is a unique c in the constrained optimal design problem that corresponds to a unique η in the compound optimal design, and conversely. For $k = \frac{1-\rho}{4}$, we have $c \leq 1 - \frac{\rho^2}{2-\rho^2}$ in the constrained optimal design problem and $\eta = 0$ in the compound optimal design problem. In other words, for every compound optimal design problem, there is a constrained optimal design problem with the same solution, and conversely.

A proof of the equivalence of the constraint optimal design and the compound optimal design was given by Cook and Wong (1994) when the criteria are convex and the design regression problem is defined on a compact convex design space. In this section, we provided a simpler proof for the equivalence of the constrained optimal design problem and the compound optimal design problem defined in this chapter.

Theorem 2.5.1: *Under model 2.1, the constrained optimal design problem and the compound optimal design problem are equivalent. That is, for any design d which is optimal in the constrained optimal design problem, there exists a compound optimal design problem in which the optimum is achieved with the same design, and vice versa.*

Proof:

We recall that we let $x = 1 - 4m/n$. First, we prove that for any x within $(0, \rho)$ is a solution to the constrained optimal design problem for some c and a solution in the compound optimal design problem for some η.

Consider $x_* \in (0, \rho)$. According to 2.8, x_* is the solution in a constrained optimal design with a certain c_* where $c_* > 1 - \frac{\rho^2}{2-\rho^2}$ and $x_* = \sqrt{(1-c_*)(2-\rho^2)}$.

Now, let $L_x(\eta)$ denote the first derivative of the linear combination of the efficiencies in the compound optimal design problem, with respect to η for any x. We have,

$$L_x(\eta) = [\eta Eff_\tau(d) + (1-\eta)Eff_\gamma(d)]'_x \tag{2.11}$$

$$= [\eta(1 - \frac{x^2}{2-\rho^2}) + (1-\eta)\frac{2-\rho^2-x^2}{2(1-\rho x)}]'_x \tag{2.12}$$

$$= \eta\frac{-2x}{2-\rho^2} + (1-\eta)\frac{\rho x^2 - 2x + 2\rho - \rho^3}{2(1-\rho x)^2}. \tag{2.13}$$

Let $x = x_*$, then $L_{x=x_*}(\eta)$ is a continuous function of η on [0,1]. It can be written as

$$L_{x_*}(\eta) = \eta\frac{-2x_*}{2-\rho^2} + (1-\eta)\frac{\rho x_*^2 - 2x_* + 2\rho - \rho^3}{2(1-\rho x_*)^2}. \tag{2.14}$$

It is easy to verify that $L_{x_*}(0) > 0$, as

$$L_{x_*}(0) = \frac{\rho x^2 - 2x + 2\rho - \rho^3}{2(1-\rho x)^2} = \frac{(1-\rho x)^2 - (1-\rho^2)^2}{2\rho(1-\rho x)^2} \tag{2.15}$$

and $L_{x_*}(1) < 0$, as

$$L_{x_*}(1) = \frac{-2x_*}{2-\rho^2}. \tag{2.16}$$

So, there exists an $\eta_* \in (0,1)$ such that $L_{x_*}(\eta_*) = 0$, meaning the x_* is a solution in the compound optimal design problem when $\eta = \eta_*$.

Therefore, we proved that any $x_* \in (0, \rho)$ is the solution in the constrained optimal design problem with $c = c_*$ and is, at the same time, the solution in the compound optimal design problem with $\eta = \eta_*$.

Next, we prove that for $x \leq 0$ and $x \geq \rho$, the theorem also holds. Actually, when $x \leq 0$, it is the constrained optimal design when $c = 1$ and the compound optimal design when $\eta = 1$; when $x \geq \rho$, it is the constrained optimal design when $c \leq 1 - \frac{\rho^2}{2-\rho^2}$ and the compound optimal design when $\eta = 0$. □

2.6 Discussion

In two-treatment two-period crossover designs, it is known that the four-sequence design AB, BA, AA and BB with equal allocation is optimal for the estimation of the direct treatment effects. When the carryover effect is present, this design is also optimal for the estimation of the carryover effect under the assumption of independent errors. However, in cases where the equal-correlated errors are assumed, the optimality of the estimation of the carryover effects depends on the within-subject correlation, which is defined as $\rho = \frac{\sigma_\xi^2}{\sigma_\epsilon^2+\sigma_\xi^2}$. It is found that the optimal design for the estimation of the carryover effects is the four sequence design which assigns $n(1-\rho)/4$ subjects to each of AB and BA and $n(1+\rho)/4$ subjects to each of AA and BB.

In applications when the estimation for both of the direct treatment effects and the carryover effects are of interest, the design can be chosen with consideration of the efficiency of both types of the effects. With small within-subject correlation (e.g. 0.3 and 0.5), the optimal design for the direct treatment effects can produce high efficiency for estimating the carryover effects When the within-subject correlation is large (e.g. 0.7 and 0.9), the optimal design for the direct treatment effects can perform poorly in estimating the carryover effects, and an alternative design which balances the efficiencies of estimating both of the effects are suggested.

In recent survey of the articles in Cochran's database (which is the leading resource for systematic reviews in health care) that utilized crossover trials, we observed 59 out of 198 studies still analyzed the data, discarding the data from the second periods mainly because the investigator planned the trial without regard to the possible carryover effects but found the data contaminated with lingering residual effects from the previous treatment effects.

Carriere (1992, 1993, 1994) discussed extensions including having more periods and also allocating an unequal number of subjects to each sequence (i.e. less to clinically undesirable sequences). However, the gap between clinicians and statisticians in deciding optimal clinical designs has not narrowed and merged to date.

In this chapter, we applied two methods to find a design that is optimal for the combined objectives of estimating both effects. We defined the constrained optimal design problem and the compound optimal design problem, and showed the produced designs for the two problems. A proof was provided to show that the two methods are equivalent.

The results proposed in this chapter provide some guidance of clinical trials

for designing in two-treatment two-period crossover experiments. When there is evidence of large within-subject correlation, the conventional four-sequence design with equal allocation may not be appropriate as it is deficient in the estimation of the carryover effects. Planning with a constrained or compound optimal design could reduce the risk of failure in clinical crossover experiments due to a potentially poor estimation on the carryover effects.

Chapter 3

Universally optimal N-of-1 designs

Abstract

The purpose of this chapter is to build universally optimal N-of-1 designs. Originally, Kiefer (1975) proposed the concept of universal optimality with zero row and column sums in the information matrices. We examine special conditions when such universally optimal designs exist. We construct universally optimal N-of-1 designs for $t \geq 2$. Special consideration is given to situations when the universal optimal designs do not exist, which occurs when $t > 2$. The one sequence N-of-1 universal optimal designs would have the number of periods to be a multiple of t when $t = 2$, while it would be a multiple of t plus 1 when $t > 2$. Recognizing possible practical difficulty in adopting the designs constructed, we also suggest N-of-1 designs with block sizes less than t.

3.1 Introduction

Multi-crossover single-patient trials, known as N-of-1 trials, are often employed when there are concerns about making the best possible treatment decision for an individual patient. Most randomized controlled trials are focused on optimizing the treatment effect for an average patient. However, individuals enrolled in a trial may be better or worse than the average patient, and the available optimal designs built to optimize treatment on average are not capable of offering such individual-based treatment effects. The simplest two-treatment N-of-1 trial uses the AB (or BA) sequence for treatments A and B. The sequence has one crossover pair over two periods. As the patient becomes

his or her own control, the N-of-1 trials provide an individual-based clinical evidence for the treatment effect, free of between-patient variations. With rising cost of patient care, N-of-1 trials have the potential to be extremely useful, as it can minimize clinic visits and time on suboptimal treatments (Greenfield et al., 2007; Kravitz, Duan and Braslow, 2004; Kravitz and Duan, 2014; Larson, 1990; Guyatt et al., 1986, 1990). When suitably planned, N-of-1 trials can improve health outcomes, cut clinical costs for certain drugs and significantly reduce health care costs (Nikles et al., 2005; Edgington, 1984; Cochrane, 1972). However, the literature is lacking in providing guidelines for optimal N-of-1 trials.

These repeated measurement designs, although practically appealing, suffer from a long-standing controversy regarding residual treatment effects. N-of-1 trials are no exception. Sometimes referred to as the carryover effect, the residual effect is the effect of a previous treatment that carries over into the subsequent treatment periods. A washout period placed between treatment periods could reduce the carryover effects, but a long washout period may increase the risk of subject loss to follow-up among other issues. Also, there is no guarantee that it completely removes the residual effects. Therefore, careful planning is important (Bose and Mukherjee, 2003; Kunert and Stufken, 2002; Rupp et al., 2008).

For two-treatment experiments, a general N-of-1 trial can have multiple AB or BA crossover pairs in a sequence of treatments for within-patient comparisons. Hence, the number of periods is a multiple of 2 and it is desirable to have more than two of these pairs for a stable estimate of a treatment effect or its contrast. Possible sequences to consider rapidly increase with the increasing p for multi-periods. For example, three pair N-of-1 trials with $p = 6$

would require considering 2^p=64 treatment sequences before we could determine the optimal treatment sequence(s). Since it is only feasible to use a small set of sequences, we aim to determine the optimal sequences, while ensuring that each pair of periods consists of two distinct treatments. Therefore, such sequences as AAABBA or AAAABA are unlikely to be used in N-of-1 trials. Eliminating unsuitable sequences among 2^p for N-of-1 trials, we are left with $2^{p/2}$ distinct sequences to consider for p-period two treatment N-of-1 trials.

In the N-of-1 trials with $t \geq 2$ treatments, the sequences consist of treatments in blocks of the same size t. Every block contains each of the t treatments exactly once. Constructing N-of-1 designs in this way, treatments are compared fairly and poor balance can be prevented when early termination of the experiments happens. For example, in a 3-treatment N-of-1 trial, a six-period design could be $ABC|BCA$, where the sign '|' divides them into blocks. This class of designs is denoted in our study as No1(t,t), representing N-of-1 trial that compares t (first t in the notation) treatments and consists of treatments in blocks of size t. Therefore, a six-period design in the above example is No1(3,3).

A design is universally optimal if (i) its information matrix is completely symmetric, and (ii) it maximizes the trace of the information matrix. Such a design is universally optimal, as it satisfies all three of A-, D-, or E-optimality criteria. To study the universal optimality of treatment effects in the t treatments N-of-1 designs, we consider the traditional model and obtain the information matrix for the parameters of interest. Then the universally optimal designs could be constructed as long as the conditions given by Kiefer (1975) are satisfied, as specified above.

We show that Kiefer's conditions could not be satisfied with the designs in

the class No1(t,t). However, if we consider a slightly different class of designs, then we can obtain universally optimal designs. We use No1(t,t+) to denote the new class of designs, which consist of designs with one extra treatment to the last period in No1(t,t). For instance, a design in No1(3,3+) could be $ABC|BCA|A$, $ABC|BCA|B$, or $ABC|BCA|C$, and similarly, some possibility in No1(2,2+) are $AB|BA|A$, $AB|BA|B$, etc.

In this chapter, we present universal optimality results in N-of-1 trials with $t \geq 2$ treatments. In section 2, we introduce the traditional model, which is typically considered in crossover trials. Upon constructing the information matrix for the parameters of interest under the traditional model, a sufficient condition of universally optimal designs is given for N-of-1 trials with t treatments in Section 3. In section 4, we build the universally optimal N-of-1 designs for $t = 2$. In section 4.1, we discuss the universally optimal properties in No1(t,t) and No1(t,t+). In section 4.2, we consider the designs in a relaxed condition where the block size can be smaller than the number of treatments, e.g., in No1(t,s) with $t > s$. By allowing the block size to be smaller than t, we can make universally optimal designs smaller than previously, thereby reduce the risk of early dropouts and the burden of treatment administration.

3.2 Model and information matrices

Crossover design models have typically assumed that the treatments assigned to subjects have lasting effects on their responses to treatments in subsequent periods. When it is assumed that the carryover effects last for only one period, they are known as a first-order residual effect model. In such a model, no interaction is assumed between the treatment administered during the current

period and the carryover effects from the previous period.

Further, the period effects cannot be accommodated as there are p responses in total in one patient trials. The traditional model with a first-order residual effect reduced for a single N-of-1 trial is written as:

$$Y_i = \mu + \tau_{d(i)} + \gamma_{d(i-1)} + \epsilon_i. \tag{3.1}$$

for $i = 1, \cdots, p$, where Y_i is the outcome in the ith period from the subject; τ and γ are, respectively, the treatment and the carryover effect contrasts; $d(i)$ $in 1, 2, \cdots, t$ is the treatment assigned to the patient in period i. Note $\gamma_{d(0)} = 0$, ϵ_i is the error term for the i period. If we define Σ as the covariance matrix of the measurement errors, we have $\Sigma = \sigma^2 I$ for the uncorrelated error assumption and $\Sigma = \sigma^2[I + \frac{\rho}{1-\rho}11']$ for the equal-correlated error assumption.

Typically, we are interested in estimating the direct and carryover treatment effects, while all others are treated as nuisance parameters. We shall construct a design for such a typical situation. To proceed, we define the design matrix under a particular model into $[\mathbf{X}_1, \mathbf{X}_2]$, where $\mathbf{X}_1$ contains the columns of the design matrix pertaining to nuisance parameters and $\mathbf{X}_2$ contains those for parameters of interest, $\theta = (\tau', \gamma')'$ with $\tau = (\tau_1, \tau_2, \cdots, \tau_t)$ and $\gamma = (\gamma_1, \gamma_2, \cdots, \gamma_t)$. Then the information matrix for θ can be written as

$$I_d(\theta) = X_2'\Sigma^{-1}X_2 - X_2'\Sigma^{-1}X_1(X_1'\Sigma^{-1}X_1)^- X_1'\Sigma^{-1}X_2 \tag{3.2}$$

First, we show a lemma in the joint information matrix of τ and γ in N-of-1 trials, which makes the later discussion simpler.

Lemma 3.2.1: *With uncorrelated errors or equal-correlated errors, the joint*

information matrix for τ and γ in the N-of-1 design can be written as

$$I_d(\theta) = X_2'X_2 - \frac{1}{p}X_2'11'X_2 \tag{3.3}$$

Proof: When uncorrelated errors are assumed, it is trivial that the Lemma 3.2.1 holds. Now we derive the joint information matrix for the traditional model with equal-correlated errors.

Under this case, we have $X_1 = 1$ and $\Sigma^{-1} = I - b11'$, where $b = \rho/((p-2)\rho + 1)$. Then according to (3.2),

$$\begin{aligned} I_d(\theta) &= X_2'(I - b11')X_2 - X_2'(I - b11')1(1'(I - b11')1)^{-}1'(I - b11')X_2 \\ &= X_2'X_2 - bX_2'11'X_2 - (1 - bp)X_2'1\ (\frac{1}{p - bp^2})\ (1 - bp)1'X_2 \\ &= X_2'X_2 - (b + \frac{(1 - bp)^2}{p - bp^2})X_2'11'X_2 \\ &= X_2'X_2 - \frac{1}{p}X_2'11'X_2 \end{aligned}$$

Thus, the Lemma 3.2.1 holds.

Therefore, the optimal designs are the same under the model in N-of-1 trials with uncorrelated and equi-correlated covariance structures.

In N-of-1 trials comparing t treatments, $A_1, A_2, \cdots, A_t$, there can be p-1 subsequences with treatments in 2 concessive periods, such as A_iA_j, which has the ith treatment in its first period and jth treatment in its second period, We define m_{ij} as the total number of subsequences of A_iA_j, in a p-period N-of-1 trial. Since there are $p-1$ subsequences of length 2, we have $1'M1 = p - 1$, where $M = (m_{ij})_{1 \le i,j \le t}$. Let $d = (m_1, m_2, \cdots, m_t)'$, where m_i is the number of assignments for the ith treatment. Define $d_1 = M1$, in which the ith

entry is the number of subsequences beginning with the ith treatment. Define $d_2 = M'1$, in which the jth entry is the number of subsequences ending with the jth treatment. Therefore, $d-d_1 = (0, \cdots, 0, 1, 0, \cdots, 0)'$ where the position of '1' represent the treatment in the pth period; $d - d_2 = (0, \cdots, 0, 1, 0, \cdots, 0)'$ where the position of '1' represent the treatment in the 1st period. Denote $D = diag(d)$, $D_1 = diag(d_1)$ and $D_2 = diag(d_2)$. We then represent 3.3 as follows using these notations.

Lemma 3.2.2: *The information matrix of the joint direct and carryover treatment effects is given by* $I_d(\theta)$.

$$I_d(\theta) = \begin{bmatrix} D - \frac{1}{p}dd' & M' - \frac{1}{p}dd_1' \\ M - \frac{1}{p}d_1d' & D_1 - \frac{1}{p}d_1d_1' \end{bmatrix}$$

Based on Lemma 3.2.1, we have the following result.

Lemma 3.2.3: *The information matrix* $I(\tau)$ *of the direct treatment effect* τ*, adjusted for the other effects in the model, is given by*

$$I_d(\tau) = D_2 - M'D_1^{-1}M \tag{3.4}$$

Proof:

If we denote the information matrix $I_d(\theta)$ in partitioned matrices,

$$I_d(\theta) = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}$$

the information matrix $I_d(\tau)$ can be obtained by $I(\tau) = I_{11} - I_{12}I_{22}^{-1}I_{21}$. Therefore, from Lemma 3.2.3, we have

$$I(\tau) = D - \frac{1}{p}dd' - (N' - \frac{1}{p}dd_1')(D_1 - \frac{1}{p}d_1d_1')^{-1})(N - \frac{1}{p}d_1d')$$

.

Since $(D_1 - \frac{1}{p}d_1d_1')(D_1 - 1 + 11') = I$, we can write

$$\begin{aligned} I(\tau) &= D - \frac{1}{p}dd' - (M' - \frac{1}{p}dd_1')(D_1^{-1} + 11')(M - \frac{1}{p}d_1d') \\ &= D - dd' - M'D_1^{-1}M + M'1d' + d1'M - M'11'M \\ &= D - M'D_1^{-1}M - (d - d_2)(d - d_2)' \\ &= D_2 - M'D_1^{-1}M \end{aligned}$$

3.3 A general condition for universal optimality in N-of-1 trials

For a p-period N-of-1 trial, m_{ij} is the number of subsequences with the ith treatment in its first period and jth treatment in its second period. Therefore, $\sum_{1 \le i,j \le t} n_{ij} = p - 1$.

Theorem 3.3.1: *Under model 3.1, if an N-of-1 design is balanced on all subsequence A_iA_j, which leads to $n_{ij} = (p-1)/t^2$ for any $1 \le i, j \le t$, then it is universally optimal for estimating the direct treatment effects with adjustment of carryover effects among all t-treatment N-of-1 designs.*

Proof:

We can rewrite the terms in equation (3.4) with m_{ij}, $m_{i\cdot} = \sum_{j=1}^{t} m_{ij}$ and

$m_{\cdot j} = \sum_{i=1}^{t} m_{ij}$. We have $D_2 = \text{diag}(m_{\cdot 1}, m_{\cdot 2}, \cdots, m_{\cdot t})$, and the element of row i and column j from $M' D_1^{-1} M$ is

$$\frac{m_{1i} m_{1j}}{m_{1\cdot}} + \frac{m_{2i} m_{2j}}{m_{2\cdot}} + \cdots + \frac{m_{ti} m_{tj}}{m_{t\cdot}}$$

.

Therefore, the trace of the information matrix for treatment effects can be written as,

$$tr(I_d(\tau)) = p - 1 - tr(M' D_1^{-1} M) \tag{3.5}$$

where $M' D_1^{-1} M = \sum_{i=1}^{t} \frac{\sum_{j=1}^{t} m_{ij}^2}{m_{i\cdot}}$.

The trace is maximized by minimizing $M' D_1^{-1} M = \sum_{i=1}^{t} \frac{\sum_{j=1}^{t} m_{ij}^2}{m_{i\cdot}}$ subject to $\sum_{1 \le i,j \le t} m_{ij} = p - 1$. The minimization is achieved by $m_{ij} = m_{ij'}$ for each i and any $j \neq j'$. A special case in the solutions is when m_{ij} are equal for all $1 \le i, j \le t$, which results in that the information matrix $I(\tau)$ is completely symmetric. Therefore, by the universal optimality proposed by Kiefer (1975), the design which is balanced on all subsequences $A_i A_j$ is universally optimal among all t-treatment N-of-1 designs.

3.4 Examples

3.4.1 Universally optimal designs in No1(2,2)

We first apply the theorem 3.3.1 to construct universally optimal designs for No1(2,2). To compare two treatments A and B in an N-of-1 trial, we consider design sequences with AB or BA crossover pairs. From the Lemma 3.2.3, the

information matrix for treatment effects can be written as,

$$
\begin{aligned}
I_d(\tau) &= D_2 - M' D_1^{-1} M \\
&= \begin{bmatrix} \frac{m_{11}m_{12}}{m_{11}+m_{12}} + \frac{m_{21}m_{22}}{m_{21}+m_{22}} & -\frac{m_{11}m_{12}}{m_{11}+m_{12}} + \frac{m_{21}m_{22}}{m_{21}+m_{22}} \\ -\frac{m_{12}m_{11}}{m_{11}+m_{12}} + \frac{m_{22}m_{21}}{m_{21}+m_{22}} & \frac{m_{12}m_{11}}{m_{11}+m_{12}} + \frac{m_{22}m_{21}}{m_{21}+m_{22}} \end{bmatrix}
\end{aligned} \tag{3.6}
$$

Theorem 3.4.1: *A treatment sequence that alternates the two crossover pairs is universally optimal for No1(2,2).*

Proof:

It is trivial to see that the information matrix 3.6 is completely symmetric. Therefore, a design is universally optimal if it maximizes the trace.

For convenience, we consider the sequences starting with A. The solution of the maximization is

$$m_{11} = \tfrac{p}{4}, m_{12} = \tfrac{p}{4}, m_{21} = \tfrac{p}{4}, m_{22} = \tfrac{p}{4} - 1$$

for $p = 4k$, and

$$m_{11} = \tfrac{p}{4} - \tfrac{1}{2}, m_{12} = \tfrac{p}{4} + \tfrac{1}{2}, m_{21} = \tfrac{p}{4} - \tfrac{1}{2}, m_{22} = \tfrac{p}{4} - \tfrac{1}{2}$$

for $p = 4k + 2$, with any positive integer k.

For the sequences starting with B, we can switch the two indices of m_{ij} in the above results to get the universally optimal designs.

In both cases, we have that $m_{11} + m_{22} = m_{12} + m_{21} - 1$. It only occurs when all the crossover pairs AB are followed by BA and all crossover pairs BA are followed by AB. Therefore, we can construct the universally optimal

N-of-1 in No1(2,2) by alternating crossover pairs, such as $ABBA$, $ABBAAB$ and $ABBAABBA$ for 4, 6, and 8 periods designs, respectively. One could switch A and B to obtain a dual sequence with the same effect.

□

3.4.2 Universally optimal designs in No1(t,t+) with t≥3

In this section, we discuss the universal optimality in No1(t,t). As mentioned, a sequence in No1(t,t) have kt periods, which can be divided into k blocks of length t. Each of the k blocks in the sequence is a permutation of the t treatments. In the implementation of N-of-1 trials, clinicians may prefer a design in No1(t,t), in which t treatments are assigned in a block of size t for obvious reasons.

Unfortunately, we can not construct universally optimal designs in No1(t,t) from Theorem 3.3.1 except when $t = 2$. We devise a new class of N-of-1 designs, denoted as No1(t,t+), where the first $p - 1$ periods construct a sequence in No1(t,t) and the last period is treated with the same treatment used as in the beginning period. Then we construct universally optimal designs as follows.

Theorem 3.4.2: *If a t-treatment N-of-1 design*

1. *uses all possible permutations of the t treatments once,*
2. *connects them into a sequence in a way that the last treatment in the previous permutation is the same as the first treatment in the following permutation, and*
3. *adds one more period to the end with the treatment used in the first period,*

then the design is universally optimal.

Proof:

The described design in 3.4.2 uses all possible permutations of the t treatments once. The number of all possible permutations of the t treatments is $t!$. Denote the t treatments by $A_1, A_2, \cdots, A_t$. Then, for any subsequence A_iA_j with $i \neq j$, the number of the permutations containing A_iA_j is $(t-1)!$. This can be derived by multiplying the number of all permutations with the other $t-2$ treatments ($(t-2)!$) and the number of possible spot to place A_iA_j into the $t-2$ treatment permutations ($(t-1)$).

The permutations of the t treatments are then connected into a sequence. For any two connected permutations, the last treatment in the previous permutation is the same as the first treatment in the following permutation. Therefore, subsequences A_iA_i for some i can be observed at the connection so that the first A_i is from the last period in the previous permutation and the second A_i is from the first period in the following permutation. Since there are $(t-1)!$ permutations starting with A_i and $(t-1)!$ permutations ending with A_i, the number of A_iA_i is $(t-1)!$, when all the permutations are connected.

There is one exception with the subsequence A_iA_i when A_i is the starting treatment in the first permutation. Suppose the first permutation is chosen as the first treatment is A_i. Then among all of the other permutations, there are $(t-1)!-1$ permutations starting with A_i but there are $(t-1)!$ permutations ending with A_i. As a result, the number of A_iA_i is $(t-1)!-1$ and there is one extra permutation ending with A_i that is not able to connect with any permutation starting with A_i. Actually, the extra permutation must be the last permutation when all the permutations are combined.

Finally, by adding one more period to the end with the treatment A_i which is used in the first period, the number of A_iA_i is $(t-1)!$. Therefore, the constructed N-of-1 sequence is balanced on all A_iA_j for any $i, j = 1, \cdots, t$. According to Theorem 3.3.1, the design is universally optimal. □

Now we use the three-treatment N-of-1 trials as an example in constructing the universally optimal design in No1(3,3+). Denoting the treatments as A,B and C, the possible three-treatment permutations are

$$ABC, BAC, CAB, ACB, BCA, CBA.$$

The universally optimal design can be constructed, for example, as

$$ABC|\, CAB|\, BCA|\, ACB|\, BAC|\, CBA|\, A\ , or$$
$$ABC|\, CBA|\, ACB|\, BAC|\, CAB|\, BCA|\, A$$

requiring the number of periods to be $p = t \cdot t! + 1 = 19$.

3.4.3 Universally optimal designs in No1(t,2) with block size $< t$

One disadvantage of the universally optimal designs for t treatments is that the length of the sequence can be unmanageable. As discussed in the previous section, a universally optimal design in No1(t,t+) requires the length of the sequence equal to $(t-1)!\, t^2 + 1$, which is 5 for $t = 2$, 19 for $t = 3$ and 97 for $t = 4$. It may be infeasible in practice because the longer the period of the experiment, the more expensive the experiment and the higher the risk of drop-outs. To shorten the length of the experiment without losing the balance

in the comparison of treatments, we now consider the design in No1(t,s) or No1(t,s+) for some $s < t$, especially when $s = 2$. To do so, we relax the restriction that the block size must be equal to the number of the treatments being compared.

In this section, we provide universally optimal designs for three-, four- and five-treatment in blocks of size 2. In each block, 2 different treatments are assigned such as a crossover pair. For t-treatment designs, there are $t(t-1)$ different kinds of crossover pairs. To construct the universally optimal design, the crossover pairs are selected such that each subsequence of A_iA_j, $1 \leq i, j \leq t$, appears only once. Therefore, in the universally optimal designs, the number of periods is $p = t^2 + 1$, for example, p is 10 for three-treatment designs, 17 for four-treatment designs and 26 for five-treatment designs.

When $t = 3$, the universally optimal design in No1(3,2) can be, for examples,

$$AB|\ BC|\ CA|\ AC|\ BA \ \ or$$
$$BC|\ AB|\ BA|\ AC|\ CB.$$

When $t = 4$, the universally optimal design in No1(4,2+) can be, for examples,

$$AB|\ BC|\ CD|\ DA|\ CB|\ DC|\ AD|\ BA|\ A \ \ or$$
$$BC|\ CD|\ DA|\ AB|\ BD|\ CA|\ DB|\ AC|\ B.$$

When $t = 5$, the universally optimal design in No1(5,2) can be, for examples,

$$AB|\ BC|\ CD|\ DE|\ EA|\ AC|\ BD|\ CE|\ DA|\ EB|\ AD|\ BE|\ CA \ \ or$$

$$CD|\ DE|\ EA|\ AB|\ BC|\ CE|\ BD|\ AC|\ BE|\ DB|\ AE|\ CA|\ DC.$$

3.5 Discussion

N-of-1 trials are extremely useful in subject-focused investigations, for example, medical experiments. As far as we are aware, no guidelines are available in the literature on how to plan such a trial optimally. In this paper, we present a sufficient condition of universally optimal N-of-1 designs under a traditional model accommodating the carryover effects. We propose universally optimal sequences for general $t \geq 2$.

We investigated a sufficient condition for universal optimality in t treatments N-of-1 designs. However, we did not find the universally optimal designs in the general class No1(t,t). The condition requires the number of periods of the designs to be $kt + 1$. For some $k = t!$, we consider a sequence of length $kt+1$ that allows the first $(p-1)$th periods are assigned in blocks of size t, but end with an additional period of treatment, which is the same as the treatment that was given in the 1st period of the sequence. We denoted such class of designs as No1(t,t+). Then, universally optimal designs can be constructed in No1(t,t+). In order to shorten the number of periods in N-of-1 trials, we also considered universally optimal designs in No1(t,2) and No1(t,2+).

Chapter 4

Optimal two-treatment N-of-1 trial designs

Abstract

N-of-1 trials are randomized multi-crossover experiments using two or more treatments on a single patient. They provide evidence-based information on an individual patient, thus optimizing the management of the individual's chronic disease. This approach is preferred in many medical experiments, as opposed to the more common statistical designs constructed to optimize treating the average patient. N-of-1 trials are also popular when the sample size is too small to adopt traditional optimal designs. However, there are very few guidelines available in the literature. We constructed optimal N-of-1 designs for two treatments under a variety of conditions about the carryover effects, the covariance structure, and the number of planned periods. Extension to optimal aggregated N-of-1 designs is also discussed.

4.1 Introduction

N-of-1 trials are single-patient crossover designs which aim to provide the best possible treatment decisions for an individual patient. When suitably designed, they can be extremely useful to minimize patients' clinic visits and time on suboptimal treatments. However, the literature is lacking in providing guidelines for optimal N-of-1 trials.

In Chapter 3, we constructed universally optimal designs, which are A-, D- and E-optimal designs under rather restricted models. In this chapter, we relax the universal optimality condition to find practically optimal designs for fixed p, as often the choice of p is not random.

The majority of clinical studies employ randomized controlled trials (RCTs) (Armitage, 1975; Kenword and Jones, 1987; Wei and Durham, 1978). One approach to designing an RCT is the use of optimal experimental designs. An optimal design is a technique designed to assist a decision maker in identifying a preferable choice among many possible alternatives. Among the many RCT designs available, the most useful and popular design is the crossover design.

To illustrate the logistics of choosing a particular design, we first note that there are a number of excellent articles on optimal designs in the RCT literature (Cheng and Wu, 1980; Carriere, 1994; Carriere Huang, 2000; Liang and Carriere, 2009; Laska and Meisner, 1985; Afsarinejed and Hedayat, 2002; Kunert and Stufken, 2002). However, most of these designs, if not all, focus on optimizing the treatment effect for an average patient. The average patient is a construct - a virtual person who responds to the intervention by the mean of population's responses. Individuals enrolled in a trial will respond better or worse than, or simply differently from the average patient. The available optimal designs are not adequate when studying individual-based treatment effects is desired.

Multi-crossover single-patient trials are often employed when the focus is to make the best possible treatment decision for an individual patient. From a clinician's perspective, having clear evidence of the value of one treatment over another (or no treatment) is far more useful than knowing the average response. The average response gives the clinician the probability that a treatment will be effective, whereas N-of-1 trials give far more certainty about whether the treatment for the patient sitting in front of them will work or not.

It is known that the two-treatment design AB, AA and their duals BA, BB is found to be universally optimal for two-period experiments, with the

duality defined as the sequence that switches A and B with the same effect. Similarly, it is known that the two-sequence design ABB and its dual BAA and the four-sequence design $ABBA$, $AABB$ and their duals $BAAB$, $BBAA$ are optimal for three- and four-period experiments, respectively (Carriere, 1994; Laska and Meisner, 1985).

Straight application of this two-treatment optimal design literature with A to AB and B to BA would suggest that optimal N-of-1 trials can use the four-sequence design with $ABBA$, $ABAB$ and their duals for two within-patient comparisons, the two-sequence design with $ABBABA$ and its dual for three within-patient comparisons, and the four-sequence design with $ABBABAAB$, $ABABBABA$ and their duals for four within-patient comparisons. It is not yet known whether all of these sequences are equivalent so that each sequence is optimal for each individual patient for 4, 6 and 8-period N-of-1 trials. Further, applying the results from the literature would require at least two patients to utilize these existing designs, as the optimal design uses at least two sequences and is unsuitable for N-of-1 trials. In this chapter, we show that not all sequences in these repeated measurement designs are optimal for N-of-1 trials for estimating individual-based treatment effects.

Ideally, when aggregated, the series of N-of-1 trials that are optimal for individual patients can also provide an optimal estimate of the treatment effects for the average patient. For example, in a multi-clinic setting in three AB pair six-period N-of-1 studies, all eight possible sequences ($2^{6/2} = 8$) have been used, i.e., $ABABAB$, $ABABBA$, $ABBAAB$, $ABBABA$ and their duals to estimate both individual-based and average treatment effects (Guyatt et al., 1990). However, it is not known whether each of these eight sequences is optimal for individual patients. Further, it is not known whether a collection

of the optimal and not-so-optimal n-of-1 trials will lead to optimal designs for estimating the average treatment effects. In the next sections, we discuss how these do not lead to optimal aggregated N-of-1 trials for estimating the treatment effects for the average patient. We first discuss issues arising due to the repeated nature of these experiments.

4.2 Models and information matrix

We consider the following model, frequently employed by repeated measures crossover data.

$$Y_{ij} = \mu + \alpha_i + \beta_j + \tau_{d(i,j)} + \gamma_{d(i-1,j)} + \epsilon_{ij}, \tag{4.1}$$

for $i = 1, \cdots, p$ and $j = 1, \cdots, N$, where Y_{ij} denotes the outcome in the ith period from the jth subject;α_i denotes the ith period effect and β_j denotes the jth subject effect; $d(i, j)$ represents the treatment assigned to the patient in period i of subject j, and $\tau_{d(i,j)}$ and $\gamma_{d(i-1,j)}$ are, respectively, the treatment effect of the treatment on the ith period and the carryover effect of the treatment on the $i-1$th period.

The model assumes that the carryover effects only depend on the treatment assigned on the previous period but not on the treatment on the current period, which may be unrealistic. Taking the interaction into account without introducing too many parameters, Kunert and Stufken (2002) presented a model with self and mixed carryover effects. The self carryover effect occurs when the treatments administered in the current and the previous period are the same; otherwise we have a mixed carryover effect. The model with the self

and mixed carryover effects is written as,

$$Y_{ij} = \begin{cases} \mu + \alpha_i + \beta_j + \tau_{d(i,j)} + \gamma_{s,d(i-1,j)} + \epsilon_{ij}, \text{ if d(i,j)=d(i-1,j);} \\ \mu + \alpha_i + \beta_j + \tau_{d(i,j)} + \gamma_{m,d(i-1,j)} + \epsilon_{ij}, \text{ if d(i,j)}\neq\text{d(i-1,j).} \end{cases} \tag{4.2}$$

where α_i, β_j, $d(i,j)$ and $\tau_{d(i,j)}$, are defined as in model (1.1); $\gamma_{s,d(i-1,j)}$ and $\gamma_{m,d(i-1,j)}$ represent the self and mixed carryover effects of the treatment assigned on the $i-1$th period, respectively.

In an N-of-1 trial with $N = 1$, the j index can be omitted. Here, as there is one subject and p responses in total, the period effects and subject effects cannot be accommodated. Therefore, we need to reduce the models for the case of $N = 1$. However, a full model can be used for $N > 1$, and aggregated optimal N-of-1 sequences can make the estimation on an average level possible.

For model (4.1) and (4.2), we define the contrast of the direct treatment effects as $\tau = (\tau_A - \tau_B)/2$, the contrast of the first-order carryover effects as $\gamma = (\gamma_A - \gamma_B)/2$, the contrast of the self carryover effects as $\gamma_s = (\gamma_{s,A} - \gamma_{s,B})/2$ and the contrast of the mixed carryover effects as $\gamma_m = (\gamma_{m,A} - \gamma_{m,B})/2$. We also define $\Phi_{d(i)}$ as the indicator of the treatment assigned to the ith period. For $i = 1, \cdot, p$, $\Phi_{d(i)} = 1$ if $d(i) = A$, $\Phi_{d(i)} = -1$ if $d(i) = B$. When $i = 0$, $\Phi_{d(i)} = 0$.

Then, the model (4.1) and (4.2) can be reduced to

$$Y_i = \mu + \tau \cdot \Phi_{d(i)} + \gamma \cdot \Phi_{d(i-1)} + \epsilon_i, \tag{4.3}$$

and

$$Y_i = \begin{cases} \mu + \tau \cdot \Phi_{d(i)} + \gamma_s \cdot \Phi_{d(i-1)} + \epsilon_i, \text{ if d(i)=d(i-1);} \\ \mu + \tau \cdot \Phi_{d(i)} + \gamma_m \cdot \Phi_{d(i-1)} + \epsilon_i, \text{ if d(i)}\neq\text{d(i-1).} \end{cases} \tag{4.4}$$

In the following discussion, we simply refer model (4.1) and (4.3) as the traditional model, and refer model (4.2) and (4.4) as the self and mixed carryover effect model.

One common assumption on the variances of the error terms is that they are independently identically distributed with mean 0 and variance σ_ϵ^2. However, because Y_i are measures from the same subject, it is appropriate to introduce dependence among the $Y_i's$. In our discussion of optimal 2-treatment N-of-1 trials, we consider 3 different covariance structures for the responses from a subject.

To construct the optimal design, we often devise an optimality criterion. We define a design to be optimal based on *A*-, *D*-, *or* E-optimality criteria, maximizing either the trace, the determinant, or the eigenvalue of the information matrix among a class of all competing designs. To do so, we would first obtain an information matrix for the parameters of interest. To proceed, we define the design matrix under a particular model into $[\mathbf{X}_1, \mathbf{X}_2]$, where $\mathbf{X}_1$ contains the columns of the design matrix pertaining to nuisance parameters and $\mathbf{X}_2$ contains those for parameters of interest, $\theta = (\tau,\gamma)'$ or $(\tau, \gamma_s, \gamma_m)'$, which denote the direct treatment effects and carryover effects. Then the information matrix can be written as

$$I_d(\theta) = X_2'\Sigma^{-1}X_2 - X_2'\Sigma^{-1}X_1(X_1'\Sigma^{-1}X_1)^{-1}X_1'\Sigma^{-1}X_2 \tag{4.5}$$

By partitioning the information matrix into a 2-by-2 matrix according to dimensions of τ and $\tilde{\gamma}$=γ or $(\gamma_s, \gamma_m)'$, we have,

$$I_d(\tau, \tilde{\gamma}) = \begin{bmatrix} I_{d11} & I_{d12} \\ I_{d21} & I_{d22} \end{bmatrix} \tag{4.6}$$

Then the information matrix of direct treatment effects adjusted by the carryover effects can be written as,

$$I_d(\tau) = I_{d11} - I_{d12} I_{d22}^{-1} I_{d21}. \tag{4.7}$$

4.3 Cycles and Sequences

We focus on the N-of-1 designs, which compare only two treatments. The main interest is the estimation of the direct treatment effect contrast, τ. We present the optimal designs in 2-treatment N-of-1 trials under model (4.3) and (4.4).

First, we define some sequence feature parameters and show the association between the feature parameters and the sequences in N-of-1 designs.

In our discussion of N-of-1 trials for two treatments, the design sequences consist of crossover pairs, AB and BA. Within each crossover pair, the two treatments are distinct. For two consecutive crossover pairs, the treatments assigned to the second period in the previous pair and the first period in the latter pair can be different or the same.

If an AB pair is followed by a BA pair, as in $ABBA$ (or $BAAB$), we define the design as having alternating pairs in the sequence. The performance of an N-of-1 trial sequence is related to how the pairs AB and BA alternate. The following feature parameters define how AB and BA alternate in a sequence.

- s: the number of subsequences of AA and BB;
- m: the number of subsequences of AB and BA;
- $h = s - m$.

When we define s and m, the subsequences can be constructed by either the treatments from a crossover pair, or be the treatments assigned to the second period in the previous pair and the first period in the latter pair. Therefore, in a p-period sequence, there are $p-1$ such subsequences with a length of 2. By the definition of feature parameters, we have $s + m = p - 1$. Determined by

how a sequence is constructed, the h can only be negative and takes the values in $-1,-3,\cdots,-(p-1)$. Table (4.1) demonstrates the relationship among h, s and m.

Table 4.1: Feature parameters of a sequence in a 2-treatment N-of-1 design

h	s	m
$-(p-1)$	0	p-1
$-(p-3)$	1	p-2
$-(p-5)$	2	p-3
$\vdots$	$\vdots$	$\vdots$
-1	$\frac{p}{2}-1$	$\frac{p}{2}$

For a particular h, we can calculate s and m by $s=(p-1+h)/2$ and $m=(p-1-h)/2$. Further, for any given p, the N-of-1 designs can be classified by h. As an example, for $p=8$, Table (4.2) shows the relationship between the design sequences and the feature parameters.

Table 4.2: Sequences for $p=8$ with corresponding design parameter values

h	Sequence	Alternation	s	m
-7	$ABABABAB$	0	0	7
-5	$ABABABBA$	1	1	6
	$ABABBABA$		1	6
	$ABBABABA$		1	6
-3	$ABABBAAB$	2	2	5
	$ABBAABAB$		2	5
	$ABBABAAB$		2	5
-1	$ABBAABBA$	3	3	4

Note: s = the number of AA and BB and m = the number of AB and BA in a treatment sequence, and $h=s-m$

In the next section, we show that the information matrix of the parameters of interest are only dependent on the feature parameters. That is, sequences with the same h values have the same information matrix. For instance, when $h=-3$, the three sequences $ABABBAAB$, $ABBAABAB$, $ABBABAAB$

and their dual sequences share the same information matrix. If this h is the optimum value, the 8 period N-of-1 trials can use any of these three sequences and their duals.

4.4 Optimal 2-treatment N-of-1 designs

We denote $\mathbf{x}_\tau$, $\mathbf{x}_\gamma$, $\mathbf{x}_s$ and $\mathbf{x}_m$ as the design vector for the parameter τ, γ, γ_s and γ_m, respectively. Then under the traditional model, the design matrix is $[\mathbf{1}_p, \mathbf{x}_\tau, \mathbf{x}_\gamma]$ for the parameters $[\mu, \tau, \gamma]$ and we let $\mathbf{X}_1 = 1_p$ and $\mathbf{X}_2$=[$\mathbf{x}_\tau$, $\mathbf{x}_\gamma$], while under the self and mixed effect model, we have $[\mathbf{1}_p\ \mathbf{x}_\tau, \mathbf{x}_s, \mathbf{x}_m]$ for parameters $[\mu, \tau, \gamma_s, \gamma_m]$ and we let $\mathbf{X}_1 = 1_p$ and $\mathbf{X}_2$=[$\mathbf{x}_\tau$, $\mathbf{x}_s$, $\mathbf{x}_m$].

In 2-treatment N-of-1 trials, the $I_d(\tau, \gamma)$ is a function of the following quantities:

$$x'_\tau x_\tau = p, \quad x'_\gamma x_\gamma = p - 1, \quad x'_\tau x_\gamma = h \tag{4.8}$$

under model (4.3) for $\theta = (\tau, \gamma)$, or

$$x'_\tau x_s = s, \quad x'_s x_s = s, \quad x'_m x_s = 0, \quad x'_\tau x_m = -m, \quad x'_m x_m = m \tag{4.9}$$

under model (4.4) for $\theta = (\tau, \gamma_s, \gamma_m)$. Hence, the information matrix can be expressed in terms of p, s, m and h or simply by h and p only.

Since the information matrices can be expressed in terms of h and p only, for a given p, the optimal $p-$period N-of-1 trial is completely determined by h, and much simpler to construct than previously. We proceeded by defining $I_d(\tau)$ and found the design that maximizes the information.

If the estimation of carryover effects is the main interest under some circumstances, we can similarly define $I_d(\gamma)$ and determine the optimal design

for carryover effects. One could also find the optimal design that simultaneously optimizes τ and γ or the one that optimizes the carryover effects under the constraint that it optimizes for the direct treatment effects. We note that the approach can also apply to find optimal designs for estimating some linear combinations of the parameters of interest, such as $\tau + \gamma$ (See also Carriere 1993). Since we are primarily interested in the optimal estimation of the direct treatment effects, we do not consider these cases here.

Under the traditional model

When we assume an independent error structure, the joint information of τ and γ is,

$$I_d(\tau, \gamma) = \frac{1}{\sigma_\epsilon^2} \begin{bmatrix} p & h \\ h & p - 1 - \frac{1}{p} \end{bmatrix}$$

Then,

$$I_d(\tau) \propto p - \frac{h^2}{p - 1 - 1/p}.$$

Maximization of the information of τ is achieved by minimizing $h^2 = (x_\tau' x_\gamma)^2$. When h equals -1, the information attains its maximum. This results in the optimal sequence consisting of pairs of AB and BA appearing alternatively throughout the trial. Therefore, we have the following result.

Result 1. Under an independent error assumption, the optimal N-of-1 trial for τ and γ is the one sequence design that consists of pairs of AB and BA appearing alternatively.

For example, the optimal N-of-1 trials for 4, 6, and 8 periods are the one sequence design, $ABBA$, $ABBAAB$ and $ABBAABBA$, respectively.

□

When we assume an equi-correlated error structure, we have,

$$I_d(\tau,\gamma) = \frac{1}{\sigma_\epsilon^2(1-\rho)} \begin{bmatrix} p & h \\ h & p-1-\frac{1}{p} \end{bmatrix}$$

Again, the maximum of the information matrix of τ is achieved by $h = -1$. Therefore, the optimal N-of-1 design for τ is still the one treatment sequence, which alternates treatment pairs as above for each level ρ of within-subject correlation.

Result 2. Under an equi-correlated error assumption, the optimal N-of-1 trial for τ and γ is the one sequence design that consists of pairs of AB and BA appearing alternatively, regardless of ρ.

Hence, similarly to the case of independent errors, the optimal N-of-1 trials for 4, 6, and 8 periods are the one sequence design, $ABBA$, $ABBAAB$ and $ABBAABBA$, respectively. One could switch A and B to obtain a dual sequence with the same effect. □

When we assume an autoregressive error structure, we can also compute the joint information of τ and γ.

If we write,

$$I_d(\tau,\gamma) = \frac{(1-\rho^2)}{\sigma_\epsilon^2} \begin{bmatrix} I_{11}\ I_{12} \\ I_{21}\ I_{22} \end{bmatrix},$$

then,

$$I_{11} = \rho^2(p-2) + p - 2\rho h - \frac{(1-\rho)\rho^2(s_1+s_p)^2}{p - p\rho + 2\rho} \tag{4.10a}$$

$$I_{12} = I_{21} = \rho + \rho^2 + (1+\rho)^2 h + \frac{(1-\rho)\rho s_p(s_1+s_p)}{p - p\rho + 2\rho} \tag{4.10b}$$

$$I_{22} = (1+\rho^2)(p-2) + 1 - 2\rho h - 2\rho - \frac{1-\rho}{p - p\rho + 2\rho} \tag{4.10c}$$

where s_1 and s_p are the 1st and pth entry of x_τ, and $s_1 + s_p$ takes values +2 or -2 or 0 depending upon the treatments given in the first and last periods in the sequence. In a sequence that starts and ends with a distinct treatment, we have $s_1 + s_p = 0$; otherwise, we have ${s_1 + s_p}^2 = 4$. This is determined by h and p. Specifically, the sum is 0 if and only if h takes values in $-(p-1)$,$-(p-1)+4$,$\cdots$,$-(p-1)+4k$. The k is a greatest integer that makes $-(p-1)+4k$ equal to -1 or -3, depending on whether the number of crossover pairs of the design is odd or even. Therefore, optimal design for τ is determined by the value of h and ρ.

The information for τ can be derived by (4.7). However, a closed form for the optimal h is complicated. To find the optimal h, we assume that h is continuous and seek a numerical solution. The optimal design can be found for possible integer values of h, which is closest to the solution. Optimal results for $p = 4$ to 12 are shown in the top portion of Table (4.3).

□

Under the model with self and mixed carryover effects

The joint information matrices are obtainable from (4.5) but have different

forms depending on whether the value of s is odd or even in the design.

When we assume an independent error structure, the joint information matrix is as following.

When s is odd,

$$I_d(\tau, \gamma_s, \gamma_m) = \frac{1}{\sigma_\epsilon^2} \begin{bmatrix} \text{p} & \text{s} & \text{-m} \\ \text{s} & s - \frac{1}{p} & 0 \\ \text{-m} & 0 & \text{m} \end{bmatrix}; \tag{4.11}$$

and when s is even

$$I_d(\tau, \gamma_s, \gamma_m) = \frac{1}{\sigma_\epsilon^2} \begin{bmatrix} \text{p} & \text{s} & \text{-m} \\ \text{s} & \text{s} & 0 \\ \text{-m} & 0 & m - \frac{1}{p} \end{bmatrix}. \tag{4.12}$$

The information of τ is,

$$I_d(\tau) \propto \begin{cases} 1 - \frac{s}{ps-1}, & \text{when } s \text{ is odd;} \\ 1 - \frac{m}{pm-1}, & \text{when } s \text{ is even.} \end{cases}$$

The maximum information for τ is obtained when $s = 0$, or $h = -(p-1)$, which means the sequence has only AB (or BA) pairs, as summarized in the Result 3.

Result 3. Under the independent errors, the optimal N-of-1 trial for estimating the direct treatment contrasts under self and mixed model is the sequence with only AB (BA) pairs with no alternation.

Therefore, the optimal design for estimating the treatment effect contrast is the sequence with only AB or BA pairs, such as $BABABA$ and $ABABABAB$.

However, based on such sequences, the self carryover effect cannot be estimated. □

When we assume an equi-correlated error structure, the information matrix is $\frac{1}{1-\rho}$ times the matrix under independent error assumption.

When s is odd,

$$I_d(\tau, \gamma_s, \gamma_m) = \frac{1}{\sigma_\epsilon^2(1-\rho)} \begin{bmatrix} \text{p} & \text{s} & \text{-m} \\ \text{s} & s - \frac{1}{p} & 0 \\ \text{-m} & 0 & \text{m} \end{bmatrix} \tag{4.13}$$

and when s is even,

$$I_d(\tau, \gamma_s, \gamma_m) = \frac{1}{\sigma_\epsilon^2(1-\rho)} \begin{bmatrix} \text{p} & \text{s} & \text{-m} \\ \text{s} & \text{s} & 0 \\ \text{-m} & 0 & m - \frac{1}{p} \end{bmatrix} \tag{4.14}$$

Again, the maximum of the information matrix of τ is achieved by $s = 0$, or $h = -(p-1)$.

Result 4. Under the equi-correlated errors, the optimal N-of-1 trial for estimating the direct treatment contrast is the sequence with only AB (BA) pairs with no alternation.

The optimal design for estimating the treatment effect contrast is the same as the design given by **Result 3**, such as $BABABA$ and $ABABABAB$. □

When we assume autoregressive error structure, the joint information can

be written as,

$$I_d(\tau, \gamma_s, \gamma_m) = \frac{(1-\rho^2)}{\sigma_\epsilon^2} \begin{bmatrix} I_{d11} & I_{d12} & I_{d13} \\ I_{d21} & I_{d22} & I_{d23} \\ I_{d31} & I_{d32} & I_{d33} \end{bmatrix},$$

where, when s is odd,

$$\begin{aligned}
I_{d11} &= \rho^2(p-2) + p - 2\rho h - \rho^2(s_1+s_p)^2 \\
I_{d12} &= I_{d21} = (1+\rho^2)s \pm (s_1+s_p)(-\rho+\rho^2) \\
I_{d13} &= I_{d31} = -\rho(4m-2p+1) - (1+\rho^2)m + (s_1+s_p)\rho^2 s_p \\
I_{d22} &= (1+\rho^2)s - (1-\rho)^2 \\
I_{d23} &= I_{d32} = \pm s_p(\rho-\rho^2) \\
I_{d33} &= (1+\rho^2)m + 1 + 2\rho(p-2) - \rho^2
\end{aligned}$$

and when s is even,

$$\begin{aligned}
I_{d11} &= \rho^2(p-2) + p - 2\rho h - \rho^2(s_1+s_p)^2 \\
I_{d12} &= I_{d21} = (1+\rho^2)s \\
I_{d13} &= I_{d31} = -\rho(4m-2p+1) - (1+\rho^2)m + (s_1+s_p)\rho s_p \\
I_{d22} &= (1+\rho^2)s \\
I_{d23} &= I_{d32} = 0 \\
I_{d33} &= (1+\rho^2)m + 2\rho(p-2)
\end{aligned}$$

The information for τ can be derived by (4.7). Similar as in the case under the traditional model, a closed form for the optimal h is complicated and numerical solution was sought. We present selected optimal results in the

bottom portion of Table (4.3), as a function of h.

Table 4.3: The value of h determining the optimal N-of-1 trials for $p = 4, 6, 8, 10, 12$.

Model	p	Indep.	Equal-correlation		Auto-correlation	
			$\rho = 0.3$	$\rho = 0.7$	$\rho = 0.3$	$\rho = 0.7$
Traditional	4	-1	-1	-1	-1	-1
	6	-1	-1	-1	-1	-1
	8	-1	-1	-1	-1	-1
	10	-1	-1	-1	-1	-1
	12	-1	-1	-1	-1	-3
Self and mixed	4	-3	-3	-3	-1	-1
	6	-5	-5	-5	-1	-1
	8	-7	-7	-7	-1	-3
	10	-9	-9	-9	-1	-1
	12	-11	-11	-11	-3	-3

To summarize, the optimal N-of-1 trials for estimating direct treatment effects are determined by the three feature parameters h, s, and m. However, specifying one of these along with p determines the design sequence, as illustrated in Table (4.2). In this section, we use h to summarize the optimal designs under both the traditional and self and mixed models for 4,6,8,10 and 12-period N-of-1 trials. As the designs are dependent on the level of correlation coefficient, we consider a low (0.3) and a high (0.7) value for ρ . The optimal trials for estimating τ are summarized in Table (4.3) in terms of the corresponding h values.

Table (4.3) shows that the optimal N-of-1 trials for the direct treatment effects depend on the assumed models and the covariance structures. Under the traditional model, the optimal trial for the direct treatment effect uses the sequence with $h = -1$ for all covariance structures. Therefore, the optimal N-of-1 trial for estimating the direct treatment effect is to alternate between

AB and BA pairs. In case that the carryover effect is of interest, it can be easily shown that these designs are also optimal for estimating the carryover effect, which can be obtained using the same technique for optimal designs in treatment effects.

Under the self and mixed effects model, the optimal N-of-1 trial for the direct treatment effect uses a sequence with $h = -(p-1)$ for both uncorrelated and equal-correlated covariances. Therefore, the optimal N-of-1 trial is to use only AB pairs throughout.

Under the auto-regressive covariance structure, however, the optimal designs depend on the value of p and the auto-regressive correlation ρ. Generally, the optimal design uses AB and BA pairs alternately, but as ρ or p increases, some abnormalities are observed.

4.5 Optimal Aggregated N-of-1 Trial Designs with $N > 1$

In addition to the interest in the patient-based evidence of a treatment contrast, it may also be desirable to obtain a population average effect of treatments. Aggregating the series of N-of-1 trials can give such an estimate of average effect (Zucker 2010). Using the one sequence that was found optimal for N-of-1 trial to all patients seems to be an obvious choice. However, it might not optimize the trial for estimating the effects on the average patient and therefore, using the one sequence that is optimal for a single individual patient to all patients might not serve this purpose.

The optimal designs for aggregated N-of-1 trials can also be derived from

the information matrices we obtained, similarly as for N-of-1 trials for one patient, by allowing $j = 1, ..., N$ with $N > 1$. We approach the problem in two steps; first we optimize single N-of-1 trials, as the primary goal is to optimize estimating the effects for each patient. Next, we optimize the overall N-of-1 trials in aggregation.

To find the optimal design, we typically choose N_k for $k = 1, ..., s$ to allocate subjects to a sequence s. The sufficient condition on N_k was given by Laska and Meisner (1985) for a design to be optimal. The condition is called a duality in the design matrices, as defined earlier. Among other things, it permits simplification of the search for the optimal choice for N_k (see also Carriere 1992).

In previous sections, we constructed the optimal N-of-1 trial for $N = 1$ patient, which is a one-sequence design for all p periods. As noted earlier for Table 1, designs with the same value of h perform equally in estimation precision. Although all or only one of those with an equally optimal h can be used in a trial, practical consideration will lead to using the least necessary number of sequences for ease of treatment administration. Further, in Section 4, we found that there is a unique N-of-1 trial sequence in all p-period experiments. Since the designation of A and B is arbitrary, the optimal N-of-1 trial can be obtained by reversing the order of treatment administration. For example, the optimal 6-period N-of-1 trial is $ABBAAB$ under the traditional model for $N = 1$. Its dual, $BAABBA$ also has the same value of $h = -1$ and is optimal. Hence, when $N = 1$, either of these two sequences will provide the maximum amount of information. When $N > 1$, we can adopt both of these sequences, as they maximize the information, and this approach also simplifies the search for the optimal design for estimating the treatment effect for the

average patient, satisfying the duality condition in Laska and Meisner (1985) . Based on this rationale, we make the following two propositions.

Proposition 1 The optimal design for aggregated N-of-1 trials under the traditional model is to allocate the same number of subjects to the optimal sequence with AB and BA alternating and its dual.

For example, the optimal design for aggregated six-period N-of-1 trials is the two-sequence design using sequences $ABBAAB$ and $BAABBA$, allocating the same number of subjects to each. For a balanced design, N must be a multiple of 2.

Proposition 2 The optimal design for aggregated N-of-1 trials under the self and mixed model is to allocate the same number of subjects to the optimal sequence with no alternation between AB and BA pairs and its dual. However, under the autocorrelation errors, the optimal design is to allocate the same number of subjects to the optimal sequence that alternates between AB and BA pairs subsequently and its dual.

For example, the optimal aggregated 6-period N-of-1 trials for multi-clinic setting is to use the two-sequence design $ABABAB$ and $BABABA$ under the equal or uncorrelated errors, and to use the two-sequence design $ABBAAB$ and $BAABBA$ under the autocorrelated errors, allocating the same number of subjects to each sequence.

From each sequence, we can obtain individual patient specific treatment

effects and by aggregating these one sequence N-of-1 trials, we can quantify the average treatment effects.

4.6 Numerical Comparisons

To appreciate the practical performance of the optimal N-of-1 trials we constructed, we compare the efficiencies of some selected designs in estimating the treatment and carryover effects under the two models. We also investigate their performance in some aggregation for estimating the average treatment effect. We limit the comparison to the cases with independent and equi-correlation errors.

Recall that the optimal N-of-1 trials are either to alternate between AB and BA pairs or simply to repeat the AB pair in a sequence. Under the traditional model, the optimal N-of-1 trial uses $ABBAAB$ and $ABBAABBA$ for 6 and 8 period experiments, respectively. We refer them to S63 and S83. Under the self and mixed effects model, the optimal N-of-1 trial is to use $ABABAB$ and $ABABABAB$ for 6 and 8 period experiments, respectively, which we refer to S61 and S81. Some other mixtures are also considered, as defined in Table 4.4. Table 4.4 shows that the optimal individual-based N-of-1 trials are S63 and S81 under the respective models, as expected. However, there are no real practical differences among various N-of-1 trials under the self and mixed model. Further, S61 and S81 cannot estimate self carryover effects, making S63 and S83 preferable. Therefore, recommendation for robust and optimal N-of-1 trials is to use a sequence alternating between AB and BA pairs, such as S63 and S83, under all models.

Based on these single sequence trials, we also consider aggregated N-of-1

trials to numerically prove Propositions 1 and 2. We constructed 5 aggregated N-of-1 trials for $p = 6, 8$ with $N = 32$ and compare their efficiencies in estimating the average treatment effects, as follows.

- A61. $ABABAB$ and its dual with 16 subjects in each sequence
- A62. $ABABBA$ and its dual with 16 subjects in each sequence
- A63. $ABBAAB$ and its dual with 16 subjects in each sequence
- A64. $ABBAAB, ABABBA$ and their duals with 8 subjects in each sequence
- A65. All 8 sequences, S61–S64 and their duals with 4 subjects in each sequence
- A81. $ABABABAB$ and its dual with 16 subjects in each sequence
- A82. $ABABBABA$ and its dual with 16 subjects in each sequence
- A83. $ABBAABBA$ and its dual with 16 subjects in each sequence
- A84. $ABABBABA, ABBABAAB$ and their duals with 8 subjects in each sequence
- A85. All 8 sequences, S81–S84 and their duals with 4 subjects in each sequence

The design A63 uses the optimal sequence S63 under the traditional model; the design A61 uses the optimal sequence S61 under the self and mixed model although the self carryover effect is not estimable; the design A62 is a slight rearrangement of designs A61 and A63; the design A64 is a combination of designs A62 and A63; the design A65 contains all 8 possible sequences of a

6-period design. Designs A81–A85 are also similarly constructed from various N-of-1 trials. We compare these designs under the traditional model and the self and mixed model as discussed in Section 4.3. These are also summarized in Table 4.4, in terms of the variances for estimating τ, γ, γ_s and γ_m, divided by their leading constants σ^2/N (under an independence error) or $\sigma^2(1-\rho)/N$ (under an equi-correlated error).

Table 4.4: Variances of the estimators of treatment and carryover effects in six- and eight-period designs

Design	h	Traditional model		Self and mixed model		
		$var(\hat{\tau})$	$var(\hat{\gamma})$	$var(\hat{\tau})$	$var(\hat{\gamma}_s)$	$var(\hat{\gamma}_m)$
S61: ABABAB	-5	1.208	1.500	1.208	NE	1.500
S62: ABABBA	-3	0.242	0.300	1.250	3.000	1.500
S63: ABBAAB	-1	0.173	0.214	1.214	1.714	1.714
S64: ABBABA	-3	0.242	0.300	1.250	3.000	1.500
A61=S61+dual		1.208	1.500	1.208	NE	1.500
A62=S62+dual		0.242	0.300	1.250	3.000	1.500
A63=S63+dual		0.173	0.214	1.214	1.714	1.714
A64=S63+S62+duals		0.193	0.240	1.210	2.063	1.563
A65=S61:S64+duals		0.242	0.300	1.214	2.535	1.521
S81: ABABABAB	-7	1.146	1.333	1.146	NE	1.333
S82: ABABBABA	-5	0.229	0.267	1.167	2.667	1.333
S83: ABBAABBA	-1	0.127	0.148	1.150	1.600	1.400
S84: ABBABAAB	-3	0.150	0.174	1.147	1.647	1.412
A81=S81+dual		1.146	1.333	1.146	NE	1.333
A82=S82+dual		0.229	0.267	1.167	2.667	1.333
A83=S83+dual		0.127	0.148	1.150	1.600	1.400
A84=S82+S84+dual		0.176	0.205	1.147	1.945	1.358
A85=S81:84+dual		0.176	0.205	1.147	1.945	1.358

Note: NE means 'Not estimable'. For $N = 1$, a six-period N-of-1 trial may consider any one of S61,$\cdots$, S64. For $N > 1$, aggregated six-period N-of-1 trials may use a combination of these, A61,$\cdots$, A65. Similarly, an eight-period N-of-1 trial may consider any one of S81,$\cdots$, S84. For N > 1, aggregated six-period N-of-1 trials may use a combination of these, A81,$\cdots$, A85. The variances reported are divided by σ_ϵ^2/N (under an independence error) or $\sigma_\epsilon^2(1-\rho)/N$ (under an equi-correlated error).

Table 4.4 shows that under the traditional model, the design D63 with the

optimal sequence $ABBAAB$ and its dual provides the best precision for estimating both the direct treatment effect and the carryover effect for the average patient. Each of the sequences optimally estimates the individual-based treatment effect. The least efficient choice would be the design A61. Design A65, which has been used in a recent multi-clinical trial (Zucker 2010), is rather inefficient as well, not to mention the unnecessarily lengthy administration time and cost required to manage many treatment groups, which requires the number of patients to be a multiple of 8.

When using the self and mixed effects model, Design A61 provides the best precision for estimating the direct treatment effect and the mixed carryover effect. However, the self carryover effect is not estimable. Overall, A63 is the optimal choice even in this case. However, all designs performed rather similarly with over 95% relative efficiency under the self and mixed effects model, as observed earlier for single N-of-1 trials.

A similar observation is possible for 8-period designs and their sequences. In summary, it appears that there is no discernable advantage to distinguish among the two models and various error structures.

Overall, S63 and S83 for single N-of-1 trials or designs A63 and A83 in aggregation of S63, S83 and their duals seem to be the best under both models. They are optimal for estimating direct treatment and mixed carryover effects. Further, they are optimal for estimating both the treatment and carryover effects under the traditional model. Hence, we conclude that the optimal six-period aggregated N-of-1 trials is $ABBAAB$ and its dual $BAABBA$, while the optimal eight-period aggregated N-of-1 trials is $ABBAABBA$ and its dual $BAABBAAB$. For an N-of-1 trial, using one of these sequences will optimize the treatment for an individual patient.

In general, we suggest that alternating AB and BA pairs in sequence will result in a nearly optimal $p-$period design, if not the optimal one, under all models we considered, for estimating both individual effects in N-of-1 trials and average effects in aggregated N-of-1 trials.

4.7 Discussion

N-of-1 trials are extremely useful in individual patient-focused medical experiments. As far as we are aware, no guidelines are available in the literature on how to plan such a trial optimally. We propose 2-treatment practically optimal N-of-1 designs, which consist of AB and BA pairs.

A straight application of the two-treatment optimal design in the literature with A to AB and B to BA can result in a design suitable only for aggregated N-of-1 trials, as it requires more than one sequence (and hence more than one patient) to be used. We demonstrated that not all of the suggested sequences are optimal in N-of-1 trials nor they are optimal for estimating effects at the average level. For example, when $p = 4$, the literature gives ABBA/AABB and their duals as the optimal design. Applying this result to 8-period N-of-1 trials, we would need to consider at least four sequences, $ABBABAAB/ABABBABA$ and their duals. However, none of these four sequences are optimal for N-of-1 trials for $p = 8$. Also, these sequences, when aggregated, are not optimal for estimating effects for the average patient.

For the first-order residual effect model with uncorrelated or equal-correlated errors, the optimal N-of-1 trial is to use the sequence consisting of alternating AB and BA pairs. We can use its dual sequence with the same effect. For the self and mixed effect model, the optimal N-of-1 trial is to use the sequence

consisting of only AB pairs. Also, its dual can be used with the same effect.

However, numerical calculation of the estimation precision using several six- and eight-period designs revealed the actual performance of a particular design, giving us practical guidelines. Overall, we conclude that alternating between AB and BA pairs in sequence will result in practically optimal N-of-1 trial for a single patient, if not the optimal, under all the models we considered without the need to guess or conduct a pilot study to conform at the correlation structure. Alternating between AB and BA pairs in a single trial is nearly robust to misspecified error structures. When an experiment has been carried out with the optimal N-of-1 trial and additional patients are accrued in the trial, we can aggregate these N-of-1 trials optimally by allocating the same number of patients to its dual sequence, thereby optimizing the trial for both the individual and average patients.

Chapter 5

Multiple objective response adaptive crossover designs with binary outcomes

Abstract

A multiple-objective allocation strategy was recently proposed for constructing response-adaptive repeated measurement designs for continuous responses. We extend the allocation strategy to constructing response-adaptive repeated measurement designs for binary responses. The approach with binary responses is quite different from the continuous case, as the information matrix is a function of responses, and it involves non-linear modeling. To deal with these problems, we first build the design based on success probabilities. Then we illustrate how various models can accommodate carryover effects based on logits of response profiles as well as any correlation structure. Through computer simulations, we find that the allocation strategy developed for continuous responses also works well for binary responses. As expected, design efficiency in terms of mean squared error drops sharply, as more emphasis is placed on increasing treatment benefit than estimation precision. However, we find that it can successfully allocate more patients to better treatment sequences without sacrificing much estimation precision.

5.1 Introduction

Standard randomized designs, such as completely randomized controlled trials or crossover designs, usually employ a simple randomization scheme that equally allocates study subjects to each treatment group or treatment sequence. Equal randomization often allows statistically efficient and powerful comparisons of treatment effects. However, equal treatment allocation may

pose ethical concerns, especially in clinical settings where a treatment begins to show clearly inferior or superior. It is unethical when the needs of patients in the study come second to a quest for balanced statistically optimal data. To cope with such situations, alternative designs with adaptive allocation schemes have been advocated (Hu and Rosenberger, 2006). For example, in a play-the-winner rule design, using a type of response-adaptive design, we can modify the treatment allocation rule based on patients' responses already accrued in the trial. This approach allows assigning more patients to beneficial treatment groups (Zelen, 1969; Wei and Durham, 1978). In addition, a response-adaptive randomization procedure can be easily implemented by adding a loop to a standard randomization routine, and the procedure can be simulated under various parameterizations to determine the appropriateness for use in clinical trials (Hu and Rosenberger, 2006).

There are various types of response-adaptive designs, depending on the goals of a particular study. Adaptive treatment allocation is typically used to fulfill a single objective such as increasing the number of patients assigned to the eventual beneficial treatment group (Zelen, 1969; Wei and Durham, 1978; Armitage, 1975; Kushner, 2003; Liang and Carriere, 2009), reducing the sample size in a trial (Liang and Carriere, 2012), or increasing the estimation precision of a treatment effect (Kunert and Stufken, 2008). One could incorporate covariates into single-objective response-adaptive designs (Kunert and Stufken, 2002). To conduct optimal designs satisfying multiple objectives, two standard approaches have been used: constrained and compound optimal designs. These designs are essentially equivalent (Afsarinejad and Hedayat, 2002; Mehtala, Auranen and Kulathinal, 2011). Some investigators have developed methods of obtaining multiple objective designs for non-

longitudinal settings (Bandyopadhyay, Biswas and Mukherjee, 2011; Biswas and Dewanji, 2004; Biswas, Park and Bhattacharya, 2010). Recently, Liang and Carriere (2009) developed a new treatment allocation scheme to construct multiple-objective response-adaptive repeated measurement designs (RMD), where study subjects can receive two or more treatments (not necessarily the same treatments) over a time period. Their adaptive allocation strategy can simultaneously achieve two objectives: potentially preventing patients from being exposed to inferior treatments and enhancing the precision of the estimates of parameters. They extensively discussed the new adaptive allocation rule for multiple-objective response-adaptive designs, and constructed practical applications for two- and three-period RMDs with continuous responses (Liang and Carriere, 2009).

The primary goal of this chapter is to investigate the applicability of the same adaptive allocation strategy for two-treatment RMDs with binary responses, and to deliberate if the same strategy would work well for binary outcomes. The utility of the new proposed multiple-objective response-adaptive RMDs will be demonstrated on several practically useful designs with two, three, and four periods.

This chapter is organized as follows. Section 2 briefly reviews recent work on multiple-objective response-adaptive RMDs with continuous responses. Section 3 illustrates the allocation approach with binary responses and the assumptions we employ. Section 4 assumes equal carryover effects model and constructs the two-treatment two-, three- and four-period multiple-objective response-adaptive designs in computer simulations. Section 5 assumes unequal carryover effects model and constructs the multiple-objective response-adaptive designs in computer simulation for two treatments, two and three

periods. A numerical example is given in Section 6. Finally, we give our conclusions and suggestions for future work in Section 7.

5.2 Multiple-Objective Response-Adaptive RMD

This section briefly introduces methodologies and assumptions used in constructing multiple-objective response-adaptive RMDs for continuous outcomes under the self- and mixed-carryover effects model with random subject effects (Liang and Carriere 2009). In a clinical trial, we desire a randomized allocation scheme to simultaneously achieve a dual objective: increasing the precision of estimation and decreasing the number of patients allocated to unfavorable treatments. We use the information matrix and an evaluation function, respectively, to assess and evaluate the two goals.

Given the assignment history and the responses of the first j patients, a treatment sequence k will be assigned to the $(j+1)^{th}$ patient such that it maximizes the following assignment criterion:

$$\Phi_{j,k} = \lambda \frac{\Delta(\widehat{\mathbf{I}}^{k}_{j+1}(\theta))}{\Delta(\widehat{\mathbf{I}}^{k'}_{j+1}(\theta))} + (1-\lambda)\frac{f_{j,k}}{f_{j,k''}}, \tag{5.1}$$

where $\widehat{\mathbf{I}}^{k}_{j+1}(\theta)$ represents the estimated information matrix for θ, a vector of parameters of interest, given the information of the first j patients' responses, and the assumption that the $(j+1)^{th}$ patient will be assigned to a treatment sequence k (Liang and Carriere, 2009). The $\Delta(.)$ can be any optimality criterion, for instance, the determinant (D-optimality), the trace (A-optimality), or the maximum eigenvalue (E-optimality) of a matrix. The $f_{j,k}$ is an evalu-

ation function for treatment sequence k based on the responses of the first j patients. Here, $k^{'}$ denotes the treatment sequence that maximizes $\Delta(.)$, and $k^{''}$ denotes the treatment sequence that maximizes $f_{j,k}$. These $k^{'}$ and $k^{''}$ are not necessarily the same.

In Equation (5.1), the first term is to find a treatment sequence that could maximize the information matrix, while the second term is to detect a treatment sequence that performs best clinically given the first j patients' responses. Thus, the coefficient λ, a constant between zero and one, can weigh and balance the two objectives. For example, if λ takes the value of one, the resulting treatment sequence will achieve the most precise estimation but ignore any treatment advantages (Kushner, 2003). If λ is set to zero, we are only concerned with the efficacy of the treatments as measured by a pre-selected evaluation function (Liang and Carriere, 2009). Therefore, by choosing a suitable value of λ prior to the trial, we can find a treatment sequence that balances the two objectives and maximizes the value of Φ, as the best treatment sequence for the new $(j+1)^{th}$ patient. This procedure will continue until all subjects are allocated. Liang and Carriere (2012) recently expanded the allocation approach to allocate multiple subjects at the same time. We will consider allocating one patient at a time to achieve the two objectives.

5.3 Allocation Scheme with Binary Responses

We begin by assuming that the patients' dichotomous responses (success or failure) are independently distributed with a Bernoulli distribution. With repeated measurements within subjects, the chances of success or failure are likely to be correlated from one period to the next. Further, some residual

treatment effects may still remain in subsequent periods. We shall describe how we deal with these two issues - correlation in repeated measures and carryover effects.

With binary responses, the approach is quite different from the continuous case. First, we recognize that the eventual data analysis involves non-linear models. Second, due to the special structure of binary data, designs for binary responses depend on the outcomes, unlike those for continuous responses. Hence, rather than trying to capture the expected mean responses by devising a suitable crossover design model, we separate the two issues of building a design and data analysis, and work with the success probabilities of treatments measured by the binary outcomes. Then any carryover effects can be accommodated in a model for the data analysis as well as any correlations among repeated measures. Basically, our approach to building an adaptive design for binary responses follows the modeling strategies illustrated in Kenward and Jones (1987).

5.3.1 Adaptive Design Construction

Let $\delta_{ti} = (\delta_{ti1}, \delta_{ti2}, ..., \delta_{tij}, ..., \delta_{tiJ})'$ be a vector of treatment indicators given in period i for all J subjects who have been treated in the study. t is the treatment index, where $t = A$ or B in a two-treatment design, i is the period index with $i = 1, 2, ...p$, where p is the number of periods, j is the subject index with $j = 1, 2, ...J$, where J is the total number of subjects treated. Let δ_{tij}=1 if the j^{th} patient receiving treatment t in period i and δ_{tij}=0 otherwise. Let y_{tij} be 1 if the treatment given in period i for subject j is a success and 0 otherwise. Let $m_{ti} = \Sigma_{j=1}^{J} \delta_{tij}$ be the number of patients receiving treatment

t in period i, and let $S_{ti} = \Sigma_{j=1}^{J} y_{tij}\delta_{tij}$ be the number of successes in period i for treatment t.

Let ν_{ti} be the maximum likelihood estimator of the probability of success for treatment t in period i. In a two-treatment p-period design, ν_{ti} would be $\nu_{A1}, \nu_{A2}, ..., \nu_{Ap}, \nu_{B1}, ...,$ or ν_{Bp}. In our model, the success probability of a given treatment in one period may differ from that in a different period because of treatment-by-period interaction effect or carryover effect, which is similar to the model that was considered by Bandyopadhyay et al.(2009) in their two-period two-treatment repeated measurement designs. Assume that ν_{ti}s are independent. The likelihood function based on all responses after the first J patients have been treated in the trial is:

$$\begin{aligned} L_J &= \Pi_t \Pi_{i=1}^{p} \Pi_{j=1}^{J} [\nu_{ti}^{y_{tij}} (1-\nu_{ti})^{1-y_{tij}}]^{\delta_{tij}} \\ &= \Pi_t \Pi_{i=1}^{p} [\nu_{ti}^{S_{ti}} (1-\nu_{ti})^{m_{ti}-S_{ti}}]. \end{aligned}$$

The log-likelihood function is written as:

$$log(L_J) = \sum_{t} \sum_{i=1}^{p} [S_{ti} log(\nu_{ti}) + (m_{ti} - S_{ti}) log(1-\nu_{ti})].$$

Equivalently, in a two-treatment p-period design, if we let $l = 1, 2, ..., 2p$, the log-likelihood can be rewritten as:

$$log(L_J) = \sum_{l=1}^{2p} [S_l log(\nu_l) + (m_l - S_l) log(1-\nu_l)],$$

where S_l denotes the l^{th} element of $\mathbf{S} = (S_{A1}, ..., S_{Ap}, S_{B1}, ..., S_{Bp})'$, the set of number of successes with treatment A and B in each period. For $l \leq p$, S_l

denotes the number of successes for treatment A in period l, and for $l > p$, S_l denotes the number of successes for treatment B in period $(l-p)$. For instance, when $l = p+1$, $S_l = S_{B1}$ represents the number of successes for treatment B in period 1. The ν_l and m_l are the corresponding success probability and the number of patients, respectively, that is, the l^{th} element from $\nu = (\nu_{A1}, ..., \nu_{Ap}, \nu_{B1}, ..., \nu_{Bp})'$ and $\mathbf{m} = (m_{A1}, ..., m_{Ap}, m_{B1}, ..., m_{Bp})'$, respectively.

Let ν (a vector of successes probabilities) be the vector of the parameters of interest. Then the expected information matrix for ν, based on the first J patients, is a $2p \times 2p$ diagonal matrix:

$$\mathbf{I}_J(\nu) = Diag[\mathbf{E}(\frac{S_l}{\nu_l^2} + \frac{m_l - S_l}{(1-\nu_l)^2})], \;\; 1 \leq l \leq 2p.$$

Under the assumption that the $(J+1)^{th}$ patient will be given treatment sequence k, the estimated information matrix conditionally on the first J patients then becomes:

$$\widehat{\mathbf{I}}_{J+1}(\widehat{\nu}) = Diag[\frac{S_l^k}{\widehat{\nu}_l^2} + \frac{m_l^k - S_l^k}{(1-\widehat{\nu}_l)^2}], \;\; 1 \leq l \leq 2p, \tag{5.2}$$

where S_l^k=$S_l + \widehat{\nu}_l {d_l}^k$, ${m_l}^k = m_l + {d_l}^k$, $\widehat{\nu}$=$(\widehat{\nu}_{A1}, ...\widehat{\nu}_{Bp})$, and ${d_l}^k$ is the l^{th} element from $\mathbf{d}^k$, a set of indicator variables of length $2p$ for treatment sequence k. As an example, $\mathbf{d}^k = (1,0,1,0,1,0)'$ corresponds to treatment sequence $k = ABA$ in a two-treatment three-period RMD with treatment A applied in periods 1 and 3 and treatment B applied in period 2.

In our setup, an evaluation function $f_{J,k}$ for a treatment sequence k based on the first J patients, is defined as the average total number of successes

over all subjects receiving treatment sequence k, as in Example 1 of Liang and Carriere (2009). For example, in a two-period design, the range of the evaluation function is 0 (no success over both periods for all subjects receiving a particular treatment sequence) to 2 (two successive successes over both periods for all subjects receiving a particular treatment sequence). That is:

$$f_{J,k} = \frac{\mathbf{d}^{k'} \times \mathbf{S}}{n_k}, \tag{5.3}$$

where n_k=$\mathbf{d}^{k'} \times \mathbf{m}$ is the number of patients receiving treatment sequence k.

We are now ready to propose and define the optimal adaptive allocation, which will be demonstrated in simulation.

Proposition 1. With the information matrix (5.2) and the evaluation function (5.3) for binary responses, the $(J+1)^{th}$ patient will be allocated to a treatment sequence that maximizes the selection criterion (5.1), while accomplishing the two objectives.

Without loss of generality, the criterion of D-optimality is used in Equation (5.1). Similar to the case of continuous responses, this procedure will continue until all subjects are allocated. Note that how we incorporate carryover effects is completely determined by ν_{ti} for $i = 2, ...p$ in $\nu = (\nu_{A1}, ..., \nu_{Ap}, \nu_{B1}, ..., \nu_{Bp})'$. For example, in a two-period design, if $\nu_{A2} \neq \nu_{B2}$, it corresponds to the model with unequal carryover effects. If $\nu_{A2} = \nu_{B2}$, carryover effects are expected to be equal or ignorable in the data analysis from the resulting design (Bandyopadhyay, Biswas and Bhattacharya 2009). Therefore, our strategy builds the desired design flexibly based on what is expected of carryover effects in the experiments. This is unlike the continuous responses, where the design is

completely dependent on a chosen model.

5.3.2 Design Efficiency

There are various ways to assess the efficiency of a constructed design. As we are concerned with the error and possible bias in the estimation of parameters of interests, θ , we compute a matrix of mean squared error (MSE) for θ as:

$$MSE = E[(\widehat{\theta} - \theta)(\widehat{\theta} - \theta)^T]$$

where $\widehat{\theta}$ is an estimator of θ. In simulation studies, it is often estimated from results of repeated B runs as:

$$\widehat{MSE} = \frac{\sum_{b=1}^{B}(\widehat{\theta_b} - \theta)(\widehat{\theta_b} - \theta)^T}{B}, \tag{5.4}$$

where $\widehat{\theta_b}$ is the maximum likelihood estimate of θ obtained in the b^{th} simulation. The relative efficiency (RE) of a design is defined relatively to a reference design by comparing their respective MSEs. By denoting MSE_1 and MSE_0 as the matrix of MSE for a proposed adaptive design and the reference design, we compare their RE in terms of A-, D-, or E-optimality criteria based on Kiefer (1975), respectively:

$$RE_A = \frac{trace(MSE_0)}{trace(MSE_1)}, \quad RE_D = \frac{|MSE_0|}{|MSE_1|}, \quad or \quad RE_E = \frac{maxeigenvalue(MSE_0)}{maxeigenvalue(MSE_1)},$$

When $RE < 1$, the adaptive design is less efficient than the reference design in terms of the MSE, but we note that this measure ignores the treatment benefit for the patients. Further, reference designs are not necessarily the

optimal designs. We do not yet know what the optimal designs are for binary responses. In this paper, we shall use the D-optimality criteria for estimating $\theta = \nu$, the success probabilities, and use the designs that are known to be optimal as the reference designs, although they were constructed for continuous responses.

5.4 Response-Adaptive RMD for Binary Responses with equal carryover effects

5.4.1 Models

In this subsection, we briefly discuss how carryover treatment effects can be accommodated in the analysis. In our adaptive design construction with binary responses, the designs are built based on the success probability specified for each period, ν_{ti}, $i = 1, ..., p$ and $t = 1(A), 2(B)$, as we specify $\nu = (\nu_{A1}, ..., \nu_{Ap}, \nu_{B1}, ..., \nu_{Bp})'$. Below, we will give an example of possible models based on an AB/BA design.

Specifically, let Y_{ijk} represent the response from the j^{th} subject in group k in period i, where $k = 1(AB), 2(BA)$, $i = 1, 2$, $j = 1, ..., n_k$, and n_k is the number of subjects in group k. Then, we can build the logit of the probabilities of a success as $logitP(Y_{1j1} = 1) = \mu+\pi_1+\tau_1$, $logitP(Y_{2i1} = 1) = \mu+\pi_2+\tau_2+\gamma_1$, $logitP(Y_{1j2} = 1) = \mu + \pi_1 + \tau_2$, and $logitP(Y_{2j2} = 1) = \mu + \pi_2 + \tau_1 + \gamma_2$ for the general mean μ, the i^{th} period effect π_i, the direct and residual treatment effects due to treatment t, τ_t and γ_t. This results in 2×4 response profiles of $(y_{1jk}, y_{2jk}) \in$ ((0,0), (0,1), (1,0), (1,1)) for all j from each treatment sequence

group and the overall likelihood becomes:

$$L = \prod_{j=1}^{n_k} \prod^{2} P(Y_{1jk} = y_{1jk}, Y_{2jk} = y_{2jk}). \tag{5.5}$$

Kenward and Jones (1987) assume an independence model for data from two periods, while Ezzet and Whitehead (1992) consider a random effects model to accommodate correlated binary responses over two periods. Hence, various forms of carryover effects and possible dependence structures can be considered for data analysis. In this section, we shall focus on constructing adaptive designs.

5.4.2 Constructing Response-Adaptive RMD

We construct two-, three-, and four-period response-adaptive designs to compare two treatments. We simulate various response-adaptive RMDs with binary responses (1=success, 0=failure) under different parameterizations. For each parameter setting, we perform 5,000 simulations to smooth out randomness. We consider λ=1, 0.9, 0.7, 0.3, and 0, and the total number of patients in a trial, N=40, 80, and 100. Table (5.1) gives the detailed parameter settings in the presence of a treatment difference between treatments A and B. When the treatment difference is absent, we assume all ν_{ti}s, the probability of success for treatment t in period i, are equal to 0.5. The D-optimality of a matrix of mean squared error for a vector of success probabilities is used to assess the efficiency of an adaptive design. A relative efficiency is computed to compare an adaptive design with a fixed reference design with a relative efficiency greater than 1, indicating that the adaptive design is more efficient than the reference design (Liang and Carriere 2009).

Table 5.1: Parameter setting for constructing adaptive designs

Design	Success Probabilities	Treatment sequence	Expected responses	Expected sum of of success probability	Order of performance
Two-treatment two-period	$\nu_{A1} = 0.5$	AA	$(0.5, 0.6)'$	1.1	1
	$\nu_{A2} = 0.6$	AB	$(0.5, 0.3)'$	0.8	3
	$\nu_{B1} = 0.4$	BB	$(0.4, 0.3)'$	0.7	4
	$\nu_{B2} = 0.3$	BA	$(0.4, 0.6)'$	1.0	2
Two-treatment three-period	$\nu_{A1} = 0.5$	AAA	$(0.5, 0.6, 0.7)'$	1.8	1
	$\nu_{A2} = 0.6$	AAB	$(0.5, 0.6, 0.2)'$	1.3	5
	$\nu_{A3} = 0.7$	ABA	$(0.5, 0.3, 0.7)'$	1.5	3
	$\nu_{B1} = 0.4$	ABB	$(0.5, 0.3, 0.2)'$	1.0	7
	$\nu_{B2} = 0.3$	BBB	$(0.4, 0.3, 0.2)'$	0.9	8
	$\nu_{B3} = 0.2$	BBA	$(0.4, 0.3, 0.7)'$	1.4	4
		BAB	$(0.4, 0.6, 0.2)'$	1.2	6
		BAA	$(0.4, 0.6, 0.7)'$	1.7	2
Two-treatment four-period	$\nu_{A1} = 0.5$	ABAA	$(0.5, 0.3, 0.7, 0.8)'$	2.3	1
	$\nu_{A2} = 0.6$	ABBA	$(0.5, 0.3, 0.2, 0.8)'$	1.8	4
	$\nu_{A3} = 0.7$	AABA	$(0.5, 0.6, 0.2, 0.8)'$	2.1	3
	$\nu_{A4} = 0.8$	AABB	$(0.5, 0.6, 0.2, 0.1)'$	1.4	7
	$\nu_{B1} = 0.4$	BABB	$(0.4, 0.6, 0.2, 0.1)'$	1.3	8
	$\nu_{B2} = 0.3$	BAAB	$(0.4, 0.6, 0.7, 0.1)'$	1.8	4
	$\nu_{B3} = 0.2$	BBAB	$(0.4, 0.3, 0.7, 0.1)'$	1.5	6
	$\nu_{B4} = 0.1$	BBAA	$(0.4, 0.3, 0.7, 0.8)'$	2.2	2

Note: The ν_{ti} denotes the probability of success (favorable outcome) for treatment t given in period i. According to Equation (3), the value of evaluation function increases by increasing the sum of successes, equivalently the sum of success probabilities. Consequently, the performance of each treatment sequence is ordered according to the second last column.

In our simulation, an evaluation function for a treatment sequence is defined as the average total number of successes over all subjects receiving a particular treatment sequence, as described in Section 3. Table (5.1) (last column) summarizes the order of the performance of each treatment sequence (with 1 being the best) under the given evaluation function for each parameter setting with unequal treatment effects. For example, in the simulated two-period design, treatment sequence AA is the best of the four treatment sequences, followed by BA, AB and BB in that order. When there are no treatment differences, the allocation is uniform to all sequences and identical to the case of $\lambda = 1$, as expected. Therefore, in this section, we present the detailed simulation results for the case of unequal treatment parameter settings

only.

5.4.3 Two-Period Designs

In two-period two-treatment designs, four different treatment sequences are available for assignment. In the simulation, the first four subjects are randomly assigned, one for each treatment sequence. The rest of the subjects are allocated adaptively. Table (5.2) shows the average number of subjects assigned to each of the four treatment sequences. When the treatment difference is absent, the strategies with all λ assign an approximately equal number of subjects to each of the four treatment sequences. Even when the treatment difference is present, the patterns for $\lambda = 1$ are similar to the equal success probability case, with equal numbers of subjects assigned to a dual block, as expected. However, with $\lambda < 1$, adaptive designs successfully recognize and assign more subjects to the best treatment sequence AA and fewer subjects to the worst treatment sequence BB. As λ approaches 0 and decreases, the proportion of subjects receiving the best treatment increases; whereas the proportion of subjects receiving the worst treatment decreases.

The estimated success probabilities of treatments in different periods, $\hat{\nu}_{ti}$s, and the estimation precision are also calculated for both absence and presence of a treatment difference. The results indicate that the spread of the estimates of ν_{ti}s decreases when the total number of patients increases, consistent with the setting (results not shown). Figure (5.1) (first plot) illustrates the relative efficiency of an adaptive design to a reference design, the four-treatment-sequence design with AA/AB and their duals, which is optimal for continuous responses. It shows that adaptive designs with a large λ are almost

Table 5.2: Results of the adaptive allocation of subjects for each treatment sequence for p=2

N	λ	N_{AA}	N_{AB}	N_{BB}	N_{BA}
40	1	10.04	9.96	10.04	9.96
	0.9	12.93	8.41	8.12	10.54
	0.7	15.76	6.87	5.25	12.12
	0.3	16.9	6.17	4.56	12.37
	0	17.5	5.9	4.39	12.22
80	1	20.47	19.53	20.47	19.53
	0.9	29.29	14.83	12.63	23.25
	0.7	38.17	10.13	6.63	25.06
	0.3	38.66	10.03	6.24	25.07
	0	38.64	9.96	6.29	25.11
100	1	25.43	24.57	25.43	24.57
	0.9	39.36	17.09	14.47	29.08
	0.7	49.01	10.82	7.09	33.08
	0.3	49.59	11.61	7.17	31.63
	0	49.99	11.15	7.3	31.56

Note: Entries are the average number of patients allocated to each sequence at the end of the trial for the situation given in Table (5.1).

as efficient as the reference design, but they allocate more subjects to better treatment sequences. As expected, adaptive designs with a small value of λ are not efficient according to the purely statistically measure of conventional efficiency based on the D-optimality of a matrix of mean squared error for $\nu = (\nu_{A1}, \nu_{A2}, \nu_{B1}, \nu_{B2})'$, but they recognize the favorable performance of a treatment sequence.

5.4.4 Three-Period Designs

In two-treatment three-period RMDs, eight different treatment sequences are available. At the initial stage, eight subjects are assumed to enter in the study, one for each treatment sequence. As with the two-period designs, we also

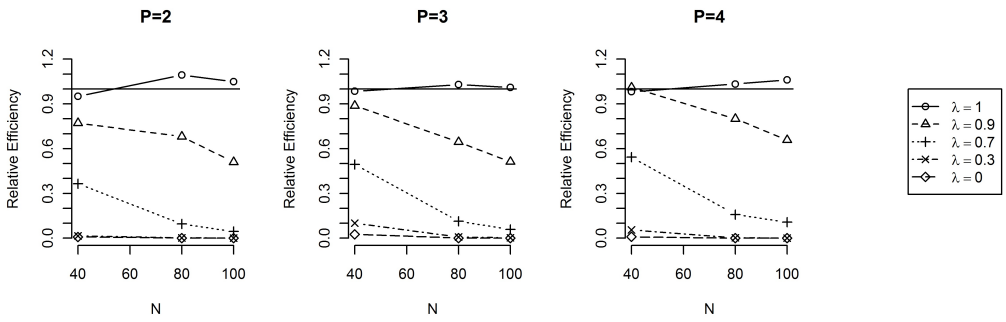

Figure 5.1: Relative efficiencies of adaptive designs to the fixed designs

Note: The solid lines (relative efficiency=1) indicate the reference design, which are fixed optimal designs for continuous responses with AA/AB and their duals for p=2; ABB/BAA for p=3, and $ABBA/AABB$ and their duals for p=4.

consider two situations - treatments A and B have an equal success probability in all periods, i.e. $\nu_{ti} = 0.5$, for $t = A$ or B and $i = 1, 2$, and 3; and there are treatment differences between the two treatments. According to the given parameter setting in Table (5.1), the treatment sequence AAA has the largest total number of successes, followed by BAA, ABA, BBA, AAB, BAB, ABB, and BBB. Therefore, we expect that our adaptive allocation approach would allocate most subjects to the best treatment sequence AAA , and the least number of subjects to the worst treatment sequence BBB.

Table (5.3) summarizes the allocation results for the situation given in Table (5.1). We note that, when $\lambda = 0$, the allocation results are consistent with the order of treatment performances as defined by the given evaluation function, as expected. In general, the allocation is successful with an approximately equal number of subjects assigned to each of the eight treatment

sequences in the absence of treatment differences and the most subjects to the best treatment sequence AAA and the least to the worst treatment sequence BBB in the presence of treatment difference. When $\lambda = 1$, the allocation pays no attention to differential treatment effects even if there were any, and it treats all sequences equally, as expected. As λ decreases, more and more subjects are assigned to the best treatment sequence AAA. The choice of λ will depend on how much the investigator is willing to sacrifice one of the study objectives (Liang and Carriere 2009). The results become more pronounced as the sample size increases.

Table 5.3: Results of the adaptive allocation of subjects for each treatment sequence for p=3

N	λ	N_{AAA}	N_{AAB}	N_{ABA}	N_{ABB}	N_{BBB}	N_{BBA}	N_{BAB}	N_{BAA}
40	1	5.00	5.01	4.97	5.02	5.01	5.00	4.97	5.03
	0.9	7.11	4.59	5.24	3.88	3.71	4.84	4.32	6.32
	0.7	9.71	4.00	5.54	2.61	2.40	4.25	3.37	8.13
	0.3	11.82	3.41	5.36	1.92	1.72	4.10	2.76	8.91
	0	12.60	3.20	5.42	1.68	1.47	4.00	2.38	9.25
80	1	10.66	9.66	9.65	10.03	10.65	9.67	9.66	10.02
	0.9	17.90	8.17	10.67	5.99	5.69	9.35	7.09	15.13
	0.7	25.91	5.94	10.63	3.19	2.73	7.56	4.58	19.46
	0.3	28.85	5.26	10.12	2.10	1.82	6.98	3.27	21.60
	0	30.02	4.65	10.27	1.83	1.54	7.01	3.27	21.40
100	1	13.16	12.17	12.06	12.61	13.13	12.19	12.09	12.59
	0.9	24.34	9.63	13.44	6.67	6.38	11.32	8.23	19.99
	0.7	35.83	6.40	12.89	3.15	2.94	8.81	4.70	25.28
	0.3	38.45	5.80	12.61	2.14	1.77	8.43	3.98	26.82
	0	38.39	5.46	12.19	1.90	1.58	8.32	3.90	28.25

Note: See notes for Table (5.2).

Figure (5.1) (second plot) illustrates the relative efficiency of an adaptive design to the fixed reference design ABB/BAA, the optimal design for con-

tinuous responses. Note that when there are no treatment differences, the adaptive design uses all available treatment sequences equally. However, the adaptive design is not better or worse than the fixed optimal design using just two sequences ABB/BAA in terms of MSE, because there are no treatment differences anyway. When there are treatment differences, a clinically beneficial approach is to allocate more subjects to the best sequence, which is AAA. Similar to the two-period designs, the adaptive designs with $\lambda < 1$ have low efficiency compared to the reference design. However, the reference design is not necessarily the optimal choice in the presence of treatment difference. Nonetheless, adaptive designs with a large $\lambda < 1$ achieves a high enough efficiency for $n = 40$. For ethical reasons, we recommend adaptive designs with a large $\lambda(< 1)$.

5.4.5 Four-Period Designs

As the number of periods increases, so does the number of possible treatment sequences to consider for inclusion in the study. However, multiple-period designs could quickly become impractical and having too many treatment sequences could become difficult to administer. For two-treatment four-period designs, there are 16 ($=2^4$) possible different treatment sequences to compose a RMD. To demonstrate the utility of the strategy, we considered eight of these 16 sequences ($ABAA/ABBA/AABA/AABB$ and their duals). The rational is that these eight treatment sequences have been identified and included in the optimal designs in a variety of settings. Kunert and Stufken (2002, 2008) noted that, for continuous responses, the design $ABBA/AABB$ and their duals is the optimal design under the traditional model with a simple

first-order carryover effect, and the design $AABB/AABA$ and their duals, the design $ABAA/ABBA$ and their duals, and the design $ABBA/AABA$ and their duals are optimal designs under the self- and mixed-carryover effects model (Afsarinejad and Hedayat 2002).

As in previous simulations, we consider two scenarios (absence and presence of the treatment difference). The values of success probabilities are chosen arbitrarily to distinguish the effectiveness of the treatment sequences included. Based on Table (5.1), we would expect $ABAA$ to perform the best followed by $BBAA, AABA, ABBA/BAAB, BBAB, AABB$ and $BABB$. The performance of treatment sequences $ABBA$ and $BAAB$ is indistinguishable because they have the same value as the evaluation function.

Table 5.4: Results of the adaptive allocation of subjects for each treatment sequence for p=4

N	λ	N_{ABAA}	N_{ABBA}	N_{AABA}	N_{AABB}	N_{BABB}	N_{BAAB}	N_{BBAB}	N_{BBAA}
40	1	4.42	4.51	5.16	5.91	4.47	4.43	6.05	5.05
	0.9	5.77	3.89	5.95	5.08	3.60	4.98	4.89	5.84
	0.7	7.74	3.57	6.38	3.79	2.88	5.08	3.81	6.75
	0.3	10.32	3.27	6.75	2.28	2.01	4.59	2.47	8.31
	0	12.71	3.12	6.95	1.51	1.39	3.23	1.76	9.33
80	1	9.33	9.22	10.77	10.88	9.12	9.22	10.93	10.72
	0.9	15.06	6.02	14.73	7.32	5.84	9.87	6.92	14.24
	0.7	22.25	5.42	14.49	4.45	3.36	7.98	4.52	17.53
	0.3	26.91	4.68	14.59	2.45	2.11	6.16	2.66	20.45
	0	29.88	4.42	14.72	1.71	1.49	4.66	1.98	21.12
100	1	11.67	11.63	13.35	13.3	11.67	11.63	13.39	13.31
	0.9	19.99	7.27	19.10	8.24	6.55	12.36	7.52	18.97
	0.7	29.71	6.21	19.03	4.55	3.49	8.81	4.75	23.45
	0.3	35.93	5.43	17.74	2.52	2.09	7.01	2.79	26.48
	0	37.74	5.07	18.42	1.68	1.47	4.73	2.02	28.88

Note: See notes for Table (5.2).

Table (5.4) summarizes the average allocation results. Similar to the two-

and three-period designs, when $\lambda = 0$, the allocation results are consistent with the order of treatment performances as indicated in Table (5.1). When $\lambda = 1$ or in the absence of treatment effect differences, it allocates an approximately equal number of subjects to each of the eight treatment sequences, as it would not matter anyway which sequences are used. When $\lambda < 1$, the allocation recognizes the treatment differences and allocates the most to the best treatment sequence $ABAA$, and the least to the worst treatment sequence $BABB$. As $\lambda < 1$ and decreases, more and more subjects are assigned to the best treatment sequence $ABAA$ in the presence of treatment difference, as expected.

Figure (5.1) (third plot) illustrates the relative efficiency of an adaptive design and the fixed reference design $ABBA/AABB$ and their duals under the D-optimality. Similar to the two- and three-period designs, the adaptive design with $\lambda < 1$ has low statistical efficiency compared to the reference design. However, adaptive designs with a large $\lambda < 1$ can be almost as efficient as the fixed reference design, but it allocates more patients to the beneficial treatment sequences.

5.5 Response-Adaptive RMD with unequal carryover effects

5.5.1 Models

In Section 5.4, we construct the adaptive design under a model with success probabilities $\nu = (\nu_{A1}, \nu_{A2}, \cdots, \nu_{Ap}, \nu_{B1}, \nu_{B2}, \cdots, \nu_{Bp})$, where ν_{ti} is the probability of success for treatment t in period i. Consider the model in a two-period

design, then $\nu = (\nu_{A1}, \nu_{A2}, \nu_{B1}, \nu_{B2})$. As ν_{t2} does not rely on the treatment assigned in the first period, the model assumes no carryover effects or assumes only identical carryover effects of treatment A and B. To extend the model so that the carryover effects can be accommodated, we consider the model $\nu = (\nu_{A1}, \nu_{A2|A1}, \nu_{A2|B1}, \nu_{B1}, \nu_{B2|A1}, \nu_{B2|B1})$ in a two-period design. In the model, $\nu_{ti|t'(i-1)}$ is defined so that the probability of success relies on the treatment assigned in the previous period, except the probability in the first period. The index $ti|t'(i-1)$ indicates that treatment t= A or B is assign at ith period with t'=A or B assigned at $(i-1)$th period.

The carryover effect contrast between treatment A and B can be evaluated by $(\nu_{A2|A1} - \nu_{A2|B1} + \nu_{B2|A1} - \nu_{B2|B1})/4$. In a special case where $\nu_{A2|A1} = \nu_{A2|B1}$ and $\nu_{B2|A1} = \nu_{B2|B1}$, the carryover effect contrast is 0 and the model is reduced to $\nu = (\nu_{A1}, \nu_{A2}, \nu_{B1}, \nu_{B2})$.

In the adaptive design construction, we need to redefine $\mathbf{S}$ and $\mathbf{m}$ to accommodate new parameters in the model. In a two-treatment p-period design, there are $p^* = 2p - 1$ parameters considered for each treatment in ν. E.g., for treatment A, we have $\nu_{A1}, \nu_{A2|A1}, \nu_{A2|B1}, \cdots, \nu_{Ap|A(p-1)}, \nu_{Ap|B(p-1)}$. We can have $\mathbf{S}$ and $\mathbf{m}$ with the same index in ν, so that accordingly, we define S_{A1} and $S_{Ai|t(i-1)}$ as the number of successes for treatment A at period 1 and for treatment A at period i which is also in a subsequence of tA at periods $i-1, i$; we define m_{A1} and $m_{Ai|t(i-1)}$ as the number of patients assigned with treatment A at period 1 and assigned with tA at period $i-1, i$.

Let ν_l, S_l and m_l are the lth elements in ν, $\mathbf{S}$ and $\mathbf{m}$, where $l = 1, 2, \cdots, 2p^*$. Similar as in Section 5.4, we can construct the log-likelihood function for ν

based on the first J patients, which can be written as:

$$log(L_J) = \sum_{l=1}^{2p^*}[S_l log\nu_l + (m_l - S_l)log(1-\nu_l)],$$

and the expected information matrix for ν, based on the first J patients, is a $2p^* \times 2p^*$ diagonal matrix:

$$\begin{aligned}\mathbf{I}_J(\nu) &= Diag[\mathbf{E}(\frac{S_l}{\nu_l^2} + \frac{m_l - S_l}{(1-\nu_l)^2})] \\ &= Diag[\frac{m_l}{\nu_l(1-\nu_l)}]. \end{aligned} \tag{5.6}$$

With the same evaluation function $f_{J,k}$, the optimal adaptive allocation can be achieved by maximizing the selection criterion (5.1), while the two objectives are accomplished.

The procedure will continue until all subjects are allocated. Note that how we incorporate carryover effects is completely determined by ν. For example, in a two-period design, if $\nu_{A2|A1} = \nu_{A2|B1}$ and $\nu_{B2|A1} = \nu_{B2|B1}$, carryover effects are expected to be equal or ignorable in the data analysis from the resulting design (Bandyopadhyay, Biswas and Bhattacharya 2009). Therefore, our strategy builds the desired design flexibly based on what is expected of carryover effects in the experiments. This is unlike the continuous responses, where the design is completely dependent on a chosen model.

5.5.2 Constructing Response-Adaptive RMD

In this section, we construct two-, three-period response-adaptive designs to compare two treatments under the success probability model considering treatment-specific carryover effects. Similarly as in Section 5.4, we simulate

response-adaptive RMDs with binary responses in which the total number of patients considered is N=40, 80, and 100. Table (5.5) gives the detailed parameter settings. We still construct the design for λ=1, 0.9, 0.7, 0.3, and 0. Since the adaptive design in the case when $\lambda = 1$ puts all weight on the estimation but no weight on performance, the design would achieve the best estimation among the five cases. A relative efficiency is computed to compare the adaptive designs with different λ's to the case when $\lambda = 1$.

An evaluation function for a treatment sequence is defined as the same as in Section 5.4. Table (5.5) (last column) summarizes the performance of each treatment sequence with the average success probability per period. For example, in the simulated two-period design, treatment sequence AA has $\nu_{A1} = 0.6$ and $\nu_{A2|A1} = 0.7$ for the success probabilities of the two periods. The success rate for AA is 0.65 per period. An average success rate is calculated for all the sequences in each adaptive design in order to compare them in the performance of design.

5.5.3 Two-Period Designs

Similar to the simulations in Section 5.4, the first four subjects are randomly assigned, one for each treatment sequence. The rest of the subjects are allocated adaptively according to (5.1). A slight change has been made in the calculation of $\mathbf{I}_J(\nu)$, where we use the expected information 5.6. This change removes the randomness in the construction of the adaptive designs.

Table (5.6) shows the expect number of subjects assigned to the four treatment sequences. As λ decreases, adaptive designs successfully recognize and assign more subjects to the best treatment sequence AA and fewer subjects to the

Table 5.5: Parameter setting for constructing adaptive designs when carryover effect is considered

Design	Success Probabilities	Treatment sequence	Expected responses	Expected success rate per period
Two-treatment two-period	$\nu_{A1} = 0.6$, $\nu_{B1} = 0.3$	AA	$(0.6, 0.7)'$	0.65
	$\nu_{A2\|A1} = 0.7$, $\nu_{B2\|A1} = 0.4$	AB	$(0.6, 0.4)'$	0.5
	$\nu_{A2\|B1} = 0.5$, $\nu_{B2\|B1} = 0.2$	BA	$(0.3, 0.5)'$	0.4
		BB	$(0.3, 0.2)'$	0.25
Two-treatment three-period	$\nu_{A1} = 0.6$, $\nu_{B1} = 0.3$	AAA	$(0.6, 0.7, 0.65)'$	0.975
	$\nu_{A2\|A1} = 0.7$, $\nu_{B2\|A1} = 0.4$	AAB	$(0.6, 0.7, 0.35)'$	0.825
	$\nu_{A2\|B1} = 0.5$, $\nu_{B2\|B1} = 0.2$	ABA	$(0.6, 0.4, 0.55)'$	0.775
	$\nu_{A3\|A2} = 0.65$, $\nu_{B3\|A2} = 0.35$	ABB	$(0.6, 0.4, 0.25)'$	0.625
	$\nu_{A3\|B2} = 0.55$, $\nu_{B3\|B2} = 0.25$	BAA	$(0.3, 0.5, 0.65)'$	0.725
		BAB	$(0.3, 0.5, 0.35)'$	0.575
		BBA	$(0.3, 0.2, 0.55)'$	0.525
		BBB	$(0.3, 0.2, 0.25)'$	0.375

Note: The ν_{t1} denotes the probability of success (favorable outcome) for treatment t given in period 1. The $\nu_{ti|t'(i-1)}$ denotes the probability of success for treatment t given in period i when treatment t' is assigned in period $i-1$. According to Equation 5.3, the value of the evaluation function increases with the sum of successes, equivalently the sum of success probabilities.

worst treatment sequence BB.

For each N, the column 'Relative efficiency' in the table illustrates the relative estimation precision of an adaptive design to the design in the case of $\lambda = 1$, where the estimation precision achieves its maximum among the 5 designs. The column 'Success rate' shows the average success probability per period for all the sequences in each design. We can see that adaptive designs with $\lambda = 0.9$ increase the success rate considerably without losing too much in the estimation efficiency. As expected, adaptive designs with a small value of λ are not efficient especially when N is large.

5.5.4 Three-Period Designs

According to the parameter settings for the success probabilities shown in Table (5.5), the treatment sequence AAA has the largest success rate, followed

Table 5.6: Results of the adaptive allocation of subjects for each treatment sequence for p=2

N	λ	N_{AA}	N_{AB}	N_{BA}	N_{BB}	Relative efficiency	Success rate
40	1	10	10	10	10	1.000	0.450
	0.9	13	10	8	9	0.962	0.476
	0.7	22	8	4	6	0.402	0.543
	0.3	36	2	1	1	0.004	0.626
	0	37	1	1	1	0.002	0.630
80	1	20	20	20	20	1.000	0.450
	0.9	32	20	12	16	0.745	0.503
	0.7	57	12	4	7	0.075	0.586
	0.3	76	2	1	1	0.000	0.638
	0	77	1	1	1	0.000	0.640
100	1	25	25	25	25	1.000	0.450
	0.9	43	25	14	18	0.642	0.512
	0.7	76	13	4	7	0.037	0.597
	0.3	96	2	1	1	0.000	0.641
	0	97	1	1	1	0.000	0.642

Note: Entries are the average number of patients allocated to each sequence at the end of the trial for the situation given in Table (5.5).

by AAB, ABA, BAA, ABB, BAB, BBA, and BBB. Therefore, we expect that our adaptive allocation approach would allocate most subjects to the best treatment sequence AAA, and the least number of subjects to the worst treatment sequence BBB.

The simulated allocation results are shown in Table (5.7). We note that, as λ decreases, increasingly subjects are assigned to the better treatment sequences. When $\lambda = 0$, the design allocates all the sequences to the best performance sequence AAA and ends up with a very low efficiency in estimation. The choice of λ will depend on how much the investigator is willing to sacrifice the estimation for the ethical improvement of the design.

Table 5.7: Results of the adaptive allocation of subjects for each treatment sequence for p=3

N	λ	N_{AAA}	N_{AAB}	N_{ABA}	N_{ABB}	N_{BAA}	N_{BAB}	N_{BBA}	N_{BBB}	Relative efficiency	Success rate
40	1	5	5	5	5	5	5	5	5	1.000	0.450
	0.9	7	5	7	4	4	5	3	5	0.961	0.468
	0.7	10	7	6	5	5	2	3	2	0.589	0.507
	0.3	27	4	2	2	2	1	1	1	0.009	0.589
	0	33	1	1	1	1	1	1	1	0.001	0.610
80	1	10	10	10	10	10	10	10	10	1.000	0.450
	0.9	13	14	11	11	12	5	9	5	0.848	0.482
	0.7	29	14	10	8	10	2	5	2	0.213	0.536
	0.3	62	6	3	2	4	1	1	1	0.001	0.611
	0	73	1	1	1	1	1	1	1	0.000	0.630
100	1	12	13	13	12	13	12	12	13	1.000	0.450
	0.9	17	19	12	15	15	5	12	5	0.779	0.486
	0.7	39	18	11	10	13	2	5	2	0.135	0.544
	0.3	80	7	3	2	5	1	1	1	0.000	0.616
	0	93	1	1	1	1	1	1	1	0.000	0.634

Note: See notes for Table (5.6).

5.6 Numerical Examples

In this section, we demonstrate our strategy in an example that Bandyopadhyay et al. (2009) considered based on the data from a three-period crossover trial of two anti-hypertensive agents. In the crossover trial, the two-treatment design with sequences ABB, BAA, ABA, and BAB were used with 17 patients in each sequence, using metroprolol (A) or metroprolol with chlorthalidone (B). We demonstrate how our strategy utilizes the responses adaptively to allocate the total of 68 patients accomplishing multiple objectives, using the data from the last two time periods of the three-period design for the purpose of comparison with Bandyopadhyay et al.'s (2009) design. Since the data are continuous, we dichotomize it so that a response >135 is a failure according to Bandyopadhyay et al. (2009). Inputting the estimates $(\hat{\nu}_{A1}, \hat{\nu}_{A2}, \hat{\nu}_{B1}, \hat{\nu}_{B2})$=(.24, .24, .24, .35), we would expect the treatment sequences AB/BB to be the best with the most allocation of patients, and

AA/BA the worst with the least number of patients. Our adaptive design is to allocate (13.13, 21.03, 13.03, 20.80) when only the treatment benefit is to be considered ($\lambda = 0$), or (14.69, 19.06, 13.64, 20.62) with $\lambda = .9$ while enjoying much of estimation precision based on the D-optimality to AA, AB, BA, and BB sequences respectively. This is an improvement to Bandyopadhyay et al. (2009) where their ethically appropriate two-period design was to allocate (15.75, 16.92, 17.01, 18.32) to AA, AB, BA, and BB sequences, respectively.

We also considered dichotomizing the responses so that a systolic blood pressure >140 is a failure to compare with the second adaptive design in Bandyopadhyay et al. (2009). Based on this cutoff, the estimated success probabilities are obtained as $\hat{\nu}$ =(.35, .5, .35, .53), and their strategy was to allocate (13.00, 16.42, 16.46, 22.12) to AA, AB, BA, and BB sequences respectively (Bandyopadhyay et al. 2009). Clearly, their adaptive design recognized BB sequence to be the best, followed by BA/AB, and AA. In contrast, our strategy allocates (7.32, 16.35, 14.88, 29.46) when $\lambda = 0$ or (12.38, 16.71, 15.80, 23.11) when $\lambda = .9$, that is, a considerably larger allocation to the BB sequence and less to the AA sequence than designs in Bandyopadhyay et al. (2009). Our strategy also recognizes that the sequence AB is slightly better than the sequence BA and can accomplish multiple objectives.

5.7 Discussion

The literature on RMDs for binary responses is limited (Mehtala, Auranen and Kulathinal 2011), and there are few studies of response-adaptive allocation designs for repeated binary responses (Biswas and Dewanji 2004). In this paper, we extended the multiple-objective allocation strategy of construct-

ing response-adaptive RMDs with continuous responses to those with binary responses. With binary responses, the strategy is quite different from the continuous case where the optimal design is completely model dependent (Liang and Carriere 2009, 2012; Kunert and Stufken 2002, 2008; Afsarinejad and Hedayat 2002). Further, unlike continuous responses, the designs for binary responses are dependent on the responses.

Initially, we built the adaptive design assuming independent binary responses, and then discussed ways to accommodate the dependence structure in a crossover model. Because the eventual data analysis will involve nonlinear models, rather than trying to capture the expected mean responses by devising a crossover design model, we separated the two issues of design and analysis. By specifying the success probabilities of treatments for binary outcomes, the first-order carryover effects are accommodated in a model for data analysis. For multi-period designs, a second order carryover effect can also be accommodated, as well as the self- and mixed-carryover effects. Therefore, our strategy is flexible, based on what form of carryover effects is expected in the experiments.

We provided a detailed allocation rule for constructing practically useful RMDs with two, three, and four periods for comparing two treatments with binary responses. Our goal was to adaptively allocate a new patient to a treatment sequence in a way that would maximize selection criteria while balancing estimation precision with treatment advantage. Through computer simulations, we demonstrated that adaptive designs may not be as efficient as fixed RMDs in terms of the MSE. However, the MSE only captures statistical optimality. We find that our strategy successfully allocates more patients to better treatment sequences, which is often the best strategy depending on the

study goals.

Similar to the continuous response case, the investigator can predetermine the value of λ to balance the two objectives of increasing estimation precision and decreasing the proportion of patients receiving inferior treatments. Design efficiency is a skewed function of λ that decreases sharply as λ decreases. As Figure (5.1) indicates, as λ decreases, the ethical benefit of the proposed adaptive designs becomes more pronounced, especially as the sample size gets larger, but the design efficiency based on the conventional measure of MSE drops sharply. Similar results were observed for constructing multiple-objective response-adaptive RMDs with continuous responses (Liang and Carriere 2009).

As evident in Tables (5.2), (5.3), (5.4), (5.6) and (5.7), when $\lambda = 1$, the presence or absence of treatment effects has no bearing on the adaptive design. However, as λ moves away from 1 and even at 0.9 with only a 10% of attention given to treatment benefit quantified by a given evaluation function, the allocation favors effective treatment sequences without compromising much statistical efficiency.

Chapter 6

Convex optimization and its applications to optimal crossover designs

Abstract

This Chapter builds optimal crossover designs via convex optimization techniques. Upon identifying the unique problems and conditions for constructing optimal designs to that of convex optimization problem, it can be shown that finding optimal designs is relatively quick and simple. The approach is especially useful when constructing some designs is not possible analytically. We first show that the technique produces identical results where analytical solution was possible. Then, we apply it to obtain numerical solutions for 6 and 8 N-of-1 trial designs under autocorrelated error structure. A number of other areas that convex optimization technique can be applicable are suggested as future work.

6.1 Introduction

Crossover designs in clinical trials are popular for comparing non-curative treatments for their efficacy. We denote a crossover design with t treatments and p periods as COD(t,p). In COD(t,p), each subject receives a treatment sequence with p periods. And for each period, any of the t treatment can be assigned. Let N denote the number of distinct sequences involved in a design. Then the maximum value that N can achieve is t^p where all possible distinct sequences are considered in the trial. We consider a linear model for analyzing the crossover design, which can be written as,

$$y_i = X_i\beta + \epsilon_i,$$

where $y_i \in R^p$ is the vector of responses from a subject assigned the ith sequence; $X_i \in R^{p\times k}$ is the design matrix for the ith sequence under the model with the parameter vector $\beta \in R^k$. Assuming the error terms have the covariance matrix Σ, then the information matrix is proportional to,

$$I(\xi) = \sum_{i=1}^{N} \xi_i X_i \Sigma^{-1} X_i^T,$$

where $\xi_i = m_i/M$; M denotes the number of subjects in the experiment and m_i denotes the number of the ith sequence assigned. Normally, there is a constraint on the values of ξ_i so that $m_i = M\xi_i$ must be non-negative integers. When M is large enough, the constraint can be relaxed to $\xi \in R_+^N$ and $\mathbf{1}^T\xi_i = 1$. This relaxed design problem is often called the approximate experimental design problem.

Finding the optimal designs concern with the optimization, with respect to ξ, of certain measurement defined on the information matrix $I(\xi)$. For example, a design is D-optimal (ξ_D) if it maximizes the determinant of $I(\xi)$, i.e.,

$$\xi_D = argmax_\xi \det I(\xi).$$

A design is A-optimal (ξ_A) if it minimizes the trace of the inverse of the information matrix, i.e.,

$$\xi_A = argmin_\xi tr(I(\xi)^{-1}).$$

Optimization problems have been studied for centuries. Some special classes of optimization problems, such as least square and linear programming problems, can be solved numerically quite efficiently. Convex optimization is a

wider class of optimization problems than linear programming problems. After interior-point methods were developed in the 1980s to solve linear programming problems, researches realized these methods could be used for convex optimization problems as well. Some classes of convex optimization problems can be solved numerically efficiently, as the linear programming problems are(Boyd and Vandenberghe, 2004).

In this chapter, we will show that convex optimization can be applied to finding the optimal approximate crossover designs, especially when a closed form of solutions is not found. In fact, building optimal designs is an optimization problem, in which the objective function is a defined measurement on $I(\xi)$ in terms of ξ subject to $\mathbf{1}^T\xi_i = 1$. As long as the objective function and the constraint functions are convex (or concave) functions, the optimization problem can be numerically solved efficiently, just like estimating parameters in linear regression. Therefore, we approach optimal crossover design problems as convex optimization problems, and use software to generate the optimal designs numerically efficiently.

In Section 2, we introduce concepts and properties in convex optimization. In Section 3, we construct a convex optimization problem to find the A-optimal crossover designs. In Section 4, we demonstrate its utility using the Matlab package 'CVX'. Section 5 provides additional potential work in the future.

6.2 Convex optimization

6.2.1 Convex sets

A set C is convex if for any two points $x, y \in C$ and any $\alpha \in [0, 1]$, we have $\alpha x + (1 - \alpha)y \in C$. We use the notation S^n to denote the set of symmetric $n \times n$ matrices, i.e.,

$$S^n = \{X \in R^{n \times n} | X = X^T\}.$$

We use the notation S^n_+ to denote the set of symmetric positive semidefinite matrices, i.e.,

$$S^n_+ = \{X \in S^n | X \succeq 0\},$$

where $X \succeq 0$ denotes X is positive semidefinite. Similarly, we denote the set of symmetric positive definite matrices as,

$$S^n_{++} = \{X \in S^n | X \succ 0\},$$

where $X \succ 0$ denotes X is positive definite. It is trivial to see that S^n, S^n_+ and S^n_{++} are all convex sets.

6.2.2 Convex functions

A function $f : R^n \to R$ is convex if the domain, $\mathbf{dom} f$, is a convex set and if for all $x, y \in \mathbf{dom} f$ and any $\alpha \in [0, 1]$, we have

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha) f(y).$$

We say f is concave if $-f$ is convex. If f is convex and also concave, then f is an affine function.

Some examples of convex or concave functions defined on S^n_{++} are,

- $f(X) = tr(X^{-1})$ is convex on S^n_{++};
- $f(X) = \log\det X$ is concave on S^n_{++}.

A function $f : R^m \to S^n$ is matrix convex if for any $x, y \in \mathbf{dom} f$ and any $\alpha \in [0, 1]$, we have

$$f(\alpha x + (1-\alpha)y) \preceq \alpha f(x) + (1-\alpha)f(y),$$

where the matrix inequality $X \preceq Y$ means $Y - X \succeq 0$ or $Y - X \in S^n_+$.

6.2.3 Conditions on composition functions to preserve convexity

Consider the function f is a composition of h and g, i.e. $f(x) = h \circ g(x)$, where $g, h : R \to R$. According to the second derivative rule, a function is convex if and only if its second derivative is non-negative, and a function is concave if and only if its second derivative is non-positive. Taking the second derivative of $f(x)$, we have,

$$f''(x) = h''(g(x))(g'(x))^2 + h'(g(x))g''(x),$$

which gives the conditions on h and g, such that the composition preserves the convexity.

- If h is convex and nondecreasing, and g is convex, then f is convex.

- If h is convex and nonincreasing, and g is concave, then f is convex.
- If h is concave and nondecreasing, and g is concave, then f is concave.
- If h is concave and nonincreasing, and g is convex, then f is concave.

The results can be generalized to more complicated composition, such as when $h : R^k \to R$ and $g : R^n \to R^k$ (S. Boyd and L. Vandenberghe 2004).

6.3 Convex optimization problem and constructing A-optimal repeated measurement designs

Now we consider crossover designs, application to in COD(t,p), to solve finding the optimal design problems. Under a certain linear model with k parameters, the design matrix $X_i \in R^{p\times k}$, $i = 1, 2, \cdots, N$ are fixed. The information matrix is proportional to,

$$I(\xi) = \sum_{i=1}^{N} \xi_i X_i \Sigma^{-1} X_i^T,$$

It can be shown that $I(\xi) \succeq 0$. In this chapter, we assume $I(\xi) \succ 0$, and the case with singular information matrix is discussed in section 5.

Suppose β can be partitioned as $(\theta', \psi')'$, where θ is the vector of parameters of interest and ψ is the vector of nuisance parameters. Then , we can partition

the information matrix $I(\xi)$ accordingly into the following form,

$$I(\xi) = \begin{bmatrix} I_1 & I_2 \\ I_2^T & I_3 \end{bmatrix}$$

where $I_1 \in R^{k' \times k'}$, $I_3 \in R^{(k-k') \times (k-k')}$. Since $I(\xi) \succ 0$, we have $I_1 \succ 0$ and $I_3 \succ 0$.

The information matrix for θ adjusted by ψ can be written as,

$$I_\theta = I_1 - I_2^T I_3^{-1} I_2.$$

I_θ is called the Schur complement of I_1 in $I(\xi)$. Since $I(\xi) \succ 0$, the property of Schur complement gives that $I_\theta \succ 0$.

Finding the A-optimal design for θ can be expressed as an optimization problem as in the following,

(i) minimize $\mathbf{tr}\ I_\theta^{-1}$

(ii) subject to $\xi \succeq 0$, $\mathbf{1}^T \xi = 1$.

In the following, we show that the above is a convex optimization problem. Let $I \in S_{++}^k$ partitioned as,

$$I = \begin{bmatrix} I_1 & I_2 \\ I_2^T & I_3 \end{bmatrix}$$

where $I_1 \in S^{k'}$. Define a function $SC(I) : S_{++}^k \to S_{++}^{k'}$,

$$SC(I) = I_1 - I_2^T I_3^{-1} I_2.$$

Lemma 1. The function $SC(I)$ is matrix concave in S^k_{++}.

Proof:

S. Boyd and L. Vandenberghe (2004) provided a convex function called matrix fractional function. The function $f : R^n \times S^n \to R$, defined as

$$f(x, Y) = x^T Y^{-1} x$$

is convex on $\mathbf{dom} f = R^n \times S^n_{++}$. We prove lemma 1 based on this result.

For any $\nu \in R^{k'}$, define a function $g : S^{k'}_{++} \times R^{k' \times (k-k')} \times S^{k-k'}_{++}$ as

$$\begin{aligned} g(I_1, I_2, I_3) &= \nu^T (I_1 - I_2^T I_3^{-1} I_2)\nu \\ &= \nu^T I_1 \nu - (I_2\nu)^T I_3^{-1}(I_2\nu) \\ &= \nu^T I_1 \nu - f(I_2\nu, I_3) \end{aligned}$$

Since $-f$ is concave, we have g is a concave function of I_2, I_3 plus an affine function of I_1. Therefore, g is concave function of I_1, I_2, I_3.

That is, for any $I_a, I_b \in S^n_{++}$, with their partitions $(I_{a1}, I_{a2}, I_{a3}), (I_{b1}, I_{b2}, I_{b3}) \in \mathbf{dom} g$ and any $\alpha \in [0, 1]$, we have,

$$g(\alpha I_{a1}+(1-\alpha)I_{b1}, \alpha I_{a2}+(1-\alpha)I_{b2}, \alpha I_{a3}+(1-\alpha)I_{b3}) \geq \alpha g(I_{a1}, I_{a2}, I_{a3})+(1-\alpha)g(I_{b1}, I_{b2}, I_{b3})$$

$$\nu^T SC(\alpha I_a + (1-\alpha)I_b)\nu \geq \nu^T(\alpha SC(I_a) + (1 - \alpha SC(I_b)))\nu$$

$$SC(\alpha I_a + (1-\alpha)I_b) \succeq \alpha SC(I_a) + (1 - \alpha SC(I_b))$$

Hence, $SC(I)$ is matrix concave of I on S^k_{++}. Lemma 1 follows. □

The information matrix $I(\xi)$ can be considered as an affine function of ξ, $I : R^N \to S^k_{++}$.

Lemma 2. The composition $SC \circ I$ is a concave function of ξ.

Proof:

$I : R^N \to S^n_{++}$ is an affine function. So, for any $\xi_1, \xi_2 \in R^N$ and any $\alpha \in [0, 1]$, we have

$$I(\alpha\xi_1 + (1-\alpha)\xi_2) = \alpha I(\xi_1) + (1-\alpha)I(\xi_2).$$

Therefore,

$$\begin{aligned} SC \circ I(\alpha\xi_1 + (1-\alpha)\xi_2) &= SC(I(\alpha\xi_1 + (1-\alpha)\xi_2)) \\ &= SC(\alpha I(\xi_1) + (1-\alpha)I(\xi_2)) \\ &\geq \alpha SC \circ I(\xi_1) + (1-\alpha)SC \circ I(\xi_2) \end{aligned}$$

The Lemma 2 holds.

Lemma 3. $h(I_\theta) = tr(I_\theta^{-1}) : S^{k'}_{++} \to R$ is convex and decreasing.

Lemma 4. The composition $h \circ SC \circ I$ is a convex function of ξ.

According to the conditions preserves the convexity introduced in section 2.3 and Lemma 3, Lemma 4 holds.

From Lemma 4, it follows that the A-optimal design is a convex optimization problem. We next discuss how such problems can be solved numerically efficiently.

6.4 Application to two-treatment optimal crossover problem using CVX package in Matlab

In this section, we show how we can solve the two-treatment optimal crossover designs by applications of convex optimization. The programs are written in Matlab with CVX, which is a package specifying and solving convex programs (Byod and Vandenberghe, 2004). The model considered is the traditional model with a first-order residual effect.

In a two-treatment crossover design, the parameters of interest can be constructed as the contrast of the two treatment effects and the contrast of the two carryover effects. Since the information matrix for treatment effect contrast or carryover effect contrast is a scaler, if a design is A-optimal, it is also D-optimal and E-optimal.

6.4.1 Optimal two-treatment two-, three- and four-period design for treatment effect

Optimal results have been well developed in two-treatment p-period crossover trials (Laska and Meisner, 1985). It is known that if $p = 2$, the design AA, AB, BB, BA with $M/4$ subjects per sequence is an optimal design for treatment effect. When $p = 3$, the design built up of the sequence ABB and its dual is optimal. For $p = 4$, according to Cheng and Wu (1980), the strongly balanced designs are universally optimal.

Using CVX, we obtain the following results for the optimal two-treatment two-, three- and four-period design. Tables (6.1), (6.2) and (6.3) show that the obtained optimal design sequences and their weights in the design for two-,

three- and four-period crossover trials. The numerical solutions confirm the theoretical results.

Table 6.1: Optimal design for treatment effect in COD(2,2)

AA	AB	BA	BB
0.2500	0.2500	0.2500	0.2500

Table 6.2: Optimal design for treatment effect in COD(2,3)

ABB	BAA
0.5000	0.5000

Table 6.3: Optimal design for treatment effect in COD(2,4)

ABBA	ABAB	AABB	BBAA	BABA	BAAB
0.1042	0.0729	0.3229	0.3229	0.0729	0.1042

Note that for two- and three-period cases, CVX produced the optimal designs expected by the theories. When $p = 4$, CVX produced a strongly balanced design which is optimal according to the theories.

6.4.2 Optimal two-objective design in COD(2,2)

In this section, we consider optimal two-treatment two-period CODs under the traditional model with equal-correlated measurement errors. When within-subject correlation is not equal to zero, the optimal design for direct treatment effect and the optimal design for carryover effect are not identical.

The optimal design for the direct treatment effect is known as the design AA, AB, BB, BA with equal allocation. The result obtained using CVX is the same as the one in Table (6.1).

The optimal design for carryover effect obtained by CVX is shown in Table (6.4).

Table 6.4: Optimal design for carryover effect in COD(2,2)

AA	AB	BA	BB
0.3750	0.1250	0.1250	0.3750

When both treatment effect and carryover effect are of interest, we find the optimal two-objective design by applying a weighted sum to the objective functions for the direct treatment effect and the carryover effect to form a new objective function. A tuning parameter is used to form a convex combination of the two objectives. Specifically, we want to

(i) minimize $\alpha \mathbf{tr} I_{\theta_1}^{-1} + (1-\alpha)\mathbf{tr} I_{\theta_2}^{-1}$

(ii) subject to $\xi \succeq 0$, $\mathbf{1}^T \xi = 1$

Since a combination of convex functions is a convex function, we apply the same technique for finding a single objective optimal design to find the two-objective optimal design.

In Table (6.5), we show the optimal result for the compound optimal design for estimating the direct treatment effect and the carryover effect.

Table 6.5: Compound optimal design for treatment and carryover effect in COD(2,2), $\rho = 0.5$ and $\alpha = 0.5$

AA	AB	BA	BB
0.2914	0.2086	0.2086	0.2914

6.4.3 Optimal N-of-1 trial designs

For two-treatment multi-crossover single-patient trials, a general N-of-1 design can have multiple AB or BA crossover pairs in a sequence of treatments for within-patient comparisons.

In addition to within patient-based evidence of a treatment contrast, it may also be desirable to obtain a population average effect of treatments. Aggregating the series of N-of-1 trials can give such an estimate of the average effect. In the following, we apply convex optimization to optimal six-, eight-period design in N-of-1 trials for treatment effect under the traditional model. The CVX solution is shown in Table (6.6) and (6.7).

Table 6.6: Optimal six-period N-of-1 trials with uncorrelated errors

ABBAAB	BAABBA
0.5000	0.5000

Table 6.7: Optimal eight-period N-of-1 trials with equal-correlated errors

ABBAABBA	BAABBAAB
0.5000	0.5000

Table 6.8: Optimal eight-period N-of-1 trials with auto-correlated errors $\rho = 0.5$

ABABABAB	0.0029
ABABABBA	0.0044
ABABBAAB	0.0087
ABABBABA	0.0044
ABBAABAB	0.0087
ABBAABBA	**0.4579**
ABBABAAB	0.0087
ABBABABA	0.0044
BAABABAB	0.0044
BAABABBA	0.0087
BAABBAAB	**0.4579**
BAABBABA	0.0087
BABAABAB	0.0044
BABAABBA	0.0087
BABABAAB	0.0044
BABABABA	0.0029

From Table (6.6) and (6.7), we can see the optimal design in N-of-1 trials

under the traditional model consists of sequences with alternating crossover pairs AB and BA, which is consistent with the theoretical results of the optimal two-treatment N-of-1 trials.

Table (6.8) gives the optimal N-of-1 design under the traditional model with auto-correlated errors. As discussed in the optimal two-treatment N-of-1 trials, constructing the optimal N-of-1 design with auto-correlated errors is analytically complicated. But the optimal design can be obtained using CVX easily. The results in the Table (6.8) suggest that $ABBAABBA$ and $BAABBAAB$ are the options for a single N-of-1 sequence. The weights for other sequences are not zero implying that the sequences other than $ABBAABBA$ and $BAABBAAB$ may also be included to obtain an optimal aggregate N-of-1 design where $N > 1$. Extended multi-period N-of-1 designs over eight periods can be similarly constructed efficiently via convex optimization strategy.

6.5 Discussion

We have shown how convex optimization can solve the optimal design problems upon properly defining the objective functions. The alternate solution via convex optimization has been verified to be identical to what were obtained analytically in the literature. In this section, we propose a number of other areas that this approach could provide uncomplicated solutions to.

1. Generating the matrix fractional function on $R^n \times S_+^n$.

   Consider the cases where information matrix is singular. If we could prove that the convexity holds for the matrix fractional function on $R^n \times S_+^n$, i.e., $f(x, Y) = x^T Y^- x$ is convex, then we can have more general

results and establish the optimality problem with positive semidefinite information matrix, instead of positive definite information matrix.

2. Constructing convex optimization with other criteria.

   The function $h : S^n \to R$ defined by $h(X) = -\log\det(-X)$ is convex and increasing on $\mathbf{dom}h = -S^n_{++}$. Therefore, D-optimal design could also be constructed.

3. Under generalized linear models.

   The approach mentioned above concerns with the information matrix under a linear model. We can extend the above to solve discrete response experimental design problems. The task would involve defining a proper objective function, which must be shown to be convex or concave.

4. Response adaptive multiple objective optimal designs.

   The approach can be used in a response adaptive design. Given the history of previous k subjects, it could not only find the best sequence assigned to the next subject, but also provide the optimal design for the rest of subjects based on the information given.

5. When the number of subjects is not large enough and the assignment on the sequences is not even.

   Suppose in a certain class of crossover designs, the optimal design is dual balanced with an equal assignment on two dual sequences, as in chapter 3. When the total number of subjects (M) is odd, the assignment of the subjects to dual sequences are not always even. Then, it is not trivial to determine the optimal design. In this situation, the convexity of the

objective function implies that the optimal design should be among the closest support points to $M \cdot (\xi_1, \cdots, \xi_N)$. The optimal design can be found upon a quick check with the neighboring support points.

Chapter 7

Concluding Remarks

We highlight the major contributions of the thesis in this chapter, and propose further research.

1. Constrained and compound optimal designs in two-treatment two-period crossover trials.

When building optimal designs, there are often competing objectives that the investigator desires to optimize. These multiple objectives can include two or more parameters or some functionals, ultimately involving simultaneous considerations. Crossover designs are well known for controversy involving non-orthogonal key parameters of direct and residual or crossover treatment effects, which can lead to completely different experimental designs depending on the primary parameter. Under the traditional model with equal-correlated measurement errors, it is known that the optimal two-treatment two-period crossover designs for the direct treatment effect and for the carryover effect are competing objectives that cannot be achieved at the same time. We revisit this controversy from the point of view of constrained and compound designs. We applied each of the two methods to find the design that is optimal with a

combined objective of estimating both of the effects. The optimal design problems and objectives were defined and the optimal designs were produced for each of the defined problem. Cook and Wong (1994) showed that constrained and compound optimal designs are basically equivalent. In this chapter, we provided an alternate simpler proof for crossover designs.

2. Universally optimal designs for N-of-1 trials.

We considered universally optimal N-of-1 designs. Originally, Kiefer (1975) proposed the concept of universal optimality with zero row and column sums in the information matrices. We examine special conditions when such universally optimal designs exist with special application to N-of-1 trial designs that will make optimal no matter what criteria are applied.

We presented a sufficient condition of universally optimal N-of-1 designs under a traditional model accommodating the carryover effects. We proposed universally optimal sequences for general $t \geq 2$. When $t = 2$, the one sequence N-of-1 universal optimal designs exist if the number of periods is a multiple of t; when $t > 2$, the one sequence N-of-1 universal optimal designs exist if the number of periods is a multiple of t plus 1. Recognizing possible practical difficulty in adopting the designs constructed, we also obtained optimal N-of-1 designs with block size less than t.

When a self and mixed carryover effects model is employed, it would be of interest to further research what the universal optimal N-of-1 designs would be.

3. Optimal two-treatment N-of-1 trial designs.

In this chapter, we closely examined N-of-1 trials to construct optimal designs for two treatments under a variety of conditions about the carryover effects,

the covariance structure, and the number of planned periods. Extension to optimal aggregated N-of-1 designs is also discussed.

It is known that the design AB/AA and their duals is optimal for two-period experiments, with the duality of a sequence defined as the sequence that switches A and B with the same effect. Similarly, it is known that the two-sequence design ABB and its dual and the four-sequence design ABBA/AABB and their duals are optimal for three- and four-period experiments (Carriere 1994; Laska and Meisner 1985). Straight application of the two-treatment optimal designs in literature with A to AB and B to BA suggests that optimal N-of-1 trials would need to use $ABBA$, $ABAB$ and their duals for two within-patient comparisons, $ABBABA$ and its dual for three within-patient comparisons, and the sequences $ABBABAAB$, $ABABBABA$ and their duals for four within-patient comparisons.

In this chapter, we proved that these designs were not optimal for N-of-1 trials for estimating individual-based treatment effects. We then constructed optimal designs for two treatments under a variety of conditions about the carryover effects, the covariance structure, and the number of planned periods. Extension to optimal aggregated N-of-1 designs is also discussed.

Numerical consideration of the estimated precision of using several six- and eight-period designs revealed the practical performance of particular designs, giving us realistic guidelines. Overall, we conclude that alternating between AB and BA pairs in sequence will result in a nearly optimal N-of-1 trial for a single patient, if not the optimal, under all the models we considered without the need to guess or conduct a pilot study to confirm the correlation structure. Alternating between AB and BA pairs in a single trial is nearly robust to misspecified error structures of the repeated measurements. This

is identical to the universally optimal designs built in previous chapter under rather restricted conditions of a traditional model.

When an experiment has been carried out with the optimal N-of-1 trial and additional patients are accrued in the trial, we can aggregate these N-of-1 trials optimally by allocating the same number of patients to its dual sequence, thereby optimizing the trial for both the individual and average patients.

4. Multiple objective response adaptive crossover designs with binary outcomes.

In this chapter, we extended the allocation strategy for continuous responses to constructing response-adaptive repeated measurement designs for binary responses.

We built the adaptive design assuming independent binary responses, and then discussed ways to accommodate the dependence structure in a crossover model.

We provided a detailed allocation rule for constructing practically useful RMDs with two, three, and four periods for comparing two treatments with binary responses. We concluded that the allocation strategy developed for continuous responses also worked well for binary responses. As expected, design efficiency in terms of mean squared error drops sharply, as more emphasis is placed on increasing treatment benefit than estimation precision. However, we showed that the design could successfully allocate more patients to better treatment sequences without sacrificing much estimation precision.

Similar to the continuous response case, the investigator can predetermine the value of weight (λ) to balance the two objectives of increasing estimation precision and decreasing the proportion of patients receiving inferior treat-

ments. Design efficiency is a skewed function of λ that decreases sharply as λ decreases. As λ decreases, the ethical benefit of the proposed adaptive designs becomes more pronounced, especially as the sample size gets larger, but the design efficiency based on the conventional measure of MSE drops sharply, as observed in multiple-objective response-adaptive RMDs with continuous responses.

5. Convex optimization and its applications in optimal crossover designs.

This Chapter proposed constructing optimal crossover designs via convex optimization techniques. Upon identifying the unique problems and conditions for constructing optimal designs to that of convex optimization problem, it can be shown that finding optimal designs is relatively quick and simple. The approach is especially useful in cases where constructing designs is not possible analytically. We first showed that the technique produced identical results where an analytical solution was possible. Then, we applied it to obtain numerical solutions for 6 and 8 N-of-1 trial designs under autocorrelated error structure. Some of other areas that convex optimization technique can be applicable are suggested as future work.

Bibliography

[1] Afsarinejad, K. (1988). Circular balanced uniform repeated measurements designs. *Statistics and probability letters, 7(3), 187–189.*

[2] Afsarinejad, K., and Hedayat, A. S. (2002). Repeated measurements designs for a model with self and simple mixed carryover effects. Journal of Statistical Planning and Inference, 106(1), 449–459.

[3] Armitage, P. (1975). Sequential Medical trials. Blackwell:Oxford.

[4] Bandyopadhyay, U., and Bhattacharya, R. (2009). Response adaptive procedures with dual optimality. Statistica Neerlandica ***63,*** *353–367.*

[5] Bandyopadhyay, U., Biswas, A, and Mukherjee, S. (2011). Some inferential procedures in randomized repeated measurement design for binary response. Journal of Korean Statistical Society ***40,*** *245–255.*

[6] Bate, S. T., and Jones, B. (2008). A review of uniform cross-over designs. Journal of Statistical Planning and Inference, 138, 336–351.

[7] Biswas, A, and Dewanji, A. (2004). A randomized longitudinal play-the-winner design for repeated binary data. Australian and New Zealand Journal of Statistics ***46,*** *675–684.*

[8] Biswas, A, Park, E. and Bhattacharya, R. (2010). Covariate-adjusted response-adaptive designs for longitudinal treatment responses: PEMF trial revisited. Statistical Methods in Medical Research. Published online before print, Oct. 23, 2010.

[9] Bose, M., and Mukherjee, B. (2003). Optimal crossover designs under a general model. Statistics & Probability Letters, 62, 413–418.

[10] Boyd, S., and Vandenberghe, L. (2004). Convex optimization. Cambridge university press.

[11] Carriere, K. C. (1994). Crossover designs for clinical trials. Statistics in medicine, 13, 1063–1069.

[12] Carriere, K. C. (1995). Optimal Two-Period Crossover Designs: A Practical Solution. CLIN. RESEARCH and REG. AFFAIRS, 12, 183–191.

[13] Carriere, K. C., and Huang, R. (2000). Crossover designs for two-treatment clinical trials. Journal of Statistical Planning and Inference, 87, 125–134.

[14] Carriere, K. C., and Reinsel, G. C. (1992). Investigation of dual-balanced crossover designs for two treatments. Biometrics, 48, 1157–1164.

[15] Carriere, K. C., and Reinsel, G. C. (1993). Optimal two-period repeated measurements designs with two or more treatments. Biometrika, 80, 924–929.

[16] Cheng, C. S., and Wu, C. F. (1980). Balanced repeated measurements designs. The Annals of Statistics, 8, 1272–1283.

[17] Clyde, M., and Chaloner, K. (1996). The Equivalence of Constrained and Weighted Designs in Multiple Objective Design Problems. Journal of the American Statistical Association, 91, 1236–1244.

[18] Cochrane, A. L. (1972). Effectiveness and Efficiency: Random Reflections on Health Service (Vol. 900574178). London: Nuffield Provincial Hospitals Trust.

[19] Cook, R. D., and Wong, W. K. (1994). On the Equivalence of Constrained and Compound Optimal Deigns. Journal of the American Statistical Association, 89, 687–692.

[20] Ebbutt, A. F. (1984). Three-Period Crossover Designs for Two Treatments. Biometrics, 40, 219–224.

[21] Edgington, E. S. (1984). Statistics and single case analysis. Progress in behavior modification, 16, 83–119.

[22] Ezzet, F, and Whitehead, J. (1992). A random effects model for binary data from crossover clinical trials. Appl. Statist. ***41,*** *117-126.*

[23] Greenfield, S., Kravits, R., Duan, N., and Kaplan, S.H. (2007) Heterogeneity of Treatment Effects: Implications for Guidelines, Payment, and Quality Assessment. The American Journal of Medicine, 120, S3–S9.

[24] Guyatt, G., Sackett, D., Taylor, D. W., Ghong, J., Roberts, R., and Pugsley, S. (1986). Determining optimal therapy-randomized trials in individual patients. New England Journal of Medicine, 314, 889–892.

[25] Guyatt, G. H., Heyting, A., Jaeschke, R., Keller, J., Adachi, J. D., and Roberts, R. S. (1990). N of 1 randomized trials for investigating new drugs. Controlled clinical trials, 11, 88–100.

[26] Hedayat, A., and Afsarinejad, K. (1978). Repeated measurements designs, II. The Annals of Statistics, 6, 619–628.

[27] Hedayat, A. S., and Yang, M. (2003). Universal optimality of balanced uniform crossover designs. The Annals of Statistics, 31, 978–983.

[28] Hedayat, A. S., and Yang, M. (2004). Universal optimality for selected crossover designs. Journal of the American Statistical Association, 99, 461–466.

[29] Hu, F., and Rosenberger, W. F. (2006). The Theory of Response-Adaptive Randomization in Clinical Trials. New York: Wiley.

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

[31] Kempton, R. A., Ferris, S. J., and David, O. (2001). Optimal change-over designs when carry-over effects are proportional to direct effects of treatments. Biometrika, 88, 391–399.

[32] Kenword, M. G, and Jones, B. (1987). A log-linear model for binary cross-over data. Appl. Statist. ***36,*** *192-204.*

[33] Kershner, R. P., and Federer, W. T. (1981). Two-Treatment Crossover Designs for Estimating a Variety of Effects. Journal of the American Statistical Association, 76, 612–619.

[34] Kiefer, J. (1975). General Equivalence theory for optimum designs (approximate theory). The Annals of Statistics, 2, 849–879.

[35] Kravits, R.L., Duan, N., and Braslow, J. (2004). Evidence-based Medicine, Heterogeneity of Treatment Effects, and the Trouble with Averages. Milbank Quarterly, 82, 661–87.

[36] Kravitz, R. L., and Duan, N. (2014). Design and Implementation of N-of-1 Trials: a user's guide. Agency for healthcare research and quality, US Department of Health and Human Services.

[37] Kunert, J. (1984). Optimality of balanced uniform repeated measurements designs. The Annals of Statistics, 12, 1006–1017.

[38] Kunert, J. (1991). Cross-over designs for two treatments and correlated errors. Biometrika, 78, 315–324.

[39] Kunert, J., and Stufken, J. (2002). Optimal crossover designs in a model with self and mixed carryover effects. Journal of the American Statistical Association, 97, 898–906.

[40] Kunert, J., and Stufken, J. (2008). Optimal crossover designs for two treatments in the presence of mixed and self-carryover effects. Journal of the American Statistical Association ***103,*** *1641–1647.*

[41] Kushner, H. B. (2003). Allocation rules for adaptive repeated measurements designs. Journal of Statistical Planning and Inference ***113,*** *293–313.*

[42] Larson, E.B. (1990). N-of-1 clinical trials: A technique for improving medical therapeutics. Western Journal of Medicine, 152, 52.

[43] Laska, E. M., and Meisner, M. (1985). A variational approach to optimal two-treatment crossover designs: application to carryover-effect models. Journal of the American Statistical Association, 80, 704–710.

[44] Laska, E., Meisner, M., and Kushner, H. B. (1983). Optimal Crossover Designs in the Presence of Carryover Effects. Biometrics, 39, 1087–1091.

[45] Lee, C. M. S. (1988). Constrained Optimal Designs. Journal of Statistical Planning and Inference, 18, 377–389.

[46] Liang, Y., and Carriere, K. C. (2009). Multiple-Objective Response-Adaptive Repeated Measurement Designs for Clinical Trials. Journal of Statistical Planning and Inference, 139, 1134–1145.

[47] Liang, Y., and Carriere, K. C. (2012). Allocating multiple patients at a time in multiple-objective response-adaptive repeated measurement designs. International Journal of Statistical and Probability ***1,*** *186–197.*

[48] Mandal, S., Torsney, B., and Carriere, K. C. (2005). Constructing optimal designs with constraints. Journal of Statistical Planning and Inference, 128, 609–621.

[49] Matthews, J. N. S. (1987). Optimal crossover designs for the comparison of two treatments in the presence of carryover effects and autocorrelated errors. Biometrika, 74, 311–320.

[50] Matthews, J. N. S. (1990). Optimal dual-balanced two-treatment crossover designs. Sankhya: The Indian Journal of Statistics, Series B, 332–337.

[51] Mehtala, J., Auranen, K., and Kulathinal, S. (2011). Optimal designs for epidemiologic longitudinal studies with binary outcomes. Statistical Methods in Medical Research. Published online before print, Dec. 13, 2011.

[52] Nikles, C.J., Clavarino, A.M., and Del Mar, C.B. (2005). Using N-of-1 Trials as a clinical tool to improve prescribing. British Journal of General Practice, 55, 175–80.

[53] Rupp, M. E., Fitzgerald, T., Puumala, S., Anderson, J. R., Craig, R., Iwen, P. C.,... Sholtz, L. (2008). Prospective, controlled, cross-over trial of alcohol-based had gel in critical care units. Infection Control & Hospital Epidemiology, 29, 8–15.

[54] Saha, R. (1983). Repeated measurements designs. Calcutta Statistical Association Bulletin, 32, 153–168.

[55] Sen, M., and Mukerjee, R. (1987). Optimal repeated measurements designs under interaction. Journal of Statistical Planning and Inference, 17, 81–91.

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

[57] Zelen, M. (1969). Play the winner rule and the controlled clinical trial. Journal of the American Statistical Association, 64, 131–146.

[58] Zucker, D. R., Ruthazer, R., and Schmid, C. H. (2010) Individual (N-of-1) trials can be combined to give population comparative treatment effect

estimates: methodologic considerations. Journal of Clinical Epidemiology, 63, 1312–1323.

Appendices

Matlab codes are presented for constructing A-optimal crossover designs via CVX package.

A.1 (Setting up parameters for obtaining Table (6.1))

```
t=2;
p=2;
Blocksize=1;
[xcand,pooldesign,tauindex,gammaindex] =
designmatgen(t,p,Blocksize,'traditional','uncorrelated','true','false')
rho=0;
K=0;
```

A.2 (Setting up parameters for obtaining Table (6.2))

```
t=2;
p=3;
Blocksize=1;
[xcand,pooldesign,tauindex,gammaindex] =
designmatgen(t,p,Blocksize,'traditional','uncorrelated','true','false')
rho=0;
K=0;
```

A.3 (Setting up parameters for obtaining Table (6.3))

```
t=2;
p=4;
Blocksize=1;
[xcand,pooldesign,tauindex,gammaindex] =
designmatgen(t,p,Blocksize,'traditional','uncorrelated','true','false')
```

```
rho=0;
K=0;
```

A.4 (Setting up parameters for obtaining Table (6.4) and (6.5))

```
t=2;
p=2;
Blocksize=1;
[xcand,pooldesign,tauindex,gammaindex] =
designmatgen(t,p,Blocksize,'traditional','uncorrelated','true','false')
rho=0.5;
K=0;
```

A.5 (Setting up parameters for obtaining Table (6.6))

```
t=2;
p=6;
Blocksize=2;
[xcand,pooldesign,tauindex,gammaindex] =
designmatgen(t,p,Blocksize,'traditional','uncorrelated','true','false')
rho=0;
K=0;
```

A.6(Setting up parameters for obtaining Table (6.7))

```
t=2;
p=8;
Blocksize=2;
[xcand,pooldesign,tauindex,gammaindex] =
designmatgen(t,p,Blocksize,'traditional','uncorrelated','true','false')
```

```
rho=0;
K=0;
```

A.7 (Setting up parameters for obtaining Table (6.8))

```
t=2;
p=8;
Blocksize=2;
[xcand,pooldesign,tauindex,gammaindex] =
designmatgen(t,p,Blocksize,'traditional','uncorrelated','true','false')
rho=0.5;
K=1;
```

B.1 (Applying CVX for obtaining Table (6.1), (6.2), (6.3), (6.6), (6.6) and (6.8))

```
sigma=eye(p);
for s_i=1:p
    for s_j=1:p
        if s_i~=s_j
            if abs(s_i-s_j)==1
                sigma(s_i,s_j)= rho;
            else
                sigma(s_i,s_j)=rho * rho ^( (abs(s_i-s_j)-1)*K);
            end
        end
    end
end
```

```
if Blocksize==1
    poolsize=t^p;
end
if Blocksize>1
    poolsize=t^(p/Blocksize);
end
Sigma=kron(eye(poolsize),sigma);
Sinv=inv(Sigma);
Sinvhalf=sqrtm(Sinv);
1/rho*(eye(p)-rho/((p-1)*rho+1)*ones(p,1)*ones(1,p))
V=xcand;
n=length(V(:,1));
tauid=[tauindex(1):tauindex(2)];
if tauindex(1)>1
    taucomplement=[1:(tauindex(1)-1)];
    if tauindex(2)<n
        taucomplement=[taucomplement,(tauindex(2)+1) : n];
    end
else
    if tauindex(2)<n
        taucomplement=[taucomplement,(tauindex(2)+1) : n];
    end
end
gammaid=[gammaindex(1):gammaindex(2)];
if gammaindex(1)>1
    gammacomplement=[1:(gammaindex(1)-1)];
```

```
    if gammaindex(2)<n
        gammacomplement=[gammacomplement,(gammaindex(2)+1) : n];
    end
else
    if gammaindex(2)<n
        gammacomplement=[gammacomplement,(gammaindex(2)+1) : n];
    end
end
% for direct treatment effects (tau)
cvx_begin
    variable m(poolsize)
    r=repmat(m,1,p)';
    r=r(:);
    M=diag(r);
    Xd=V*Sinvhalf*M*Sinvhalf*V';
    A=Xd(tauid,tauid);
    B=Xd(tauid,taucomplement)';
    C=Xd(taucomplement,taucomplement);
    Schur=A-matrix_frac(B,C);
%    maximize(Schur + min(m))
    maximize(Schur)
    subject to
       m>=0;
       sum(m)==1;
cvx_end
m
```

```
epsilon=10^(-4);
count=0;
clear L
for mi=1:length(m)
    if m(mi)>epsilon
        count=count+1;
        L(count,:)=pooldesign(mi,:);
    end
end
char('A'+L-1)
```

B.2 (Applying CVX for obtaining Table (6.4))

```
% Replace the CVX definition in Part B.1 with the
% following codes between 'cvx_begin' and 'cvx_end'.
cvx_begin
    variable m(poolsize)
    r=repmat(m,1,p)';
    r=r(:);
    M=diag(r);
    Xd=V*Sinvhalf*M*Sinvhalf*V';
    A=Xd(gammaid,gammaid);
    B=Xd(gammaid,gammacomplement)';
    C=Xd(gammacomplement,gammacomplement);
    Schur=A-matrix_frac(B,C);
%    maximize(Schur + min(m))
    maximize(Schur)
```

```
    subject to
       m>=0;
       sum(m)==1;
cvx_end
```

B.3 (Applying CVX for obtaining Table (6.5))

```
%Replace the CVX definition in Part B.1 with the
%following codes between 'cvx_begin' and 'cvx_end'.
 cvx_begin
    variable m(poolsize)
    r=repmat(m,1,p)';
    r=r(:);
    M=diag(r);
    Xd=V*Sinvhalf*M*Sinvhalf*V';
    A=Xd(tauid,tauid);
    B=Xd(tauid,taucomplement)';
    C=Xd(taucomplement,taucomplement);
    Schur_tau=A-matrix_frac(B,C);

    A=Xd(gammaid,gammaid);
    B=Xd(gammaid,gammacomplement)';
    C=Xd(gammacomplement,gammacomplement);
    Schur_gamma=A-matrix_frac(B,C);
%    maximize(Schur + min(m))
    maximize(Schur_tau + Schur_gamma)
    subject to
```

```
        m>=0;
        sum(m)==1;
cvx_end
```

C.1 (Function: designmatgen.m)

```
%Revised function of 'designmatgen.m' in CVX package
% so that it can also be applied to N-of-1 trials
function [xcand,pooldesign,tauindex,gammaindex] =
designmatgen(T,P,Blocksize,model,error,Periodeff,SubjectRNDeff,varargin)
% This function is revised from 'designmatgen' function in CVX package
% so that it can also be applied to N-of-1 trials.
%  CANDGEN Generate candidate set for D-optimal design.
%   XCAND = CANDGEN(NFACTORS,MODEL) generates a candidate set
%   appropriate for a D-optimal design with NFACTORS factors and
%   the model MODEL.  The output matrix XCAND is N-by-NFACTORS,
%   with each row representing the coordinates of one of the N
%   candidate points.  MODEL can be any of the following strings:
%
%     'linear'          constant and linear terms (the default)
%     'interaction'     constant, linear, and cross product terms
%     'quadratic'       interactions plus squared terms
%     'purequadratic'   constant, linear, and squared terms
%
%   Alternatively MODEL can be a matrix of term definitions as
%   accepted by the X2FX function.
%
```

```
%   [XCAND,FXCAND] = CANDGEN(NFACTORS,MODEL) returns both the
%   matrix of factor values XCAND and the matrix of term values
%   FXCAND.  The latter can be input to CANDEXCH to generate the
%   D-optimal design.
%
%   [...] = CANDGEN(NFACTORS,MODEL,'PARAM1',VALUE1,'PARAM2',VALUE2,...)
%   provides more control over the candidate set generation through a set
%   of parameter/value pairs.  Valid parameters are the following:
%
%      Parameter     Value
%      'bounds'      Lower and upper bounds for each factor, specified
%                    as a 2-by-NFACTORS matrix.  Alternatively, this value
%                    can be a cell array containing NFACTORS elements, each
%                    element specifying the vector of allowable values for
%                    the corresponding factor.
%      'levels'      Vector of number of levels for each factor.
%      'categorical' Indices of categorical predictors.
%
%   The ROWEXCH automatically generates a candidate set using the
%   CANDGEN function, and creates a D-optimal design from it using
%   the CANDEXCH function.  You may prefer to call these functions
%   separately if you want to modify the default candidate set.
%
%   See also ROWEXCH, CANDEXCH, X2FX.

%   Copyright 1993-2005 The MathWorks, Inc.
```

```
%   $Revision: 1.2.2.2 $  $Date: 2005/11/18 14:27:48 $

% Get default values for optional arguments
% if nargin < 3 || isempty(model)
%    model = 'traditional';
% end




t=T;
p=P;
if isempty(model)
   model = 'traditional';
end
if ischar(model)
   nchars = length(model);
   istraditional = strncmpi(model,'traditional',nchars);
   isselfmixed = strncmpi(model,'selfmixed',nchars);
end

if isempty(error)
    error = 'uncorrelated';
end
if ischar(error)
    nchars=length(error);
    isun = strncmpi(error,'uncorrelated',nchars);
```

```
    isequi=strncmpi(error,'equicorrelated',nchars);
    isauto=strncmpi(error,'autocorrelated',nchars);
end

if isempty(Periodeff)
    Periodeff = 'false';
end
if ischar(Periodeff)
    nchars=length(Periodeff);
    isPeriodeff = strncmpi(Periodeff,'true',nchars);
end

if isempty(SubjectRNDeff)
    SubjectRNDeff = 'false';
end
if ischar(SubjectRNDeff)
    nchars=length(SubjectRNDeff);
    isSubjectRNDeff = strncmpi(SubjectRNDeff,'true',nchars);
end

if isempty(Blocksize)
    Blocksize = 1;
end


%ly if Blocksize==1 and traditional model
```

```
if Blocksize==1

    %ly generating design matrix for direct and carryover effect
    poolsize=t^p;
    poolid=repmat((1:p)',[poolsize,1]);
    pooldesign=poolgen(t,p);



    periodmat=periodmatgen(t,p);
    mumat=ones(t^p*p,1);

    taumat=taumatgen(pooldesign);
    if istraditional
        gammamat=gammamatgen(pooldesign,t);
    end
    if isselfmixed
        [gammamat_s,gammamat_m]=gammamatgen_sm(pooldesign,t);
    end
end

if Blocksize>1
        %ly generating design matrix for direct and carryover effect
    poolsize=t^(p/2);
    poolid=repmat((1:p)',[poolsize,1]);
    pooldesign=zeros(poolsize,p);
    pooldesign_temp=poolgen(t,p/2);
```

```
    for i=1:length(pooldesign_temp(1,:))
        pooldesign(:,(2*i-1))= pooldesign_temp(:,i);
        pooldesign(:,(2*i))= 3-pooldesign_temp(:,i);
    end

    periodmat=periodmatgen_no1(t,p,Blocksize);
    mumat=ones(t^(p/2)*p,1);
    taumat=taumatgen_no1(pooldesign,t);

    if istraditional
        gammamat=gammamatgen_no1(pooldesign,t);
    end
    if isselfmixed
        [gammamat_s,gammamat_m]=gammamatgen_no1sm(pooldesign,t);
    end
end


V=[mumat];
if isPeriodeff
    V=[V, periodmat(:,(2:p))-repmat(periodmat(:,1),[1,p-1])];
end


beforetau=length(V(1,:));
V=[V,taumat(:,(1:(t-1)))-repmat(taumat(:,t),[1,t-1])];
```

```
aftertau=length(V(1,:));

if istraditional
    V=[V,gammamat(:,(1:(t-1)))-repmat(gammamat(:,t),[1,t-1])];
end

if isselfmixed
    V=[V,gammamat_s(:,(1:(t-1)))-repmat(gammamat_s(:,t),[1,t-1])];
    V=[V,gammamat_m(:,(1:(t-1)))-repmat(gammamat_m(:,t),[1,t-1])];
end
aftergamma=length(V(1,:));

V=V';
tauindex=[beforetau+1,aftertau];
gammaindex=[aftertau+1,aftergamma];

%output
%xcand=[T,P,model,error,Periodeff,SubjectRNDeff,Blocksize];

sprintf('poolsize=%d',poolsize)
xcand=V;
end
```