University of Alberta

Response-Adaptive Repeated Measurement Designs for Clinical Trials

by

Yuanyuan Liang

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**.

Department of Mathematical and Statistical Sciences

Edmonton, Alberta
Fall 2006

Your file Votre référence
ISBN: 978-0-494-23067-1
Our file Notre référence
ISBN: 978-0-494-23067-1

Abstract

This thesis focuses on issues arising from repeated measurement designs for clinical trials. We construct repeated measurement designs under general models compared to those constructed in previous studies. We also study the influence of baseline measurements on repeated measurement designs, and propose two new design strategies to construct response-adaptive repeated measurement designs.

We study the optimal design problem under both the traditional and a more general model (self and mixed carryover effects model). We also explore the baseline measurement effect on constructing optimal designs, and give recommendations on constructing two-treatment p-period ($p = 2, 3, 4$) repeated measurement designs.

For dichotomous responses, a new response-adaptive allocation rule, called the stratified and randomized play-the-winner rule (SRPWR), is developed. SRPWR is a modification of the play-the-winner rule (PWR) that skews the allocation pattern in favor of a better treatment. SRPWR is applicable to clinical trials with more than two treatments. In addition, SRPWR allows for treatment comparisons among homogenous patients by stratifying them based on possible confounders (age, sex, disease status, etc.).

One of the main contributions of this thesis is to extend the single-objective designs to multiple-objective designs. We develop a new adaptive allocation rule, that can provide good estimates of the parameters of interest, and assign more patients to a better treatment. The basic idea is to modify the allocation rule based on the observed data from previous patients. We assume that patients enter the study sequentially, as is typically the case in clinical trials. The first m patients are assigned using the optimal design

suggested in the literature, or a completely randomized design. Then the information matrix can be calculated based on the observed data. We introduce the concept of an evaluation function to evaluate the performance of each treatment sequence. Among all possible treatment sequences, we choose the one that maximizes the allocation criteria. The criteria have two components: the first component determines a treatment sequence that maximizes the information matrix; the second determines a treatment sequence that gives the best performance based on the observed data. The new strategy is demonstrated by simulations using dichotomous and continuous responses.

Acknowledgements

I would like to thank my supervisor, Dr. Keumhee Chough (K.C.) Carrière, for leading me to the research area of repeated measurement designs, and for her many helpful and crucial comments and advices on my research during my preparation for this thesis. Without her invaluable suggestions, advice and support, this thesis would have been impossible.

I thank Dr. Peter Hooper, Dr. N.G.Narasimha Prasad, Dr. Giseon Heo and Dr. Mary E. Putt, the members of my supervisory and examining committee for reading my thesis with many insightful comments.

For all these years, my parents were always there to share every piece of my anxiety and joy. I thank them for their encouragement, understanding and patience. This thesis is dedicated to my mother, Jingyu Wang, and my father, Jin Liang.

Finally, I would like to thank my husband, Guang Li, for his love, his caring, and his computer assistance. He has encouraged and supported me throughout this program.

Contents

List of Tables

List of Figures

List of Notations

Design Parameters

t: number of treatment sequences

p: number of period

s: number of different treatment sequences

N: total number of patients in the study

N_k: total number of patients in treatment sequence k, and $\sum_k N_k = N$

Model Parameters

i: period index

j: subject index

k: treatment sequence

y_{ijk}: observation from subject j in treatment sequence k in period i, where $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, N_k$, and $k = 1, 2, \ldots, s$

μ: overall mean effect

$\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_p)^T$: period effects

$\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_t)^T$: direct treatment effects

$\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_t)^T$: first-order carryover effects (or called residual effects) under the traditional model; mixed carryover effects under the self and mixed carryover effects model

$\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \ldots, \varphi_t)^T$: self carryover effects under the self and mixed carryover effects model

ξ: random subject effect following a normal distribution with mean 0 and variance σ_ξ^2

ε: random error term following a normal distribution with mean 0 and variance σ_ε^2

$\rho = \frac{\sigma_\xi^2}{\sigma_\epsilon^2 + \sigma_\xi^2}$: with-in subject correlation

$\mathbf{C}$: variance matrix of the $\mathbf{y}_{jk} = (y_{1jk}, y_{2jk}, \ldots, y_{pjk})^T$

$\mathbf{X}_k$: design matrix in treatment sequence k

$\boldsymbol{\omega}_k$: weights of the observations in treatment sequence k for the linear estimator of the parameter of interest

$\mathbf{A}$: information matrix

Stratified and Randomized Play-the-Winner Rule

i: treatment index

t: number of treatments

s: number of stratifications

μ: number of balls of each type in the urn at the initial stage and $\mu \geq 0$

α and β: design parameters. α and β are both multiples of $(t-1)$, and $\beta \geqslant \alpha(t-1) \geqslant 0$.

$R_i(n)$: number of type i balls in the urn after n responses

$S_i(n)$: numbers of successes with treatment i after n assignments

$F_i(n)$: numbers of failures with treatment i after n assignments

T_n: total number of balls after n responses

$N_i(n)$: number of patients assigned to treatment i after n assignments

p_i: probability of a single trial success for treatment i

$q_i = 1 - p_i$: probability of a single trial failure for treatment i

Response-Adaptive Allocation Rule

$g_k(.)$: evaluation function for treatment sequence k based on the existing data

Λ: criterion for the new allocation rule in Chapter 5

$\Theta(.)$: optimality criteria function such as the determinant (D-optimality), the trace (A-optimality) or maximum eigenvalue (E-optimality) of the information matrix

$\mathbb{H}_j$: allocation-and-response history up to the j^{th} patients

$A_j^k(\mathbb{H}_{j-1})$: (expected) Fisher information matrix after the j^{th} observation given the history of the first $j-1$ patients, $\mathbb{H}_{j-1}$, and the assumption that

j^{th} patient will be treated by treatment sequence k

Adaptive Two-treatment p-period Repeated Measurement Designs with Dichotomous Responses

T_k: k^{th} treatment sequence among all possible treatment sequences T, $k = 1, 2, \ldots, 2^p$.

π_r: when $r = 1, 2, \ldots, p$, π_r is the success probability of treatment A in the r^{th} period; when $r = p+1, p+2, \ldots, 2p$, π_r is the success probability of treatment B in the $(r-p)^{th}$ period.

N_{ki}: number of subjects receiving treatment sequence T_k up to the i^{th} patient, where $k = 1, 2, \ldots, 2^p$.

$S_i = (S_{1Ai}, \ldots, S_{pAi}, S_{1Bi}, \ldots, S_{pBi})^T$: S_{qti} denotes the number of successes of treatment t in the q^{th} period, where $q = 1, 2, \ldots, p$ and $t = A$ or B.

$S_i[r]$: r^{th} element of S_i.

$N_{L_i}[r]$: if $1 \leq r \leq p$, $N_{L_i}[r]$ denotes the total number of patients receiving treatment A in the r^{th} period; if $p+1 \leq r \leq 2p$, $N_{L_i}[r]$ denotes the total number of patients receiving treatment B in the $(r-p)^{th}$ period.

Chapter 1

Introduction

1.1 Background

This PhD dissertation research focuses on issues that arise in the use of repeated measurement designs for clinical trials. An optimal design for any given situation is strongly model dependent. Many researchers have constructed optimal designs based on certain models. For example, Bate and Jones (2003) considered a particular subset of crossover designs that are *uniform*. They proved that under the *traditional model*, and the assumption of independence of the error terms (Hedayat and Afsarinejad, 1978), a *uniform strongly balanced design* (Cheng and Wu, 1980) is universally optimal (Kiefer, 1975) for estimating the treatment and carryover effects.

However, since responses are measured on the same subject over several periods, the independence assumption of the error terms is often violated, and unreliable estimates of the regression parameters will be obtained. Because the error terms are correlated, the standard errors of the regression coefficients will be smaller than they should be. Hence, the statistical tests of these parameters will be misleading, and they will suggest that the estimates of the parameters are more precise than they really are. Matthews (1987) discussed this problem and proposed a method of generating optimal repeated measurement designs for the comparison of two treatments in the presence of carryover effects and autocorrelated errors. To deal with the problem of specifying an efficient design when little is known about the covariance matrix of responses, researchers have also used adaptive designs

to establish appropriate rules for assigning subjects to treatment sequences (Silvey, 1980, p. 61).

The main objective of most clinical designs is to compare the effectiveness of treatments efficiently. In these trials, we not only wish to improve the precision of treatment effect contrasts, but also to treat each patient in the best way possible. In addition, it may not be ethical to prolong a trial longer than necessary, because it may happen that an excessive number of patients might receive poor treatments. Those subjects who receive poor treatments might drop out before the experiment is complete, which results in a serious problem in statistical analysis: missing/incomplete data. Many researchers have constructed optimal designs focusing on achieving one of the above goals. However, research on adaptive designs for longitudinal and repeated responses has not received much attention. Pocock (1979) stated that any procedure must be simple, fast, objective and foolproof to be useful in practice. In this thesis, we have developed new appropriate response-adaptive rules that optimize these goals and are easily accessible to users.

With *response-adaptive designs*(RAD), we modify the trial on the basis of outcomes/responses in the previous observations in order to achieve a specific goal (Kushner 2003). Rosenberger and Lachin (1993) give a nice review of various types of RADs. The classical sequential trial is an RAD in which the decision to terminate the accession of new subjects is based on minimizing the expected sample size (Armitage, 1975). In play-the-winner designs, the goal is to minimize the number of subjects receiving an inferior treatment. This strategy is supported largely on ethical grounds (Zelen, 1969; Simon, 1977; Wei and Durham, 1978; Pocock, 1979 and Bartlett et al., 1985). Alternatively, the designs using the randomized play-the-winner rule (Wei and Durham, 1978) have been adopted in major clinical trials. *Covariate-adaptive allocation* is a sequential stratification rule used to achieve balance in the study (Pocock and Simon, 1975). In a simple case, if, at some point, treatment A is being given to more old patients than treatment B, the remaining old patients can be given treatment B until " balance" is achieved. It is clear that both covariate-adaptive designs and response-adaptive designs

require access to the history of the trial.

On the other hand, adaptive designs may have a primary goal of improving the precision of estimators of unknown parameters. Schwabe (1987) has studied the problem of estimating regression coefficients in an experimental situation, in which a fixed (classical or deterministic) optimal design can be specified. He showed that the adaptive designs are superior with respect to the A-optimality criterion to any fixed design. Kushner (2003) considered multivariate responses, for which a fixed, optimal crossover design is not available due to an unknown covariance matrix. He proposed adaptive rules for symmetric designs that specify how to assign future subjects to sequences on the basis of updated estimates of the covariance matrix. This method relaxes the assumption of a known error structure and can be generalized to other designs. Huang (2001) extended it to situations when within-subject covariance matrices are unknown and heterogeneous. However, all of these investigations focused on continuous responses and emphasized increasing the precision of treatment comparisons rather than assigning more patients to better treatments using the traditional model.

In this thesis, we improve the current design construction strategies in three directions:

1. by developing the strategy for both continuous and discrete responses;

2. by increasing both the estimation precision and the proportion of patients assigned to a better treatment, to construct multiple-objective designs;

3. by using a more general model considering two types of carryover effects and random subject effects, where the direct treatment effect will manifest itself no matter where and when the treatment is applied.

1.2 Thesis Overview

In chapter 2, we review response-adaptive design rules that have been considered to date.

In Chapter 3, we first discuss the types of carryover effects and introduce models for repeated measurement designs. We then apply the Lagrange multiplier method to solve the optimal design problem under the traditional model and the self and mixed carryover effects model. We also study the influence of baseline measurements on constructing optimal designs for two-treatment p-period ($p = 2, 3, 4$) repeated measurement designs, under the traditional model and the self and mixed carryover effects model, respectively. Overall conclusions and recommendations are given.

In Chapter 4, we develop a new allocation rule for treatment assignments in stratified and randomized sequential clinical trials. The new rule is a modified scheme in the spirit of the play-the-winner rule that skews the allocation pattern in favor of superior treatments. The results of the simulation studies are also discussed.

In chapter 5, we propose a new multiple-objective response-adaptive design strategy for constructing repeated measurement designs. This new design construction method improves the current response-adaptive design strategy, which has only used a single objective criterion. In addition, it is applicable to both dichotomous responses and continuous responses.

In chapter 6, we implement the adaptive allocation rule proposed in Chapter 5 for repeated measurement designs with dichotomous responses. We provide detailed allocation rules for constructing adaptive two-treatment p-period repeated measurement designs. The allocation results and efficiency of designs based on the simulation studies are also presented.

In chapter 7, we use the adaptive allocation rule proposed in Chapter 5 for trials with continuous responses. Under the self and mixed carryover effects model, we construct adaptive two-treatment two-period repeated measurement designs first, and then extend it to two-treatment three-period repeated measurement designs. In simulation studies, we compare the designs constructed under the new proposed allocation rule with fixed optimal designs available in the literature. We also discuss the challenges and difficulties in generalizing the implementations of the adaptive allocation rule to construct multi-treatment multi-period repeated measurement designs.

Finally, Chapter 8 summarizes the main contributions of this thesis to the literature. We also discuss possible future research to expand and improve the design strategies proposed in this thesis.

Chapter 2

Review of Allocation Rules for Response-Adaptive Designs

In this chapter, we review the basic principles concerning the design of clinical trials and the sample size in an attempt to maximize the information gained and to optimize the treatment benefit. We classify the existing adaptive allocation rules into three categories.

2.1 Minimizing the Sample Size

In sequential trials, data are analyzed as they become available, and the total number of subjects to enter the trial is not predetermined. These decisions depend on the results to be accumulated in an effort to avoid unnecessary use of inferior treatments. The trials often come to an early termination if an important difference can be established (Armitage, 1975).

Cook (1995, 1996) provided interim analyses for continuous responses in 2×2 crossover trials with serial patient entry. The goal is to allow early termination, minimize the cost and shorten the duration of the trials. Cook (1995) studied the properties of a two-stage crossover design in which patients are entered simultaneously and a single interim analysis is planned at the end of the first period. Alternately, Cook (1996) considered the case of a single interim analysis taking place after m patients have been observed on both treatments for similar trials.

Let y_{ijk} denote the response from patient j randomized to sequence k

during period i, $j = 1, 2, \ldots, m$, $i = 1, 2$, and $k = 1(AB)$ and $2(BA)$. Let N denote the total number of analyses, with n indexing the analysis stage. Since patients enter the study sequentially, at analysis stage n some patients may be randomized to group k, but are observed for a single period, while some are observed for both periods. Let $\mathbf{S}_{kn}$ and $\mathbf{T}_{kn}$ represent the set of individuals observed for one and two periods respectively, in sequence group k at analysis stage n, with $\mathbf{S}_n = \mathbf{S}_{1n} \cup \mathbf{S}_{2n}$ and $\mathbf{T}_n = \mathbf{T}_{1n} \cup \mathbf{T}_{2n}$.

Individuals in $\mathbf{S}_n$ contribute information on the effects of treatments A and B, in the same way as if they were in a complete randomized trial. Based on these individuals, Cook(1995) considered a discrepancy measure,

$$D_{1n} = \Sigma_{j \in \mathbf{S}_{2n}} y_{1j2} / \|\mathbf{S}_{2n}\| - \Sigma_{j \in \mathbf{S}_{1n}} y_{1j1} / \|\mathbf{S}_{1n}\|$$

for testing

$$H_0 : \tau = 0 \text{ vs } H_a : \tau \neq 0$$

where $\tau = \tau_2 - \tau_1$ is the difference in the efficacy of the two treatments and $\|\mathbf{S}_{kn}\|$ denotes the number of subjects in $\mathbf{S}_{kn}$.

Similarly, at stage n, individuals in $\mathbf{T}_n$ provide responses for both treatments and hence generate paired data that can be analyzed in the standard fashion for 2×2 crossover trials. Let D_{2n} represent the usual discrepancy measure based on individuals in $\mathbf{T}_n$ given by

$$D_{2n} = \{\Sigma_{j \in \mathbf{T}_{1n}} (y_{2j1} - y_{1j1}) / \|\mathbf{T}_{1n}\| - \Sigma_{j \in \mathbf{T}_{2n}} (y_{2j2} - y_{1j2}) / \|\mathbf{T}_{2n}\|\} / 2.$$

In order to make maximal use of the data provided by individuals in both $\mathbf{S}_n$ and $\mathbf{T}_n$, Cook (1995) considered a linear summary discrepancy measure consisting of a weighted combination of D_{1n} and D_{2n},

$$D_n = \omega_{1n} D_{1n} + \omega_{2n} D_{2n}.$$

If $\mathbf{T}_n$ or $\mathbf{S}_n$ are null sets, then one would naturally choose weights given by $(\omega_{1n}, \omega_{2n}) = (0, 1)$ and $(1, 0)$, respectively. In general, however, the weighting of these components is obtained by taking coefficients consisting of the inverse of the corresponding variances. Regardless of the particular weighting

scheme, we can consider $Z_n = D_n[var(D_n)]^{-1/2}$. If there is no differential carryover effect, under $H_0 : \tau = 0$, we have $Z_n \sim N(0, 1)$. If the test leads to rejection of no treatment effects, then the trial is terminated and the minimal possible sample size is attained at stage n.

2.2 Play-the-Winner Rule

Zelen (1969) developed play-the-winner rules (PWR) for a clinical trial to allocate more patients to the treatment that appears to be beneficial, based on the responses from patients already treated. Assume that patients enter the trial one at a time; the outcome of a trial is a success or failure (i.e. binary response) and only depends on the treatment given. He proposed a basic idea for the play-the-winner rule. That is, success with a current treatment generates a future trial with the same treatment, while a failure generates a trial with an alternative treatment.

In the following steps, one can easily implement this rule.

1. Place a ball marked with an "A" in a box whenever a success is obtained with treatment A or a failure with treatment B, and vice versa.

2. When a new patient is available for assignment, draw a ball randomly from the box, without replacement.

3. If the box is empty, then the assignment is determined by tossing a coin.

However, in practice, the time required to observe a patient's response to treatment may be much longer than the time between patient entries. It follows that most assignments are determined by tossing a coin. Then the PWR assigns approximately equal numbers of patients to each treatment, thus not achieving the goal.

To improve the PWR, Zelen (1969) proposed the modified play-the-winner rule (MPWR) under the assumption that patients respond immediately to treatments. Under the MPWR, after each "success" we continue to use the

same treatment. After each "failure" we switch to the other treatment. This method assigns more patients to a better treatment. However, the process overlooks past history except for the immediate past. It is not applicable to delayed responses to treatments. It also suffers from selection bias, because the response of one patient determines the allocation of the next patient, and it is evident what the next assignment will be.

Following Zelen (1969), various researchers proposed improved allocation rules. Hoel and Sobel (1971) extended the idea of the MPWR to comparing more than two treatments in a trial. Under the same assumption as MPWR, they introduced the cyclic-play-the-winner rule (PWC). The basic idea of PWC is to order the given treatments in a cyclic manner. After each "success" we continue to use the same treatment, and after each "failure" we switch to the next treatment in the ordering scheme. Again, this PWC is completely deterministic after the first assignment, and not applicable when patients have delayed responses to treatments.

Wei and Durham (1978) developed the randomized play-the-winner rule (RPWR), using all past information on allocations and responses. It removes the restriction about the immediate response and rescued the selection bias. However, RPWR considers only two treatments in a trial. The RPWR $(RPWR(\mu, \alpha, \beta), \beta \geqslant \alpha \geqslant 0)$ can be easily implemented by the following steps.

1. In a box, place two different types of balls marked A and B with μ balls of each type.

2. When a patient is available for an assignment, draw a ball from the box at random, with replacement. If it is type A, then treatment A is assigned to this patient, and vice versa.

3. When the response of a previous patient to the treatment $A(B)$ is available, change the structure of the box based on the following rule. If the response is a success, then an additional β balls of type $A(B)$ and an additional α balls of type $B(A)$ are put in the box; if this response

is a failure, then an additional α balls of type $A(B)$ and an additional β balls of type $B(A)$ are put in the box.

4. If the box is empty, then the assignment is determined by tossing a coin.

Note that the difference between the PWR and the $RPWR(0,0,1)$ is that the balls are drawn without replacement in the former case and with replacement in the latter case. Wei and Durham (1978) showed that the $RPWR(\mu,\alpha,\beta)$ introduced more randomization when β/α is small, but tended to put more patients on the better treatment when β/α is large. They also pointed out that a theoretical comparison of the $RPWR(0,0,1)$ and the PWR is quite difficult.

Wei (1979) proposed the Generalized Polya's urn design ($GUPD(\mathbf{W},\alpha,\beta)$), which was an extension of RPWR to $k(k \geq 2)$ treatments case. Where $\mathbf{W}$ is a vector that indicates how many balls of different type are in the urn at the beginning. If the response of a treatment is a "success," one can add α balls of the same type; while the response is a "failure," add β balls of each other type of balls. It can be easily implemented in clinical trials based on a generalized polya's urn model. If there is no information about the relative effectiveness of these k treatments at the outset of the trial, the author suggested that $GPUD(\mathbf{1}, k-1, 1)$ should be used. However, the scheme $GPUD(\mathbf{1}, k-1, 1)$ appears a little drastic in its early stages, especially when k is large (Wei 1979). Therefore, other alternatives may be considered. In addition, it will be more reasonable to consider possible confounders such that the treatment comparison can be made between comparable patients.

Recent efforts to generalize the principal of the PWR have been made in three directions. The first major generalization is to allow the ball selected not to be replaced, or to allow some balls to be removed from the urn (Durham and Yu 1990, Smythe 1996, Durham et al. 1998, Ivanova et al. 2000, Ivanova and Durham 2000, and Ivanova and Flournoy 2001). The second major generalization is to add different expected numbers of balls to the urn across draws (Bai and Hu 1999). The third generalization is to add

the number of balls obtained as a function of the previous draws (Andersen et al. 1994, Bai et al. 2002). Rosenberfer (2002) gave a review on the main properties and some recent developments of the urn models.

In a clinical trial comparing two treatments, Bandyopadhyay and Biswas (1999) introduced heterogeneity in patient characteristics through a discrete ordinal covariate, and Bandyopadhyay and Biswas (2000) considered the case when the response was a discrete ordinal variable.

Wei (1978) and Smith (1984) proposed adaptive biased coin designs, which are special cases of the doubly adaptive biased coin design (DBCD) proposed by Eisele (1994). The goal is to assign a predetermined proportion of patients to one of the two treatments, applicable to both continuous and discrete responses. Hu and Zhang (2004) generalized the DBCD for k-treatment clinical trials.

2.3 Maximum Likelihood (ML) Based Allocation Rules

Simon, Weiss and Hoel (1975) proposed a nondeterministic allocation plan based on the likelihood function, with the goal of reducing the use of a poorer treatment. They used assumptions similar to those in MPWR for binary responses. The allocation rule and likelihood ratio stopping rule are given as follows.

Assume that when the success probabilities satisfy $|P_A - P_B| \geqslant \Delta^*$, the probability of selecting a poorer treatment for a patient will be no greater than Δ^* . The maximum number of patients to be treated is N^*. At any point in the trial, let s_i and f_i be the number of successes and failures, respectively, with treatment $i(i = A \text{ or } B)$. Define a likelihood function

$$L(a, b) = a^{s_A} b^{s_B} (1 - a)^{f_A} (1 - b)^{f_B}$$

where $0 \leq a, b \leq 1$.

Further, define

$$L_A = \max_{\Delta^* \leqslant p \leqslant 1} L(p, p - \Delta^*)$$

$$L_B = \max_{\Delta^* \leqslant p \leqslant 1} L(p - \Delta^*, p)$$

where L_A is the maximum value of the likelihood given that treatment A is the better one, while L_B is calculated under the contrary assumption. Termination of the sequential procedure is based on the ratio $\Lambda = L_A/L_B$. The adaptive allocation scheme is based on the quantity $\theta = \Lambda/(1+\Lambda)$. In this adaptive procedure, a patient is given treatment A with probability θ and treatment B with probability $1-\theta$.

The adaptive assignment continues as long as the number of tests is less than N^* and as long as $1/k \leqslant \Lambda \leqslant k$, where k is a stopping parameter. If Λ exceeds k at any point in the trial, treatment A is deemed a better one. The stopping rule constrains θ to satisfy $1/(1+k) \leqslant \theta \leqslant k/(1+k)$.

It is a substantial improvement over PWR and MPWR, except in the case of competing treatments with high success probabilities. However, it is rather complicated for practical use.

Kushner (2003) proposed a method of constructing repeated measurement designs adaptively when little is known about the covariance matrix of responses. The rules specify the assignment of subjects to treatment sequences on the basis of updated estimates of the covariance matrix. The goal of his allocation rule is to increase the precision of treatment effect estimators, i.e., to increase the power of the design. Based on the traditional model with fixed subject effects, the N independent error vectors, $\boldsymbol{\varepsilon}_j = (\varepsilon_{ij})$, $1 \leq i \leq p$, and $1 \leq j \leq N$, are multivariate normal with mean $\mathbf{0}$, and $p \times p$ covariance matrix, $\boldsymbol{V}$, where N is the number of subjects, and p is the number of periods. A very important matrix, $\boldsymbol{V}^*$ is defined as

$$\boldsymbol{V}^* = (\nu^{ij} - \nu^i \nu^j / \nu^{tot}), \qquad 1 \leq i, j \leq p, \tag{2.1}$$

where ν^{ij} is the ij^{th} element, ν^i the i^{th} row sum, ν^j the j^{th} row sum, and ν^{tot} the total sum, of $\boldsymbol{V}^{-1}$.

He suggested starting the experiment with initial subjects using the optimal or "nearly" optimal design suggested in the literature. He then computed the maximum likelihood (ML) estimators of $\mathbf{V}^*$ and $\boldsymbol{\beta} = (\tau_1, \ldots, \tau_t, \gamma_1, \ldots,$

$\gamma_t)^T$, for which $0 = \sum_{k=1}^{t} \tau_k = \sum_{k=1}^{t} \gamma_k$, $\hat{\boldsymbol{V}}^*$ and $\hat{\boldsymbol{\beta}}$, by solving the following equations:

$$\sum_{j=1}^{N}(\mathbf{X}_j - \bar{\mathbf{X}})^T\hat{\mathbf{V}}^*(\mathbf{X}_j - \bar{\mathbf{X}})\hat{\boldsymbol{\beta}} = \sum_{j=1}^{N}(\mathbf{X}_j - \bar{\mathbf{X}})^T\hat{\mathbf{V}}^*(\mathbf{y}_j - \bar{\mathbf{y}}) \qquad (2.2)$$

$$\mathbf{B}_p\left(\sum_{j=1}^{N}(\mathbf{y}_j - \bar{\mathbf{y}} - (\mathbf{X}_j - \bar{\mathbf{X}})\hat{\boldsymbol{\beta}})(\mathbf{y}_j - \bar{\mathbf{y}} - (\mathbf{X}_j - \bar{\mathbf{X}})\hat{\boldsymbol{\beta}})^T\right)\mathbf{B}_p = (N-1)(\hat{\mathbf{V}}^*)^+ \qquad (2.3)$$

where $\mathbf{y}_j = (y_{j1}, \ldots, y_{jp})^T$, $\bar{\mathbf{y}} = (\sum_{j=1}^{N}\mathbf{y}_j)/N$, $\mathbf{X}_j = [\mathbf{T}_j; \tilde{\mathbf{T}}_j]$, $\bar{\mathbf{X}} = [\bar{\mathbf{T}}; \bar{\tilde{\mathbf{T}}}]$, $\mathbf{T}_j$ (respectively, $\tilde{\mathbf{T}}_j$) is the j^{th} subject's $p \times t$ design matrix of treatment effects (respectively, carryover effects), $\bar{\mathbf{T}} = (\sum_{j=1}^{N}\mathbf{T}_j)/N$, $\bar{\tilde{\mathbf{T}}} = (\sum_{j=1}^{N}\tilde{\mathbf{T}}_j)/N$, $\mathbf{B}_p = \mathbf{I}_p - \mathbf{J}_p/p$ and $(\hat{\mathbf{V}}^*)^+$ denotes the Moore-Penrose inverse of $\hat{\mathbf{V}}^*$.

The equations (2.2) and (2.3) are obtained by maximizing the density (2.4), defined as below, over all $2t$-dimensional vectors $\boldsymbol{\beta}$ and over all non-negative $p \times p$ matrices $\mathbf{V}^*$ such that $\mathbf{V}^*\mathbf{1}_p = \mathbf{0}$ and $\text{rank}(\mathbf{V}^*) = p - 1$

$$c(Tr_{p-1}(\mathbf{V}^*))^{(N-1)/2}exp\left(-\sum_{j=1}^{N}\frac{(\mathbf{y}_j - \bar{\mathbf{y}} - (\mathbf{X}_j - \bar{\mathbf{X}})\boldsymbol{\beta})^T\mathbf{V}^*(\mathbf{y}_j - \bar{\mathbf{y}} - (\mathbf{X}_j - \bar{\mathbf{X}})\boldsymbol{\beta})}{2}\right) \qquad (2.4)$$

where $c = (2\pi)^{-(p-1)(N-1)/2}$ and $Tr_{p-1}[\mathbf{V}^*] = p|\mathbf{V}^*_{p-1}|$.

New subjects were assigned to sequence k and its dual sequences, such that the information matrix of treatment effect, $\mathbf{C}_d(\tau)$ defined as

$$\mathbf{C}_d(\tau) = \mathbf{C}_{d11} - \mathbf{C}_{d21}^T\mathbf{C}_{d22}^{-1}\mathbf{C}_{d21},$$

will be maximized under A-, D- or E-optimality criteria. Where

$$\mathbf{C}_d(\tau, \gamma) = \begin{pmatrix} \mathbf{C}_{d11} & \mathbf{C}_{d12} \\ \mathbf{C}_{d21} & \mathbf{C}_{d22} \end{pmatrix},$$

$\mathbf{C}_{d12} = \mathbf{C}_{d21}$, and $\mathbf{C}_d(\tau, \gamma) = \sum_{j=1}^{N}(\mathbf{X}_j - \bar{\mathbf{X}})^T\mathbf{V}^*(\mathbf{X}_j - \bar{\mathbf{X}})$ is the joint information matrix of treatment and carryover effects,

Update the estimates of $\hat{\boldsymbol{V}}^*$ and $\hat{\boldsymbol{\beta}}$ using the subjects from previous steps, repeating until all subjects have been allocated.

Chapter 3

Constructing Optimal Designs for Repeated Measures Data

In this chapter, we construct optimal designs under both the traditional model and a more general model that includes the self and mixed carryover effects with random subject effects. We also study the baseline measurement effect on constructing optimal designs, and give recommendations on constructing two-treatment p-period ($p = 2, 3, 4$) repeated measurement designs.

3.1 Introduction

We first discuss models for repeated measures data, as optimal designs are strongly model dependent. The model for repeated measurement designs is a special case of mixed effects models, where both fixed effects such as treatment effects, period effects, and carryover effects (or sometimes called residual effects), and random effects like subject effects and measurement errors are considered. Ideally, we would like the residual effects to be washed out completely by the time the next treatment is applied (Figure 3.1). In Figure 3.1, the horizontal axis defines the treatment periods and the vertical axis measures the effects of treatments. The dotted lines are the carryover treatment effects lasting beyond the period of treatment application. The solid lines indicate the treatment effects increasing in a peak to relieve the symptoms or disease under consideration.

In general, it is more than likely that the treatment effects do not wash out at the same speed as they reach their peak effect. An extra washout period for the treatment effects may make it possible to proceed with the planned investigation. However, in practice it is not always known what constitutes a "sufficiently long" washout period. Even if it is the case, it may not be feasible to have sufficiently long washout period due to budget limitations, time constraints, dropout problems, etc. Figure 3.2 shows the situation when carryover effects are not washed out completely. Even when the wash-out is completely effective, the physiological or psychological state induced by the first treatment may to some extend persist, so that the subjects are no longer comparable in their clinical state at the start of the next period (Hills and Armitage 1979 and Putt 2006). In sequence, we introduce the carryover effects to the models because carryover effects are not negligible between two consecutive treatment periods, which we refer to as the "first order carryover effect." Usually we assume that the carryover effects are washed out completely or are negligible after two treatment periods. Therefore, models considering only the first order carryover effects have been used in building optimal designs and analyzing the data.

It is ideal to measure the treatment effect when it reaches its peak. However, in practice, it can be an aspect of clinical research that it is difficult or impossible to know when a treatment effect peaks, such that the treatment effect is actually measured before it reaches its peak (Figure 3.3 (I)). Sometimes, the treatment has a lasting and curative treatment effect (Figure 3.3 (II)). We do not consider this situation in this thesis because repeated measurement design is not a good experimental tool in this case.

The traditional model for repeated measurement designs (for details see Section 3.2.1) assumes that each treatment has a carryover effect which does not interact with the direct effect of the treatment in the following period. This has often been criticized as unrealistic. When a treatment follows itself, the carryover effect for the preceding period may not be identical to the carryover effect when a treatment follows the other treatments. Especially in the situation that a drug given in one period may still be present

in the body in the next period (Hills and Armitage, 1979), the assumption that a treatment's carryover effect is the same no matter which treatment follows seems more likely to be violated. To cope with this problem, Sen and Mukerjee (1987) introduced a model with interactions between direct and carryover effects, such that each treatment has a different carryover effect for every treatment in the next period. However, Sen and Mukerjee's model generally contains too many parameters to be practically useful. For example, in a clinical trial comparing three treatments, we will need to consider 9 different interaction effects between all three direct and carryover effects:$AA, AB, AC, BA, BB, BC, CA, CB$ and CC, where the interaction effect xy means the effect due to a treatment y given that there is a treatment x in the previous period.

A compromise was proposed by Afsarinejad and Hedayat (2002, and see also Section 3.2.2). They considered that each treatment has only two different carryover effects, one, called *self carryover effect* if it is followed by itself, and the other one, called *mixed carryover effect* if it is followed by any other treatment. In a clinical trial with three treatments, we would have self carryover effects A, B and C, and mixed carryover effects A, B and C.

We consider subject effects as random instead of fixed when analyzing the data, as the subjects in the study often represent a sample from a larger population.

In this chapter, we first introduce models for repeated measurement designs and discuss the type of carryover effects. Then, optimal designs will be constructed under these models with detailed discussion on some special designs. In Section 3.2, we introduce two specific types of models. Section 3.3 discusses the Lagrange multiplier solution to the optimal design problems. In Section 3.4, we consider the use of the baseline measurements for the optimal design construction under the two types of repeated measurement design models described in Section 3.2, and present the optimal designs for two-treatment p-period ($p = 2, 3, 4$) repeated measurement designs.

This chapter aims to unify all optimal design results, expanding to include more complex models and baseline measurements.

3.2 Models

3.2.1 Traditional Model

In a repeated measurement design with t treatments, p periods and N subjects, denoted by $RMD(N,p,t)$, let $\mathbf{y}_{jk} = (y_{ijk})^T$ be the vector of observations from subject j in treatment sequence k, where $i = 1,2,\ldots,p$, $j = 1,2,\ldots,N_k$, $k = 1,2,\ldots,s$, N_k is the number of subjects in sequence k, s is the total number of treatment sequences, and $\sum_k N_k = N$. A traditional model for the response $\mathbf{y}_{jk}$ is

$$\mathbf{y}_{jk} = \mathbf{X}_{jk}\,\boldsymbol{\beta} + \boldsymbol{\xi}_{jk}\mathbf{1}_{[p]} + \boldsymbol{\varepsilon}_{jk} \tag{3.1}$$

where $\mathbf{1}_{[p]}$ is a $p \times 1$ vector of ones. The parameter vector $\boldsymbol{\beta} = (\mu, \boldsymbol{\pi}^T, \boldsymbol{\tau}^T, \boldsymbol{\gamma}^T)^T$ consists of the overall mean effect μ, the period effects $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_p)^T$, the direct treatment effects $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_t)^T$ and the first-order carryover or residual effect of the treatment given in the previous period $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_t)^T$ (Laska, Meisner and Kushner, 1983; Matthews, 1987). Subject effects $\boldsymbol{\xi}_{jk}$ can be assumed fixed (Hedayat and Afsarinejad, 1978; Cheng and Wu, 1980; Laska, Meisner and Kushner, 1983; Kunert, 1983 and 1984; Hedayat and Zhao, 1990) or random (Laska and Meisner, 1985; Carriere and Reinsel, 1992 and 1993; Kushner 2003). If they are treated as random, they are typically assumed to have a multi-normal distribution with mean $\mathbf{0}$ and variance-covariance $\sigma^2_{\xi}\mathbf{1}_{[p]}\mathbf{1}^T_{[p]}$, mutually independent of the random errors $\boldsymbol{\varepsilon}_{jk} = (\varepsilon_{1jk}, \varepsilon_{2jk}, \ldots, \varepsilon_{pjk})^T$, which also follow a multi-normal distribution with mean $\mathbf{0}$ and variance-covariance $\sigma^2_{\varepsilon}\mathbf{I}_{[p]}$. In this thesis, the traditional model refers to the model defined in 3.1 with random subject effects.

3.2.2 Self and Mixed Carryover Effects Model

To address the criticisms of the modeling of carryover effects in the traditional model, Afsarinejad and Hedayat (2002) proposed an alternative model that allows for two different types of carryover effects from each treatment, which is a slight variation of Carrière (1994b). In their paper (Afsarinejad

and Hedayat, 2002), they also studied the optimal two-period repeated measurements designs with two or more treatments based on the self and mixed carryover effects model with fixed subject effects. In this thesis, we incorporate the random subject effects into the model. In sequence, a self and mixed carryover effects model is defined as

$$y_{ijk} = \mu + \pi_i + \tau_{d[i,j]} + (1 - \delta_{ij})\gamma_{d[i-1,j]} + \delta_{ij}\varphi_{d[i-1,j]} + \xi_{jk} + \varepsilon_{ijk} \qquad (3.2)$$

where y_{ij} denotes the response variable for subject j in period i, μ is an overall mean, π_i and ξ_j are the period and subject effects, respectively, $d(i,j)$ denotes the treatment used for subject j in period i, $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, N_k$, $k = 1, 2, \ldots, s$, N_k is the number of subjects in sequence k, s is the total number of treatment sequences, and $\sum_k N_k = N$. Both $\gamma_{d[i-1,j]}$ and $\varphi_{d[i-1,j]}$ represent carryover effects, while δ_{ij} is an indicator variable, taking 1 if $d(i,j) = d(i-1,j)$ and 0 otherwise. Thus $\gamma_{d[i-1,j]}$ is the carryover effect of one treatment on a different treatment, called *mixed carryover effect*, while $\varphi_{d[i-1,j]}$ is the carryover effect from a treatment onto itself, called *self carryover effect*, with $\gamma_{d[0,j]} = \varphi_{d[0,j]} = 0$. ξ_{jk} and ε_{ijk} are random effects, mutually independent, with mean 0 and variance σ_ξ^2 and σ_ε^2, respectively.

Note that, when $\sigma_\xi^2 = 0$, the model 3.2 becomes the fixed effects model with no subject effect.

3.3 Lagrange Multiplier Solution to the Optimal Design Problem

The optimal design involves determining the number of subjects to allocate to each treatment sequence in order to achieve a specific goal. It is well known that the optimal design problem is strongly model dependent. In consequence, some optimal designs, which are optimal under certain model assumptions, are not optimal under other models. Therefore, in this section, we study the optimal design problem under the two models introduced in Section 3.2.

3.3.1 General Repeated Measures Model

First, let us review the Lagrange Multiplier solution to the optimal-design problem based on a general repeated measures model.

Let $\mathbf{y}_{jk} = (y_{ijk})^T$ be the vector of observations from subject j in treatment sequence k, where $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, N_k$, and N_k is the number of subjects in treatment sequence k. In a typical experiment, among the t^p possible sequences, only a few sequences are administered. Let $N_k = 0$ for unused treatment sequences, then we have

$$N = \sum_k N_k \tag{3.3}$$

Assume that p and the total number of subjects N are fixed, and the $p \times 1$ response-vector $\mathbf{y}_{jk}$ has a constant variance-covariance matrix $\mathbf{C}$. The mean vector $\mathbf{E}[\mathbf{y}_{jk}]$ is modelled as

$$E[\mathbf{y}_{jk}] = \mathbf{X}_k \, \boldsymbol{\beta} \tag{3.4}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)^T$ is a $q \times 1$ column vector of unknown parameters, and $\mathbf{X}_k$ is a $p \times q$ design matrix for treatment sequence k. We are interested in finding designs that yield minimum-variance best linear unbiased estimators (BLUE) of any linear combination of the unknown parameters, $\theta = \mathbf{m}^T \, \boldsymbol{\beta}$, where $\mathbf{m} = (m_1, m_2, \ldots, m_q)^T$.

The linear estimator of θ is given by

$$\hat{\theta} = \sum_{jk} \boldsymbol{\omega}_k^T \mathbf{y}_{jk} \tag{3.5}$$

where $\boldsymbol{\omega}_k$ $(k = 1, 2, \ldots, t^p)$ are p-dimensional vectors, and they are the weights of the observations.

For $\hat{\theta}$ to be unbiased, we have

$$E(\hat{\theta}) = \theta$$

i.e.,

$$E(\sum_{jk} \boldsymbol{\omega}_k^T \mathbf{y}_{jk}) = \mathbf{m}^T \, \boldsymbol{\beta}$$

And also,

$$E(\sum_{jk} \omega_k^T \mathbf{y}_{jk}) = \sum_{jk} \omega_k^T E(\mathbf{y}_{jk}) = \sum_{jk} \omega_k^T \mathbf{X}_k \boldsymbol{\beta}$$
$$= (\sum_{k} N_k \, \omega_k^T \mathbf{X}_k) \boldsymbol{\beta} = \mathbf{m}^T \, \boldsymbol{\beta}$$

Therefore, ω_k must satisfy the q linear constraints

$$\sum_{k} N_k \, \omega_k^T \mathbf{X}_k^s = m_s, \; s = 1, 2, \ldots, q \tag{3.6}$$

where, for each k, $\mathbf{X}_k^s$ is the s^{th} column of the matrix $\mathbf{X}_k$.

And

$$Var(\hat{\theta}) = Var(\sum_{jk} \omega_k^T \mathbf{y}_{jk}) = \sum_{jk} Var(\omega_k^T \mathbf{y}_{jk}) = \sum_{jk} \omega_k^T Var(\mathbf{y}_{jk}) \omega_k$$
$$= \sum_{jk} \omega_k^T \mathbf{C} \omega_k = \sum_{k} N_k \omega_k^T \mathbf{C} \omega_k$$

Therefore, the variance of $\hat{\theta}$ is given by

$$Var(\hat{\theta}) = \sum_{k} N_k \omega_k^T \mathbf{C} \omega_k \tag{3.7}$$

For particular values of N_k, the BLUE of θ and its corresponding variance are easily computed by solving Equations 3.5, 3.6 and 3.7. The optimal design problem we consider is the determination of the number of subjects, N_k, to allocate to each sequence k, under the constraint (3.3), that yields the BLUE of θ with minimum variance. Note that the constraints (3.3) are for integers N_k and fixed N. According to Kiefer and Wolfowitz (1959), we define a discrete probability measure $p_k = N_k/N$. The fact that p_k can only take on multiples of $1/N$ makes the optimality problem very difficult in general. However, the minimum of (3.7) does not depend on N. Thus, if we choose N beforehand such that Np_k takes on only integral values, it yields an exact solution to the original optimal design problem. We shall see in Sections 3.3.3 and 3.4, in all examples we considered in this thesis, it is feasible to choose such a value of N. In the sequel we treat N_k as continuous variables.

Introduce Lagrange multipliers, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_q)^T$ and λ_0, corresponding respectively to constraints (3.6) and (3.3), and minimized the function

$$f(N_1, N_2, \ldots, N_k; \lambda_0, \lambda_1, \lambda_2, \ldots, \lambda_q)$$

$$= \textstyle\sum_k N_k \boldsymbol{\omega}_k^T \mathbf{C} \boldsymbol{\omega}_k - 2 \sum_s \lambda_s (\sum_k N_k \ \boldsymbol{\omega}_k^T \mathbf{X}_k^s - m_s) - \lambda_0 (\sum_k N_k - N)$$

Then we set the differentials to zero to get

$$\frac{\partial f}{\partial \boldsymbol{\omega}_k} = 2N_k \mathbf{C} \boldsymbol{\omega}_k - 2 \sum_s \lambda_s N_k \mathbf{X}_k^s = 0, \text{ for each k}$$

Therefore, for given $N_k \neq 0$, the weights of the observations in the BLUE of any repeated measurement design are given by

$$\boldsymbol{\omega}_k = \sum_s \lambda_s \mathbf{C}^{-1} \mathbf{X}_k^s \tag{3.8}$$

or

$$\boldsymbol{\omega}_k = \mathbf{C}^{-1}(\mathbf{X}_k^1, \mathbf{X}_k^2, \cdots, \mathbf{X}_k^q)(\lambda_1, \lambda_2, \cdots, \lambda_q)^T = \mathbf{C}^{-1} \mathbf{X}_k \boldsymbol{\lambda} \tag{3.9}$$

Taking the left product with $\mathbf{X}_k^T$ on both sides of Equation 3.9, multiplying by N_k and summing over k, we get

$$\sum_k N_k {\mathbf{X}_k}^T \boldsymbol{\omega}_k = \sum_k N_k {\mathbf{X}_k}^T \mathbf{C}^{-1} \mathbf{X}_k \boldsymbol{\lambda}$$

i.e.,

$$\begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_q \end{pmatrix} = (\sum_k N_k \mathbf{X}_k^T \mathbf{C}^{-1} \mathbf{X}_k) \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_q \end{pmatrix}$$

Let

$$\mathbf{A} = \sum_k N_k \mathbf{X}_k^T \mathbf{C}^{-1} \mathbf{X}_k \tag{3.10}$$

then

$$\boldsymbol{\lambda} = \mathbf{A}^{-1} \mathbf{m} \tag{3.11}$$

For given N_k, the variance of the BLUE is

$$
\begin{aligned}
Var(\hat{\theta}) &= \sum_k N_k \boldsymbol{\omega}_k^T \mathbf{C} \boldsymbol{\omega}_k = \sum_k N_k \boldsymbol{\omega}_k^T \mathbf{C} (\sum_s \lambda_s \mathbf{C}^{-1} \mathbf{X}_k^s) \\
&= \sum_k N_k \boldsymbol{\omega}_k^T (\sum_s \lambda_s \mathbf{X}_k^s) = \sum_s \lambda_s (\sum_k N_k \boldsymbol{\omega}_k^T \mathbf{X}_k^s) \\
&= \sum_s \lambda_s m_s = \mathbf{m}^T \boldsymbol{\lambda} = \mathbf{m}^T \mathbf{A}^{-1} \mathbf{m}
\end{aligned}
$$

In order to increase the estimation precision, we need to minimize the variance of the estimation, or equivalently, maximize the matrix **A**. In fact, the matrix **A** is the information matrix of θ.

3.3.2 Two-Treatment Repeated Measures Data

In two-treatment repeated measurement designs, a duality in the design matrices permits simplification of the search for the optimal choice of N_k.

Let $d(i,k)$ denote the treatment, A or B, given in the i^{th} period in sequence k. The treatment sequence k^* is the **dual** of k if, for all i, $d(i,k^*)$ is not equal to $d(i,k)$. For example, $BBAB$ is the dual of $AABA$.

The parameters of models, in which the concept of duals is useful, satisfy specific conditions. Some parameters, such as the general effect (mean effect and period effects in a traditional repeated measurement design model), are absent from the contrast of interest. For such parameters, $m_s = 0$ in the equation $\theta = \sum_s m_s \beta_s$, and the s^{th} column of each of the design matrices $\mathbf{X}_k$ are equal, $\mathbf{X}_k^s = \mathbf{X}^s$. For the remaining parameters, corresponding column vectors of the design matrices in general depend on the specific treatment sequence, but in some cases they may be the negatives of the corresponding column vectors of their dual. When these conditions are satisfied, the lemma on duals gives sufficient conditions on N_k and the weights of the BLUE for a design to be optimal.

Lemma 3.3.2.1. *Let Γ be a nonempty subset of the integers. Suppose that for $s \in \Gamma$, we have $\boldsymbol{X}_k^s = -\boldsymbol{X}_{k^*}^s$, where k^* is the dual of k; and for $s \notin \Gamma$, we have $m_s = 0$ and $\boldsymbol{X}_k^s = \boldsymbol{X}^s$. Then, a design with weights, $\boldsymbol{\omega}_k$, satisfying $\boldsymbol{\omega}_k = -\boldsymbol{\omega}_{k^*}$ and $N_{k^*} = N_k$, is optimal.*

Proof. Suppose that the optimal design has an allocation of subjects and weights of the BLUE, denoted, respectively, by $N_k^{'}$ and $\boldsymbol{\nu}_k$.

For each subject j, such that $N_j^{'} + N_{j^*}^{'} \neq 0$, introduce new allocation N_j and weights $\boldsymbol{\omega}_j$ defined by

$$N_j = N_{j^*} = (N_j^{'} + N_{j^*}^{'})/2$$

and

$$\boldsymbol{\omega}_j = -\boldsymbol{\omega}_{j^*} = (N_j^{'}\,\boldsymbol{\nu}_j - N_{j^*}^{'}\,\boldsymbol{\nu}_{j^*})/(N_j^{'} + N_{j^*}^{'})$$

Then

$$N_j + N_{j^*} = 2 \times \frac{(N_j^{'} + N_{j^*}^{'})}{2} = N_j^{'} + N_{j^*}^{'}$$

For $s \in \Gamma$,

$$\begin{aligned}
& N_j\,\boldsymbol{\omega}_j^T\mathbf{X}_j^s + N_{j^*}\,\boldsymbol{\omega}_{j^*}^T\mathbf{X}_{j^*}^s \\
= \; & \frac{(N_j^{'} + N_{j^*}^{'})}{2} \times \frac{N_j^{'}\,\boldsymbol{\nu}_j^T - N_{j^*}^{'}\,\boldsymbol{\nu}_{j^*}^T}{N_j^{'} + N_{j^*}^{'}} \times \mathbf{X}_j^s \\
+ \; & \frac{(N_j^{'} + N_{j^*}^{'})}{2} \times \left(-\frac{N_j^{'}\,\boldsymbol{\nu}_j^T - N_{j^*}^{'}\,\boldsymbol{\nu}_{j^*}^T}{N_j^{'} + N_{j^*}^{'}}\right)(-\mathbf{X}_j^s) \\
= \; & N_j^{'}\,\boldsymbol{\nu}_j^T\mathbf{X}_j^s - N_{j^*}^{'}\,\boldsymbol{\nu}_{j^*}^T\mathbf{X}_j^s \\
= \; & N_j^{'}\,\boldsymbol{\nu}_j^T\mathbf{X}_j^s + N_{j^*}^{'}\,\boldsymbol{\nu}_{j^*}^T\mathbf{X}_{j^*}^s
\end{aligned}$$

For $s \notin \Gamma$,

$$\begin{aligned}
& N_j\,\boldsymbol{\omega}_j^T\mathbf{X}_j^s + N_{j^*}\,\boldsymbol{\omega}_{j^*}^T\mathbf{X}_{j^*}^s \\
= \; & \frac{(N_j^{'} + N_{j^*}^{'})}{2} \times \frac{N_j^{'}\,\boldsymbol{\nu}_j^T - N_{j^*}^{'}\,\boldsymbol{\nu}_{j^*}^T}{N_j^{'} + N_{j^*}^{'}} \times \mathbf{X}_j^s \\
+ \; & \frac{(N_j^{'} + N_{j^*}^{'})}{2} \times \left(-\frac{N_j^{'}\,\boldsymbol{\nu}_j^T - N_{j^*}^{'}\,\boldsymbol{\nu}_{j^*}^T}{N_j^{'} + N_{j^*}^{'}}\right) \times \mathbf{X}_j^s \\
= \; & 0
\end{aligned}$$

Since the original allocations and weights satisfy the constraints (3.3 and 3.6), so do the new allocations and weights.

And,

$$
\begin{aligned}
& (N_j' + N_{j^*}')[(N_j' \boldsymbol{\nu}_j^T \mathbf{C} \boldsymbol{\nu}_j + N_{j^*}' \boldsymbol{\nu}_{j^*}^T \mathbf{C} \boldsymbol{\nu}_{j^*}) - (N_j \boldsymbol{\omega}_j^T \mathbf{C} \boldsymbol{\omega}_j + N_{j^*} \boldsymbol{\omega}_{j^*}^T \mathbf{C} \boldsymbol{\omega}_{j^*})] \\
= \; & (N_j')^2 (\boldsymbol{\nu}_j^T \mathbf{C} \boldsymbol{\nu}_j) + (N_{j^*}')^2 (\boldsymbol{\nu}_{j^*}^T \mathbf{C} \boldsymbol{\nu}_{j^*}) + N_j' N_{j^*}' (\boldsymbol{\nu}_j^T \mathbf{C} \boldsymbol{\nu}_j) + N_j' N_{j^*}' (\boldsymbol{\nu}_{j^*}^T \mathbf{C} \boldsymbol{\nu}_{j^*}) \\
& (N_j' + N_{j^*}') \times (-2) \times \frac{(N_j' + N_{j^*}')}{2} \times \frac{N_j' \boldsymbol{\nu}_j^T - N_{j^*}' \boldsymbol{\nu}_{j^*}^T}{N_j' + N_{j^*}'} \times \mathbf{C} \times \frac{N_j' \boldsymbol{\nu}_j - N_{j^*}' \boldsymbol{\nu}_{j^*}}{N_j' + N_{j^*}'} \\
= \; & N_j' N_{j^*}' (\boldsymbol{\nu}_j^T \mathbf{C} \boldsymbol{\nu}_j) + N_j' N_{j^*}' (\boldsymbol{\nu}_{j^*}^T \mathbf{C} \boldsymbol{\nu}_{j^*}) - 2 N_j' N_{j^*}' (\boldsymbol{\nu}_j^T \mathbf{C} \boldsymbol{\nu}_{j^*}) \\
= \; & N_j' N_{j^*}' (\boldsymbol{\nu}_j^T \mathbf{C}_2^1 - \boldsymbol{\nu}_{j^*}^T \mathbf{C}_2^1)(\boldsymbol{\nu}_j^T \mathbf{C}_2^1 - \boldsymbol{\nu}_{j^*}^T \mathbf{C}_2^1)^T \\
\geq \; & 0
\end{aligned}
$$

Clearly, the variance of $\hat{\theta}$ made by the new allocations and weights for sequences j and j^* is not more than that made by the original allocation. Therefore, the new allocation cannot be worse. □

If the conditions of *Lemma 3.3.2.1* are satisfied, the unbiasedness constraints (3.6) become

$$\sum_{d(1,k)=A} N_k \, \boldsymbol{\omega}_k^T \mathbf{X}_k^s + \sum_{d(1,k^*)=B} N_{k^*} \, \boldsymbol{\omega}_{k^*}^T \mathbf{X}_{k^*}^s = m_s, \; s = 1, 2, \ldots, q$$

For $s \in \Gamma$:

$$\sum_{d(1,k^*)=B} N_{k^*} \, \boldsymbol{\omega}_{k^*}^T \mathbf{X}_{k^*}^s = \sum_{d(1,k)=A} N_k \, (-\boldsymbol{\omega}_k^T)(-\mathbf{X}_k^s) = \sum_{d(1,k)=A} N_k \, \boldsymbol{\omega}_k^T \mathbf{X}_k^s$$

i.e.,

$$2 \sum_{d(1,k)=A} N_k \, \boldsymbol{\omega}_k^T \mathbf{X}_k^s = m_s$$

$$\sum_{d(1,k)=A} N_k \, \boldsymbol{\omega}_k^T \mathbf{X}_k^s = \frac{m_s}{2}$$

For $s \notin \Gamma$:

$$\sum_{d(1,k^*)=B} N_{k^*} \, \boldsymbol{\omega}_{k^*}^T \mathbf{X}_{k^*}^s = \sum_{d(1,k)=A} N_k \, (-\boldsymbol{\omega}_k^T) \mathbf{X}_k^s = - \sum_{d(1,k)=A} N_k \, \boldsymbol{\omega}_k^T \mathbf{X}_k^s$$

Then the unbiasedness constraints (3.6) are satisfied automatically and independent of N_k and $\boldsymbol{\omega}_k$.

Therefore, we only need to consider the unbiasedness constraints for $s \in \Gamma$. The duality effectively halves the number of sequences that need to be considered.

Then the optimal design problem becomes that of determining the number of subjects, N_k, to allocate to each sequence k satisfying $d(1,k) = A$, under the constraints

$$\sum_{d(1,k)=A} N_k = \frac{N}{2} \tag{3.12}$$

and

$$\sum_{d(1,k)=A} N_k \, \boldsymbol{\omega}_k^T \mathbf{X}_k^s = \frac{m_s}{2}, s \in \Gamma \tag{3.13}$$

that minimize

$$Var(\hat{\theta}) = 2 \sum_{d(1,k)=A} N_k \boldsymbol{\omega}_k^T \mathbf{C} \boldsymbol{\omega}_k \tag{3.14}$$

Upon introducing Lagrange multipliers, $\boldsymbol{\lambda}^\Gamma = (\lambda_s)^T$, where $s \in \Gamma$ and λ_0, correspond respectively to constraints (3.13) and (3.12), and minimize the function

$$\begin{aligned} f \quad & (N_1, N_2, \ldots, N_k; \lambda_0, \lambda_s, s \in \Gamma) \\ = \; & 2 \sum_{d(1,k)=A} N_k \boldsymbol{\omega}_k^T \mathbf{C} \boldsymbol{\omega}_k - 2 \sum_{s \in \Gamma} \lambda_s \Big(\sum_{d(1,k)=A} N_k \, \boldsymbol{\omega}_k^T \mathbf{X}_k^s - \frac{m_s}{2} \Big) \\ & - \lambda_0 \Big(\sum_{d(1,k)=A} N_k - \frac{N}{2} \Big), \end{aligned} \tag{3.15}$$

We get

$$\boldsymbol{\omega}_k = \frac{1}{2} \sum_{s \in \Gamma} \lambda_s \mathbf{C}^{-1} \mathbf{X}_k^s = \frac{1}{2} \mathbf{C}^{-1} \mathbf{X}_k^\Gamma \boldsymbol{\lambda}^\Gamma, \tag{3.16}$$

where $\mathbf{X}_k^\Gamma$ is a submatrix of the design matrix $\mathbf{X}_k$ including all $\mathbf{X}_k^s$ with $s \in \Gamma$.

We have

$$\boldsymbol{\lambda}^\Gamma = \mathbf{A}^{\Gamma^{-1}} \mathbf{m}^\Gamma \tag{3.17}$$

where

$$\mathbf{A}^\Gamma = \sum_{d(1,k)=A} N_k \mathbf{X}_k^{\Gamma^T} \mathbf{C}^{-1} \mathbf{X}_k^\Gamma \tag{3.18}$$

and

$$\mathbf{m}^\Gamma = (m_s)^T, s \in \Gamma$$

Also,

$$\begin{aligned} Var(\hat{\theta}) &= 2\sum_{d(1,k)=A} N_k\boldsymbol{\omega}_k^T\mathbf{C}\boldsymbol{\omega}_k = 2\sum_{d(1,k)=A} N_k\boldsymbol{\omega}_k^T\mathbf{C}(\frac{1}{2}\sum_{s\in\Gamma}\lambda_s\mathbf{C}^{-1}\mathbf{X}_k^s) \\ &= \sum_{d(1,k)=A} N_k\boldsymbol{\omega}_k^T\sum_{s\in\Gamma}\lambda_s\mathbf{X}_k^s = \sum_{s\in\Gamma}\lambda_s(\sum_{d(1,k)=A} N_k\boldsymbol{\omega}_k^T\mathbf{X}_k^s) \\ &= \sum_{s\in\Gamma}\lambda_s\frac{m_s}{2} = \frac{1}{2}\mathbf{m}^{\Gamma T}\lambda^\Gamma = \frac{1}{2}\mathbf{m}^{\Gamma T}\mathbf{A}^{\Gamma -1}\mathbf{m}^\Gamma \end{aligned} \tag{3.19}$$

Then the optimal design is obtained by minimizing the variance in Equation 3.19, or maximizing the information matrix $\mathbf{A}^\Gamma$ defined in Equation 3.18.

3.3.3 Example

We now consider a traditional two-treatment model

$$E[y_{ijk}] = \mu + \pi_i + \tau_{d(i,j)} + \gamma_{d(i-1,j)} \tag{3.20}$$

where μ, π_i, $\tau_{d(i,j)}$ and $\gamma_{d(i-1,j)}$ are the general effect, the *ith* period effect, the direct effect of treatment $d(i,j)$, and the carryover effect of treatment $d(i-1,j)$, respectively, subject to $\tau_A+\tau_B = \gamma_A+\gamma_B = 0$, where $d(i,j)$ denotes the treatment used for subject j in period i, $i = 1,2,\ldots,p$, $j = 1,2,\ldots,N_k$, $k = 1,2,\ldots,2^p$ and N_k is the number of subjects in sequence k.

Let $\tau = (\tau_A - \tau_B)/2$ and $\gamma = (\gamma_A - \gamma_B)/2$. Then, in effect, $\tau = \tau_A = -\tau_B$ and $\gamma = \gamma_A = -\gamma_B$. The model becomes

$$E[y_{ijk}] = \mu + \pi_i + \Phi_{d(i,j)}\tau + \Phi_{d(i-1,j)}\gamma \tag{3.21}$$

where

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

and

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

Under the equicorrelated assumption, the covariance matrix of the vector $\mathbf{y}_{jk}$ is

$$\mathbf{C} = \sigma_\varepsilon^2\mathbf{I}_p + \sigma_\xi^2\mathbf{1}_p\mathbf{1}_p^T \tag{3.22}$$

and the correlation between y_{ijk} and $y_{i'jk}$ $(i \neq i')$, called the within-subject correlation, is

$$\rho = \frac{\sigma_\xi^2}{\sigma_\varepsilon^2 + \sigma_\xi^2} \tag{3.23}$$

For notational simplicity, divide Equation 3.22 by σ_ε^2, but continue to denote the resulting matrix as $\mathbf{C}$. To obtain the true variance, all of the following expressions for variance need to be multiplied by σ_ε^2.

For $\rho \neq 1$,

$$\mathbf{C} = \mathbf{I}_p + \frac{\rho}{1-\rho}\mathbf{1}_p\mathbf{1}_p^T \tag{3.24}$$

and

$$\mathbf{C}^{-1} = \mathbf{I}_p - \frac{\rho}{1+(p-1)\rho}\mathbf{1}_p\mathbf{1}_p^T \tag{3.25}$$

The Case of $p = 2$

For a two-period design ($p = 2$), Equation 3.24 and Equation 3.25 become

$$\mathbf{C} = \begin{pmatrix} 1+\frac{\rho}{1-\rho} & \frac{\rho}{1-\rho} \\ \frac{\rho}{1-\rho} & 1+\frac{\rho}{1-\rho} \end{pmatrix}$$

and

$$\mathbf{C}^{-1} = \begin{pmatrix} \frac{1}{\rho+1} & -\frac{\rho}{\rho+1} \\ -\frac{\rho}{\rho+1} & \frac{1}{\rho+1} \end{pmatrix}$$

Since the conditions of the *Lemma 3.3.2.1* are satisfied, we need only to consider two sequences, AB and AA. Assume m patients receive AB treatment sequence and then $(N/2 - m)$ receive AA treatment sequence.

First, let us find the optimal design for estimating of τ, and let $\mathbf{m}^\Gamma = (1,0)^T$ and $\boldsymbol{\beta}^\Gamma = (\tau, \gamma)^T$.

Then according to Equation 3.18, we have

$$\begin{aligned}
\mathbf{A}^{\Gamma} &= \sum_{d(1,k)=A} N_k \mathbf{X}_k^{\Gamma T} \mathbf{C}^{-1} \mathbf{X}_k^{\Gamma} \\
&= m \times \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}^T \begin{pmatrix} \frac{1}{\rho+1} & -\frac{\rho}{\rho+1} \\ -\frac{\rho}{\rho+1} & \frac{1}{\rho+1} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \\
&+ (\frac{N}{2} - m) \times \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}^T \begin{pmatrix} \frac{1}{\rho+1} & -\frac{\rho}{\rho+1} \\ -\frac{\rho}{\rho+1} & \frac{1}{\rho+1} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \\
&= \begin{pmatrix} \frac{4m\rho+N-N\rho}{\rho+1} & -\frac{4m\rho-N+N\rho}{2(\rho+1)} \\ -\frac{4m\rho-N+N\rho}{2(\rho+1)} & \frac{N}{2(\rho+1)} \end{pmatrix}
\end{aligned} \tag{3.26}$$

And also, according to Equation 3.17, we have

$$\boldsymbol{\lambda}^{\Gamma} = \mathbf{A}^{\Gamma -1} \mathbf{m}^{\Gamma} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} \frac{2N(\rho+1)}{N^2-16m^2+8mN-N^2\rho^2} \\ \\ \frac{2(4m-N+N\rho)(\rho+1)}{N^2-16m^2+8mN-N^2\rho^2} \end{pmatrix} \tag{3.27}$$

And according to Equation 3.19,

$$var(\hat{\tau}) = \frac{1}{2} \mathbf{m}^{\Gamma T} \boldsymbol{\lambda}^{\Gamma} = \frac{1}{2} \lambda_1 = \frac{N(\rho+1)}{N^2 - 16m^2 + 8mN - N^2\rho^2}$$

Clearly, the minimum of $var(\hat{\tau})$ is achieved at $m = N/4$, which means the optimal design is AA, AB, BA and BB with an equal number of subjects per sequence.

Plug $m = N/4$ into Equation 3.27, and we have

$$\boldsymbol{\lambda}^{\Gamma} = \left(\frac{2(\rho+1)}{N(2-\rho^2)}, \frac{2\rho(\rho+1)}{N(2-\rho^2)} \right)^T$$

and

$$var(\hat{\tau}) = \frac{(\rho+1)\sigma_{\varepsilon}^2}{N(2-\rho^2)}$$

Based on Equation 3.16, the weights of the BLUE are

$$\boldsymbol{\omega}_1 = \frac{1}{2} \mathbf{C}^{-1} \mathbf{X}_1^{\Gamma} \boldsymbol{\lambda}^{\Gamma} = \left(\frac{-\rho - 1 + \rho^2}{N(-2+\rho^2)}, \frac{1}{N(-2+\rho^2)} \right)^T$$

and

$$\boldsymbol{\omega}_2 = \frac{1}{2} \mathbf{C}^{-1} \mathbf{X}_2^{\Gamma} \boldsymbol{\lambda}^{\Gamma} = \left(\frac{\rho - 1 + \rho^2}{N(-2+\rho^2)}, \frac{-1}{N(-2+\rho^2)} \right)^T$$

We can see that the weights of the BLUE for the optimal design depend on ρ. Letting $\rho \to 1$ (i.e. $\sigma_\xi^2 \to \infty$) yields the weights $(1/N, -1/N)^T$ for the sequence AB, and the weights $(-1/N, 1/N)^T$ for sequence AA, and the $var(\hat{\tau}) = 2\sigma_\varepsilon^2/N$. And letting $\rho = 0$ (i.e. $\sigma_\xi^2 = 0$) yields the weights $(1/2N, -1/2N)^T$ for the sequence AB, and the weights $(1/2N, 1/2N)^T$ for sequence AA, and the $var(\hat{\tau}) = \sigma_\varepsilon^2/2N$.

Now, let us find the optimal design for estimating γ.

We have $\mathbf{m}^\Gamma = (0,1)^T$, $\boldsymbol{\beta}^\Gamma = (\tau, \gamma)^T$, and

$$\mathbf{A}^\Gamma = \sum_{d(1,k)=A} N_k \mathbf{X}_k^{\Gamma T} \mathbf{C}^{-1} \mathbf{X}_k^\Gamma = \begin{pmatrix} \frac{4m\rho+N-N\rho}{\rho+1} & -\frac{4m\rho-N+N\rho}{2(\rho+1)} \\ -\frac{4m\rho-N+N\rho}{2(\rho+1)} & \frac{N}{2(\rho+1)} \end{pmatrix}$$

and

$$\boldsymbol{\lambda}^\Gamma = \mathbf{A}^{\Gamma -1} \mathbf{m}^\Gamma = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} \frac{2(4m-N+N\rho)(\rho+1)}{N^2-16m^2+8mN-N^2\rho^2} \\ \frac{4(4m\rho+N-N\rho)(\rho+1)}{N^2-16m^2+8mN-N^2\rho^2} \end{pmatrix}$$

Hence,

$$var(\hat{\gamma}) = \frac{1}{2}\mathbf{m}^{\Gamma T}\boldsymbol{\lambda}^\Gamma = \frac{1}{2}\lambda_2 = \frac{2(4m\rho + N - N\rho)(\rho+1)}{N^2 - 16m^2 + 8mN - N^2\rho^2}$$

The minimum of $var(\hat{\gamma})$ is $(\rho+1)\sigma_\varepsilon^2/N$, which is achieved at $m = N(1-\rho)/4$. So the optimal design for estimating γ depends on the value of ρ. One can implement the optimal design by obtaining the value of ρ from the observed data or literature.

We have

$$\boldsymbol{\lambda}^\Gamma = \left(0, \frac{2(\rho+1)}{N}\right)^T$$

The weights of the BLUE are

$$\boldsymbol{\omega}_1 = \frac{1}{2}\mathbf{C}^{-1}\mathbf{X}_1^\Gamma \boldsymbol{\lambda}^\Gamma = \left(\frac{-\rho}{N}, \frac{1}{N}\right)^T$$

and

$$\boldsymbol{\omega}_2 = \frac{1}{2}\mathbf{C}^{-1}\mathbf{X}_2^\Gamma \boldsymbol{\lambda}^\Gamma = \left(\frac{-\rho}{N}, \frac{1}{N}\right)^T$$

When $\rho \to 1$, the optimal design for estimation of γ is AA, BB with $N/2$ subjects per sequence; and the weights for the sequence AA are $(-1/N, 1/N)^T$;

and the weights for the sequence BB are $(1/N, -1/N)^T$; and the variance of γ is $2\sigma_\varepsilon^2/N$.

When $\rho = 0$, the optimal design for estimation of γ is AA,AB, BA and BB with $N/4$ subjects per sequence; and the weights for the sequence AB are $(0, 1/N)^T$; and the weights for the sequence AA are $(0, 1/N)^T$; and the variance of γ is σ_ε^2/N. These results are consistent with these of Laska and Meisner (1985) and Carriere (1994), as expected.

The Case of $p = 3$

Applying the approach to a three-period design ($p = 3$), Equation 3.24 and Equation 3.25 become

$$\mathbf{C} = \begin{pmatrix} 1+\frac{\rho}{1-\rho} & \frac{\rho}{1-\rho} & \frac{\rho}{1-\rho} \\ \frac{\rho}{1-\rho} & 1+\frac{\rho}{1-\rho} & \frac{\rho}{1-\rho} \\ \frac{\rho}{1-\rho} & \frac{\rho}{1-\rho} & 1+\frac{\rho}{1-\rho} \end{pmatrix}$$

and

$$\mathbf{C}^{-1} = \begin{pmatrix} \frac{\rho+1}{2\rho+1} & -\frac{\rho}{2\rho+1} & -\frac{\rho}{2\rho+1} \\ -\frac{\rho}{2\rho+1} & \frac{\rho+1}{2\rho+1} & -\frac{\rho}{2\rho+1} \\ -\frac{\rho}{2\rho+1} & -\frac{\rho}{2\rho+1} & \frac{\rho+1}{2\rho+1} \end{pmatrix}$$

Although there are eight different treatment sequences in a two-treatment three-period design, according to the *Lemma 3.3.2.1*, we need only to consider four of them: AAA, AAB, ABA and ABB. Assume N_k patients receive the k^{th} treatment sequence, $k = 1, 2, 3, 4$ and $\sum_{k=1}^{4} N_k = N/2$.

And also, the design matrices including the direct and carryover treatment contrast columns for AAA, AAB, ABA and ABB are given below respectively.

$$\mathbf{X}_1^\Gamma = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}; \quad \mathbf{X}_2^\Gamma = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ -1 & 1 \end{pmatrix};$$
$$\mathbf{X}_3^\Gamma = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 1 & -1 \end{pmatrix}; \quad \mathbf{X}_4^\Gamma = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ -1 & -1 \end{pmatrix}.$$

To find the optimal design for estimating the direct treatment effect τ, let $\mathbf{m}^\Gamma = (1, 0)^T$ and $\boldsymbol{\beta}^\Gamma = (\tau, \gamma)^T$.

Then according to Equation 3.18, we have

$$\mathbf{A}^{\Gamma} = \sum_{d(1,k)=A} N_k \mathbf{X}_k^{\Gamma T} \mathbf{C}^{-1} \mathbf{X}_k^{\Gamma} \tag{3.28}$$

$$= \begin{pmatrix} \frac{3N_1-3N_1\rho+5N_2\rho+3N_2+5N_3\rho+3N_3+5N_4\rho+3N_4}{2\rho+1} & \frac{2(N_1-N_1\rho-N_2\rho-2N_3\rho-N_3)}{2\rho+1} \\ \frac{2(N_1-N_1\rho-N_2\rho-2N_3\rho-N_3)}{2\rho+1} & \frac{2(N_1+N_2+2N_3\rho+N_3+2N_4\rho+N_4)}{2\rho+1} \end{pmatrix}$$

Plug Equation 3.2 into Equation 3.19 and do a little algebra. Under the constraint (3.12), the minimum of $var(\hat{\tau})$ is $(2\rho+1)\sigma_\varepsilon^2/(N(5\rho+3))$, achieved at $N_1 = N_2 = N_3 = 0$, and $N_4 = N/2$, which means the optimal two-treatment three-period design under the model 3.20 is ABB/BAA with an equal number of subjects per sequence.

Based on Equation 3.16 and *Lemma 3.3.2.1*, the weights of the BLUE of τ are

$$\begin{aligned} \boldsymbol{\omega}_{ABB} &= -\boldsymbol{\omega}_{BAA} \\ &= \frac{1}{2}\mathbf{C}^{-1}\mathbf{X}_4^{\Gamma}\mathbf{A}^{\Gamma-1}\mathbf{m}^{\Gamma} \\ &= \left(\frac{3\rho+1}{N(5\rho+3)}, \; -\frac{\rho+1}{N(5\rho+3)}, \; -\frac{\rho+1}{N(5\rho+3)} \right)^T \end{aligned}$$

Once again, the weights of the BLUE for the optimal design depend on ρ. When $\rho \to 1$, the weights for the sequence ABB are $(1/2N, -1/4N, -1/4N)^T$, and the weights for sequence BAA are $(-1/2N, 1/4N, 1/4N)^T$, and the variance of τ is $3\sigma_\varepsilon^2/8N$. When $\rho = 0$, we have the weights $(1/3N, -1/3N, -1/3N)^T$ for the sequence ABB, and the weights $(-1/3N, 1/3N, 1/3N)^T$ for sequence BAA, and the variance of τ is $\sigma_\varepsilon^2/3N$.

Without much difficulty, one can show that the design ABB/BAA is the optimal design for estimation of γ as well. However, the variance and the weights of the BLUE of γ are independent of the value of ρ. They are

$$var(\hat{\gamma}) = \sigma_\varepsilon^2/2N$$

and

$$\boldsymbol{\omega}_{ABB} = -\boldsymbol{\omega}_{BAA} = \left(0, \; \frac{1}{2N}, \; -\frac{1}{2N} \right)^T$$

In fact, under the traditional model with an equi-correlated covariance structure, the design ABB/BAA is known to be the universally optimal design (Laska, Meisner and Kushner 1983, Kershner 1986).

The Case of $p = 4$

Similarly, we study the optimal four-period design. There are 16 different treatment sequences in a two-treatment four-period design. Under the *Lemma 3.3.2.1*, we need to consider eight of them: $ABBB$, $ABBA$, $ABAB$, $ABAA$, $AABB$, $AABA$, $AAAB$, and $AAAA$. Under the same notation as before, we assume N_k patients receive k^{th} treatment sequence, $k = 1, \ldots, 8$ and $\sum_{k=1}^{8} N_k = N/2$.

And also, the direct and carryover treatment contrast columns of the design matrices for each treatment sequence are given below:

$$\mathbf{X}_1^\Gamma = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ -1 & -1 \\ -1 & -1 \end{pmatrix}; \quad \mathbf{X}_2^\Gamma = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ -1 & -1 \\ 1 & -1 \end{pmatrix}; \quad \mathbf{X}_3^\Gamma = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 1 & -1 \\ -1 & 1 \end{pmatrix};$$

$$\mathbf{X}_4^\Gamma = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 1 & -1 \\ 1 & 1 \end{pmatrix}; \quad \mathbf{X}_5^\Gamma = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ -1 & 1 \\ -1 & -1 \end{pmatrix}; \quad \mathbf{X}_6^\Gamma = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ -1 & 1 \\ 1 & -1 \end{pmatrix};$$

$$\mathbf{X}_7^\Gamma = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ -1 & 1 \end{pmatrix}; \quad \mathbf{X}_8^\Gamma = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

To find the optimal design for estimating of the direct treatment effect τ, let $\mathbf{m}^\Gamma = (1, 0)^T$ and $\boldsymbol{\beta}^\Gamma = (\tau, \gamma)^T$.

Then according to Equation 3.18, we have

$$\mathbf{A}^\Gamma = \sum_{d(1,k)=A} N_k \mathbf{X}_k^{\Gamma T} \mathbf{C}^{-1} \mathbf{X}_k^\Gamma = \begin{pmatrix} \frac{a_{11}}{3\rho+1} & \frac{a_{12}}{3\rho+1} \\ \frac{a_{21}}{3\rho+1} & \frac{a_{22}}{3\rho+1} \end{pmatrix} \tag{3.29}$$

where $a_{11} = 4[(2\rho+1)(N_1+N_4+N_6+N_7)+(3\rho+1)(N_2+N_3+N_5)+(1-\rho)N_8]$; $a_{12} = a_{21} = (\rho+1)N_1 - (3\rho+1)(N_2+3N_3-N_5) - (5\rho+1)(N_4+N_6) - (3\rho-1)N_7+3(1-\rho)N_8$; and $a_{22} = (8\rho+3)(N_1+N_2+N_3+N_4+N_5+N_6)+3(N_7+N_8)$.

Plug Equation 3.29 into Equation 3.19 and do a little algebra. Under the constraint 3.12, the minimum of $var(\hat{\tau})$ is $\sigma_\varepsilon^2/4N$, achieved at two situations:

1. **Result 1**: $N_1 = N_3 = N_4 = N_6 = N_7 = N_8 = 0$, and $N_2 = N_5 = N/4$. Hence, the optimal two-treatment four-period design under the model (3.20) is $ABBA/BAAB$ and $AABB/BBAA$ with an equal number of subjects per sequence;

2. **Result 2**: $N_1 = N_4 = N_6 = N_7 = N_8 = 0$, $N_2 = N/6$, $N_3 = N/24$, and $N_5 = 7N/24$. Hence, the optimal two-treatment four-period design under the model (3.20) is $ABBA/BAAB$, $ABAB/BABA$ and $AABB/BBAA$ with $N/6$, $N/24$, and $7N/24$ number of subjects per sequence, respectively.

Based on Equation 3.16, under both optimal designs mentioned in results 1 and 2, the weights of the BLUE of τ are

$$\begin{aligned}
\omega_{ABBA} &= -\omega_{BAAB} \\
&= \frac{1}{2}\mathbf{C}^{-1}\mathbf{X}_2^{\Gamma}\mathbf{A}^{\Gamma-1}\mathbf{m}^{\Gamma} \\
&= \left(\tfrac{1}{4N}, \; -\tfrac{1}{4N}, \; -\tfrac{1}{4N}, \; \tfrac{1}{4N} \right)^T \qquad (3.30) \\
\omega_{ABAB} &= -\omega_{BABA} \\
&= \frac{1}{2}\mathbf{C}^{-1}\mathbf{X}_3^{\Gamma}\mathbf{A}^{\Gamma-1}\mathbf{m}^{\Gamma} \\
&= \left(\tfrac{1}{4N}, \; -\tfrac{1}{4N}, \; \tfrac{1}{4N}, \; -\tfrac{1}{4N} \right)^T \qquad (3.31) \\
\omega_{AABB} &= -\omega_{BBAA} \\
&= \frac{1}{2}\mathbf{C}^{-1}\mathbf{X}_5^{\Gamma}\mathbf{A}^{\Gamma-1}\mathbf{m}^{\Gamma} \\
&= \left(\tfrac{1}{4N}, \; \tfrac{1}{4N}, \; -\tfrac{1}{4N}, \; -\tfrac{1}{4N} \right)^T \qquad (3.32)
\end{aligned}$$

Note that the weights of the BLUE of τ for both designs do not depend on ρ in this case.

Similarly, one can show that the above designs are optimal for estimation of γ as well. The minimized variance of the estimation of γ is

$$var(\hat{\gamma}) = (3\rho + 1)\sigma_{\varepsilon}^2/(N(8\rho + 3)).$$

Further, the weights of the BLUE of γ are

$$\begin{aligned}
\omega_{ABBA} &= -\omega_{BAAB} \\
&= \frac{1}{2}\mathbf{C}^{-1}\mathbf{X}_2^{\Gamma}\mathbf{A}^{\Gamma-1}\mathbf{m}^{\Gamma} \\
&= \left(\tfrac{\rho}{N(8\rho+3)},\ \tfrac{4\rho+1}{N(8\rho+3)},\ -\tfrac{2\rho+1}{N(8\rho+3)},\ -\tfrac{2\rho+1}{N(8\rho+3)} \right)^T \qquad (3.33) \\
\omega_{ABAB} &= -\omega_{BABA} \\
&= \frac{1}{2}\mathbf{C}^{-1}\mathbf{X}_3^{\Gamma}\mathbf{A}^{\Gamma-1}\mathbf{m}^{\Gamma} \\
&= \left(-\tfrac{\rho}{N(8\rho+3)},\ \tfrac{2\rho+1}{N(8\rho+3)},\ -\tfrac{4\rho+1}{N(8\rho+3)},\ \tfrac{2\rho+1}{N(8\rho+3)} \right)^T \qquad (3.34) \\
\omega_{AABB} &= -\omega_{BBAA} \\
&= \frac{1}{2}\mathbf{C}^{-1}\mathbf{X}_5^{\Gamma}\mathbf{A}^{\Gamma-1}\mathbf{m}^{\Gamma} \\
&= \left(-\tfrac{\rho}{N(8\rho+3)},\ \tfrac{2\rho+1}{N(8\rho+3)},\ \tfrac{2\rho+1}{N(8\rho+3)},\ -\tfrac{4\rho+1}{N(8\rho+3)} \right)^T \qquad (3.35)
\end{aligned}$$

When $\rho \to 1$, the weights for $ABBA$ are $(1/11N, 5/11N, -3/11N, -3/11N)^T$; the weights for $ABAB$ are $(-1/11N, 3/11N, -5/11N, 3/11N)^T$; the weights for $AABB$ are $(-1/11N, 3/11N, 3/11N, -5/11N)^T$; and the variance of the estimation of γ is $4\sigma_\varepsilon^2/11N$.

When $\rho = 0$, the weights for $ABBA$ are $(0, 1/3N, -1/3N, -1/3N)^T$; the weights for $ABAB$ are $(0, 1/3N, -1/3N, 1/3N)^T$; the weights for $AABB$ are $(0, 1/3N, 1/3N, -1/3N)^T$; and the variance of the estimation of γ is $\sigma_\varepsilon^2/3N$.

Note that both designs mentioned in Results 1 and 2 are optimal in terms of minimizing the variance of the estimation. However, the design $ABBA/BAAB$ and $AABB/BBAA$, also recommended by other researchers (Cheng and Wu 1980, Laska and Meisner 1985 and Carriere 1994), is more popular in practice, because it utilizes less treatment sequences and requires the total number of patients to be a multiplier of 4 instead of 24 as in the other optimal design.

Using the same approach, we can explore the optimal designs under the self and mixed carryover effects model as well. Due to the difficulty of the complex non-linear optimization problem, we use Maple software to handle it. The optimal designs are summarized in Section 3.4.2.

3.4 Optimal Designs Utilizing Baseline Measurements

Baseline measurements are taken at the outset of an experiment, before implementing different treatments. Baseline measurements are commonly used in trials of chronic conditions where clinicians want to see whether a treatment can reduce pre-existing levels of pain, anxiety, hypertension, and so on. In some situations baseline measurements may not be required in a study, however, they can be useful to improve the efficiency of the study design. Grizzle (1965, 1968) observed that when there were unequal carryover effects in the two-period crossover design AB/BA, treatment effects were not estimable using both first- and second-period data. However, Wallenstein (1979) found that with baseline observations in each period, estimators and tests for treatment effects were obtainable using all of the data, which was therefore more efficient because no information was ignored. This phenomenon has motivated many researchers to study the effect of adding baseline observations to a study (Laska and Meisner 1985, Carriere 1989). It is interesting to see whether baseline observations improve the design to some extent. In this section, we study the influence of baseline measurements on constructing the optimal two-treatment p-period ($p = 2, 3, 4$) repeated measurement designs.

There are different ways to obtain the baseline measurements and model the data. For example, Fleiss, Wallenstein and Rosenfeld (1985) studied the 2-period 2-treatment crossover design with baseline measurements measured at the start of both periods and they modeled the changes from the baseline in each period. Researchers may obtain only one baseline measurement in the beginning of the study and include it as an explanatory variable in the model. In this thesis, we discuss the effect of baseline measurements under the traditional model and under the self and mixed carryover effects model, assuming that baseline measurements are obtained in each period and the carryover effects are the same as those on the post treatment in the same period. For example, under the self and mixed carryover effects model described

in Section 3.2.2, the columns, regarding the direct treatment contrast (τ), mixed and self carryover effects (γ and φ), of the design matrix for treatment sequence $bAbBbB$ is

$$\mathbf{X}^{\Gamma}_{bAbBbB} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & -1 \\ -1 & 0 & -1 \end{pmatrix}$$

where b represents the baseline, $\tau = (\tau_A - \tau_B)/2$, $\gamma = (\gamma_A - \gamma_B)/2$ and $\varphi = (\varphi_A - \varphi_B)/2$.

3.4.1 Two-Treatment p-Period Optimal Designs Based on the Traditional Model

The traditional model defined in Section 3.2.1 satisfies the conditions of the *Lemma 3.3.2.1* in Section 3.3.2. Therefore, an optimal design will assign an equal number of subjects to a treatment sequence and its dual.

Suppose N subjects are enrolled in the study. Let $\rho = \sigma^2_\xi/(\sigma^2_\varepsilon + \sigma^2_\xi)$. If $\sigma^2_\xi = 0$, then $\rho = 0$, the model becomes a fixed-effect model with no subject effect. If $\sigma^2_\xi \to \infty$ or $\sigma^2_\xi \gg \sigma^2_\varepsilon$, then $\rho \to 1$ and the model is equivalent to the model with fixed subject effects (Afsarinejad and Hedayat, 2002).

In addition, the relative efficiency between design 1 and design 2 is defined as $var(\theta_{Design2})/var(\theta_{Design1})$.

Maple codes for solving the optimal design problems are available upon request. The program is user friendly in that one can specify the total number of subjects, N, and the within-subject correlation, ρ, to obtain the optimal two-treatment design for any period ($p = 2, 3, 4$) under either of the two models described in Section 3.2, with or without baseline measurements. The weights of the observations for estimating the treatment effect contrast, $\tau = (\tau_A - \tau_B)/2$, are also provided. What follows is a summary of some results on practically useful optimal designs.

The Case of $p = 2$

In a two-treatment two-period ($p = 2$) repeated measurement design, 4 different treatment sequences are available (AB, AA, BA and BB). Let m be the number of patients receiving treatment sequence AB, $0 \leq m \leq N/2$. Then $(N/2 - m)$ patients receive AA treatment sequence.

Table 3.1 shows that for estimation of τ, the design AA/BB and AB/BA with an equal number of subjects per sequence is optimal, no matter whether the baseline measurements are included or not. The efficiency of the design with the baseline measurements is 1 to 2.5 times the efficiency of the design without, as the value of ρ increases from 0 to 1, while the optimal design for estimation of γ depends on the value of ρ, as shown in Section 3.3.3. Baseline measurements do improve the efficacy of the design by 78% to 100%.

The Case of $p = 3$

In a two-treatment three-period ($p = 3$) repeated measurement design, 8 different treatment sequences are available. Let N_1 be the number of patients receiving treatment sequence AAA, N_2 be the number of patients receiving treatment sequence AAB, N_3 be the number of patients receiving treatment sequence ABA, and N_4 be the number of patients receiving treatment sequence ABB. Under the constraint (3.12), we have $\Sigma_{i=1}^{4} N_i = N/2$.

Table 3.2 shows that for estimation of treatment effect contrast, the design consisting of the sequences ABB and its dual is optimal. Baseline measurements improve the efficiency only slightly. The relative efficiency between the design with the baseline measurements vs. the design without is equal to 1 to 1.0625 when the value of ρ increases from 0 to 1. Therefore, use of the baseline measurement does not appear to be very helpful in improving the design efficiency.

The Case of $p = 4$

In a two-treatment four-period ($p = 4$) repeated measurement design, 16 different treatment sequences are available. Table 3.3 shows that for estimation of treatment difference, two designs both produce the minimum variance of BLUE of τ: the design $ABBA/BAAB$ and $AABB/BBAA$ with an equal number of subjects per sequence; and the design $ABBA/BAAB$,

$ABAB/BABA$ and $AABB/BBAA$ with $N/6$, $N/24$ and $7N/24$ subjects per sequence, respectively. In both situations, adding the baseline measurements does not reduce the variance of the parameter of interest. Actually, according to the weights provided in the table, one can clearly see that the BLUE of τ does not use the baseline data at all. Hence, in a two-treatment four-period repeated measurement design, under the traditional model, the baseline measurements do not improve the efficiency of the design. Therefore, it is not recommended to use baseline measurements in this case.

3.4.2 Two-Treatment p-Period Optimal Designs Based on the Self and Mixed Carryover Effects Model

Under the same notations defined in Section 3.4.1, we consider optimal designs under the self and mixed carryover effects model defined in Section 3.2.2.

The Case of $p = 2$

In a two-treatment two-period ($p = 2$) repeated measurement design, under the self and mixed carryover effects model, Table 3.4 presents the summary of the optimal design results. All designs for estimating the treatment effect provide the same variance of the treatment contrast. Without the baseline measurements, the estimation of τ is used only for the data in the first period, no matter how many patients enrolled and what is the value of ρ. Failure to use the second period data is a major drawback. However, as shown in Table 3.4, after adding the baseline measurements, we can estimate the treatment effect by using all data available. In addition, the design with the baseline measurements increases the efficiency at least 50%. When within subject correlation, ρ, is larger, more benefit is achieved from using baseline measurements. For example, when $\rho = 0.5$, the design efficiency will increase by 130%; when $\rho = 0.8$, the design efficiency will increase by 428%.

Consider estimating the mixed carryover effect, γ, the optimal design is AB/BA with an equal number of patients per sequence. The efficiency of the design with the baseline measurements is at least 3 times of that of the

design without baseline measurements as ρ increases from 0 to 1.

The optimal design for estimation of the self carryover effect is AA/BB. The efficiency of the design with the baseline measurements is 2 to 3 times of that of the design without baseline measurements as ρ decreases from 1 to 0.

The Case of $p = 3$

In a two-treatment three-period ($p = 3$) repeated measurement design, under the self and mixed carryover effects model, the optimal design for estimation of the the treatment effect, τ, is ABA/BAB. However, there are no self carryover effects with this design since a treatment is never immediately followed by itself. Therefore, in this section, we will draw our attention to those designs that are able to estimate all parameters in the self and mixed carryover effects model, and provide good estimates of the treatment differences as well. Hence, not surprisingly, there is a price to be paid for allowing different types of carryover effects in the model.

Three designs are considered: 1) design AAB/BBA and ABA/BAB with an equal number of subjects per sequence; 2) design ABA/BAB and ABB/BAA with an equal number of subjects per sequence; and 3) design AAB/BBA, ABA/BAB and $ABB\ /BAA$ with an equal number of subjects per sequence. These three designs are also recommended by Hedayat and Stufken (2003).

Table 3.5 summarizes the relative efficiencies (RE) of the selected designs compared with the design ABA/BAB, and the weights of the observations in the estimation of the treatment effect contrast, τ, without the baseline measurements. Similar results for these designs with baseline measurements are displayed in Table 3.6. We can see that without baseline measurements, the design AAB/BBA and ABA/BAB is as at least 80% as efficient as the optimal design; so is the design AAB/BBA and ABA/BAB; and the design AAB/BBA, ABA/BAB and ABB/BAA, which utilizes more treatment sequences, is as at least 93.3% as efficient as the optimal design. With the baseline measurements, all three designs are almost as efficient as the optimal

design, with relative efficiency between 96.4% to 98.8%.

Table 3.7 presents the variance of the estimator of the treatment difference, τ, under the designs with or without baseline measurements, and the relative efficiencies of the designs with the baseline measurement and without. It shows that under all three designs, the baseline measurements improve the efficiency 2 to 3 times when the value of ρ increases from 0 to 1. Therefore, baseline measurements should be recommended in this case.

The Case of $p = 4$

In a two-treatment four-period ($p = 4$) repeated measurement design, the procedure to theoretically identify the optimal design for estimation of treatment difference becomes very complex. However, we were able to numerically prove that the lower bound of the $\mathrm{Var}(\tau)$ is $(3\rho + 1)\sigma_\varepsilon^2/N(2\rho + 1)$. Due to the multiple solutions of the optimal problem, there are several designs which can achieve the lower bound, for example, the design ABBA/BAAB(N/4), ABAA/BABB(N/8) and AABA/BBAB(N/8), and the design ABBA/BAAB(3N/8) and AAAA/BBBB(N/8), and the design ABBA /BAAB (N/4) and AABA/BBAB(N/4).

In this section, we studied the baseline measurements influences under the design ABBA/BAAB, AABA/BBAB. The reasons that we choose this design are: 1) it is an optimal design; 2) it includes an equal number of subjects per sequence; and 3) there is no simpler design as efficient as this design. Actually, the designs with only one treatment sequence and its dual are, at most, 89% as efficient as the design ABBA/BAAB, AABA/BBAB, when $\rho \to 1$.

Table 3.8 shows that the relative efficiency between the design with the baseline measurements vs. the design without equals 2.5 to 3.167 when the value of ρ changes from 0 to 1. Hence, it is worthwhile to add the baseline measurements in order to improve the design efficiency, under the assumption that patients will comply with the study.

3.5 Conclusion

In this chapter, we applied the Lagrange multiplier method to solve the optimal design problems under the traditional and the self and mixed carryover effects models. In addition, we studied the influence of baseline measurements on constructing optimal designs for two-treatment p-period ($p = 2, 3, 4$) repeated measures data. Our findings are that:

1. For two-treatment repeated measurement designs, the *Lemma 3.3.2.1* proves that optimal designs allocate an equal number of subjects to a treatment sequence and its dual.

2. Optimal designs are strongly model dependent. However, when estimating the treatment difference, having or not having baseline measurements will not change the optimal design results.

   Under the traditional model, the results are consistent with those obtained by other researchers (Cheng and Wu 1980, Laska, Meisner and Kushner 1983, Laska and Meisner 1985, Kershner 1986, Carriere 1994).

   - $p = 2$: The design AA/BB and AB/BA with an equal number of subjects per sequence is optimal for estimating the treatment effect contrast. For estimation of carryover effect, the optimal design depends on the value of ρ: when $\rho = 0$, the optimal design is AA/BB and AB/BA with an equal number of subjects per sequence; when $\rho \to 1$ the optimal design is AA/BB with an equal number of subjects per sequence.
   - $p = 3$: The design ABB/BAA with an equal number of subjects per sequence is optimal for estimations of both direct and carryover treatment effects.
   - $p = 4$: The design $ABBA/BAAB$ and $AABB/BBAA$ with an equal number of subjects per sequence is optimal. The design $ABBA/BAAB$, $ABAB/BABA$, and $AABB/BBAA$ with 1/6, 1/24 and 7/24 of the total subjects per sequence, respectively, also

gives the optimal minimum variance. The former design is more popular in practice because it utilizes less treatment sequences, uses an equal number of subjects per treatment, and requires the total number of subjects is a multiplier of 4 instead of 24, as in the latter design.

Under the self and mixed carryover effects model,

- $p = 2$: Without baseline measurements, the estimation of the treatment difference uses only the data in the first period, hence is inefficient. For estimation of the mixed carryover effect, the design AB/BA is optimal. For estimation of the self carryover effect, the design AA/BB is optimal. Those findings are consistent with Afsarinejad and Hedayat (2002), where they studied two-period optimal designs under the self and mixed carryover effects model with fixed subject effects.
- $p = 3$: The optimal design for estimation of the treatment difference is ABA/BAB, however, there are no self carryover effects with this design. Other almost equally efficient designs are recommended, including design AAB/BBA and ABA/BAB, design ABB/BAA and ABA/BAB, and design AAB/BBA, ABA/BAB and ABB/BAA, especially when within-subject correlation, ρ, is small. Therefore, there is a price to pay for allowing different types of carryover effects in the model.
- $p = 4$: The simplest optimal design for estimating the treatment effect contrast in this case is the design $ABBA/BAAB$ and $AABA/BBAB$ with an equal number of subjects per sequence.

3. The influence of baseline measurements should be discussed in each specific situation. Under the assumptions as given in Section 3.4.1, we conclude:

   Under the traditional model,

- $p = 2$: The efficiency of the design with baseline measurements is 1 to 2.5 times that of the design without baseline measurements. Therefore, it is recommended to use the baseline measurements.
- $p = 3$: The baseline measurements improve the design efficiency only slightly. The relative efficiency between the design with baseline measurements vs. the design without is equal to 1 to 1.0625 when ρ increases from 0 to 1. Therefore, the use of the baseline measurements does not appear to be helpful in improving the design efficiency.
- $p = 4$: The baseline measurements do not improve the design efficiency at all. In addition, by using the baseline measurement in each period, we extend a four-period design to an eight-period design. Due to the difficulty of having all subjects comply until the termination of the experiment and the degree of difficulty increases as the number of periods gets larger, long period designs should be avoided in practice. Therefore, it is not recommended to use baseline measurements in this case.

Under the self and mixed carryover effects model,

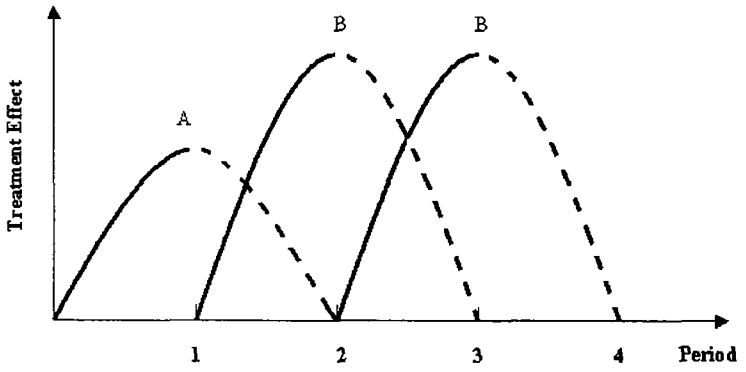
- $p = 2$: The baseline measurements improve the efficiency of the design measurements at least 1.5 times. Therefore, it is strongly recommended to use the baseline measurements.
- $p = 3$: The baseline measurements improve the design efficiency significantly: the improvement in relative efficiency between the design with and without baseline measurements ranges from 2 to 3 times as ρ increases from 0 to 1. Therefore, it is recommended to use the baseline measurements.
- $p = 4$: The efficiency of the design with baseline measurements is 2.5 to 3 times of the efficiency of the design without. Therefore, it is worthwhile to add the baseline measurements under the more complex models, such as those including self and mixed carryover

effects.

4. Carriere (1994b) found that there is a dramatic reduction in variability for estimating the direct treatment effect contrast in three-period designs compared to two-period designs under various types of models. Our study confirm Carriere's result and extend it to four-period designs under more general models utilizing baseline measurements. Table 3.9 summarizes the variances of the estimators for the treatment difference with $\rho = 0, 0.5$ or $\rightarrow 1$ under two types of models discussed in Section 3.2 having or not having baseline measurements. One can see that

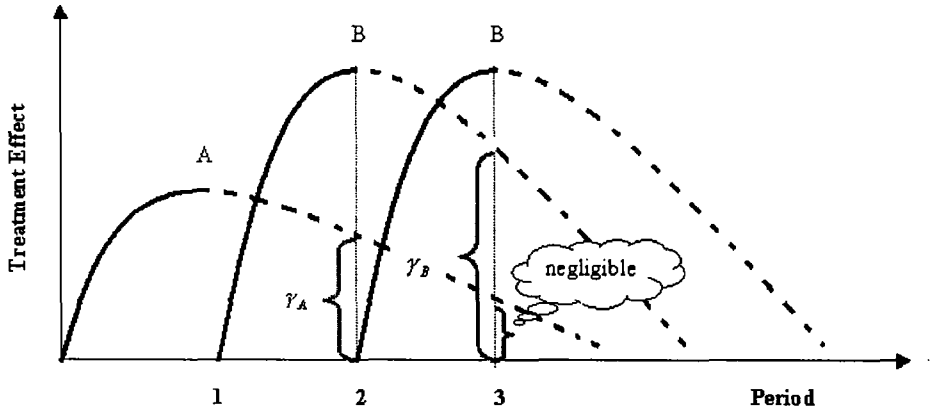
   - Under the traditional model, no matter using baseline measurements or not, the three-period designs achieve at least a 33% reduction in variance compared to the two-period designs, and the four-period designs achieve at least a 25% reduction in variance compared to the three-period designs.
   - Similar patterns are found under the self and mixed carryover effects model. For the designs utilizing baseline measurements in each period, there is at least a 25% reduction in variance in three-period designs compared to two-period designs, and at least a 20% reduction in four-period designs compared to three-period designs. Without baseline measurements, when within subject correlation is 0.5 or more, compared to the two-period designs, the three-period designs achieve a 27% or more reduction in variability. In addition, a 14% or more reduction in variability is achieved when add one more period after the third period.

Figure 3.1: Experiment without Carryover Effects



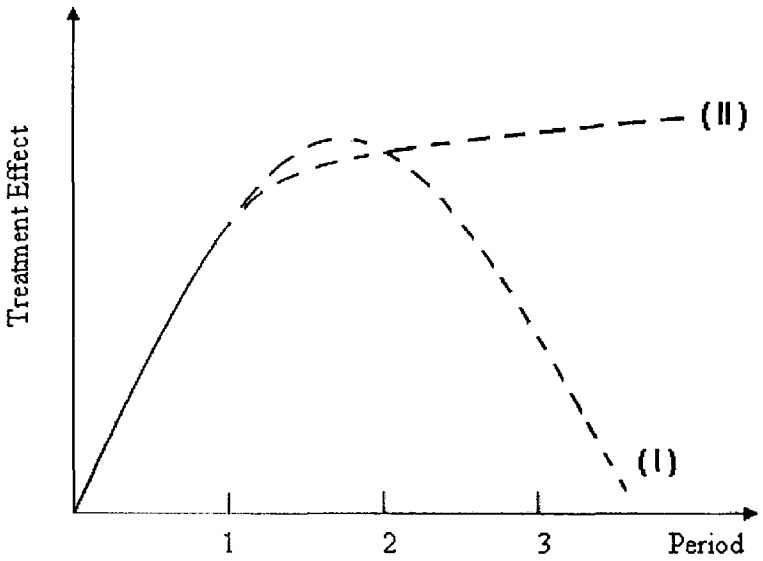
Note: In this situation, carryover effects are washed out completely by the next treatment period. Therefore, no carryover effect needs to be considered in the model

Figure 3.2: Experiment with Carryover Effects



Note: In this situation, carryover effects are not washed out completely, therefore, not negligible, especially those between two consecutive treatment periods (γ_A, and γ_B in the Figure).

Figure 3.3: Some Patterns of Treatment Effects



Note: In (I), the treatment effects are measured before it reaches its peak, while in (II), they have lasting and curative effects.

Table 3.1: Two-Treatment Two-Period Optimal Designs Based on Traditional Model

θ	Baseline	$m^\star$	Var(θ)	Weights	
				AB	AA
τ	No	$\frac{N}{4}$	$\frac{(\rho+1)\sigma_\varepsilon^2}{N(2-\rho^2)}$	$\begin{pmatrix} \frac{\rho^2-\rho-1}{N(\rho^2-2)} \\ \frac{1}{N(\rho^2-2)} \end{pmatrix}$	$\begin{pmatrix} \frac{\rho^2+\rho-1}{N(\rho^2-2)} \\ \frac{-1}{N(\rho^2-2)} \end{pmatrix}$
	Yes	$\frac{N}{4}$	$\frac{(\rho+1)(3\rho+1)\sigma_\varepsilon^2}{2N(1+\rho^2+3\rho)}$	$\begin{pmatrix} -\frac{\rho^2}{N(1+\rho^2+3\rho)} \\ \frac{\rho^2+4\rho+1}{2N(1+\rho^2+3\rho)} \\ \frac{\rho^2+\rho}{2N(1+\rho^2+3\rho)} \\ -\frac{(2\rho+1)(\rho+1)}{2N(1+\rho^2+3\rho)} \end{pmatrix}$	$\begin{pmatrix} -\frac{\rho(2\rho+1)}{N(1+\rho^2+3\rho)} \\ -\frac{\rho^2-2\rho-1}{2N(1+\rho^2+3\rho)} \\ -\frac{\rho(\rho+1)}{2N(1+\rho^2+3\rho)} \\ \frac{(2\rho+1)(\rho+1)}{2N(1+\rho^2+3\rho)} \end{pmatrix}$
γ	No	$\frac{N(1-\rho)}{4}$	$\frac{(\rho+1)\sigma_\varepsilon^2}{N}$	$\begin{pmatrix} -\frac{\rho}{N} \\ \frac{1}{N} \end{pmatrix}$	$\begin{pmatrix} -\frac{\rho}{N} \\ \frac{1}{N} \end{pmatrix}$
	Yes	$\frac{N(1-\rho)}{4(1+\rho)}$	$\frac{(3\rho+1)\sigma_\varepsilon^2}{2N(\rho+1)}$	$\begin{pmatrix} -\frac{\rho}{N(\rho+1)} \\ -\frac{\rho}{N(\rho+1)} \\ \frac{1}{2N} \\ \frac{1}{2N} \end{pmatrix}$	$\begin{pmatrix} -\frac{\rho}{N(\rho+1)} \\ -\frac{\rho}{N(\rho+1)} \\ \frac{1}{2N} \\ \frac{1}{2N} \end{pmatrix}$

$^\star m$: The number of patients receive treatment sequence AB, $0 \leq m \leq N/2$. Then $(N/2 - m)$ patients receive AA treatment sequence. $\rho = \sigma_\xi^2/(\sigma_\varepsilon^2 + \sigma_\xi^2)$ is the intraclass correlation coefficient.

Table 3.2: Two-Treatment Three-Period Optimal Designs for Estimating Treatment Effect Based on Traditional Model

Baseline	Optimal Design	$\text{Var}(\tau)$	Weights	
			ABB	BAA
No	ABB/BAA $(N_1 = N_2 = N_3 = 0,$ $N_4 = N/2)$	$\frac{(2\rho+1)\sigma_\epsilon^2}{N(5\rho+3)}$	$\begin{pmatrix} \frac{3\rho+1}{N(5\rho+3)} \\ -\frac{\rho+1}{N(5\rho+3)} \\ -\frac{\rho+1}{N(5\rho+3)} \end{pmatrix}$	$\begin{pmatrix} -\frac{3\rho+1}{N(5\rho+3)} \\ \frac{\rho+1}{N(5\rho+3)} \\ \frac{\rho+1}{N(5\rho+3)} \end{pmatrix}$
Yes	ABB/BAA $(N_1 = N_2 = N_3 = 0,$ $N_4 = N/2)$	$\frac{(5\rho+1)\sigma_\epsilon^2}{N(14\rho+3)}$	$\begin{pmatrix} \frac{\rho}{N(14\rho+3)} \\ \frac{6\rho+1}{N(14\rho+3)} \\ \frac{\rho}{N(14\rho+3)} \\ -\frac{4\rho+1}{N(14\rho+3)} \\ \frac{\rho}{N(14\rho+3)} \\ -\frac{4\rho+1}{N(14\rho+3)} \end{pmatrix}$	$\begin{pmatrix} -\frac{\rho}{N(14\rho+3)} \\ -\frac{6\rho+1}{N(14\rho+3)} \\ -\frac{\rho}{N(14\rho+3)} \\ \frac{4\rho+1}{N(14\rho+3)} \\ -\frac{\rho}{N(14\rho+3)} \\ \frac{4\rho+1}{N(14\rho+3)} \end{pmatrix}$

Table 3.3: Two-Treatment Four-Period Optimal Designs for Estimating Treatment Effect Based on Traditional Model

Baseline	Optimal Design	$\text{Var}(\tau)$	Weights		
			ABBA	ABAB	AABB
No	$ABBA/BAAB$ $AABB/BBAA$ $(N_2 = N_5 = N/4)$	$\frac{\sigma_\epsilon^2}{4N}$	$\begin{pmatrix} \frac{1}{4N} \\ -\frac{1}{4N} \\ -\frac{1}{4N} \\ \frac{1}{4N} \end{pmatrix}$		$\begin{pmatrix} \frac{1}{4N} \\ \frac{1}{4N} \\ -\frac{1}{4N} \\ -\frac{1}{4N} \end{pmatrix}$
	$ABBA/BAAB$ $ABAB/BABA$ $AABB/BBAA$ $(N_2 = N/6, N_3 = N/24,$ and$N_5 = 7N/24)$	$\frac{\sigma_\epsilon^2}{4N}$	$\begin{pmatrix} \frac{1}{4N} \\ -\frac{1}{4N} \\ -\frac{1}{4N} \\ \frac{1}{4N} \end{pmatrix}$	$\begin{pmatrix} \frac{1}{4N} \\ -\frac{1}{4N} \\ \frac{1}{4N} \\ -\frac{1}{4N} \end{pmatrix}$	$\begin{pmatrix} \frac{1}{4N} \\ \frac{1}{4N} \\ -\frac{1}{4N} \\ -\frac{1}{4N} \end{pmatrix}$
Yes	$ABBA/BAAB$ $AABB/BBAA$ $(N_2 = N_5 = N/4)$	$\frac{\sigma_\epsilon^2}{4N}$	$\begin{pmatrix} 0 \\ \frac{1}{4N} \\ 0 \\ -\frac{1}{4N} \\ 0 \\ -\frac{1}{4N} \\ 0 \\ \frac{1}{4N} \end{pmatrix}$		$\begin{pmatrix} 0 \\ \frac{1}{4N} \\ 0 \\ \frac{1}{4N} \\ 0 \\ -\frac{1}{4N} \\ 0 \\ -\frac{1}{4N} \end{pmatrix}$
	$ABBA/BAAB$ $ABAB/BABA$ $AABB/BBAA$ $(N_2 = N/6, N_3 = N/24,$ and$N_5 = 7N/24)$	$\frac{\sigma_\epsilon^2}{4N}$	$\begin{pmatrix} 0 \\ \frac{1}{4N} \\ 0 \\ -\frac{1}{4N} \\ 0 \\ -\frac{1}{4N} \\ 0 \\ \frac{1}{4N} \end{pmatrix}$	$\begin{pmatrix} 0 \\ \frac{1}{4N} \\ 0 \\ -\frac{1}{4N} \\ 0 \\ \frac{1}{4N} \\ 0 \\ -\frac{1}{4N} \end{pmatrix}$	$\begin{pmatrix} 0 \\ \frac{1}{4N} \\ 0 \\ \frac{1}{4N} \\ 0 \\ -\frac{1}{4N} \\ 0 \\ -\frac{1}{4N} \end{pmatrix}$

Table 3.4: Two-Treatment Two-Period Optimal Designs Based on Self and Mixed Carryover Effects Model

θ	Baseline	m	Var(θ)	Weights	
				AB	AA
τ	No	m	$\frac{1}{N(1-\rho)}\sigma_\varepsilon^2$	$\begin{pmatrix} \frac{1}{N} \\ 0 \end{pmatrix}$	$\begin{pmatrix} \frac{1}{N} \\ 0 \end{pmatrix}$
	Yes	m	$\frac{2(\rho+1)}{N(\rho+3)}\sigma_\varepsilon^2$	$\begin{pmatrix} -\frac{2\rho}{N(\rho+3)} \\ \frac{2}{N(\rho+3)} \\ \frac{\rho+1}{N(\rho+3)} \\ -\frac{\rho+1}{N(\rho+3)} \end{pmatrix}$	$\begin{pmatrix} -\frac{2\rho}{N(\rho+3)} \\ \frac{2}{N(\rho+3)} \\ -\frac{\rho+1}{N(\rho+3)} \\ \frac{\rho+1}{N(\rho+3)} \end{pmatrix}$
γ	No	$\frac{N}{2}$	$\frac{2(\rho+1)}{N(1-\rho)}\sigma_\varepsilon^2$	$\begin{pmatrix} \frac{1}{N} \\ \frac{1}{N} \end{pmatrix}$	
	Yes	$\frac{N}{2}$	$\frac{2(3\rho+1)}{N(\rho+3)}\sigma_\varepsilon^2$	$\begin{pmatrix} -\frac{4\rho}{N(\rho+3)} \\ -\frac{\rho-1}{N(\rho+3)} \\ \frac{2(\rho+1)}{N(\rho+3)} \\ -\frac{\rho-1}{N(\rho+3)} \end{pmatrix}$	
φ	No	0	$\frac{2}{N}\sigma_\varepsilon^2$		$\begin{pmatrix} -\frac{1}{N} \\ \frac{1}{N} \end{pmatrix}$
	Yes	0	$\frac{2(\rho+1)}{N(\rho+3)}\sigma_\varepsilon^2$		$\begin{pmatrix} -\frac{2\rho}{N(\rho+3)} \\ -\frac{\rho+1}{N(\rho+3)} \\ \frac{2}{N(\rho+3)} \\ \frac{\rho+1}{N(\rho+3)} \end{pmatrix}$

Note: See notes for Table 3.1.

Table 3.5: Efficiencies and Weights for Selected Two-Treatment Three-Period Designs under the Self and Mixed Carryover Effects Model without Baseline Measurements

Design	Relative Efficiency (ρ changes from 1 to 0)	Weights		
		AAB	ABA	ABB
AAB/BBA ABA/BAB	$80\% \sim 100\%$	$\begin{pmatrix} -\frac{2\rho^2-5\rho-3}{N(5\rho+3)} \\ 0 \\ -\frac{2\rho(2\rho+1)}{N(5\rho+3)} \end{pmatrix}$	$\begin{pmatrix} \frac{(2\rho+3)(\rho+1)}{N(5\rho+3)} \\ -\frac{2\rho}{N(5\rho+3)} \\ -\frac{4\rho(\rho+1)}{N(5\rho+3)} \end{pmatrix}$	
ABA/BAB ABB/BAA	$80\% \sim 100\%$		$\begin{pmatrix} \frac{(2\rho+3)(\rho+1)}{N(5\rho+3)} \\ -\frac{2\rho}{N(5\rho+3)} \\ -\frac{4\rho(\rho+1)}{N(5\rho+3)} \end{pmatrix}$	$\begin{pmatrix} -\frac{2\rho^2-5\rho-3}{N(5\rho+3)} \\ -\frac{2\rho(2\rho+1)}{N(5\rho+3)} \\ 0 \end{pmatrix}$
AAB/BBA ABA/BAB ABB/BAA	$93.3\% \sim 100\%$	$\begin{pmatrix} \frac{3(2\rho^2+5\rho+2)}{N(7\rho^2+15\rho+6)} \\ -\frac{6\rho(2\rho+1)}{N(7\rho^2+15\rho+6)} \\ -\frac{3\rho(2\rho+1)}{N(7\rho^2+15\rho+6)} \end{pmatrix}$	$\begin{pmatrix} \frac{3(3\rho^2+5\rho+2)}{N(7\rho^2+15\rho+6)} \\ -\frac{3\rho(\rho+1)}{N(7\rho^2+15\rho+6)} \\ -\frac{3\rho(5\rho+3)}{N(7\rho^2+15\rho+6)} \end{pmatrix}$	$\begin{pmatrix} \frac{3(2\rho^2+5\rho+2)}{N(7\rho^2+15\rho+6)} \\ -\frac{3\rho(2\rho+1)}{N(7\rho^2+15\rho+6)} \\ -\frac{6\rho(2\rho+1)}{N(7\rho^2+15\rho+6)} \end{pmatrix}$

Table 3.6: Efficiencies and Weights for Selected Two-Treatment Three-Period Designs under the Self and Mixed Carryover Effects Model with Baseline Measurements

Design	Relative Efficiency (ρ changes from 1 to 0)	Weights		
		AAB	ABA	ABB
AAB/BBA ABA/BAB	$96.4\% \sim 100\%$	$\begin{pmatrix} -\frac{3\rho(5\rho+1)}{N(59\rho^2+41\rho+6)} \\ \frac{20\rho^2+19\rho+3}{N(59\rho^2+41\rho+6)} \\ -\frac{35\rho^2+22\rho+3}{2N(59\rho^2+41\rho+6)} \\ \frac{35\rho^2+22\rho+3}{2N(59\rho^2+41\rho+6)} \\ \frac{3(5\rho^2+6\rho+1)}{2N(59\rho^2+41\rho+6)} \\ -\frac{55\rho^2+26\rho+3}{2N(59\rho^2+41\rho+6)} \end{pmatrix}$	$\begin{pmatrix} -\frac{\rho(7\rho+3)}{N(59\rho^2+41\rho+6)} \\ \frac{(7\rho+3)(4\rho+1)}{N(59\rho^2+41\rho+6)} \\ \frac{31\rho^2+18\rho+3}{2N(59\rho^2+41\rho+6)} \\ -\frac{39\rho^2+26\rho+3}{2N(59\rho^2+41\rho+6)} \\ -\frac{59\rho^2+30\rho+3}{2N(59\rho^2+41\rho+6)} \\ \frac{11\rho^2+14\rho+3}{2N(59\rho^2+41\rho+6)} \end{pmatrix}$	
ABA/BAB ABB/BAA	$96.4\% \sim 100\%$		$\begin{pmatrix} -\frac{\rho(7\rho+3)}{N(59\rho^2+41\rho+6)} \\ \frac{(7\rho+3)(4\rho+1)}{N(59\rho^2+41\rho+6)} \\ \frac{31\rho^2+18\rho+3}{2N(59\rho^2+41\rho+6)} \\ -\frac{39\rho^2+26\rho+3}{2N(59\rho^2+41\rho+6)} \\ -\frac{59\rho^2+30\rho+3}{2N(59\rho^2+41\rho+6)} \\ \frac{11\rho^2+14\rho+3}{2N(59\rho^2+41\rho+6)} \end{pmatrix}$	$\begin{pmatrix} -\frac{3\rho(5\rho+1)}{N(59\rho^2+41\rho+6)} \\ \frac{20\rho^2+19\rho+3}{N(59\rho^2+41\rho+6)} \\ \frac{3(5\rho^2+6\rho+1)}{2N(59\rho^2+41\rho+6)} \\ -\frac{55\rho^2+26\rho+3}{2N(59\rho^2+41\rho+6)} \\ \frac{35\rho^2+22\rho+3}{2N(59\rho^2+41\rho+6)} \\ -\frac{35\rho^2+22\rho+3}{2N(59\rho^2+41\rho+6)} \end{pmatrix}$
AAB/BBA ABA/BAB ABB/BAA	$98.8\% \sim 100\%$	$\begin{pmatrix} -\frac{3(5\rho+1)}{N(106\rho^2+51\rho+6)} \\ \frac{3(15\rho^2+8\rho+1)}{N(106\rho^2+51\rho+6)} \\ -\frac{3(30\rho^2+11\rho+1)}{2N(106\rho^2+51\rho+6)} \\ \frac{3(10\rho^2+7\rho+1)}{2N(106\rho^2+51\rho+6)} \\ \frac{3(15\rho^2+8\rho+1)}{2N(106\rho^2+51\rho+6)} \\ -\frac{3(25\rho^2+10\rho+1)}{2N(106\rho^2+51\rho+6)} \end{pmatrix}$	$\begin{pmatrix} -\frac{3\rho(4\rho+1)}{N(106\rho^2+51\rho+6)} \\ \frac{3(4\rho+1)^2}{N(106\rho^2+51\rho+6)} \\ \frac{3(17\rho^2+8\rho+1)}{2N(106\rho^2+51\rho+6)} \\ -\frac{3(23\rho^2+10\rho+1)}{2N(106\rho^2+51\rho+6)} \\ -\frac{3(33\rho^2+12\rho+1)}{2N(106\rho^2+51\rho+6)} \\ \frac{3(7\rho^2+6\rho+1)}{2N(106\rho^2+51\rho+6)} \end{pmatrix}$	$\begin{pmatrix} -\frac{3(5\rho+1)}{N(106\rho^2+51\rho+6)} \\ \frac{3(15\rho^2+8\rho+1)}{N(106\rho^2+51\rho+6)} \\ \frac{3(15\rho^2+8\rho+1)}{2N(106\rho^2+51\rho+6)} \\ -\frac{3(25\rho^2+10\rho+1)}{2N(106\rho^2+51\rho+6)} \\ \frac{3(10\rho^2+7\rho+1)}{2N(106\rho^2+51\rho+6)} \\ -\frac{3(30\rho^2+11\rho+1)}{2N(106\rho^2+51\rho+6)} \end{pmatrix}$

Table 3.7: Two-Treatment Three-Period Design for Estimating Treatment Effect Based on Self and Mixed Carryover Effects Model

Design	Baseline	Var(τ)	Relative Efficiency ρ increases from 0 to 1
AAB/BBA ABA/BAB	No	$\frac{(2\rho+1)(2\rho+3)\sigma_\varepsilon^2}{N(5\rho+3)}$	$2 \sim 3.3125$
	Yes	$\frac{(5\rho+1)(7\rho+3)\sigma_\varepsilon^2}{N(59\rho^2+41\rho+6)}$	
ABA/BAB ABB/BAA	No	$\frac{(2\rho+1)(2\rho+3)\sigma_\varepsilon^2}{N(5\rho+3)}$	$2 \sim 3.3125$
	Yes	$\frac{(5\rho+1)(7\rho+3)\sigma_\varepsilon^2}{N(59\rho^2+41\rho+6)}$	
AAB/BBA ABA/BAB ABB/BAA	No	$\frac{3(2\rho+1)(3\rho+2)\sigma_\varepsilon^2}{N(7\rho^2+15\rho+6)}$	$2 \sim 2.9107$
	Yes	$\frac{3(5\rho+1)(4\rho+1)\sigma_\varepsilon^2}{N(106\rho^2+51\rho+6)}$	

Table 3.8: Two-Treatment Four-Period Design for Estimating Treatment Effect Based on Self and Mixed Carryover Effects Model

Design	Baseline	Var(τ)	Weights	
			$ABBA$	$AABA$
$ABBA/BAAB$ $AABA/BBAB$	No	$\frac{(3\rho+1)\sigma_\varepsilon^2}{N(2\rho+1)}$	$\begin{pmatrix} \frac{1}{N} \\ -\frac{\rho}{N(2\rho+1)} \\ -\frac{\rho}{N(2\rho+1)} \\ -\frac{\rho}{N(2\rho+1)} \end{pmatrix}$	$\begin{pmatrix} \frac{1}{N} \\ -\frac{\rho}{N(2\rho+1)} \\ -\frac{\rho}{N(2\rho+1)} \\ -\frac{\rho}{N(2\rho+1)} \end{pmatrix}$
	Yes	$\frac{2(7\rho+1)\sigma_\varepsilon^2}{N(33\rho+5)}$	$\begin{pmatrix} -\frac{2\rho}{N(33\rho+5)} \\ \frac{2(6\rho+1)}{N(33\rho+5)} \\ \frac{5\rho+1}{N(33\rho+5)} \\ -\frac{9\rho+1}{N(33\rho+5)} \\ \frac{5\rho+1}{N(33\rho+5)} \\ -\frac{9\rho+1}{N(33\rho+5)} \\ -\frac{9\rho+1}{N(33\rho+5)} \\ \frac{5\rho+1}{N(33\rho+5)} \end{pmatrix}$	$\begin{pmatrix} -\frac{2\rho}{N(33\rho+5)} \\ \frac{2(6\rho+1)}{N(33\rho+5)} \\ \frac{5\rho+1}{N(33\rho+5)} \\ -\frac{9\rho+1}{N(33\rho+5)} \\ \frac{5\rho+1}{N(33\rho+5)} \\ -\frac{9\rho+1}{N(33\rho+5)} \\ -\frac{9\rho+1}{N(33\rho+5)} \\ \frac{5\rho+1}{N(33\rho+5)} \end{pmatrix}$

Table 3.9: Variance (divided by σ_ϵ^2/N) of the Estimator for Direct Treatment Effect Contrast for Two-Treatment Designs

Period	Design	Model	Baseline	$Var(\tau)$		
				$\rho = 0$	$\rho = 0.5$	$\rho \to 1$
2	AB/BA AA/BB	1	No	0.5	0.857	2
			Yes	0.5	0.682	0.8
	All	2	No	1	2	∞
			Yes	0.667	0.857	1
3	ABB/BAA	1	No	0.333	0.364	0.375
			Yes	0.333	0.35	0.353
	ABB/BAA ABA/BAB	2	No	1	1.455	1.875
			Yes	0.5	0.552	0.566
	ABA/BAB ABB/BAA	2	No	1	1.455	1.875
			Yes	0.5	0.552	0.566
	AAB/BBA ABA/BAB ABB/BAA	2	No	1	1.377	1.607
			Yes	0.5	0.543	0.552
4	$ABBA/BAAB$ $AABB/BBAA$	1	No	0.25	0.25	0.25
			Yes	0.25	0.25	0.25
	$ABBA/BAAB(N/6)$ $ABAB/BABA(N/24)$ $AABB/BBAA(7N/24)$	1	No	0.25	0.25	0.25
			Yes	0.25	0.25	0.25
	$ABBA/BAAB$ $AABA/BBAB$	2	No	1	1.25	1.333
			Yes	0.4	0.419	0.421

Note: Model 1 refers to the traditional model and model 2 refers to the self and mixed carryover effects model.

Chapter 4

The Stratified and Randomized Play-the-Winner Rule (SRPWR)

(SRPWR)

4.1 Introduction

In most clinical trials, patients participate in the study sequentially as they arrive. Suppose that the responses of the patient to treatment are dichotomous (e.g. success/failure, positive/negative). Due to ethical issues on studies involving human subjects, clinicians strive to treat patients in the best way possible. To meet such ethical requirements, Zelen (1969) introduced the play-the-winner rule (PWR) for comparing two treatments in clinical trials.

The goal of the PWR is to allocate more patients to a better treatment. The basic motivation for PWR is that a success with a current treatment should involve a future trial with the same treatment, while a failure generates a trial with an alternative treatment. Zelen assumed that as patients enter the trial one at a time, the outcome of a trial only depends on the treatment given.

One can easily implement the PWR, using a box. A type A ball is placed in the box when a success is obtained with treatment A or a failure with treatment B. A type B ball is placed in the box when a success is obtained with treatment B or a failure with treatment A. When a new patient enters the study, the treatment assignment would be determined by drawing a ball randomly from the box without replacement. When the box is empty, the

assignment is determined by tossing a coin. However, in practice, the time required to observe a patient's response to treatment may be much longer than the time between patient entries. Then the PWR assigns approximately equal numbers of patients to each treatment. Therefore, it may not achieve the goal of allocating patients in favor of the better treatment. In addition, confounding variables have not been considered in this assignment rule.

To improve the PWR, Zelen (1969) also proposed the modified play-the-winner rule (MPWR) under the assumption that patients respond immediately to treatments. Under the MPWR, after each "success" the same treatment continues to be used, and after each "failure" patients will be switched to the other treatment. MPWR assigns more patients to the better treatment. However, the process overlooks past history except for the immediate past. It is not applicable to situations with delayed response to treatments. It also suffers from selection bias, because the response of one patient determines the allocation of the next patient, and it is evident what the next assignment will be.

In this chapter, a new, simple allocation rule is proposed for treatment assignments in stratified and randomized sequential clinical trials. The stratified and randomized play-the-winner rule (SRPWR) is a modified scheme in the spirit of the play-the-winner rule, in that it skews the allocation pattern in favor of a better treatment. It is applicable to clinical trials with more than two treatments. The SRPWR also allows for delayed responses. The randomization can be applied when we have no information about the superiority of one treatment over another at the beginning of the trial. In addition, the SRPWR allows for treatment comparisons among homogenous patients by stratifying them based on possible confounders, for example, age, sex and some comorbidity measures. This new allocation rule works for both fixed and random sample sizes.

In the next section, we propose a new allocation rule and discuss its properties in Section 4.2. The results of the simulation studies will be presented in Section 4.3. Conclusions will follow in Section 4.4.

4.2 SRPWR Allocation Rule

SRPWR is an extension of the randomized play-the-winner rule (RPWR) proposed by Wei and Durham (1978), and it can be applied to trials with more than two treatments. The success probability of a treatment may depend on both the efficiency of the treatment and known confounders (for example, patient's age, gender and disease status). We assume that patients enter the trial one at a time, sequentially, and the outcome of a trial is a "success" or a "failure" (i.e., dichotomous response). In order to adjust treatment comparisons for confounding variables, stratification is used, while patient homogeneity is assumed within each stratum.

The SRPWR is denoted by $SRPWR(\mu, \alpha, \beta, t, s)$, where t is the number of treatments, s is the number of stratifications, $\mu \geqslant 0$, both α and β are multiples of $(t-1)$ and satisfies

$$0 \leq \alpha(t-1) \leq \beta.$$

A $SRPWR(\mu, \alpha, \beta, t, s)$ can be easily implemented as follows.

Step 1: Define s strata based on known confounders.

Step 2: Prepare s boxes, one for each stratum. In each box, place t different types of balls marked i $(i = 1, 2, \ldots, t)$, with μ balls of each type.

Step 3: When a patient belonging to the k^{th} $(k = 1, 2, \ldots, s)$ stratum is available for an assignment, draw a ball from the k^{th} box at random and with replacement. If the ball is type i , then treatment i is assigned to this patient. If the box is empty, then the assignment is determined by generating a uniform random number. If it is smaller than i/t and larger than or equal to $(i-1)/t$, then treatment i is assigned to the patient.

Step 4: When the response of a previous patient to treatment i is available, change the structure of the corresponding box based on the following rule. Assume this patient belongs to the k^{th} stratus. If this response is a success, then additional β balls of type i and additional α balls of other types with $\alpha/(t-1)$ for each type are put in the k^{th} box. If this response is a failure, then additional α balls of type i and additional β balls of other types

with $\beta/(t-1)$ for each type are put in the k^{th} box.

Step 5: Repeat steps 3 to 4 until all patients have been allocated (for fixed sample size) or investigators decide to terminate the study.

In the allocation rule described above, we see that after each assignment, exactly $(\alpha+\beta)$ additional balls are added to the corresponding box. If the response to treatment i is a success, there is a higher probability of assigning the next patient to the same treatment ($\beta \geq \alpha$). If the response to treatment i is a failure, there is a higher probability of assigning the next patient to one of the alternative treatments ($\alpha \leq \beta/(t-1)$). Note that the $RPWR$ is a special case of the $SRPWR$ with $t=2$ and $s=1$.

In the next subsections, we assess the performance of the SRPWR by obtaining the expected number of patients to each treatment.

4.2.1 Performance of SRPWR for Three Treatments

Now let us consider a clinical trial comparing three treatments ($t=3$) among homogenous patients ($s=1$). To study the properties of $SRPWR(\mu,\alpha,\beta,3,1)$, we assume that the response of a patient is instantaneous. Let $R_A(n)$, $R_B(n)$ and $R_C(n)$ be the numbers of balls of types A, B and C in the box, respectively, after n assignments. Let $S_i(n)$ and $F_i(n)$ be the numbers of successes and failures with treatment i after n responses, respectively, where $i=A$, B or C. Then we have

$$\begin{aligned}
R_A(n) &= \mu+\beta S_A(n)+\alpha F_A(n)+\frac{\alpha}{2}[S_B(n)+S_C(n)]+\frac{\beta}{2}[F_B(n)+F_C(n)] \\
R_B(n) &= \mu+\beta S_B(n)+\alpha F_B(n)+\frac{\alpha}{2}[S_A(n)+S_C(n)]+\frac{\beta}{2}[F_A(n)+F_C(n)] \\
R_C(n) &= \mu+\beta S_C(n)+\alpha F_C(n)+\frac{\alpha}{2}[S_A(n)+S_B(n)]+\frac{\beta}{2}[F_A(n)+F_B(n)]
\end{aligned} \tag{4.1}$$

Note that the total number of balls after n responses, T_n, is a constant, i.e.,

$$T_n = R_A(n)+R_B(n)+R_C(n) = 3\mu+n(\alpha+\beta) \tag{4.2}$$

Let $N_i(n)$ be the number of patients assigned to treatment i after n assignments, p_i be the probability of a single trial success for treatment i,

and $q_i = 1 - p_i$ where $i = A$, B or C. Then

$$N_A(n) + N_B(n) + N_C(n) = n \tag{4.3}$$

And

$$E[S_i(n)] = E[N_i(n)]p_i \tag{4.4}$$

where $i = A$, B or C.

Result 4.2.1.1: *In $SRPWR(\mu, \alpha, \beta, 3, 1)$, the expected numbers of balls of types A and B in the box after the $(n+1)$ assignments, given the first n assignments, are*

$$\begin{aligned} E[R_A(n+1)] &= \frac{1}{2}(\alpha p_C + \beta q_C) + \left[1 + \frac{\alpha(q_A - \frac{1}{2}p_C) + \beta(p_A - \frac{1}{2}q_C)}{T_n}\right] E[R_A(n)] \\ &\quad + \frac{(\beta - \alpha)(q_B - q_C)}{2T_n} E[R_B(n)] \end{aligned} \tag{4.5}$$

$$\begin{aligned} E[R_B(n+1)] &= \frac{1}{2}(\alpha p_C + \beta q_C) + \left[1 + \frac{\alpha(q_B - \frac{1}{2}p_C) + \beta(p_B - \frac{1}{2}q_C)}{T_n}\right] E[R_B(n)] \\ &\quad + \frac{(\beta - \alpha)(q_A - q_C)}{2T_n} E[R_A(n)] \end{aligned} \tag{4.6}$$

and

$$E[R_C(n+1)] = 3\mu + (n+1)(\alpha + \beta) - E[R_A(n+1)] - E[R_B(n+1)] \tag{4.7}$$

Proof: Given the first n responses, $R_A(n+1)$ has four possible values: $R_A(n) + \beta$, $R_A(n) + \alpha$, $R_A(n) + (\alpha/2)$ or $R_A(n) + (\beta/2)$ with probabilities

$$\begin{aligned} Prob[R_A(n+1) = R_A(n) + \beta | H_n] &= p_A \frac{R_A(n)}{T_n} \\ Prob[R_A(n+1) = R_A(n) + \alpha | H_n] &= q_A \frac{R_A(n)}{T_n} \\ Prob[R_A(n+1) = R_A(n) + \frac{\alpha}{2} | H_n] &= p_B \frac{R_B(n)}{T_n} + p_c \frac{T_n - (R_A(n) + R_B(n))}{T_n} \\ Prob[R_A(n+1) = R_A(n) + \frac{\beta}{2} | H_n] &= q_B \frac{R_B(n)}{T_n} + q_c \frac{T_n - (R_A(n) + R_B(n))}{T_n} \end{aligned} \tag{4.8}$$

where H_n contains the information $R_A(n)$, $R_B(n)$ and T_n upon the previous n assigned subjects.

Then,

$$
\begin{aligned}
& E[R_A(n+1)] \\
= \; & E[E[R_A(n+1)|H_n]] \\
= \; & E\left[(R_A(n)+\beta)p_A\frac{R_A(n)}{T_n}\right] \\
& +E\left[(R_A(n)+\alpha)q_A\frac{R_A(n)}{T_n}\right] \\
& +E\left[\left(R_A(n)+\frac{\alpha}{2}\right)\left(p_B\frac{R_B(n)}{T_n}+p_c\frac{T_n-(R_A(n)+R_B(n))}{T_n}\right)\right] \\
& +E\left[\left(R_A(n)+\frac{\beta}{2}\right)\left(q_B\frac{R_B(n)}{T_n}+q_c\frac{T_n-(R_A(n)+R_B(n))}{T_n}\right)\right] \\
= \; & E\left[\frac{p_AR_A^2(n)}{T_n}+\frac{\beta p_AR_A(n)}{T_n}+\frac{q_AR_A^2(n)}{T_n}+\frac{\alpha q_AR_A(n)}{T_n}\right] \\
& +E\left[\frac{p_BR_A(n)R_B(n)}{T_n}+\frac{\alpha p_BR_B(n)}{2T_n}+R_A(n)p_C-\frac{p_CR_A^2(n)}{T_n}-\frac{p_CR_A(n)R_B(n)}{T_n}\right] \\
& +E\left[\frac{\alpha}{2}p_C-\frac{\alpha p_CR_A(n)}{2T_n}-\frac{\alpha p_CR_B(n)}{2T_n}+\frac{q_BR_A(n)R_B(n)}{T_n}+\frac{\beta q_BR_B(n)}{2T_n}\right] \\
& +E\left[R_A(n)q_C-\frac{q_CR_A^2(n)}{T_n}-\frac{q_CR_A(n)R_B(n)}{T_n}+\frac{\beta}{2}q_C-\frac{\beta q_CR_A(n)}{2T_n}-\frac{\beta q_CR_B(n)}{2T_n}\right] \\
= \; & \frac{R_A^2(n)}{T_n}(p_A+q_A)-\frac{R_A^2(n)}{T_n}(p_C+q_C)+\frac{R_A(n)R_B(n)}{T_n}(p_B+q_B) \\
& -\frac{R_A(n)R_B(n)}{T_n}(p_C+q_C)+R_A(n)\left[\frac{\beta p_A}{T_n}+\frac{\alpha q_A}{T_n}+p_C-\frac{\alpha p_C}{2T_n}+q_C-\frac{\beta q_C}{2T_n}\right] \\
& +R_B(n)\left[\frac{\alpha p_B}{2T_n}-\frac{\alpha p_C}{2T_n}+\frac{\beta q_B}{2T_n}-\frac{\beta q_C}{2T_n}\right]+\frac{\alpha}{2}p_C+\frac{\beta}{2}q_C \\
= \; & \frac{1}{2}(\alpha p_C+\beta q_C)+\left[1+\frac{\alpha(q_A-\frac{1}{2}p_C)+\beta(p_A-\frac{1}{2}q_C)}{T_n}\right]E[R_A(n)] \\
& +\frac{(\beta-\alpha)(q_B-q_C)}{2T_n}E[R_B(n)]
\end{aligned}
$$

Similarly, we can prove that Equation 4.6 holds.

According to Equation 4.2, after $(n+1)$ assignments, we have

$$E[R_A(n+1)+R_B(n+1)+R_C(n+1)]=3\mu+(n+1)(\alpha+\beta)$$

Therefore,

$$E[R_C(n+1)]=3\mu+(n+1)(\alpha+\beta)-E[R_A(n+1)]-E[R_B(n+1)]$$

□

According to *Result 4.2.1.1*, we can obtain the expectations $E[R_A(n)]$, $E[R_B(n)]$ and $E[R_C(n)]$, recursively. For example, since

$$E[R_A(0)] = E[R_B(0)] = E[R_C(0)] = \mu \tag{4.9}$$

Then we have

$$\begin{aligned}
E[R_A(1)] &= \mu + \frac{1}{3}(\beta p_A + \alpha q_A) + \frac{\alpha}{6}(p_B + p_C) + \frac{\beta}{6}(q_B + q_C) \\
E[R_B(1)] &= \mu + \frac{1}{3}(\beta p_B + \alpha q_B) + \frac{\alpha}{6}(p_A + p_C) + \frac{\beta}{6}(q_A + q_C) \\
E[R_C(1)] &= \mu + \alpha + \beta - \frac{\alpha + 2\beta}{6}(p_A + p_B) - \frac{2\alpha + \beta}{6}(q_A + q_B) \\
&\quad -\frac{1}{3}(\alpha p_C + \beta q_C)
\end{aligned} \tag{4.10}$$

To assess the performance of $SRPWR(\mu, \alpha, \beta, 3, 1)$, we need to calculate the expected numbers of patients treated by treatments A, B and C, respectively, i.e., $E[N_i(n)]$ where $i = A, B,$ or C.

Result 4.2.1.2: *In $SRPWR(\mu, \alpha, \beta, 3, 1)$, the expected number of patients treated by each treatment after the n assignments is*

$$E[N_A(n)] = \frac{(\nu_4 E[R_A(n)] - \nu_3 E[N_B(n)]) + \nu_1(\nu_3 - \nu_4)}{\nu_2\nu_4 - \nu_3\nu_5} \tag{4.11}$$

$$E[N_B(n)] = \frac{(\nu_5 E[R_A(n)] - \nu_2 E[N_B(n)]) + \nu_1(\nu_2 - \nu_5)}{\nu_3\nu_5 - \nu_2\nu_4} \tag{4.12}$$

$$E[N_C(n)] = n - E[N_A(n)] - E[N_B(n)] \tag{4.13}$$

where $\nu_1 = \mu + \frac{n}{2}(\alpha p_C + \beta q_c)$, $\nu_2 = \alpha(q_A - \frac{1}{2}p_C) + \beta(p_A - \frac{1}{2}q_C)$, $\nu_3 = \frac{1}{2}(\beta - \alpha)(q_B - q_C)$, $\nu_4 = \alpha(q_B - \frac{1}{2}p_C) + \beta(p_B - \frac{1}{2}q_C)$, *and* $\nu_5 = \frac{1}{2}(\beta - \alpha)(q_A - q_C)$.

Proof: By (4.1), (4.3) and (4.4), we have

$$
\begin{aligned}
E[R_A(n)] &= E\left[\mu + \beta S_A(n) + \alpha F_A(n) + \frac{\alpha}{2}[S_B(n) + S_C(n)] + \frac{\beta}{2}[F_B(n) + F_C(n)]\right] \\
&= \mu + \beta p_A E[N_A(n)] + \alpha q_A E[N_A(n)] + \frac{\alpha}{2}\left(p_B E[N_B(n)] + p_C E[N_C(n)]\right) \\
&\quad + \frac{\beta}{2}(q_B E[N_B(n)] + q_C E[N_C(n)]) \\
&= \mu + \beta p_A E[N_A(n)] + \alpha q_A E[N_A(n)] + \frac{\alpha}{2} p_B E[N_B(n)] \\
&\quad + \frac{\alpha}{2} p_C (n - E[N_A(n)] - E[N_B(n)]) + \frac{\beta}{2} q_B E[N_B(n)] \\
&\quad + \frac{\beta}{2} q_C (n - E[N_A(n)] - E[N_B(n)]) \\
&= \mu + \frac{n}{2}(\alpha p_C + \beta q_c) + \left[\alpha(q_A - \frac{1}{2}p_C) + \beta(p_A - \frac{1}{2}q_C)\right] E[N_A(n)] \\
&\quad + \frac{1}{2}(\beta - \alpha)(q_B - q_C) E[N_B(n)] \qquad (4.14)
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
E[R_B(n)] &= \mu + \frac{n}{2}(\alpha p_C + \beta q_c) + \left[\alpha(q_B - \frac{1}{2}p_C) + \beta(p_B - \frac{1}{2}q_C)\right] E[N_B(n)] \\
&\quad + \frac{1}{2}(\beta - \alpha)(q_A - q_C) E[N_A(n)] \qquad (4.15)
\end{aligned}
$$

Let $\nu_1 = \mu + \frac{n}{2}(\alpha p_C + \beta q_c)$, $\nu_2 = \alpha(q_A - \frac{1}{2}p_C) + \beta(p_A - \frac{1}{2}q_C)$, $\nu_3 = \frac{1}{2}(\beta - \alpha)(q_B - q_C)$, $\nu_4 = \alpha(q_B - \frac{1}{2}p_C) + \beta(p_B - \frac{1}{2}q_C)$, and $\nu_5 = \frac{1}{2}(\beta - \alpha)(q_A - q_C)$, Equations 4.14 and 4.15 become

$$
\begin{aligned}
E[R_A(n)] &= \nu_1 + \nu_2 E[N_A(n)] + \nu_3 E[N_B(n)] \qquad (4.16) \\
E[R_B(n)] &= \nu_1 + \nu_4 E[N_B(n)] + \nu_5 E[N_A(n)] \qquad (4.17)
\end{aligned}
$$

Without much difficulty, we will obtain (4.11) and (4.12) by solving Equations 4.16 and 4.17. In addition, we can prove that Equation 4.13 holds according to Equation 4.3. □

To summary above, based on *Result 4.2.1.1*, we can obtain the value of $E[R_A(n)]$, $E[R_B(n)]$ and $E[R_C(n)]$ recursively from the values of $E[R_A(n-1)]$ and $E[R_B(n-1)]$, where $E[R_A(n)]$ represents the expected number of type A balls in the box after n assignments, and $E[R_B(n)]$ represents the expected number of type B balls in the box after n assignments. Based on *Result*

4.2.1.2, the expected numbers of patients treated by each treatment can be calculated. If the new allocation rule works well, we will have $E[N_A(n)] \geq E[N_B(n)] \geq E[N_C(n)]$, assuming that treatment A is the best and treatment C is the worst.

Let us consider the simplest $SRPWR(\mu, \alpha, \beta, 3, 1)$ with $\mu = \alpha = 0$ and $\beta = 2$. The expected number of patients treated by each of the three treatments using $SRPWR$ (0, 0, 2, 3, 1) is provided in Table 4.1. Without loss of generality, we assume that treatment A has the highest success probability (best treatment), and then treatment B, and treatment C has the lowest success probability. Tables 4.1 shows that $SRPWR(0, 0, 2, 3, 1)$ in a clinical trial tends to assign more patients to the better treatment no matter how small the sample size is. As an example, when $p_A = 0.8$, $p_B = 0.5$, $p_C = 0.3$ and the total number of subjects (n) is 30, the expected numbers of patients treated by treatment A, B or C are 16.19, 8.1 and 5.71, respectively.

In comparison with existing rules, Table 4.2 shows that $SRPWR(0,$ 0, 2, 3, 1) tends to allocate more patients to better treatments than the $GPUD(1, 2, 1)$ (Wei, 1979) and PWC rule (Hoel and Sobel, 1971). For example, when $p_A = 0.4$, $p_B = 0.2$, $p_C = 0.1$ and the total number of subjects (n) is 6, the expected number of patients assigned to treatment A by the new rule is 2.4081, while it is 2.2581 under $GPUD(\mathbf{1}, 2, 1)$ and 2.2292 under PWC rule. Such result holds true regardlessly of the sample size. Therefore, the new rule is superior on ethical grounds.

4.2.2 Generalization to t Treatments

The generalization to t-treatment clinical trials is rather straightforward, although the derivation is quite tedious.

Based on the same notation developed in the previous section and letting $i = 1, 2, ..., t$, we have

$$R_i(n) = \mu + \beta S_i(n) + \alpha F_i(n) + \frac{\alpha}{t-1} \sum_{i' \neq i} S_{i'}(n) + \frac{\beta}{t-1} \sum_{i' \neq i} F_{i'}(n) \tag{4.18}$$

$$T_n = \sum_i R_i(n) = t\mu + n(\alpha + \beta) \tag{4.19}$$

$$\sum_i N_i(n) = n \tag{4.20}$$

and

$$E[S_i(n)] = E[N_i(n)]p_i \tag{4.21}$$

In addition, the transition probabilities for $R_i(\cdot)$ from stage n to $n+1$ are

$$\begin{aligned}
Prob[R_i(n+1) = R_i(n) + \beta | H_n] &= p_i \frac{R_i(n)}{T_n} \\
Prob[R_i(n+1) = R_i(n) + \alpha | H_n] &= q_i \frac{R_i(n)}{T_n} \\
Prob[R_i(n+1) = R_i(n) + \frac{\alpha}{t-1} | H_n] &= \sum_{i' \neq i} p_{i'} \frac{R_{i'}(n)}{T_n} \\
Prob[R_i(n+1) = R_i(n) + \frac{\beta}{t-1} | H_n] &= \sum_{i' \neq i} q_{i'} \frac{R_{i'}(n)}{T_n}
\end{aligned} \tag{4.22}$$

for $i = 1, 2, ..., t$.

Based on the transition probabilities given in (4.22), analog to the three-treatment case discussed earlier, we can obtain the expectations $E[R_i(n)]$, recursively.

We then take the expectation on both sides of Equation 4.18 to obtain t equations. Solving the equations, along with Equations (4.19), (4.20) and (4.21), we are able to calculate $E[N_i(n)]$, the expected number of patients treated by treatment i after n assignments.

4.3 Simulation Study

To simplify, let us first consider an example of the $SRPWR$ for assigning patients to three treatments with one stratum: $SRPWR(0, 0, 2, 3, 1)$, where $t = 3$, $s = 1$, $\mu = \alpha = 0$ and $\beta = 2$.

The average number of patients from 1,000 simulations treated by each treatment using $SRPWR(0, 0, 2, 3, 1)$ is provided in Table 4.3. The average

probability of patients treated by each treatment using $SRPWR(0, 0, 2, 3, 1)$ is provided in Table 4.4. The C++ program code is available upon request.

Similar to the result in Table 4.1, Tables 4.3 and 4.4 clearly show that $SRPWR(0, 0, 2, 3, 1)$ assign more patients to the better treatment no matter how small the sample size is. As an example, when $p_A = 0.8$, $p_B = 0.5$, $p_C = 0.3$ and the total number of subjects (n) is 30, the estimated numbers of patients treated by treatment A, B or C are 15.993 (53.3%), 8.399 (28%) and 5.608 (18.7%) respectively.

We then consider the case of two strata ($s = 2$) defined according to the patients' characteristics. Table 4.5 shows the simulation result given different probabilities in each stratum, where "Total" is the total number of patients assigned to each treatment in the trial. When a treatment works well for one group of patients but not for the other group (the last case in Table 4.5), $SRPWR(0, 0, 2, 3, 2)$ successfully adjusts the allocations according to the patients' characteristics and therefore treats each patient in the best possible way.

4.4 Conclusion

Since Zelen (1969) proposed the PWR for controlled clinical trials, various researchers have proposed and investigated allocation rules for better patient treatment. The main contribution of this chapter has been in proposing a simple allocation rule which considers the heterogeneity of subjects. Simulation studies show that, on average, the SRPWR tends to assign more patients to the better treatment. SRPWR is superior to existing allocation rules, and it successfully adjusts the allocation results while accommodating the heterogeneity of the patients, leading to a better allocation strategy and better patient treatment.

Table 4.1: Expected Number of Patients Treated by each Treatment for $SRPWR$ (0, 0, 2, 3, 1)

p_A	p_B	p_C	n	A	B	C
0.8	0.5	0.3	10	5.02	2.89	2.09
			15	7.78	4.22	3.01
			20	10.56	5.52	3.91
			30	16.19	8.10	5.71
			40	21.87	10.64	7.49
			50	27.58	13.17	9.26
			100	56.34	25.63	18.03
			500	289.78	123.15	87.18
			1000	583.44	243.78	172.78
0.8	0.7	0.6	10	4.03	3.28	2.68
			15	6.174	4.896	3.930
			20	8.333	6.504	5.163
			30	12.69	9.706	7.604
			40	17.078	12.896	10.026
			50	21.49	16.08	12.44
			100	43.702	31.914	24.384
0.8	0.2	0.1	10	5.879	2.195	1.926
			15	9.120	3.132	2.748
			20	12.392	4.0512	3.557
			30	18.984	5.862	5.154
			40	25.615	7.651	6.734
			50	32.27	9.43	8.30
			100	65.723	18.202	16.075
0.3	0.3	0.1	10	3.741	3.308	2.951
			15	5.63	4.96	4.41
			20	7.513	6.607	5.880
			30	11.284	9.905	8.811
			40	15.055	13.203	11.742
			50	18.825	16.502	14.673
			100	37.675	32.994	29.331
0.8	0.8	0.8	50	16.67	16.67	16.67
0.8	0.2	0.2	50	31.54	9.23	9.23

Note: p_A, p_B and p_C are success probabilities for treatment A, B and C respectively, and n is the sample size. Without loss of generality, we assume that $p_A \geq p_B \geq p_C$.

Table 4.2: Comparisons of $SRPWR(0,0,2,3,1)$, $GPUD(\mathbf{1},2,1)$ and PWC

				$SRPWR(0,0,2,3,1)$			$GPUD(\mathbf{1},2,1)$			PWC rule		
p_A	p_B	p_C	n	A	B	C	A	B	C	A	B	C
0.4	0.2	0.1	6	2.4081	1.8961	1.6958	2.2581	1.9399	1.8021	2.2292	1.9249	1.8458
			12	4.8923	3.76	3.3478	4.671	3.833	3.496	4.4677	3.8581	3.6742
			18	7.3758	5.6228	5.0014	7.1104	5.7155	5.1741	6.7093	5.7892	5.5016
			27	11.1007	8.4166	7.4827	10.7933	8.5286	7.6781	10.0713	8.6857	8.243
0.6	0.3	0.2	6	2.6565	1.7764	1.5672	2.3913	1.8751	1.7333	2.4108	1.8321	1.7571
			12	5.5103	3.4597	3.0301	5.0679	3.6407	3.2914	4.8604	3.6591	3.4805
			18	8.3788	5.1329	4.4882	7.8129	5.3747	4.8124	7.3087	5.4867	5.2046
			27	12.6931	7.6353	6.6716	11.9885	7.947	7.0645	10.9811	8.2281	7.7908
0.8	0.4	0.2	6	3.0534	1.6785	1.2682	2.5982	1.8419	1.5599	2.7568	1.673	1.5702
			12	6.5576	3.1286	2.3138	5.7051	3.4897	2.8052	5.651	3.3047	3.0443
			18	10.1337	4.5311	3.3352	8.967	5.0595	3.9735	8.5457	4.9363	4.518
			27	15.561	6.5932	4.8458	14.0068	7.3372	5.656	12.8878	7.3837	6.7285
0.9	0.5	0.3	6	3.1449	1.6498	1.2053	2.6278	1.8338	1.5384	2.9476	1.5604	1.492
			12	6.8994	2.998	2.1027	5.8504	3.4415	2.7045	6.1829	3.0068	2.8103
			18	10.7874	4.2623	2.9503	9.2827	4.9459	3.7714	9.4169	4.4536	4.1295
			27	16.7468	6.0784	4.1748	14.6445	7.091	5.2645	14.268	6.6238	6.1082

Note: Entries are the expected number of patients treated by each Treatment.

Table 4.3: Average Number of Patients Treated by each Treatment for $SRPWR$ (0, 0, 2, 3, 1)

p_A	p_B	p_C	n	A	B	C
0.8	0.5	0.3	10	4.817	3.079	2.104
			15	7.548	4.442	3.01
			20	10.431	5.496	4.073
			30	15.993	8.399	5.608
			40	21.663	10.754	7.583
			50	28.071	12.848	9.081
			100	57.454	25.057	17.489
			500	289.716	123.535	86.749
			1000	584.881	243.834	171.285
0.8	0.7	0.6	10	3.939	3.435	2.626
			15	6.25	5.11	3.64
			20	8.285	6.528	5.187
			30	12.98	9.978	7.042
			40	17.112	13.553	9.335
			50	22.605	16.205	11.19
			100	44.797	31.837	23.366
0.8	0.2	0.1	10	5.778	2.197	2.025
			15	8.934	3.125	2.941
			20	12.604	3.729	3.667
			30	18.925	5.609	5.466
			40	25.865	7.082	7.053
			50	32.949	8.392	8.659
			100	66.447	16.955	16.598
0.3	0.2	0.1	10	3.661	3.322	3.017
			15	5.43	4.896	4.674
			20	7.284	6.47	6.246
			30	11.177	9.493	9.33
			40	15.005	12.722	12.273
			50	18.682	16.078	15.24
			100	37.365	32.379	30.256
0.8	0.8	0.8	50	18.337	16.777	14.886
0.8	0.2	0.2	50	31.854	8.645	9.501

Note: See notes for Table 4.1, and the entries are average from 1,000 simulations.

Table 4.4: Average Probability of Patients Treated by each Treatment for $SRPWR$ (0, 0, 2, 3, 1)

p_A	p_B	p_C	n	A	B	C
0.8	0.5	0.3	10	0.482	0.308	0.210
			15	0.503	0.296	0.201
			20	0.522	0.275	0.204
			30	0.533	0.280	0.187
			40	0.542	0.269	0.190
			50	0.561	0.257	0.182
			100	0.575	0.251	0.175
			500	0.579	0.247	0.173
			1000	0.585	0.244	0.171
0.8	0.7	0.6	10	0.394	0.344	0.263
			15	0.417	0.341	0.243
			20	0.414	0.326	0.259
			30	0.433	0.333	0.235
			40	0.428	0.339	0.233
			50	0.452	0.324	0.224
			100	0.448	0.318	0.234
0.8	0.2	0.1	10	0.578	0.220	0.203
			15	0.596	0.208	0.196
			20	0.630	0.186	0.183
			30	0.631	0.187	0.182
			40	0.647	0.177	0.176
			50	0.659	0.168	0.173
			100	0.664	0.170	0.166
0.3	0.2	0.1	10	0.366	0.332	0.302
			15	0.362	0.326	0.312
			20	0.364	0.324	0.312
			30	0.373	0.316	0.311
			40	0.375	0.318	0.307
			50	0.374	0.322	0.305
			100	0.374	0.324	0.303
0.8	0.8	0.8	50	0.367	0.336	0.298
0.8	0.2	0.2	50	0.637	0.173	0.190

Note: See notes for Table 4.3.

Table 4.5: Average Number of Patients Treated by each Treatment by Stratification for $SRPWR(0,0,2,3,2)$

Success Probability		Sample Size	Stratum 1			Stratum 2			Total		
Stratum 1	Stratum 2	$n(n_1,n_2)$	A	B	C	A	B	C	A	B	C
$\begin{pmatrix} p_{1A} \\ p_{1B} \\ p_{1C} \end{pmatrix} = \begin{pmatrix} 0.4 \\ 0.2 \\ 0.1 \end{pmatrix}$	$\begin{pmatrix} p_{2A} \\ p_{2B} \\ p_{2C} \end{pmatrix} = \begin{pmatrix} 0.3 \\ 0.3 \\ 0.4 \end{pmatrix}$	6 (3,3)	1.165	0.983	0.852	1.003	0.919	1.078	2.168	1.902	1.93
		12(6,6)	2.366	1.924	1.71	1.952	1.971	2.077	4.318	3.895	3.787
		18 (9,9)	3.559	2.842	2.599	3.01	2.836	3.154	6.569	5.678	5.753
		27 (13, 14)	5.169	4.171	3.66	4.648	4.474	4.878	9.817	8.645	8.538
		100 (50, 50)	20.4	15.163	14.437	16.088	16.01	17.902	36.488	31.173	32.339
$\begin{pmatrix} p_{1A} \\ p_{1B} \\ p_{1C} \end{pmatrix} = \begin{pmatrix} 0.6 \\ 0.3 \\ 0.2 \end{pmatrix}$	$\begin{pmatrix} p_{2A} \\ p_{2B} \\ p_{2C} \end{pmatrix} = \begin{pmatrix} 0.8 \\ 0.4 \\ 0.2 \end{pmatrix}$	6 (3,3)	1.204	0.929	0.867	1.316	0.902	0.782	2.52	1.831	1.649
		12(6,6)	2.603	1.822	1.575	2.957	1.8	1.243	5.56	3.622	2.818
		18 (9,9)	3.916	2.731	2.353	4.554	2.465	1.981	8.47	5.196	4.334
		27 (13, 14)	5.99	3.754	3.256	7.506	3.723	2.771	13.496	7.477	6.027
		100 (50, 50)	23.601	13.741	12.658	29.687	11.463	8.85	53.288	25.204	21.508
$\begin{pmatrix} p_{1A} \\ p_{1B} \\ p_{1C} \end{pmatrix} = \begin{pmatrix} 0.9 \\ 0.5 \\ 0.3 \end{pmatrix}$	$\begin{pmatrix} p_{2A} \\ p_{2B} \\ p_{2C} \end{pmatrix} = \begin{pmatrix} 0.3 \\ 0.5 \\ 0.9 \end{pmatrix}$	6 (3,3)	1.297	0.914	0.789	0.789	0.914	1.297	2.086	1.828	2.086
		12(6,6)	3.118	1.635	1.247	1.247	1.635	3.118	4.365	3.27	4.365
		18 (9,9)	4.83	2.434	1.736	1.736	2.434	4.83	6.566	4.868	6.566
		27 (13, 14)	7.202	3.425	2.373	2.596	3.729	7.675	9.798	7.154	10.048
		100 (50, 50)	32.544	10.369	7.087	7.087	10.369	32.544	39.631	20.738	39.631

Note: Entries are average from 1,000 simulations.

Chapter 5

Response-Adaptive Repeated Measurement Designs

In this chapter, we develop a new adaptive allocation rule, which can accurately provide good estimates of the treatment effect and assign more patients to a better treatment. To achieve this goal, we introduce the concept of an evaluation function to evaluate the performance of each treatment sequence, and define a new optimality criteria, which has two components: the first component determines a treatment sequence that maximizes the information matrix; the second determines a treatment sequence that gives the best performance based on the observed data. This new design strategy is applicable to trials with both dichotomous and continuous responses.

5.1 Introduction

There has been a growing interest in the development of clinical trials to help ensure that the allocation strategy is better, being informed from all available sources. This interest has been fueled by the fact that some health interventions are largely ineffective and even harmful, and thus both a waste of public resources and unethical. In *response-adaptive designs* (RAD), we modify the trial on the basis of outcomes in previous observations in order to achieve a specific goal. Optimal designs are usually constructed under a single optimality criterion. For example, Kushner(2003) proposed an adaptive allocation rule, which was found by replacing the classical optimal design strategy with

a response adaptive allocation method to maximize the resulting information matrix on treatment effects after first few subjects have been observed. In the classical sequential trial, the decision to terminate the accession of new subjects is based on minimizing the expected sample size (Armitage, 1975). In play-the-winner designs, the goal is to minimize the number of subjects receiving inferior treatments (Zelen, 1969). However, such designs with a single objective are less intuitive. Clinical investigators may need to deal with more than one objective when designing an experiment (Moerbeek and Wong, 2002).

In this chapter we consider improving the current response-adaptive designs in three directions: 1) by developing the strategy for continuous as well as discrete responses, 2) by constructing optimal multiple-objective designs to increase both the estimation precision and the proportion of patients treated by better treatments, and 3) by using a more general model including the model with self and simple mixed carryover effects (Afsarinejad and Hedayat, 2002) and random subject effects as discussed in Chapter 3. Some applications to dichotomous and continuous responses will be discussed in Chapter 6 and Chapter 7, respectively.

Since the new allocation rule for response-adaptive RMD aims to address ethical issues, we will first define a way to evaluate the performance of a treatment sequence in Section 5.2. We then present the new allocation rule in Section 5.3.

5.2 Evaluation Function for a Treatment Sequence

In order to assign more patients to better performing treatment sequences, we need an objective way to evaluate the performance of each treatment sequence. An *Evaluation Function* for a treatment sequence will be defined.

Properties of an Evaluation Function: An evaluation function for treatment sequence k based on the existing data, $g_k(\cdot)$, follows the following properties:

1) it is non-negative;

2) it is monotonic.

Defining an evaluation function for a given treatment sequence is not unique. As long as the above two properties are satisfied, one can define various types of evaluation functions. Following an evaluation function, when two treatment sequences have the same values, we say that the performances of these two treatment sequences are *indistinguishable.*

Following are two examples of how to define an evaluation function.

Example 1: Consider a synthetic, two-treatment, two-period repeated measurement design, where drug A and drug B are randomly assigned to 10 patients and dichotomous responses (1 if success, 0 if failure) are collected. The data are given in Table 5.1.

We advocate the idea from the play-the-winner rule and evaluate the performance of a treatment sequence by calculating the probability of success over all subjects, given that particular treatment sequence. Thus, an evaluation function can be defined as

$$g_k = \frac{\|S_K\|}{\|K\|} \tag{5.1}$$

where $\|S_K\|$ denotes the total number of successes for treatment sequence k in all periods, and $\|K\|$ denotes the total number of patients given treatment sequence k

In our example data, the corresponding value for each possible treatment sequence, based on the given evaluation function is:

$$g_{AA} = \frac{5}{3} = 1.67 \quad ; \quad g_{AB} = \frac{2}{2} = 1;$$
$$g_{BA} = \frac{2}{3} = 0.67 \quad ; \quad g_{BB} = \frac{2}{2} = 1.$$

Therefore, treatment sequence AA is the best among these four possible treatment sequences. The performances of the treatment sequence AB and BB are indistinguishable at the current stage.

Example 2: Let us consider another two-treatment two-period repeated measurement design to study the treatment effect of reducing the fever temperature. The data are given in Table 5.2.

The normal body temperature is considered to be 37°C. When the body temperature rises above 37°C, the patient is said to have a fever. A better treatment sequence could reduce the fever to normal body temperature.

An evaluation function can be defined as a sample deviation from the chosen value of 37°C of a normal body temperature. In this case, a smaller value of g_k indicates a better treatment sequence, because it indicates the treatment sequence that successfully reduces the body temperature of a patient to the normal level

$$g_k = \frac{\sum_{i=1}^{2} \sum_j (y_{ijk} - 37)^2}{\|K\|} \tag{5.2}$$

where $i(= 1, 2)$ is the index of the period, j represents those patients given treatment sequence k and y_{ijk} is the body temperature for the j^{th} subject in the i^{th} period given treatment sequence k, and $\|K\|$ is the total number of patients given treatment sequence k

In our example data,

$$\begin{aligned}
g_{AA} &= \frac{(37.5-37)^2 + (37-37)^2 + (37.8-37)^2 + (37-37)^2}{3} \\
&+ \frac{(38-37)^2 + (37-37)^2}{3} = 0.63 \\
g_{AB} &= \frac{(37.5-37)^2 + (38-37)^2 + (37.5-37)^2 + (37.5-37)^2}{2} = 0.875 \\
g_{BA} &= \frac{(37.5-37)^2 + (37-37)^2 + (38.5-37)^2 + (37.5-37)^2}{3} \\
&+ \frac{(38.5-37)^2 + (37.5-37)^2}{3} = 1.75 \\
g_{BB} &= \frac{(38.5-37)^2 + (39-37)^2 + (39-37)^2 + (38.5-37)^2}{2} = 6.25
\end{aligned}$$

Therefore, treatment sequence AA performs the best among the four possible treatment sequences.

5.3 Allocation Rule

A new allocation rule for setting up an adaptive design with total N subjects can be conducted as follows.

Step 1: The first $m(m < N)$ patients will be assigned using the optimal design suggested in the literature or a completely randomized design.

Step 2: To allocate the j^{th} patient, $j \geqslant m+1$, calculate the observed information matrix based on the data available from the first $(j-1)$ patients, denoted by $\widehat{A}_{j-1}(\mathbb{H}_{j-1})$, and the evaluation function $g_{j-1,k}(\mathbb{H}_{j-1})$,where $k = 1, 2, \ldots, s$, s is the total number of treatment sequences, $\mathbb{H}_{j-1}$ is the allocation-and-response history of all first $(j-1)$ patients, including the following information: 1) how many patients have been assigned to each treatment sequence, 2) the values of the response variable at each time period for each patient.

For simplicity, we omit $\mathbb{H}_{j-1}$ from $\widehat{A}_{j-1}(\mathbb{H}_{j-1})$ and $g_{j-1,k}(\mathbb{H}_{j-1})$.

For example, based on model 3.4 in Chapter 3, the observed information matrix given the data from the first $j-1$ patients is defined in Equation 3.10, i.e.,

$$\widehat{A}_{j-1} = \sum_{k \in \mathbb{H}_{j-1}} N_k \mathbf{X}_k^T \widehat{\mathbf{C}}^{-1} \mathbf{X}_k$$

where $\mathbf{X}_k$ is the design matrix for treatment sequence k, and $\widehat{\mathbf{C}}$ is the estimated variance-covariance matrix for the response vector $\mathbf{y}_{jk}$.

Under the equicorrelated covariance assumption,

$$\mathbf{C} = \sigma_\varepsilon^2 \mathbf{I} + \sigma_\xi^2 \mathbf{1}\mathbf{1}^T \tag{5.3}$$

one can estimate the variance-covariance matrix using

$$\widehat{\mathbf{C}} = \hat{\sigma}_\varepsilon^2 \mathbf{I} + \hat{\sigma}_\xi^2 \mathbf{1}\mathbf{1}^T \tag{5.4}$$

where $\hat{\sigma}_\varepsilon^2$ and $\hat{\sigma}_\xi^2$ can be estimated using $\mathbb{H}_{j-1}$

Step 3: Choose a treatment sequence k^* for the j^{th} patient to maximize the criterion Λ in (5.5). In situations where more than one treatment sequence achieves the maximum criterion score, one can randomly assign one treatment sequence to the j^{th} patient.

Here, without loss of generality, where we assume a higher value of $g_{j-1,k}$ indicates a better treatment sequence, the criterion Λ is defined as

$$\Lambda = \lambda \frac{\Theta(\widehat{A_j^k}(\mathbb{H}_{j-1}))}{\Theta(\widehat{A_j^{k^{(a)}}}(\mathbb{H}_{j-1}))} + (1-\lambda)\frac{g_{j-1,k}}{g_{j-1,k^{(b)}}} \tag{5.5}$$

where λ is a constant between zero and one.

Note that the criterion Λ has two components. The first component deals with choosing a treatment sequence to maximize the information matrix. The second deals with choosing a treatment sequence that gives the best performance based on the observed data. The λ is used to balance the two objectives, and can be chosen by investigators prior to the experiment. Different values of λ will give different weights to these two elements. The choice of λ is often driven by which of these components researchers want to emphasize. We will discuss the effect of using different λ values along with specific applications in Chapter 6 and Chapter 7.

$A_j^k(\mathbb{H}_{j-1})$ is the (expected) Fisher information matrix after the j^{th} observation, given the history of the first $(j-1)$ patients, $\mathbb{H}_{j-1}$, and the assumption that j^{th} patient will be treated by treatment sequence k. For example, $A_j^{AA}(\mathbb{H}_{j-1})$ indicates the expected Fisher information matrix, which is calculated using the first j patients and involves the data from the first $(j-1)$ patients and uses the assumption that the j^{th} patient will be treated by the treatment sequence AA.

The unknown parameters in the expected Fisher information matrix, $A_j^k(\mathbb{H}_{j-1})$, can be estimated based on the observed data from the first $(j-1)$ patients. After using the plug-in method, the estimated Fisher information matrix is obtained, denoted by $\widehat{A}_j^k(\mathbb{H}_{j-1})$. For example, under the model (3.4) in Chapter 3, the estimated Fisher information matrix for $k = AA$ is

$$\widehat{A}_j^{AA}(\mathbb{H}_{j-1}) = \sum_{k \in \mathbb{H}_{j-1}} N_k \mathbf{X}_k^T \widehat{\mathbf{C}}^{-1} \mathbf{X}_k + \mathbf{X}_{AA}^T \widehat{\mathbf{C}}^{-1} \mathbf{X}_{AA} = \widehat{A}_{j-1} + \mathbf{X}_{AA}^T \widehat{\mathbf{C}}^{-1} \mathbf{X}_{AA} \tag{5.6}$$

$\Theta(\cdot)$ is an optimality criteria function such as the determinant (D-optimality), the trace (A-optimality) or the maximum eigenvalue (E-optimality) of the information matrix. Under the D-optimality criteria, treatment sequence $k^{(a)}$ satisfies

$$\Theta(\widehat{A_j^{k^{(a)}}}(\mathbb{H}_{j-1})) = \max_k |\widehat{A_j^k}(\mathbb{H}_{j-1}))| \tag{5.7}$$

Treatment sequence $k^{(b)}$ is the best treatment sequence based on the observed data $\mathbb{H}_{j-1}$ under the evaluation function.

Step 4: Repeat steps 2 to 3 until all patients have been allocated.

Note that the above adaptive approach is applicable for both discrete and continuous responses, under suitable model assumptions.

In the next two chapters, we will discuss the implementation of the new allocation rule to trials with dichotomous responses and continuous responses, respectively.

5.4 Conclusion

In this chapter, we proposed a response-adaptive design strategy for constructing repeated measurement designs to increase both the estimation precision and the proportion of patients treated by better treatment sequences measured by a predefined evaluation function. This strategy improves the current response-adaptive designs, which have been constructed under a single objective. In addition, this new allocation rule is applicable to both continuous and dichotomous responses, and applicable to any type of repeated measurement design models. We will discuss the applications of the new allocation rule in Chapters 6 and 7.

Table 5.1: Dichotomous Response Data Example

Subject ID	Treatment Sequence	Resp 1	Resp 2
1	AA	1	1
2	AB	0	0
3	AA	1	1
4	BA	0	1
5	BB	1	0
6	BA	0	0
7	AB	1	1
8	AA	0	1
9	BA	0	1
10	BB	1	0

Table 5.2: Continuous Response Data Example

Subject ID	Treatment Sequence	Resp 1	Resp 2
1	AA	37.5	37
2	AB	37.5	38
3	AA	37.8	37
4	BA	37.5	37
5	BB	38.5	39
6	BA	38.5	37.5
7	AB	37.5	37.5
8	AA	38	37
9	BA	38.5	37.5
10	BB	39	38.5

Chapter 6

Adaptive Repeated Measurement Designs for Dichotomous Responses

In this chapter, we implement the adaptive allocation rule proposed in Chapter 5 for repeated measurement designs with dichotomous responses. The evaluation function is defined in the spirit of the play-the-winner rule. We provide the detailed allocation rule for constructing adaptive two-treatment two-period repeated measurement designs, and then extend it to two-treatment p-period repeated measurement designs. Simulations are carried out to study the performance of the allocation rule.

6.1 Adaptive Two-Treatment Repeated Measurement Design

6.1.1 Allocation Rule for Two-Period Repeated Measures Data

Consider a two-treatment two-period repeated measurement design, where investigators want to compare the effectiveness of two drugs, A and B. There are four possible treatment sequences AA, AB, BA and BB. Suppose that N patients were randomly selected from a well-defined population, and the first i patients were assigned using a design suggested in the literature, a completely randomized design, for example.

According to our new allocation rule, in order to allocate the $(i+1)^{th}$

patient, the observed information based on the data available from the first i patients should be calculated. Up to the i^{th} patient, let N_{1i}, N_{2i}, N_{3i} and N_{4i} be the number of patients who were allocated to treatment sequence AA, AB, BA and BB, respectively, with $\sum_{k=1}^{4} N_{ki} = i$. Let us further denote that S_{1Ai} is the number of successes by A in the first period; S_{2Ai} is the number of successes by A in the second period; S_{1Bi} is the number of successes by B in the first period; and S_{2Bi} is the number of successes by B in the second period.

Obviously, S_{1Ai} comes from two possible sources: 1) patients who received treatment sequence AA and succeeded in the first period, denoted by S_{1Ai}^{AA}, and 2) those who received treatment sequence AB and succeeded in the first period, denoted by S_{1Ai}^{AB}. Hence, $S_{1Ai} = S_{1Ai}^{AA} + S_{1Ai}^{AB}$.

Similarly, we have $S_{2Ai} = S_{2Ai}^{AA} + S_{2Ai}^{BA}$, $S_{1Bi} = S_{1Bi}^{BA} + S_{1Bi}^{BB}$ and $S_{2Bi} = S_{2Bi}^{AB} + S_{2Bi}^{BB}$, where S_{2Ai}^{AA} denotes the number of successes by AA in the second period; S_{2Ai}^{BA} denotes the number of successes by BA in the second period; S_{1Bi}^{BA} denotes the number of successes by BA in the first period; S_{1Bi}^{BB} denotes the number of successes by BB in the first period; S_{2Bi}^{AB} denotes the number of successes by AB in the second period; and S_{2Bi}^{BB} denotes the number of successes by BB in the second period.

According to *Example 1* in Chapter 5, an evaluation function for a given treatment sequence can be defined as

$$\begin{aligned} g_{AA,i} &= \frac{S_{1Ai}^{AA} + S_{2Ai}^{AA}}{N_{1i}} \\ g_{AB,i} &= \frac{S_{1Ai}^{AB} + S_{2Bi}^{AB}}{N_{2i}} \\ g_{BA,i} &= \frac{S_{1Bi}^{BA} + S_{2Ai}^{BA}}{N_{3i}} \\ g_{BB,i} &= \frac{S_{1Bi}^{BB} + S_{2Bi}^{BB}}{N_{4i}} \end{aligned} \tag{6.1}$$

The likelihood up to the i^{th} patient is

$$\begin{aligned} L_i \propto\ & \pi_1^{S_{1Ai}} (1-\pi_1)^{(N_{1i}+N_{2i}-S_{1Ai})} \pi_2^{S_{1Bi}} (1-\pi_2)^{(N_{3i}+N_{4i}-S_{1Bi})} \times \\ & \pi_3^{S_{2Ai}} (1-\pi_3)^{(N_{1i}+N_{3i}-S_{2Ai})} \pi_4^{S_{2Bi}} (1-\pi_4)^{(N_{2i}+N_{4i}-S_{2Bi})} \end{aligned} \tag{6.2}$$

where π_1 is the success probability of A in the first period; π_2 is the success probability of B in the first period; π_3 is the success probability of A in the second period; and π_4 is the success probability of B in the second period. Here, for simplicity, we assume there are no covariates, and π_1, π_2, π_3 and π_4 are fixed but unknown parameters of interest.

The log-likelihood l_i is then

$$\begin{aligned} l_i \propto\ & S_{1Ai}\log\pi_1 + (N_{1i} + N_{2i} - S_{1Ai})\log(1-\pi_1) \\ +\ & S_{1Bi}\log\pi_2 + (N_{3i} + N_{4i} - S_{1Bi})\log(1-\pi_2) \\ +\ & S_{2Ai}\log\pi_3 + (N_{1i} + N_{3i} - S_{2Ai})\log(1-\pi_3) \\ +\ & S_{2Bi}\log\pi_4 + (N_{2i} + N_{4i} - S_{2Bi})\log(1-\pi_4) \end{aligned} \tag{6.3}$$

The (expected) Fisher information matrix up to the i^{th} patient is

$$A_i = -E\begin{pmatrix} \frac{\partial^2 l_i}{\partial\pi_1^2} & \frac{\partial^2 l_i}{\partial\pi_1\partial\pi_2} & \frac{\partial^2 l_i}{\partial\pi_1\partial\pi_3} & \frac{\partial^2 l_i}{\partial\pi_1\partial\pi_4} \\ \frac{\partial^2 l_i}{\partial\pi_2\partial\pi_1} & \frac{\partial^2 l_i}{\partial\pi_2^2} & \frac{\partial^2 l_i}{\partial\pi_2\partial\pi_3} & \frac{\partial^2 l_i}{\partial\pi_2\partial\pi_4} \\ \frac{\partial^2 l_i}{\partial\pi_3\partial\pi_1} & \frac{\partial^2 l_i}{\partial\pi_3\partial\pi_2} & \frac{\partial^2 l_i}{\partial\pi_3^2} & \frac{\partial^2 l_i}{\partial\pi_3\partial\pi_4} \\ \frac{\partial^2 l_i}{\partial\pi_4\partial\pi_1} & \frac{\partial^2 l_i}{\partial\pi_4\partial\pi_2} & \frac{\partial^2 l_i}{\partial\pi_4\partial\pi_3} & \frac{\partial^2 l_i}{\partial\pi_4^2} \end{pmatrix} \tag{6.4}$$

In particular, we have

$$\frac{\partial^2 l_i}{\partial\pi_r\partial\pi_r'} = 0,$$

for any $r \neq r'$ and $r, r' = 1, 2, 3, 4$.

Therefore, the expected Fisher information matrix becomes

$$\begin{aligned} A_i = diag(\ & E(\tfrac{S_{1Ai}}{\pi_1^2} + \tfrac{N_{1i}+N_{2i}-S_{1Ai}}{(1-\pi_1)^2}), E(\tfrac{S_{1Bi}}{\pi_2^2} + \tfrac{N_{3i}+N_{4i}-S_{1Bi}}{(1-\pi_2)^2}), \\ & E(\tfrac{S_{2Ai}}{\pi_3^2} + \tfrac{N_{1i}+N_{3i}-S_{2Ai}}{(1-\pi_3)^2}), E(\tfrac{S_{2Bi}}{\pi_4^2} + \tfrac{N_{2i}+N_{4i}-S_{2Bi}}{(1-\pi_4)^2})) \end{aligned} \tag{6.5}$$

If the treatment sequence AA is assigned to the $(i+1)^{th}$ patient, at the $(i+1)^{th}$ stage, the number of patients who were allocated to treatment sequence AA, $N_{1,i+1}$, will increase by 1, and the number of patients who were allocated to other treatment sequences will remain the same. Therefore, we have

$$E(N_{k,i+1}|AA) = \begin{cases} N_{ki}+1 & \text{if} \quad \text{k=1} \\ N_{ki} & \text{if} \quad \text{k=2, 3, 4} \end{cases} \tag{6.6}$$

Also, the expected number of successes by treatment A/B in the first/second period up to the $(i+1)^{th}$ patient, given the history of $\mathbb{H}_i$ and the assumption that the $(i+1)^{th}$ patient will be allocated to treatment sequence AA, will become

$$\begin{aligned} E(S_{1A,i+1}|AA) &= (S_{1Ai}+1)\cdot\pi_1 + S_{1Ai}\cdot(1-\pi_1) = S_{1Ai}+\pi_1 \\ E(S_{2A,i+1}|AA) &= (S_{2Ai}+1)\cdot\pi_3 + S_{2Ai}\cdot(1-\pi_3) = S_{2Ai}+\pi_3 \\ E(S_{1B,i+1}|AA) &= S_{1Bi} \\ E(S_{2B,i+1}|AA) &= S_{2Bi} \end{aligned} \tag{6.7}$$

Then, by plugging Equation 6.6 and 6.7 into Equation 6.5, we have

$$\begin{aligned} A_{i+1}^{AA}(\mathbb{H}_i) = diag(& \tfrac{S_{1Ai}+\pi_1}{\pi_1^2} + \tfrac{N_{1i}+N_{2i}+1-(S_{1Ai}+\pi_1)}{(1-\pi_1)^2}, \tfrac{S_{1Bi}}{\pi_2^2} + \tfrac{N_{3i}+N_{4i}-S_{1Bi}}{(1-\pi_2)^2}, \\ & \tfrac{S_{2Ai}+\pi_3^2}{\pi_3^2} + \tfrac{N_{1i}+N_{3i}+1-(S_{2Ai}+\pi_3)}{(1-\pi_3)^2}, \tfrac{S_{2Bi}}{\pi_4^2} + \tfrac{N_{2i}+N_{4i}-S_{2Bi}}{(1-\pi_4)^2}) \end{aligned} \tag{6.8}$$

If the treatment sequence for the $(i+1)^{th}$ patient is AB, we have the similar derivation results:

$$E(N_{k,i+1}|AB) = \begin{cases} N_{ki}+1 & \text{if} \quad \text{k=2} \\ N_{ki} & \text{if} \quad \text{k=1, 3, 4} \end{cases} \tag{6.9}$$

and

$$\begin{aligned} E(S_{1A,i+1}|AB) &= S_{1Ai}+\pi_1 \\ E(S_{2A,i+1}|AB) &= S_{2Ai} \\ E(S_{1B,i+1}|AB) &= S_{1Bi} \\ E(S_{2B,i+1}|AB) &= S_{2Bi}+\pi_4 \end{aligned} \tag{6.10}$$

Also,

$$\begin{aligned} A_{i+1}^{AB}(\mathbb{H}_i) = diag(& \tfrac{S_{1Ai}+\pi_1}{\pi_1^2} + \tfrac{N_{1i}+N_{2i}+1-(S_{1Ai}+\pi_1)}{(1-\pi_1)^2}, \tfrac{S_{1Bi}}{\pi_2^2} + \tfrac{N_{3i}+N_{4i}-S_{1Bi}}{(1-\pi_2)^2}, \\ & \tfrac{S_{2Ai}}{\pi_3^2} + \tfrac{N_{1i}+N_{3i}-S_{2Ai}}{(1-\pi_3)^2}, \tfrac{S_{2Bi}+\pi_4}{\pi_4^2} + \tfrac{N_{2i}+N_{4i}+1-(S_{2Bi}+\pi_4)}{(1-\pi_4)^2}) \end{aligned} \tag{6.11}$$

Similarly, if the treatment sequence for the $(i+1)^{th}$ patient is BA, then

$$E(N_{k,i+1}|BA) = \begin{cases} N_{ki}+1 & \text{if} \quad \text{k=3} \\ N_{ri} & \text{if} \quad \text{k=1, 2, 4} \end{cases} \tag{6.12}$$

and

$$\begin{aligned} E(S_{1A,i+1}|BA) &= S_{1Ai} \\ E(S_{2A,i+1}|BA) &= S_{2Ai} + \pi_3 \\ E(S_{1B,i+1}|BA) &= S_{1Bi} + \pi_2 \\ E(S_{2B,i+1}|BA) &= S_{2Bi} \end{aligned} \tag{6.13}$$

Then

$$\begin{aligned} A^{BA}_{i+1}(\mathbb{H}_i) = diag(& \tfrac{S_{1Ai}}{\pi_1^2} + \tfrac{N_{1i}+N_{2i}-S_{1Ai}}{(1-\pi_1)^2}, \tfrac{S_{1Bi}+\pi_2}{\pi_2^2} + \tfrac{N_{3i}+N_{4i}+1-(S_{1Bi}+\pi_2)}{(1-\pi_2)^2} \\ & \tfrac{S_{2Ai}+\pi_3}{\pi_3^2} + \tfrac{N_{1i}+N_{3i}+1-(S_{2Ai}+\pi_3)}{(1-\pi_3)^2}, \tfrac{S_{2Bi}}{\pi_4^2} + \tfrac{N_{2i}+N_{4i}-S_{2Bi}}{(1-\pi_4)^2}) \end{aligned} \tag{6.14}$$

Finally, if the treatment sequence for the $(i+1)^{th}$ patient is BB, then

$$E(N_{k,i+1}|BB) = \begin{cases} N_{ki}+1 & \text{if} \quad \text{k=4} \\ N_{ri} & \text{if} \quad \text{k=1, 2, 3} \end{cases} \tag{6.15}$$

and

$$\begin{aligned} E(S_{1A,i+1}|BB) &= S_{1Ai} \\ E(S_{2A,i+1}|BB) &= S_{2Ai} \\ E(S_{1B,i+1}|BB) &= S_{1Bi} + \pi_2 \\ E(S_{2B,i+1}|BB) &= S_{2Bi} + \pi_4 \end{aligned} \tag{6.16}$$

Then

$$\begin{aligned} A^{BB}_{i+1}(\mathbb{H}_i) = diag(& \tfrac{S_{1Ai}}{\pi_1^2} + \tfrac{N_{1i}+N_{2i}-S_{1Ai}}{(1-\pi_1)^2}, \tfrac{S_{1Bi}+\pi_2}{\pi_2^2} + \tfrac{N_{3i}+N_{4i}+1-(S_{1Bi}+\pi_2)}{(1-\pi_2)^2} \\ & \tfrac{S_{2Ai}}{\pi_3^2} + \tfrac{N_{1i}+N_{3i}-S_{2Ai}}{(1-\pi_3)^2}, \tfrac{S_{2Bi}+\pi_4}{\pi_4^2} + \tfrac{N_{2i}+N_{4i}+1-(S_{2Bi}+\pi_4)}{(1-\pi_4)^2}) \end{aligned} \tag{6.17}$$

Result 6.1.1: *The unknown parameters π_1, π_2, π_3 and π_4 in equations 6.8, 6.11, 6.14 and 6.17, are estimated using the data up to the i^{th} patients, as below.*

$$\begin{aligned} \hat{\pi}_1 &= \frac{S_{1Ai}}{N_{1i}+N_{2i}} \\ \hat{\pi}_2 &= \frac{S_{1Bi}}{N_{3i}+N_{4i}} \\ \hat{\pi}_3 &= \frac{S_{2Ai}}{N_{1i}+N_{3i}} \\ \hat{\pi}_4 &= \frac{S_{2Bi}}{N_{2i}+N_{4i}} \end{aligned} \tag{6.18}$$

Proof: According to the log-likelihood function given in Equation 6.3, we have

$$\frac{\partial l_i}{\partial \pi_r} = \frac{a_r}{\pi_r} - \frac{m_r - a_r}{1 - \pi_r}, \qquad r = 1, 2, 3, 4 \tag{6.19}$$

where $a_1 = S_{1Ai}$, $a_2 = S_{1Bi}$, $a_3 = S_{2Ai}$, $a_4 = S_{2Bi}$, $m_1 = N_{1i} + N_{2i}$, $m_2 = N_{3i} + N_{4i}$, $m_3 = N_{1i} + N_{3i}$ and $m_4 = N_{2i} + N_{4i}$.

If we let Equation 6.19 equal zero, then we have

$$\pi_r = \frac{a_r}{m_r}, \qquad r = 1, 2, 3, 4 \tag{6.20}$$

Therefore Equation 6.18 holds. □

For a given value of λ, under the optimality criteria of the estimated information matrix, $\Theta(.)$, we allocate a treatment sequence k^* to the $(i+1)^{th}$ patient by maximizing the Λ as defined earlier in Chapter 5, Equation 5.5, i.e.,

$$\Lambda = \lambda \frac{\Theta(\widehat{A^k_{i+1}}(\mathbb{H}_i))}{\Theta(\widehat{A^{k^{(a)}}_{i+1}}(\mathbb{H}_i))} + (1 - \lambda)\frac{g_{k,i}}{g_{k^{(b)},i}}, \; k = AA, AB, BA \text{ or } BB$$

where $k^{(a)}$ is a treatment sequence that maximizes $\Theta(A^k_{i+1}(\mathbb{H}_i))$, and $k^{(b)}$ is a treatment sequence that maximizes $g_{k,i}$.

We continue applying the same technique until all N patients are assigned.

6.1.2 Allocation Rule for Three-Period Repeated Measures Data

In three-period repeated measurement designs comparing the effectiveness of two drugs, drug A and drug B, there are $8(= 2^3)$ possible treatment sequences. Let T be the set of all possible treatment sequences, i.e., $T = \{AAA, AAB, ABA, ABB, BBB, BBA, BAB, BAA\}$, and let T_k be the k^{th} element/treatment sequence in the set T, $k = 1, 2, \ldots, 8$. For simplicity, we sometimes use the treatment sequence k to represent the treatment sequence T_k. As in Section 6.1.1, we assume that N patients were randomly selected

from a well-defined population, and the first i patients were assigned using the optimal design suggested in the literature, a completely randomized design, for example.

Let π_r (where $r = 1, 2$ and 3) be the success probability of treatment A in the r^{th} period and π_r (where $r = 4, 5$ and 6) be the success probability of treatment B in the $(r-3)^{th}$ period. Up to the i^{th} patient, N_{ki} denotes the number of subjects receiving the treatment sequence k, where $k = 1, 2, \ldots, 8$. S_{qti} denotes the number of successes of treatment t in the q^{th} period, where $q = 1, 2$ and 3 and $t = A$ or B. For example, S_{1Ai} represents the number of successes of A in the first period up to the i^{th} patients. Let $S_i = (S_{1Ai}, S_{2Ai}, S_{3Ai}, S_{1Bi}, S_{2Bi}, S_{3Bi})^T$, and let $S_i[r]$ be the r^{th} element of S_i, $r = 1, 2, \ldots, 6$.

Under the same assumptions as in Section 6.1.1, the likelihood function up to the i^{th} patient is then

$$L_i = \prod_{r=1}^{6} \pi_r^{S_i[r]}(1-\pi_r)^{(N_{L_i}[r]-S_i[r])} \tag{6.21}$$

where $N_{L_i}[1] = N_{1i} + N_{2i} + N_{3i} + N_{4i}$, $N_{L_i}[2] = N_{1i} + N_{2i} + N_{7i} + N_{8i}$, $N_{L_i}[3] = N_{1i} + N_{3i} + N_{6i} + N_{8i}$, $N_{L_i}[4] = N_{5i} + N_{6i} + N_{7i} + N_{8i}$, $N_{L_i}[5] = N_{3i} + N_{4i} + N_{5i} + N_{6i}$, and $N_{L_i}[6] = N_{2i} + N_{4i} + N_{5i} + N_{7i}$, with $N_{L_i} = (N_{L_i}[1], N_{L_i}[2], \ldots, N_{L_i}[6])^T$.

The log-likelihood function l_i becomes

$$l_i \propto \sum_{r=1}^{6} (S_i[r] log \pi_r + (N_{L_i}[r] - S_i[r]) log(1-\pi_r)) \tag{6.22}$$

The expected Fisher information matrix up to the i^{th} patient, A_i, which is a 6×6 diagonal matrix, becomes

$$A_i = Diag\left(E(\frac{S_i[r]}{\pi_r^2} + \frac{N_{L_i}[r] - S_i[r]}{(1-\pi_r)^2})\right) \tag{6.23}$$

where $r = 1, 2, \ldots, 6$.

Similar to the *Result 6.1.1* for adaptive two-treatment two-period RMDs, the unknown parameters π_r at the current stage i are estimated as below

$$\hat{\pi_r} = \frac{S_i[r]}{N_{L_i}[r]} \tag{6.24}$$

where $r = 1, 2, \ldots, 6$

Result 6.1.2: *In two-treatment three-period repeated measurement designs with 8 possible treatment sequences:* $T = \{AAA, AAB, ABA, ABB, BBB, BBA, BAB, BAA\}$, *the expected Fisher information matrix on the* $(i+1)^{th}$ *stage, given the history of* $\mathbb{H}_i$ *under the assumption that the* $(i+1)^{th}$ *patient receiving treatment sequence* T_k, *where* T_k *is the* k^{th} *element in the treatment sequence set, and* $k = 1, 2, \ldots, 8$, *is*

$$A_{i+1}^{T_k}(\mathbb{H}_i) = Diag\left(\frac{S_{i+1}[r]}{\pi_r^2} + \frac{N_{L_{i+1}}[r] - S_{i+1}[r]}{(1-\pi_r)^2}\right) \tag{6.25}$$

where $S_{i+1} = S_i + Diag(\boldsymbol{\pi} \times \boldsymbol{u}_k)$, $\boldsymbol{\pi} = (\pi_1, \pi_2, ..., \pi_6)^T$, $N_{L_{i+1}} = N_{L_i} + \boldsymbol{u}_k^T$, S_i *and* N_{L_i} *are defined in Equation 6.21,* $S_{i+1}[r]$ *and* $N_{L_{i+1}}[r]$ *are the* r^{th} *element of* S_{i+1} *and* $N_{L_{i+1}}$, *respectively, and* $\boldsymbol{u}_k$ *is the* k^{th} *row vector of the matrix* U *defined below*

$$U = (\boldsymbol{u}_1, \boldsymbol{u}_2, ..., \boldsymbol{u}_8)^T = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 \end{pmatrix}$$

$k = 1, 2, \ldots, 8$, $r = 1, 2, \ldots, 6$.

Proof: First let us consider the case when $k = 1$, that is, the $(i+1)^{th}$ patient will receive treatment sequence $AAA(=T_1)$.

Similar to the case for adaptive two-treatment two-period RMDs in Section 6.1.1, the expected number of patients receiving the treatment sequence k up to the $(i+1)^{th}$ patient, given the history $\mathbb{H}_i$ and the assumption that the $(i+1)^{th}$ patient will be allocated to treatment sequence AAA will become

$$E(N_{k,i+1}|AAA) = \begin{cases} N_{ki} + 1 & k = 1 \\ N_{ki} & k = 2, 3, \ldots, 8 \end{cases}$$

and the expected number of successes by treatment A/B in the $1^{st}/2^{nd}/3^{rd}$ period up to the $(i+1)^{th}$ patient, given the history $\mathbb{H}_i$ and the assumption that the $(i+1)^{th}$ patient will be allocated to treatment sequence AAA will become

$$E\left(\begin{pmatrix} S_{1A,i+1} \\ S_{2A,i+1} \\ S_{3A,i+1} \\ S_{1B,i+1} \\ S_{2B,i+1} \\ S_{3B,i+1} \end{pmatrix} | AAA\right)$$

$$= \begin{pmatrix} (S_{1Ai}+1)\pi_1 + S_{1Ai}(1-\pi_1) \\ (S_{2Ai}+1)\pi_2 + S_{2Ai}(1-\pi_2) \\ (S_{3Ai}+1)\pi_3 + S_{3Ai}(1-\pi_3) \\ S_{1Bi} \\ S_{2Bi} \\ S_{3Bi} \end{pmatrix} = \begin{pmatrix} S_{1Ai}+\pi_1 \\ S_{2Ai}+\pi_2 \\ S_{3Ai}+\pi_3 \\ S_{1Bi} \\ S_{2Bi} \\ S_{3Bi} \end{pmatrix} \quad (6.26)$$

Let

$$S_{i+1}[r] = \begin{cases} S_{rAi}+\pi_r & r=1,2,3 \\ S_{(r-3)Bi} & r=4,5,6 \end{cases}$$

and

$$N_{L_{i+1}}[r] = N_{L_i}[r] + \boldsymbol{u}_{AAA}[r], r = 1,2,\ldots 6$$

with

$$\boldsymbol{u}_{AAA} = (\boldsymbol{u}_{AAA}[1],\ldots,\boldsymbol{u}_{AAA}[6])^T = (1,1,1,0,0,0)^T$$

Therefore, we have

$$A_{i+1}^{AAA}(\mathbb{H}_i) = Diag(\frac{S_{i+1}[r]}{\pi_r^2} + \frac{N_{L_{i+1}}[r] - S_{i+1}[r]}{(1-\pi_r)^2})$$

Similarly, we can show that Equation 6.25 holds true for other treatment sequences ($k = 2,3,\ldots,8$). □

Similar to the result for adaptive two-treatment two-period RMDs, the evaluation function (Equation 6.1) for treatment sequence k up to the i^{th} patient is given below

$$g_{ki} = \frac{\boldsymbol{u}_k \times S_i}{N_{ki}} \quad (6.27)$$

where $k = 1,2,\ldots,8$, and $\boldsymbol{u}_k$ is defined in *Result 6.1.2.*

For a given value of λ, under the optimality criteria of the estimated information matrix, $\Theta(.)$, we choose a treatment sequence k^* to the $(i+1)^{th}$ patient by maximizing Λ, defined in Equation 5.5, i.e.,

$$\Lambda = \lambda \frac{\Theta(\widehat{A_{i+1}^k}(\mathbb{H}_i))}{\Theta(\widehat{A_{i+1}^{k^{(a)}}}(\mathbb{H}_i))} + (1-\lambda)\frac{g_{k,i}}{g_{k^{(b)},i}},$$

where $\Theta(.)$, $k^{(a)}$ and $k^{(b)}$ are defined the same as before.

6.1.3 Simulation Study

In this section, we first apply the allocation rule described in Section 6.1.1 to construct two-treatment two-period response-adaptive RMDs, and then extend it to two-treatment three-period response-adaptive RMDS.

Two-treatment Two-period Response-Adaptive RMDs

Suppose at the initial stage, four patients were assigned, one for each type of treatment sequence. We then consider how to allocate the rest of the patients adaptively.

To assess the efficiency of an adaptive design, a matrix of mean squared error was computed

$$MSE = E[(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T] \tag{6.28}$$

In the simulation study, the **MSE** is estimated by

$$MSE = \sum_{b=1}^{B} (\widehat{\boldsymbol{\theta}}^{(b)} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}^{(b)} - \boldsymbol{\theta})^T / B \tag{6.29}$$

where $\widehat{\boldsymbol{\theta}}^{(b)}$ is the MLE obtained in the b^{th} simulation run, B is the total number of simulations.

Denote MSE_1 as the matrix of mean squared error for the proposed adaptive design and MSE_0 for the reference design. Based on $A-$, $D-$ or $E-$optimality criteria (Kiefer 1975), the relative efficiency (RE) of the adaptive design compared with the reference design is defined as below, respectively

$$\begin{aligned} RE_A &= \frac{\text{trace}(MSE_0)}{\text{trace}(MSE_1)} \\ RE_D &= \frac{|MSE_0|}{|MSE_1|} \\ RE_E &= \frac{\text{maxeigenvalue}(MSE_0)}{\text{maxeigenvalue}(MSE_1)} \end{aligned} \tag{6.30}$$

When $RE = a > 1$, the adaptive design is $(a-1) \times 100\%$ more efficient than the reference design. When $RE = a < 1$, the adaptive design is only $a \times 100\%$ as efficient as the reference design.

We investigated the performance of the allocation rule under various conditions by choosing combinations of adjusted weight (λ) and total number of patients in the study (N). Five values of λ ($1, 0.9, 0.7, 0.3$ and 0), and four values of N ($10, 20, 40$ and 100) were used. When $\lambda = 1$, we only consider the objective of increasing the estimation precision, i.e., maximizing the information matrix and this will result in the usual response adaptive designs as considered by others (Kushner, 2003). When $\lambda = 0$, the only objective of the design is to increase the proportion of patients assigned to a better treatment as considered in designs for the typical play-the-winner rule (Zelen, 1969). When $0 < \lambda < 1$, both objectives are taken into consideration, and the adaptive design balances these two objectives according to the specified goal. In this section, 1,000 simulation data were generated for each condition. The R program code is available upon request.

The Case of Equal Success Probability

We first consider the situation when treatments A and B perform equally well in both periods 1 and 2, i.e., $\pi_1 = \pi_2 = \pi_3 = \pi_4$. Table 6.1 summarizes the estimated average number of patients receiving each treatment sequence based on the 1,000 simulation results with $\pi_i = 0.5$, $i = 1, 2, 3, 4$. We can see that in this case, for all combinations of N and λ, an approximately equal number of patients will be assigned to each of the four treatment sequences. As the total number of patients increases, the distribution of the allocation is skewed a little, it is due to the limited data available in the initial stage. The estimation of the success probabilities of treatments in different periods, $\hat{\pi}_i$, and the precision of the estimations are displayed in Table 6.2 and Table 6.3.

Table 6.2 presents the mean and standard deviation (SD) of the 1,000 simulation samples. It shows that, for a fixed value of λ, when the total number of patients enrolled in the study increases, the standard deviation of the estimation decreases. When $\lambda < 1$, the estimated success probabilities are systematically lower than the true values. It is because some simulation samples produce extremely small (close to zero) estimations of the success probabilities. Therefore, mean and standard deviation may not be good

measures to characterize the center and the variability of the estimates when $\lambda < 1$.

Table 6.3 presents the median and median absolute deviation (MAD) of the estimates from the 1,000 simulation data samples, where MAD is defined as

$$\text{MAD} = \text{median}(|x_i - \tilde{x}|) \tag{6.31}$$

where $\tilde{x}$ is the median of the data. It shows that the median of the 1,000 simulated results are close to the true values. MAD measures a variation of the average absolute deviation. It is also less affected by extremes in the tail. Not surprisingly, for a fixed value of λ, when the total number of patients enrolled in the study increases, the precision of the estimation also increases. For a fixed value of N, when λ decreases, the MAD increases, that is, the precision of the estimation decreases as expected. The design with $\lambda = 1$ produces estimations with better precision than the designs with $\lambda < 1$, as expected.

Table 6.4 summarizes the characteristics of the mean squared error (MSE) of each design under the A-, D- and E-optimality, respectively. We find the similar patterns as observed in Tables 6.2 and 6.3 as a function of N and λ.

Figures 6.1, 6.2 and 6.3 illustrate the relative efficiency between the adaptive designs with $\lambda < 1$ and the reference design (the design with $\lambda = 1$) under the A-, D- and E-optimality, respectively. Once again, we can see that the design with $\lambda = 1$ produces a more precise estimation than designs with $\lambda < 1$. This is as expected, and it is because more weight is given to considering the performance of the treatments rather than the precision of the estimates as the value of λ decreases. In addition, the simulation demonstrates that the adaptive design with λ close to 1 both provides estimates with relatively high precision and favors the allocation results to more effective treatments.

The Case of Unequal Success Probability

Similarly, we consider the situation when the success probabilities, $\pi_i s$, are not all the same, for example, $\pi_1 = 0.6$, $\pi_2 = 0.3$, $\pi_3 = 0.7$ and $\pi_4 = 0.5$. For each combination of λ and N, 1,000 data samples were simulated.

Table 6.5 shows that when $\lambda = 1$, on average, adaptive designs assign approximately equal numbers of patients to each treatment sequence; however when $\lambda < 1$, the adaptive design constructed under the new allocation rule successfully assigns more patients to the better treatment sequence, which is treatment sequence AA in this case, even when sample size is small (e.g. $N = 10$). In addition, as N increases, the proportion of patients receiving the best treatment (AA) increases; whereas the proportion of patients receiving the worst treatment (BB) decreases.

Tables 6.6 and 6.7 report the mean (standard deviation) and the median (median absolute deviation) of the point estimates of the success probabilities from the 1,000 simulated data samples, respectively. They both show that the spread of the estimate of π_i decreases when the total number of patients increases.

Table 6.8 summarizes the characteristics of the mean squared error (MSE) of each design under the A-, D- and E-optimality, respectively. Figures 6.4, 6.5 and 6.6 illustrate the relative efficiency between the adaptive designs with $\lambda < 1$ and the reference design (the design with $\lambda = 1$) under the A-, D- and E-optimality, respectively. Similar to the equal success probability situation, although the design with $\lambda = 1$ has the highest efficiency in terms of MSE, the design with $\lambda = 0.9$ takes treatment performances into account and offers relatively high estimation precision.

Two-treatment Three-period Response-Adaptive RMDs

Similar to the simulation study carried out for adaptive two-treatment two-period RMDs, we then apply the allocation strategy to construct two-treatment three-period RMDs adaptively. Suppose at the initial stage, eight patients were assigned, one for each type of treatment sequence. We then consider how to allocate the rest of the patients adaptively. As before, we consider the cases when $\lambda = 1, 0.9, 0.7, 0.3$ and 0, and $N = 40, 80$ and 120, and we assess the efficiency of a design using the mean squared errors.

The Case of Equal Success Probability

Firstly, we consider the case when all treatments perform equally well. Assume that $\pi_i = 0.5$, $i = 1, 2, \ldots, 6$. Table 6.9 summarizes the estimated

average number of patients who received each of 8 possible treatment sequences based on the 1,000 simulated samples. It shows that when $\lambda < 1$, an approximately equal number of patients was assigned to each treatment sequence. When $\lambda = 1$, an equal number of patients was assigned to a treatment sequence and its dual treatment sequence, and treatment sequences ABB/BAA slightly have more patients than other treatment sequences. Note that the design with ABB/BAA is recommended by several researchers (Laska, Meisner and Kushner 1983, Kershner 1986) in the class of three-period designs with two treatments.

Tables 6.10 and 6.11 characterize the center and spread of the point estimates (based on Equation 6.24) of success probabilities, $\pi_i s$, using mean/SD and median/MAD respectively. Both tables clearly indicate that when the total number of patients involved in the study increases, the precision of the estimates increases accordingly.

Table 6.12 summarizes the mean squared error of estimates for $\boldsymbol{\theta}=(\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6)^T$ under the A-, D- and E-optimality, respectively, for each adaptive design. The smaller value indicates a design with more efficiency in terms of MSE. In addition, Figures 6.7, 6.8 and 6.9 illustrate the relative efficiency between the designs with $\lambda < 1$ and the reference design (the design with $\lambda = 1$) under the A-, D- and E-optimality, respectively. They demonstrate that the design with $\lambda = 1$ has the highest efficiency in terms of MSE. However, the design with $\lambda = 0.9$ offers relatively high precision of the estimation, and more importantly, it takes the treatment benefits into account.

The Case of Unequal Success Probability

Secondly, we consider the case when the success probabilities, $\pi_i s$, are not equal. Assume $\pi_1 = 0.5$, $\pi_2 = 0.6$, $\pi_3 = 0.7$, $\pi_4 = 0.5$, $\pi_5 = 0.4$ and $\pi_6 = 0.3$. The estimated average number of patients for each treatment sequence, based on the 1,000 simulation study, is given in Table 6.13; the center and spread of the point estimates of $\pi_i s$ are presented in Tables 6.14 and 6.15. In addition the assessment of the efficiency of adaptive designs with various combinations of λ and N are illustrated in Table 6.16, and Figures 6.10, 6.11 and 6.12.

To summarize the above, when $\lambda = 1$, an approximately equal number of subjects was assigned to a treatment sequence and its dual treatment sequence, and slightly more patients were given ABB/BAA, which is the optimal 3-period design for two treatments recommended by several researchers (Laska, Meisner and Kushner 1983, Kershner 1986). However, when $\lambda < 1$, adaptive designs assign more patients to the best treatment, AAA, and less subjects to the worst treatment, BBB, as λ decreases. In addition, when the total number of patients involved in the study increases, the precision of the estimates increases accordingly. The design with $\lambda = 1$ has the highest efficiency in terms of MSE. However, the designs with $\lambda < 1$ take the treatment advantage into account. In practice, these two objectives should be balanced out.

6.2 Adaptive Two-Treatment p-Period Repeated Measurement Design

Now we consider an adaptive two-treatment multiple-period repeated measurement design ($t = 2$ and $p > 2$). Let A and B denote the two different treatments. There are 2^p possible treatment sequences.

Suppose that N patients were randomly selected from a well-defined population, and the first i patients were assigned using the optimal design suggested in the literature, a completely randomized design, for example.

Let π_r, when $r = 1, 2, \ldots, p$, be the success probability of treatment A in the r^{th} period; when $r = p+1, p+2, \ldots, 2p$, π_r is the success probability of treatment B in the $(r-p)^{th}$ period.

Up to the i^{th} patient (or sometimes called the i^{th} stage), N_{ki} denotes the number of subjects receiving treatment sequence k, where $k = 1, 2, \ldots, 2^p$. $S_i = (S_{1Ai}, \ldots, S_{pAi}, S_{1Bi}, \ldots, S_{pBi})^T$, where S_{qti} denotes the number of successes of treatment t in the q^{th} period, where $q = 1, 2, \ldots, p$ and $t = A$ or B. For example, S_{1Ai} represents the number of successes of A in the first period at the i^{th} stage.

The likelihood function up to the i^{th} patient is then

$$L_i = \prod_{r=1}^{2p} \pi_r^{S_i[r]}(1-\pi_r)^{(N_{L_i}[r]-S_i[r])} \tag{6.32}$$

where $S_i[r]$ is the r^{th} element of S_i, and if $1 \leq r \leq p$, $N_{L_i}[r]$ denotes the total number of patients receiving treatment A in the r^{th} period; if $p+1 \leq r \leq 2p$, $N_{L_i}[r]$ denotes the total number of patients receiving treatment B in the $(r-p)^{th}$ period. Finally, let $N_{L_i} = (N_{L_i}[1], N_{L_i}[2], \ldots, N_{L_i}[2p])^T$

The log-likelihood function l_i becomes

$$l_i \propto \sum_{r=1}^{2p} (S_i[r] log \pi_r + (N_{L_i}[r] - S_i[r]) log(1-\pi_r)) \tag{6.33}$$

and for each r, $r = 1, 2, \ldots, 2p$, we have

$$\begin{aligned} \frac{\partial l_i}{\partial \pi_r} &= \frac{S_i[r]}{\pi_r} + \frac{N_{L_i}[r] - S_i[r]}{-(1-\pi_r)} \\ \frac{\partial^2 l_i}{\partial \pi_r^2} &= -\frac{S_i[r]}{\pi_r^2} - \frac{N_{L_i}[r] - S_i[r]}{(1-\pi_r)^2} \\ \frac{\partial^2 l_i}{\partial \pi_r \partial \pi_{r'}} &= 0, \text{for } r \neq r' \end{aligned} \tag{6.34}$$

The expected Fisher information matrix, A_i, which is a $2p \times 2p$ diagonal matrix, becomes

$$A_i = Diag\left(E(\frac{S_i[1]}{\pi_1^2} + \frac{N_{L_i}[1] - S_i[1]}{(1-\pi_1)^2}), \ldots, E(\frac{S_i[2p]}{\pi_{2p}^2} + \frac{N_{L_i}[2p] - S_i[2p]}{(1-\pi_{2p})^2})\right) \tag{6.35}$$

Similar to *Result 6.1.1*, the maximum likelihood estimation of unknown parameter π_r at the stage i is obtained as

$$\hat{\pi_r} = \frac{S_i[r]}{N_{L_i}[r]} \tag{6.36}$$

where $r = 1, 2, \ldots, 2p$.

Similar to *Result 6.1.2*, we have the following result for calculating the conditional expected information matrix for the next stage.

Result 6.2.1: *In two-treatment p-period repeated measurement designs, the expected information matrix on the* $(i+1)^{th}$ *stage, given the history of* $\mathbb{H}_i$

and the assumption that the $(i+1)^{th}$ *patient is receiving treatment sequence* k, *where* $k = 1, 2, \ldots, 2^p$, *is*

$$A_{i+1}^{k}(\mathbb{H}_i) = Diag(\frac{S_{i+1}[1]}{\pi_1^2} + \frac{N_{L_{i+1}}[1] - S_{i+1}[1]}{(1-\pi_1)^2}, \ldots, \frac{S_{i+1}[2p]}{\pi_{2p}^2} + \frac{N_{L_{i+1}}[2p] - S_{i+1}[2p]}{(1-\pi_{2p})^2}) \quad (6.37)$$

where

$$S_{i+1}[r] = S_i[r] + \alpha_r,\ r = 1, 2, \ldots, 2p$$

and

$$N_{L_{i+1}}[r] = N_{L_i}[r] + \beta_r,\ r = 1, 2, \ldots, 2p$$

where α_r *is the* r^{th} *element of* $Diag(\boldsymbol{\pi} \times \boldsymbol{u}_k)$, $\boldsymbol{\pi} = (\pi_1, \pi_2, ..., \pi_{2p})^T$, *and* β_r *is the* r^{th} *element of* $\boldsymbol{u}_k$, *and* $\boldsymbol{u}_k = (d(1,k), \ldots, d(2p,k))$. *If* $1 \leq r \leq p$, $d(r,k) = 1$ *if the treatment in the* r^{th} *period of the treatment sequence* k *is the treatment* A; $d(r,k) = 0$, *otherwise. If* $p+1 \leq r \leq 2p$, $d(r,k) = 1$ *if the treatment in the* $(r-p)^{th}$ *period of the treatment sequence* k *is the treatment* B; $d(r,k) = 0$, *otherwise. And* $k = 1, 2, \ldots, 2^p$, $r = 1, 2, \ldots, 2p$.

Based on the play-the-winner rule, the evaluation function for treatment sequence k at stage i is given below

$$g_{ki} = \frac{\boldsymbol{u}_k \times S_i}{N_{ki}}$$

where $k = 1, 2, \ldots, 2^p$.

For a given value of λ, under the optimality criteria of the estimated information matrix, $\Theta(.)$, we choose a treatment sequence k to the $(i+1)^{th}$ patient by maximizing the Λ

$$\Lambda = \lambda \frac{\Theta(\widehat{A_{i+1}^{k}}(\mathbb{H}_i))}{\Theta(\widehat{A_{i+1}^{k^{(a)}}}(\mathbb{H}_i))} + (1-\lambda)\frac{g_{k,i}}{g_{k^{(b)},i}},$$

where $k^{(a)}$ is a treatment sequence among 2^P possible treatment sequences that maximizes $\Theta(A_{i+1}^{k}(\mathbb{H}_i))$, and $k^{(b)}$ is a treatment sequence that maximizes $g_{k,i}$, where $k = 1, 2, \ldots, 2^p$. Continue the application until all N patients are assigned.

6.3 Adaptive t-Treatment p-Period Repeated Measurement Design

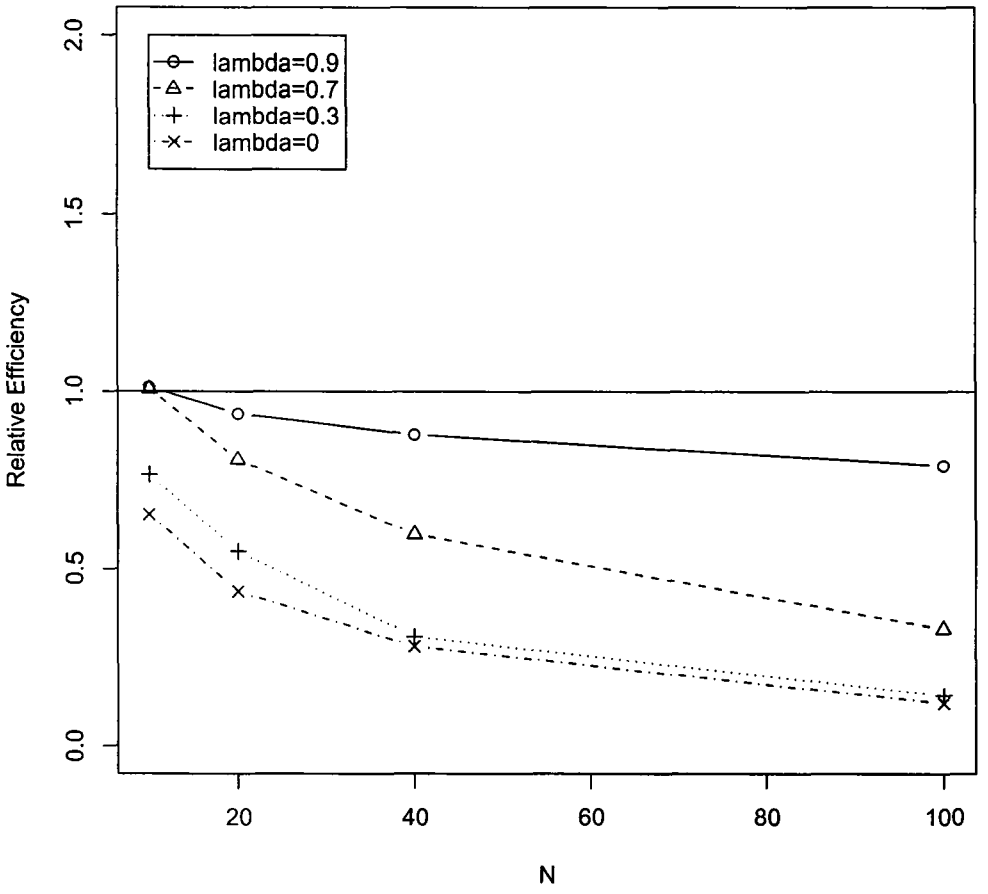
The allocation rule is also applicable to construct adaptive t-Treatment p-Period repeated measurement designs. The major difficulty is that, in a general t-treatment p-period design, there are t^p possible treatment sequences. The number of possible treatment sequences increases substantially as the values of t and p increase. In this case, one should narrow down the number of treatment sequences of interest, and then apply the allocation rule to construct an adaptive design. The resulting adaptive design may not be optimal mathematically, but it is somewhat manageable to construct. Also, due to the difficulty of having all subjects comply until the termination of the experiment and the degree of difficulty increases as the number of periods gets larger, long period designs should be avoided in practice.

6.4 Conclusion

In this chapter, we utilized the allocation strategy proposed in Chapter 5 to construct adaptive repeated measurement designs with dichotomous responses/outcomes. We provide the detailed allocation rule for constructing adaptive two-treatment two-period repeated measurement designs, and then extend it to two-treatment p-period repeated measurement designs. In simulation studies, we demonstrate that the designs with $\lambda < 1$ constructed under the new proposed allocation rule are not as efficient as the design with $\lambda = 1$ in terms of the mean squared error, but those designs successfully put more patients into the better treatment sequence. The value of λ can be pre-determined by researchers, which is used to balance the two objectives of increasing the estimation precision and decreasing the proportion of patients receiving inferior treatments. A large value of λ will place more emphasis on the estimation precision. When $\lambda = 1$ the allocation rule becomes the usual response adaptive design as considered by other researchers (Kushner, 2003). A small value of λ will emphasize the performance/benefit of the treatment.

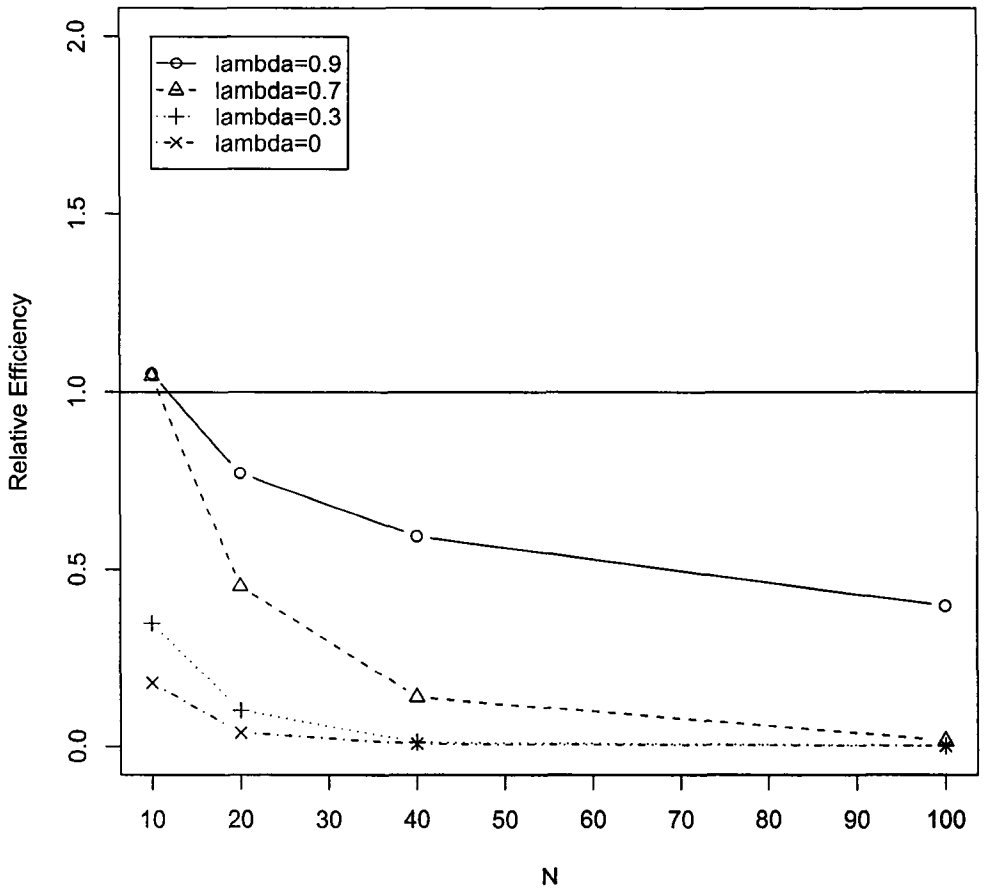
When $\lambda = 0$, the allocation rule becomes a typical play-the-winner rule (Zelen, 1969). In addition, simulation studies show that the design with a high value of λ significantly favors the allocation results toward more effective treatment sequences without loss of much estimation precision.

Figure 6.1: Relative Efficiency of $\boldsymbol{\theta}$ under A-optimality: $p = 2$, equal success probabilities



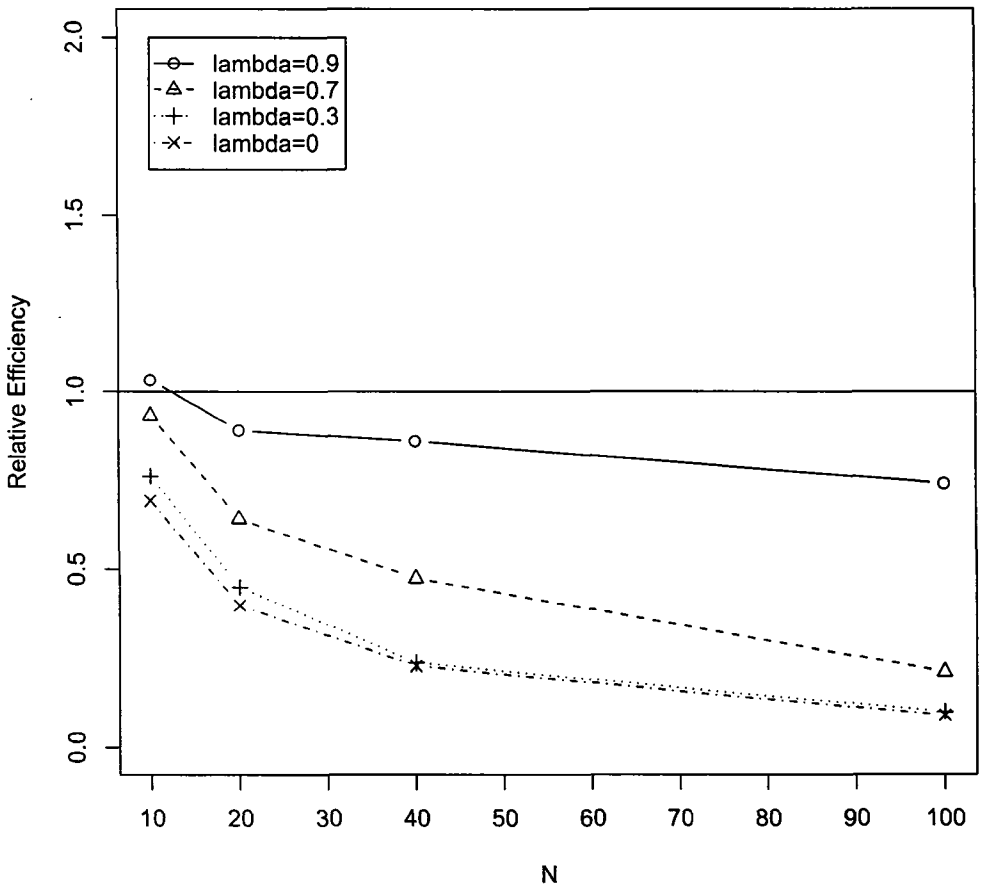
Note: $\boldsymbol{\theta}=(\pi_1, \pi_2, \pi_3, \pi_4)^T$. The design with $\lambda = 1$ is the reference design. Relative efficiencies are calculated based on 1,000 simulations under 2-treatment 2-period RMDs with $\pi_i = 0.5$, $i = 1, 2, 3, 4$

Figure 6.2: Relative Efficiency of $\boldsymbol{\theta}$ under D-optimality: $p = 2$, equal success probabilities



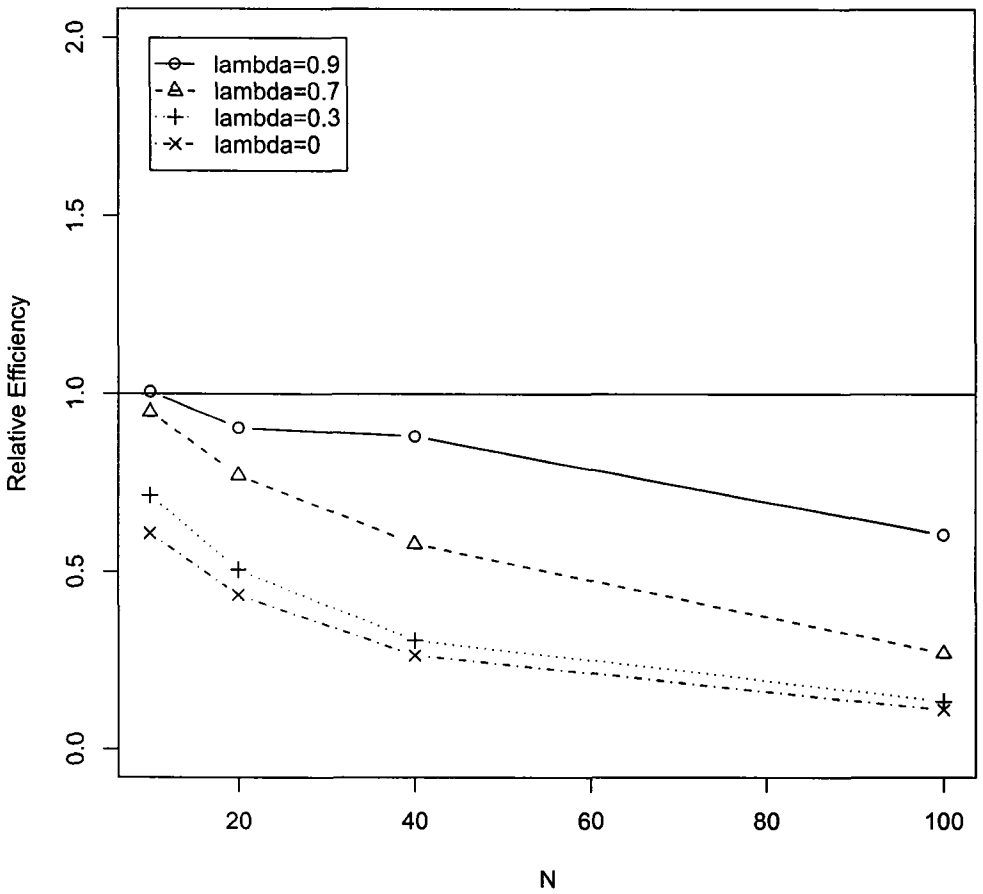
Note: See notes for Figure 6.1.

Figure 6.3: Relative Efficiency of $\boldsymbol{\theta}$ under E-optimality: $p = 2$, equal success probabilities



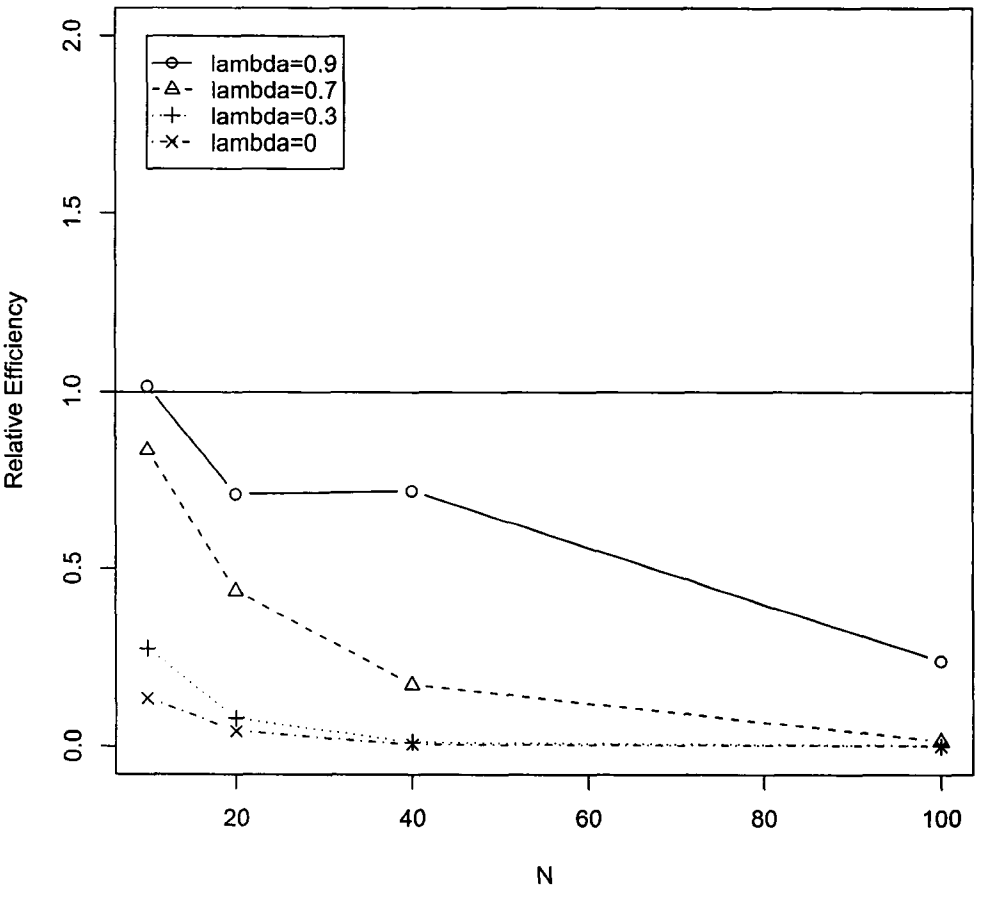
Note: See notes for Figure 6.1.

Figure 6.4: Relative Efficiency of $\boldsymbol{\theta}$ under A-optimality: $p = 2$



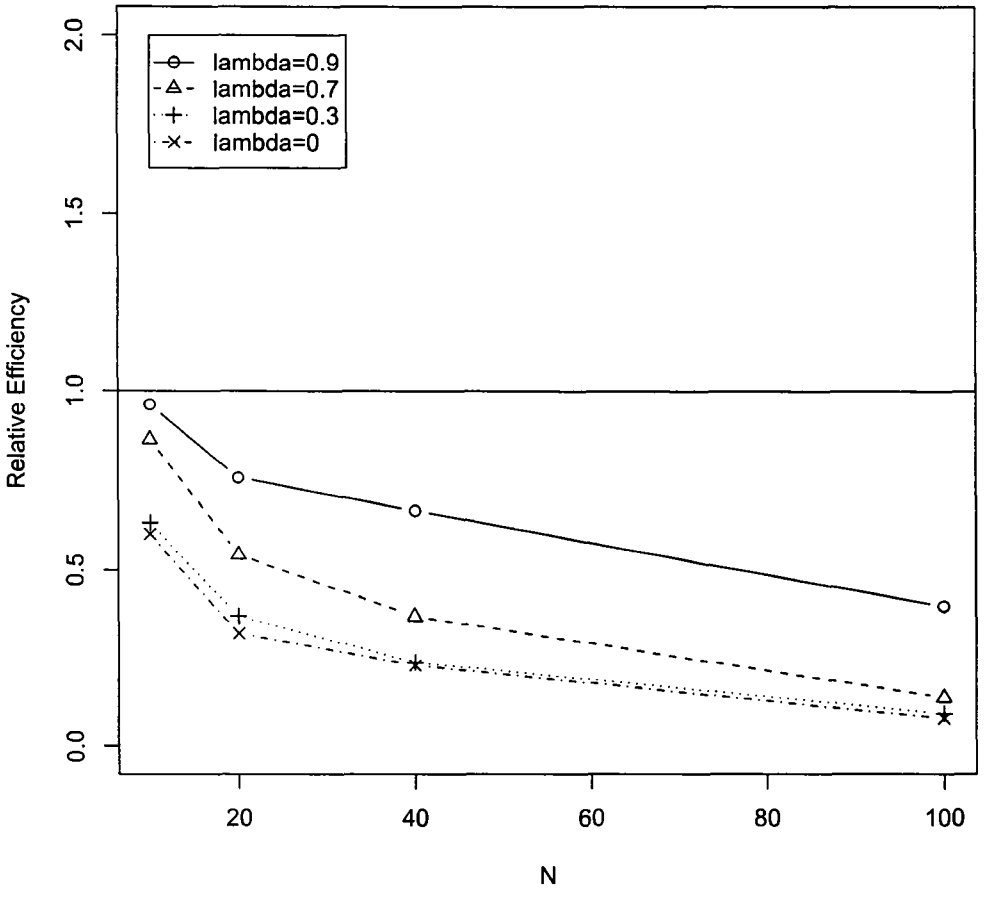
Note: $\boldsymbol{\theta}=(\pi_1, \pi_2, \pi_3, \pi_4)^T$. The design with $\lambda = 1$ is the reference design. Relative efficiencies are calculated based on 1,000 simulations under 2-treatment 2-period RMDs with $\pi_1 = 0.6$, $\pi_2 = 0.3$, $\pi_3 = 0.7$, and $\pi_4 = 0.5$.

Figure 6.5: Relative Efficiency of $\boldsymbol{\theta}$ under D-optimality: $p = 2$



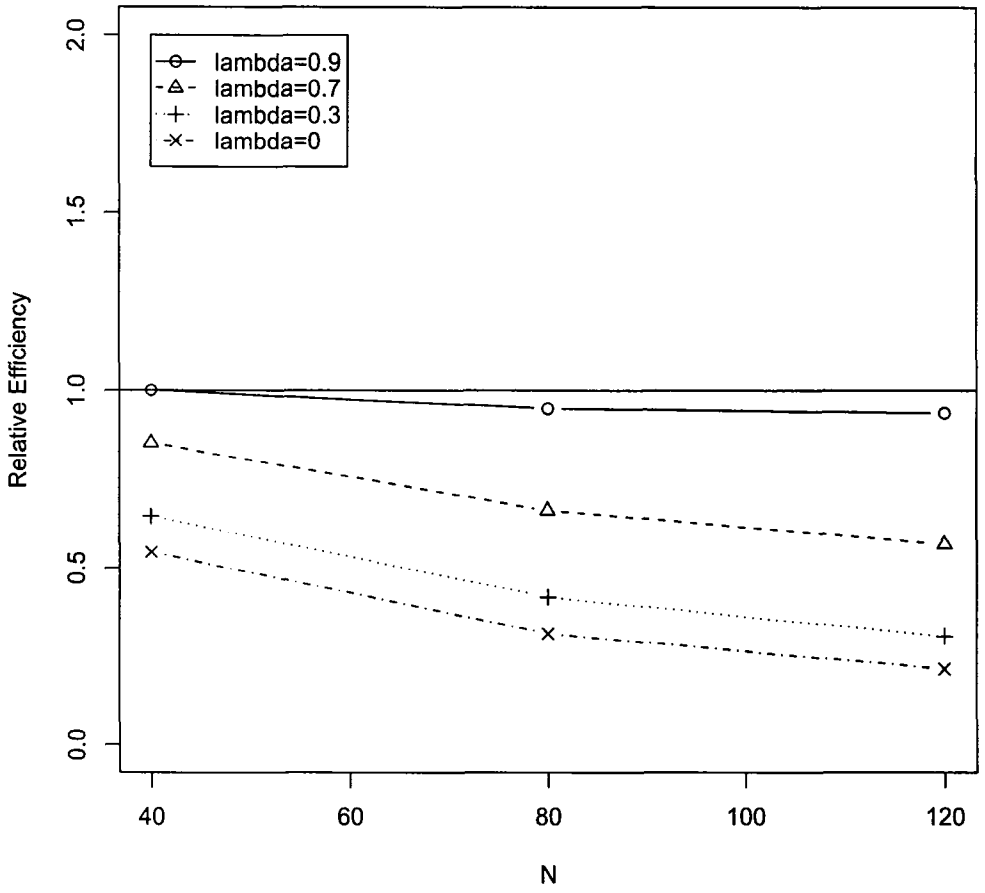
Note: See notes for Figure 6.4.

Figure 6.6: Relative Efficiency of $\boldsymbol{\theta}$ under E-optimality: $p = 2$



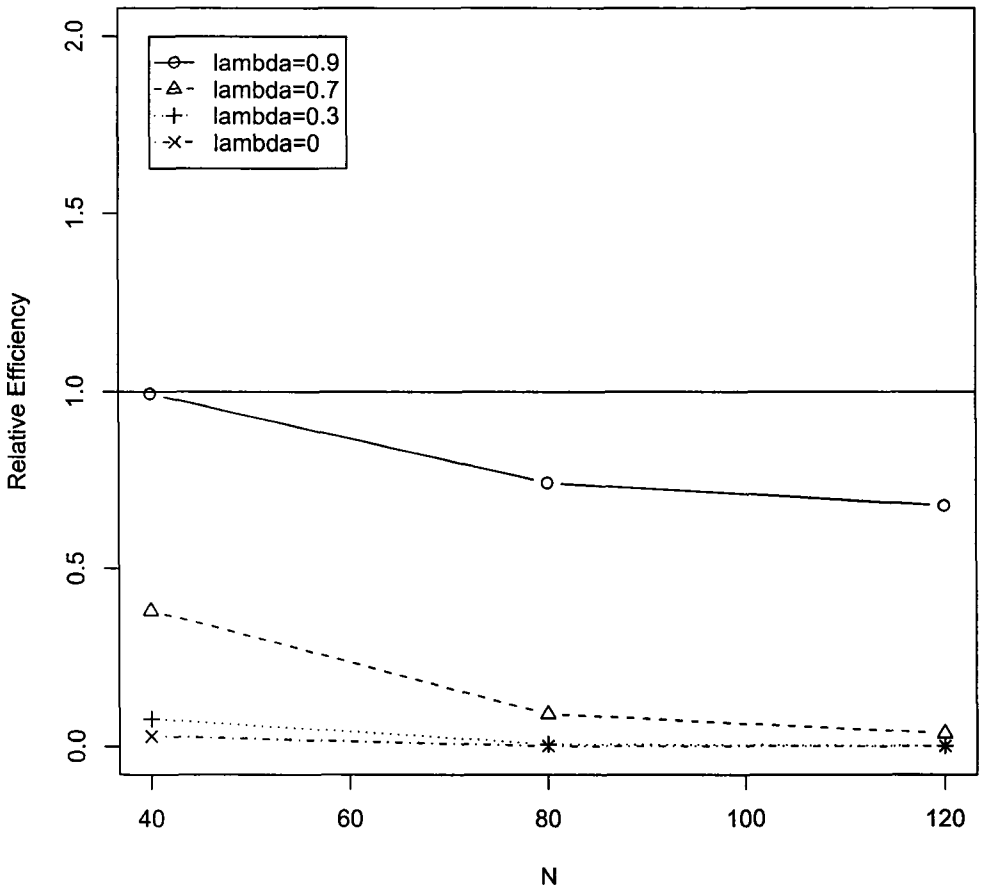
Note: See notes for Figure 6.4.

Figure 6.7: Relative Efficiency of $\boldsymbol{\theta}$ under A-optimality: $p = 3$, equal success probabilities



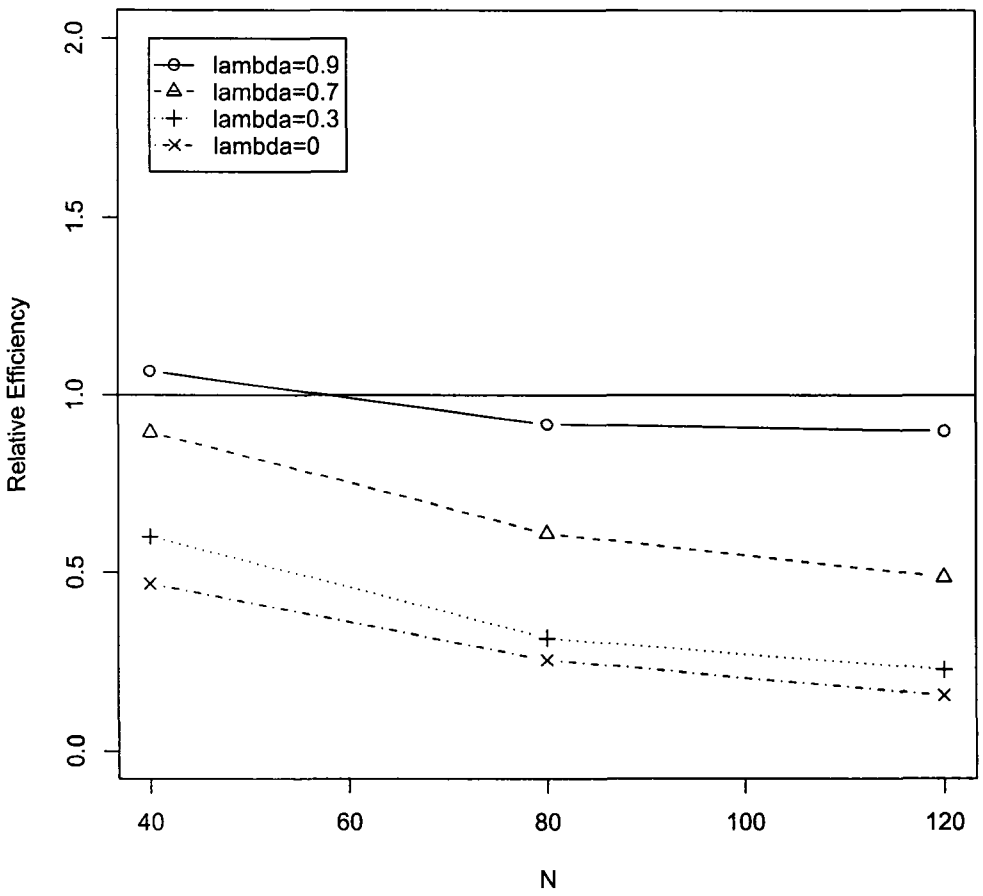
Note: $\boldsymbol{\theta}=(\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6)^T$. The design with $\lambda = 1$ is the reference design. Relative efficiencies are calculated based on 1,000 simulations under 2-treatment 3-period RMDs with $\pi_i = 0.5$, $i = 1, 2, \ldots, 6$

Figure 6.8: Relative Efficiency of $\boldsymbol{\theta}$ under D-optimality: $p = 3$, equal success probabilities



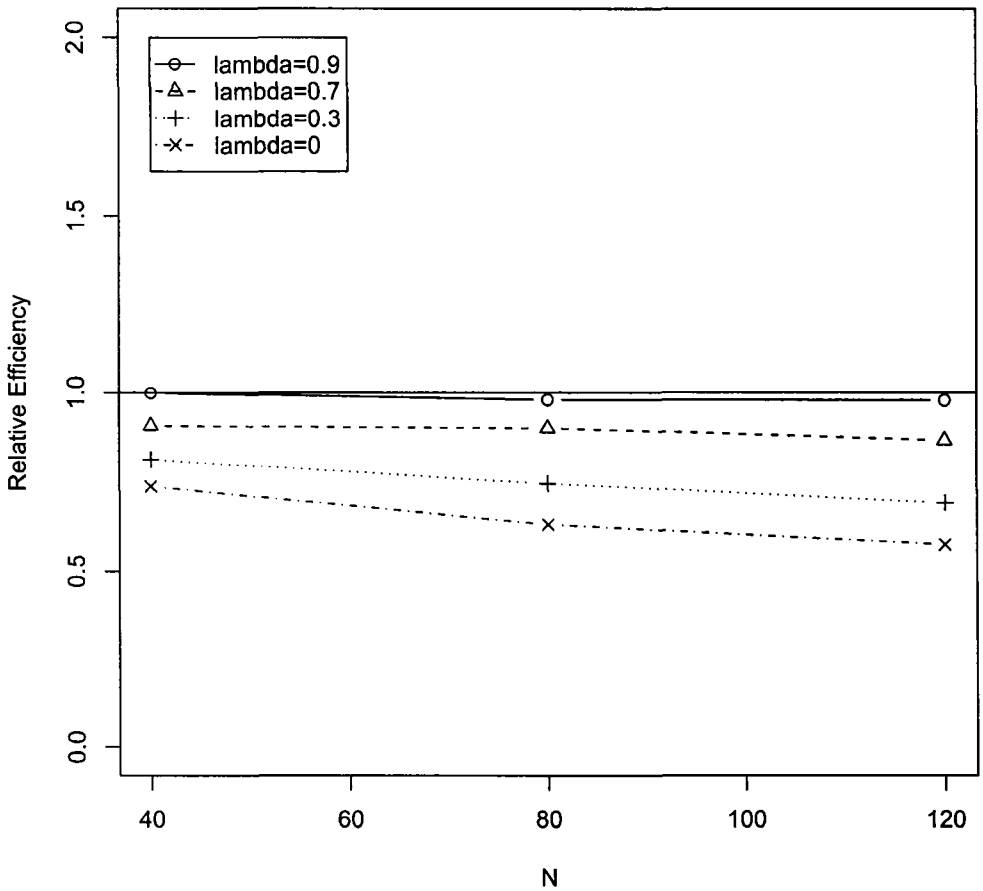
Note: See notes for Figure 6.7.

Figure 6.9: Relative Efficiency of $\boldsymbol{\theta}$ under E-optimality: $p = 3$, equal success probabilities



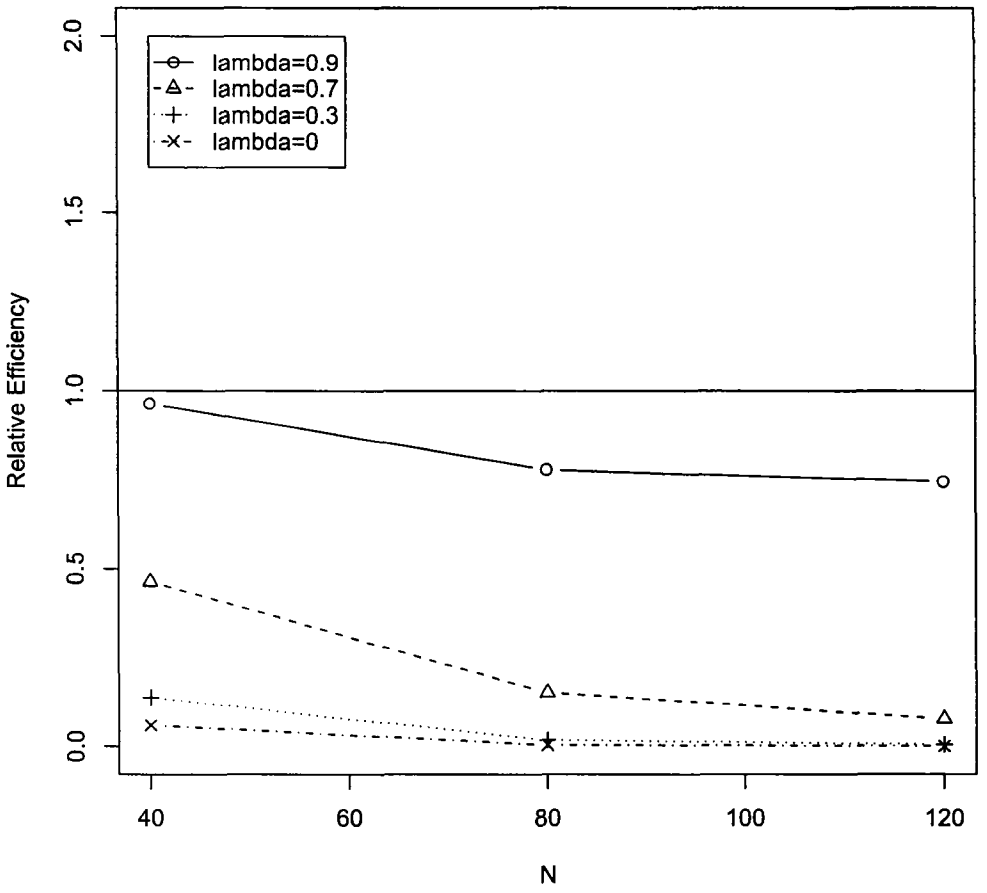
Note: See notes for Figure 6.7.

Figure 6.10: Relative Efficiency of $\boldsymbol{\theta}$ under A-optimality: $p = 3$



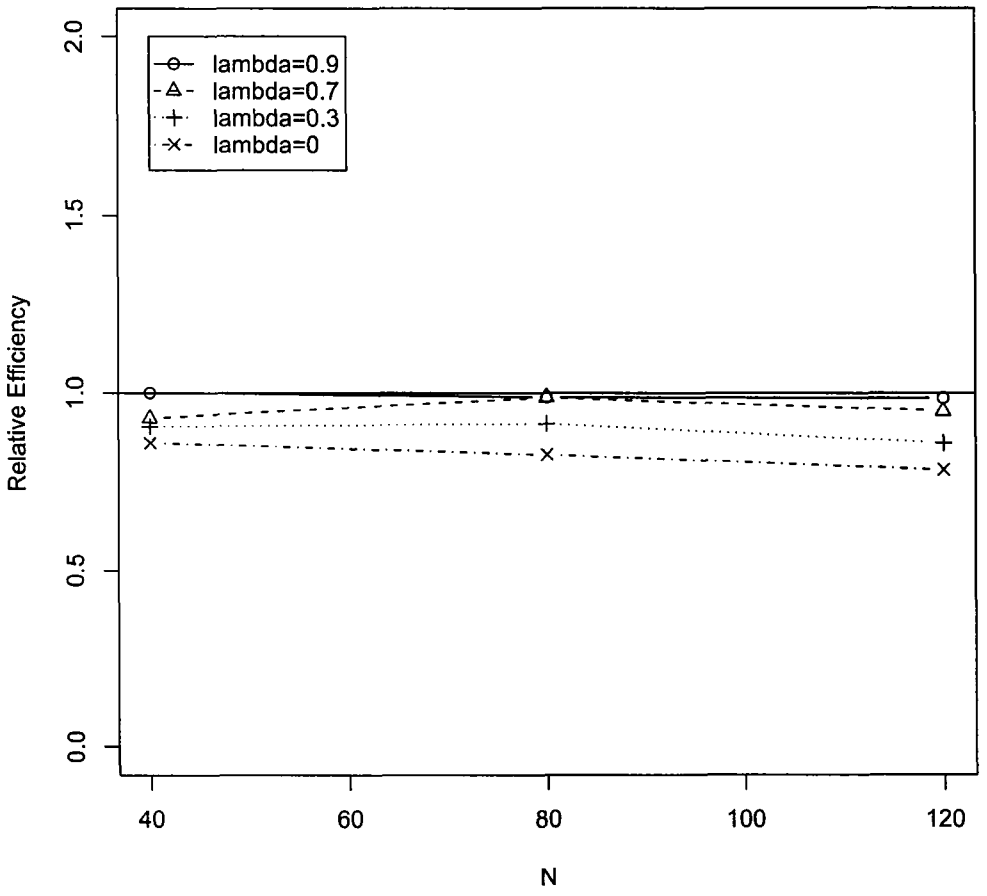
Note: $\boldsymbol{\theta}=(\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6)^T$. The design with $\lambda = 1$ is the reference design. Relative efficiencies are calculated based on 1,000 simulations under 2-treatment 3-period RMDs with $\pi_1 = 0.5$, $\pi_2 = 0.6$, $\pi_3 = 0.7$, $\pi_4 = 0.5$, $\pi_5 = 0.4$, and $\pi_6 = 0.3$.

Figure 6.11: Relative Efficiency of $\boldsymbol{\theta}$ under D-optimality: $p = 3$



Note: See notes for Figure 6.10.

Figure 6.12: Relative Efficiency of $\boldsymbol{\theta}$ under E-optimality: $p = 3$



Note: See notes for Figure 6.10.

Table 6.1: Estimated Numbers of Patients for Each Treatment Sequence: $p = 2$, equal success probabilities

N	λ	N_{AA}	N_{AB}	N_{BA}	N_{BB}
10	1	2.308	2.411	2.858	2.423
10	0.9	2.581	2.315	2.693	2.411
10	0.7	2.541	2.408	2.7	2.351
10	0.3	2.582	2.414	2.49	2.514
10	0	2.355	2.526	2.482	2.637
20	1	4.933	4.94	5.279	4.848
20	0.9	5.252	4.721	5.128	4.899
20	0.7	5.764	4.748	4.688	4.8
20	0.3	5.461	4.724	4.752	5.063
20	0	4.626	4.805	4.947	5.622
40	1	9.956	9.868	10.113	10.063
40	0.9	11.233	9.051	9.693	10.023
40	0.7	12.011	9.376	9.149	9.464
40	0.3	10.757	10.084	8.904	10.255
40	0	10.168	9.178	9.97	10.684
100	1	24.973	24.799	25.141	25.087
100	0.9	26.218	24.679	24.711	24.392
100	0.7	31.736	23.866	21.512	22.886
100	0.3	27.429	24.049	23.112	25.41
100	0	24.241	26.554	25.172	24.033

Note: Entries are based on 1,000 simulations under 2-treatment 2-period RMDs with $\pi_i = 0.5$, $i = 1, 2, 3, 4$.

Table 6.2: Mean and Standard Deviation of the Parameters of Interest: $p = 2$, equal success probabilities

N	λ	$\hat{\pi_1}$	$\hat{\pi_2}$	$\hat{\pi_3}$	$\hat{\pi_4}$
10	1	0.507 (0.208)	0.496 (0.205)	0.495 (0.212)	0.503 (0.224)
10	0.9	0.487 (0.213)	0.491 (0.202)	0.490 (0.207)	0.479 (0.220)
10	0.7	0.481 (0.206)	0.491 (0.204)	0.472 (0.201)	0.477 (0.231)
10	0.3	0.468 (0.239)	0.470 (0.236)	0.452 (0.243)	0.456 (0.241)
10	0	0.459 (0.260)	0.443 (0.257)	0.442 (0.262)	0.438 (0.251)
20	1	0.499 (0.159)	0.507 (0.155)	0.506 (0.154)	0.498 (0.157)
20	0.9	0.487 (0.161)	0.492 (0.162)	0.482 (0.154)	0.488 (0.168)
20	0.7	0.477 (0.164)	0.474 (0.164)	0.466 (0.158)	0.452 (0.195)
20	0.3	0.452 (0.197)	0.448 (0.205)	0.434 (0.199)	0.431 (0.210)
20	0	0.425 (0.226)	0.445 (0.220)	0.428 (0.242)	0.435 (0.221)
40	1	0.500 (0.112)	0.502 (0.110)	0.500 (0.114)	0.496 (0.112)
40	0.9	0.490 (0.116)	0.486 (0.123)	0.491 (0.120)	0.480 (0.116)
40	0.7	0.468 (0.131)	0.467 (0.137)	0.466 (0.135)	0.444 (0.155)
40	0.3	0.431 (0.177)	0.424 (0.193)	0.415 (0.192)	0.426 (0.187)
40	0	0.428 (0.197)	0.416 (0.193)	0.429 (0.199)	0.410 (0.197)
100	1	0.503 (0.072)	0.497 (0.071)	0.500 (0.072)	0.502 (0.070)
100	0.9	0.488 (0.079)	0.489 (0.077)	0.493 (0.079)	0.487 (0.083)
100	0.7	0.468 (0.100)	0.459 (0.120)	0.464 (0.105)	0.444 (0.139)
100	0.3	0.416 (0.170)	0.421 (0.169)	0.415 (0.171)	0.418 (0.171)
100	0	0.418 (0.183)	0.412 (0.189)	0.415 (0.186)	0.413 (0.189)

Note: See notes for Table 6.1

Table 6.3: Median and Median Absolute Deviation of the Parameters of Interest: $p = 2$, equal success probabilities

N	λ	$\hat{\pi_1}$	$\hat{\pi_2}$	$\hat{\pi_3}$	$\hat{\pi_4}$
10	1	0.500 (0.148)	0.500 (0.148)	0.500 (0.148)	0.500 (0.247)
10	0.9	0.500 (0.148)	0.500 (0.148)	0.500 (0.247)	0.500 (0.247)
10	0.7	0.500 (0.247)	0.500 (0.247)	0.500 (0.247)	0.500 (0.247)
10	0.3	0.500 (0.247)	0.500 (0.247)	0.500 (0.247)	0.500 (0.247)
10	0	0.500 (0.247)	0.500 (0.247)	0.500 (0.247)	0.500 (0.247)
20	1	0.500 (0.148)	0.500 (0.148)	0.500 (0.148)	0.500 (0.148)
20	0.9	0.500 (0.148)	0.500 (0.148)	0.500 (0.148)	0.500 (0.171)
20	0.7	0.500 (0.148)	0.500 (0.185)	0.471 (0.159)	0.464 (0.194)
20	0.3	0.500 (0.165)	0.500 (0.185)	0.471 (0.168)	0.471 (0.192)
20	0	0.500 (0.185)	0.500 (0.165)	0.500 (0.185)	0.500 (0.171)
40	1	0.500 (0.117)	0.500 (0.106)	0.500 (0.117)	0.500 (0.106)
40	0.9	0.500 (0.114)	0.500 (0.117)	0.500 (0.114)	0.477 (0.108)
40	0.7	0.482 (0.122)	0.484 (0.130)	0.486 (0.128)	0.470 (0.134)
40	0.3	0.471 (0.136)	0.474 (0.151)	0.467 (0.155)	0.474 (0.132)
40	0	0.486 (0.127)	0.469 (0.129)	0.485 (0.136)	0.473 (0.140)
100	1	0.500 (0.076)	0.500 (0.059)	0.500 (0.073)	0.500 (0.073)
100	0.9	0.500 (0.076)	0.497 (0.070)	0.500 (0.074)	0.492 (0.073)
100	0.7	0.490 (0.077)	0.488 (0.089)	0.488 (0.085)	0.480 (0.091)
100	0.3	0.474 (0.098)	0.478 (0.103)	0.471 (0.106)	0.480 (0.094)
100	0	0.479 (0.099)	0.479 (0.098)	0.479 (0.090)	0.480 (0.100)

Note: See notes for Table 6.1

Table 6.4: Characteristics of Mean Squared Error (MSE) of $\boldsymbol{\theta}=(\pi_1, \pi_2, \pi_3, \pi_4)^T$: $p = 2$, equal success probabilities

N	λ	Trace(MSE)	Det(MSE)	Eigen(MSE)
10	1	0.18019919	4.07E-06	0.050600461
10	0.9	0.1780056	3.87E-06	0.04894022
10	0.7	0.17903574	3.90E-06	0.054245484
10	0.3	0.23608263	1.18E-05	0.066660262
10	0	0.27697381	2.29E-05	0.073366209
20	1	0.09771915	3.55E-07	0.025904089
20	0.9	0.10446655	4.59E-07	0.029074776
20	0.7	0.12130864	7.85E-07	0.040481772
20	0.3	0.17835594	3.55E-06	0.057635957
20	0	0.22469509	9.63E-06	0.065059258
40	1	0.05016585	2.45E-08	0.013747111
40	0.9	0.05722919	4.13E-08	0.015987678
40	0.7	0.084193	1.78E-07	0.028976599
40	0.3	0.16325074	2.22E-06	0.057483292
40	0	0.17933922	3.61E-06	0.059754418
100	1	0.02034397	6.68E-10	0.005340812
100	0.9	0.02579329	1.69E-09	0.007233284
100	0.7	0.06192835	4.39E-08	0.025031049
100	0.3	0.14284705	1.18E-06	0.053077945
100	0	0.16899279	2.66E-06	0.059161278

Note: See notes for Table 6.1

Table 6.5: Estimated Numbers of Patients for Each Treatment Sequence: $p = 2$

N	λ	N_{AA}	N_{AB}	N_{BA}	N_{BB}
10	1	2.28	2.27	3.06	2.39
10	0.9	2.70	2.05	2.95	2.31
10	0.7	2.97	2.10	2.80	2.13
10	0.3	3.35	2.38	2.43	1.84
10	0	3.42	2.60	2.20	1.78
20	1	4.78	4.87	5.64	4.71
20	0.9	7.06	3.94	4.83	4.17
20	0.7	7.97	4.57	4.49	2.97
20	0.3	8.69	4.70	4.06	2.55
20	0	8.80	4.95	3.78	2.48
40	1	9.87	9.93	10.33	9.88
40	0.9	17.22	7.97	7.93	6.88
40	0.7	20.49	8.58	6.74	4.19
40	0.3	20.74	9.69	6.26	3.31
40	0	18.60	11.09	6.92	3.39
100	1	25.05	24.73	25.07	25.16
100	0.9	53.87	19.91	15.89	10.32
100	0.7	63.15	20.66	11.29	4.91
100	0.3	59.45	23.58	12.17	4.80
100	0	56.86	22.95	14.85	5.34

Note: Entries are based on 1,000 simulations under 2-treatment 2-period RMDs with $\pi_1 = 0.6$, $\pi_2 = 0.3$, $\pi_3 = 0.7$, and $\pi_4 = 0.5$.

Table 6.6: Mean and Standard Deviation of the Parameters of Interest: $p = 2$

N	λ	$\hat{\pi_1}$	$\hat{\pi_2}$	$\hat{\pi_3}$	$\hat{\pi_4}$
10	1	0.575 (0.213)	0.314 (0.185)	0.683 (0.187)	0.506 (0.218)
10	0.9	0.586 (0.204)	0.313 (0.189)	0.684 (0.188)	0.492 (0.222)
10	0.7	0.578 (0.208)	0.324 (0.178)	0.666 (0.199)	0.474 (0.233)
10	0.3	0.569 (0.213)	0.273 (0.221)	0.649 (0.235)	0.441 (0.267)
10	0	0.558 (0.241)	0.245 (0.241)	0.664 (0.260)	0.450 (0.276)
20	1	0.601 (0.154)	0.305 (0.141)	0.697 (0.138)	0.497 (0.158)
20	0.9	0.584 (0.150)	0.297 (0.145)	0.690 (0.141)	0.476 (0.180)
20	0.7	0.587 (0.146)	0.303 (0.144)	0.677 (0.160)	0.451 (0.207)
20	0.3	0.573 (0.180)	0.279 (0.189)	0.653 (0.188)	0.409 (0.242)
20	0	0.569 (0.187)	0.227 (0.204)	0.660 (0.200)	0.410 (0.262)
40	1	0.601 (0.109)	0.299 (0.104)	0.699 (0.100)	0.501 (0.118)
40	0.9	0.602 (0.099)	0.293 (0.114)	0.691 (0.097)	0.476 (0.141)
40	0.7	0.586 (0.105)	0.294 (0.126)	0.674 (0.123)	0.441 (0.184)
40	0.3	0.568 (0.144)	0.257 (0.182)	0.641 (0.191)	0.415 (0.220)
40	0	0.553 (0.168)	0.222 (0.203)	0.642 (0.195)	0.420 (0.228)
100	1	0.600 (0.072)	0.300 (0.065)	0.702 (0.063)	0.503 (0.069)
100	0.9	0.598 (0.058)	0.283 (0.096)	0.693 (0.065)	0.474 (0.110)
100	0.7	0.594 (0.073)	0.289 (0.126)	0.679 (0.094)	0.428 (0.178)
100	0.3	0.565 (0.131)	0.253 (0.177)	0.645 (0.158)	0.403 (0.212)
100	0	0.559 (0.146)	0.223 (0.190)	0.645 (0.180)	0.393 (0.227)

Note: See notes for Table 6.5

Table 6.7: Median and Median Absolute Deviation of the Parameters of Interest: $p = 2$

N	λ	$\hat{\pi_1}$	$\hat{\pi_2}$	$\hat{\pi_3}$	$\hat{\pi_4}$
10	1	0.600 (0.297)	0.286 (0.169)	0.714 (0.169)	0.500 (0.247)
10	0.9	0.600 (0.297)	0.250 (0.185)	0.714 (0.169)	0.500 (0.247)
10	0.7	0.600 (0.222)	0.333 (0.198)	0.714 (0.212)	0.500 (0.247)
10	0.3	0.600 (0.222)	0.286 (0.318)	0.667 (0.247)	0.500 (0.247)
10	0	0.600 (0.222)	0.250 (0.371)	0.714 (0.238)	0.500 (0.247)
20	1	0.600 (0.148)	0.300 (0.148)	0.700 (0.148)	0.500 (0.148)
20	0.9	0.583 (0.124)	0.286 (0.169)	0.700 (0.127)	0.500 (0.185)
20	0.7	0.600 (0.148)	0.286 (0.169)	0.700 (0.148)	0.500 (0.202)
20	0.3	0.600 (0.148)	0.300 (0.198)	0.688 (0.134)	0.467 (0.198)
20	0	0.611 (0.165)	0.250 (0.371)	0.688 (0.134)	0.500 (0.247)
40	1	0.600 (0.109)	0.300 (0.074)	0.700 (0.092)	0.500 (0.117)
40	0.9	0.607 (0.088)	0.297 (0.113)	0.700 (0.094)	0.500 (0.148)
40	0.7	0.594 (0.092)	0.286 (0.127)	0.692 (0.097)	0.500 (0.148)
40	0.3	0.595 (0.094)	0.266 (0.183)	0.688 (0.105)	0.500 (0.148)
40	0	0.579 (0.116)	0.250 (0.371)	0.686 (0.111)	0.486 (0.165)
100	1	0.600 (0.073)	0.300 (0.059)	0.700 (0.059)	0.500 (0.073)
100	0.9	0.600 (0.059)	0.286 (0.095)	0.699 (0.058)	0.483 (0.099)
100	0.7	0.600 (0.058)	0.281 (0.120)	0.698 (0.055)	0.473 (0.145)
100	0.3	0.593 (0.053)	0.265 (0.163)	0.691 (0.061)	0.478 (0.138)
100	0	0.596 (0.059)	0.250 (0.222)	0.690 (0.061)	0.476 (0.142)

Note: See notes for Table 6.5

Table 6.8: Characteristics of Mean Squared Error (MSE) of $\boldsymbol{\theta}=(\pi_1, \pi_2, \pi_3, \pi_4)^T$: $p = 2$

N	λ	Trace(MSE)	Det(MSE)	Eigen(MSE)
10	1	0.16301349	2.63E-06	0.047964887
10	0.9	0.16218494	2.59E-06	0.049870342
10	0.7	0.17198315	3.15E-06	0.055518706
10	0.3	0.22817374	9.59E-06	0.076045226
10	0	0.26811325	1.96E-05	0.079994734
20	1	0.08739959	2.22E-07	0.025007511
20	0.9	0.09665183	3.12E-07	0.03306143
20	0.7	0.11333011	5.08E-07	0.046129547
20	0.3	0.17281864	2.83E-06	0.06788199
20	0	0.20118257	5.24E-06	0.078267402
40	1	0.04642482	1.75E-08	0.013890562
40	0.9	0.05272822	2.44E-08	0.020892127
40	0.7	0.08027354	1.01E-07	0.037778329
40	0.3	0.15216839	1.55E-06	0.058318031
40	0	0.17710033	3.28E-06	0.060463731
100	1	0.01813612	4.12E-10	0.005239642
100	0.9	0.02997347	1.72E-09	0.013215
100	0.7	0.06735765	2.74E-08	0.037849566
100	0.3	0.13439885	8.68E-07	0.056416272
100	0	0.16316947	1.96E-06	0.066333508

Note: See notes for Table 6.5

Table 6.9: Estimated Numbers of Patients for Each Treatment Sequence: $p = 3$, equal success probabilities

N	λ	N_{AAA}	N_{AAB}	N_{ABA}	N_{ABB}	N_{BBB}	N_{BBA}	N_{BAB}	N_{BAA}
40	1	3.782	4.356	5.263	6.599	3.782	4.356	5.263	6.599
40	0.9	5.058	4.957	4.857	5.057	5.118	4.975	4.904	5.074
40	0.7	4.704	5.108	5.027	5.014	5.095	4.975	5.230	4.847
40	0.3	4.701	4.816	5.370	5.194	4.792	4.864	5.163	5.100
40	0	4.949	4.691	4.880	4.619	4.931	5.709	4.892	5.329
80	1	7.714	8.698	10.117	13.471	7.715	8.697	10.116	13.472
80	0.9	10.309	10.337	9.006	9.991	10.514	10.483	9.290	10.070
80	0.7	10.528	9.701	10.108	9.568	10.246	9.674	10.039	10.136
80	0.3	10.633	10.016	10.338	9.507	9.922	9.556	9.318	10.710
80	0	9.879	9.906	10.266	10.010	9.506	9.816	10.866	9.751
120	1	11.570	13.311	15.175	19.944	11.575	13.306	15.170	19.949
120	0.9	14.580	15.294	15.921	13.732	14.895	15.386	16.205	13.987
120	0.7	15.172	15.462	13.146	16.140	14.413	15.481	14.822	15.364
120	0.3	15.579	14.301	14.807	13.386	15.279	15.327	14.209	17.112
120	0	15.202	14.835	15.184	14.824	14.801	14.414	14.521	16.219

Note: Entries are based on 1,000 simulations under 2-treatment 3-period RMDs with $\pi_i = 0.5$, $i = 1, 2, \ldots, 6$.

Table 6.10: Mean and Standard Deviation of the Parameters of Interest: $p = 3$, equal success probabilities

N	λ	$\hat{\pi_1}$	$\hat{\pi_2}$	$\hat{\pi_3}$	$\hat{\pi_4}$	$\hat{\pi_5}$	$\hat{\pi_6}$
40	1	0.495 (0.111)	0.501 (0.109)	0.493 (0.112)	0.503 (0.110)	0.499 (0.114)	0.496 (0.120)
40	0.9	0.494 (0.110)	0.499 (0.113)	0.489 (0.110)	0.496 (0.111)	0.496 (0.115)	0.493 (0.116)
40	0.7	0.484 (0.123)	0.484 (0.121)	0.481 (0.122)	0.481 (0.123)	0.484 (0.116)	0.495 (0.120)
40	0.3	0.469 (0.140)	0.462 (0.134)	0.474 (0.137)	0.474 (0.133)	0.469 (0.139)	0.470 (0.137)
40	0	0.464 (0.149)	0.451 (0.153)	0.469 (0.145)	0.463 (0.145)	0.463 (0.147)	0.462 (0.147)
80	1	0.498 (0.078)	0.500 (0.080)	0.503 (0.079)	0.498 (0.080)	0.504 (0.078)	0.504 (0.079)
80	0.9	0.490 (0.083)	0.493 (0.079)	0.493 (0.079)	0.499 (0.084)	0.494 (0.081)	0.497 (0.080)
80	0.7	0.487 (0.094)	0.490 (0.091)	0.481 (0.095)	0.484 (0.097)	0.482 (0.097)	0.479 (0.100)
80	0.3	0.464 (0.122)	0.466 (0.116)	0.470 (0.111)	0.465 (0.117)	0.462 (0.117)	0.461 (0.121)
80	0	0.455 (0.133)	0.460 (0.136)	0.460 (0.139)	0.455 (0.132)	0.459 (0.139)	0.461 (0.132)
120	1	0.499 (0.066)	0.500 (0.064)	0.501 (0.066)	0.497 (0.064)	0.497 (0.063)	0.497 (0.063)
120	0.9	0.497 (0.066)	0.495 (0.069)	0.497 (0.066)	0.493 (0.066)	0.493 (0.067)	0.495 (0.064)
120	0.7	0.486 (0.083)	0.482 (0.083)	0.484 (0.080)	0.479 (0.084)	0.474 (0.085)	0.483 (0.085)
120	0.3	0.458 (0.115)	0.466 (0.103)	0.467 (0.112)	0.470 (0.111)	0.459 (0.107)	0.460 (0.115)
120	0	0.450 (0.132)	0.465 (0.124)	0.448 (0.130)	0.458 (0.129)	0.449 (0.139)	0.449 (0.134)

Note: See notes for Table 6.9

Table 6.11: Median and Median Absolute Deviation of the Parameters of Interest: $p = 3$, equal success probabilities

N	λ	$\hat{\pi_1}$	$\hat{\pi_2}$	$\hat{\pi_3}$	$\hat{\pi_4}$	$\hat{\pi_5}$	$\hat{\pi_6}$
40	1	0.500 (0.074)	0.500 (0.148)	0.500 (0.074)	0.500 (0.148)	0.500 (0.148)	0.500 (0.148)
40	0.9	0.500 (0.106)	0.500 (0.117)	0.500 (0.106)	0.500 (0.117)	0.500 (0.117)	0.500 (0.117)
40	0.7	0.500 (0.117)	0.500 (0.106)	0.500 (0.117)	0.500 (0.114)	0.500 (0.117)	0.500 (0.106)
40	0.3	0.500 (0.128)	0.484 (0.124)	0.500 (0.124)	0.500 (0.124)	0.500 (0.124)	0.500 (0.124)
40	0	0.484 (0.127)	0.479 (0.117)	0.500 (0.116)	0.485 (0.126)	0.486 (0.127)	0.500 (0.124)
80	1	0.500 (0.074)	0.500 (0.074)	0.500 (0.074)	0.500 (0.074)	0.500 (0.074)	0.500 (0.074)
80	0.9	0.490 (0.080)	0.500 (0.078)	0.490 (0.076)	0.500 (0.086)	0.500 (0.078)	0.500 (0.078)
80	0.7	0.500 (0.082)	0.500 (0.087)	0.489 (0.083)	0.500 (0.082)	0.500 (0.084)	0.500 (0.089)
80	0.3	0.489 (0.092)	0.486 (0.086)	0.493 (0.090)	0.486 (0.090)	0.485 (0.094)	0.484 (0.091)
80	0	0.485 (0.097)	0.491 (0.093)	0.500 (0.093)	0.486 (0.091)	0.493 (0.095)	0.487 (0.087)
120	1	0.500 (0.074)	0.500 (0.074)	0.500 (0.074)	0.500 (0.074)	0.500 (0.074)	0.500 (0.074)
120	0.9	0.500 (0.064)	0.500 (0.067)	0.500 (0.062)	0.500 (0.065)	0.500 (0.067)	0.500 (0.062)
120	0.7	0.500 (0.067)	0.494 (0.071)	0.494 (0.066)	0.491 (0.066)	0.487 (0.067)	0.498 (0.069)
120	0.3	0.482 (0.082)	0.488 (0.075)	0.495 (0.070)	0.500 (0.073)	0.485 (0.079)	0.486 (0.084)
120	0	0.484 (0.082)	0.500 (0.077)	0.483 (0.084)	0.491 (0.077)	0.491 (0.086)	0.486 (0.077)

Note: See notes for Table 6.9

Table 6.12: Characteristics of Mean Squared Error (MSE) of $\boldsymbol{\theta}=(\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6)^T$: $p = 3$, equal success probabilities

N	λ	Trace(MSE)	Det(MSE)	Eigen(MSE)
40	1	0.0760775	4.04E-12	0.014800874
40	0.9	0.07599807	4.06E-12	0.013851653
40	0.7	0.0893677	1.06E-11	0.01656812
40	0.3	0.11764665	5.32E-11	0.024628863
40	0	0.13984216	1.47E-10	0.031615421
80	1	0.03736813	5.75E-14	0.006886318
80	0.9	0.03935808	7.74E-14	0.007510049
80	0.7	0.05636365	6.42E-13	0.011299752
80	0.3	0.0900547	9.75E-12	0.021702422
80	0	0.11987797	5.75E-11	0.026890976
120	1	0.02482528	4.91E-15	0.004631228
120	0.9	0.02649665	7.24E-15	0.00515708
120	0.7	0.04382803	1.40E-13	0.009511979
120	0.3	0.08165085	5.17E-12	0.020276189
120	0	0.11699973	4.42E-11	0.029867605

Note: See notes for Table 6.9

Table 6.13: Estimated Numbers of Patients for Each Treatment Sequence: $p = 3$

N	λ	N_{AAA}	N_{AAB}	N_{ABA}	N_{ABB}	N_{BBB}	N_{BBA}	N_{BAB}	N_{BAA}
40	1	4.287	4.455	5.039	6.219	4.276	4.460	5.037	6.227
40	0.9	6.598	4.456	5.257	3.829	4.155	5.126	4.633	5.946
40	0.7	8.168	3.661	5.193	2.945	2.681	5.506	3.760	8.086
40	0.3	8.754	3.480	5.167	2.301	2.233	5.408	3.127	9.530
40	0	9.439	2.923	5.321	2.033	1.958	5.200	3.266	9.860
80	1	8.367	9.042	10.171	12.420	8.382	9.027	10.156	12.435
80	0.9	15.035	8.258	10.429	6.753	6.773	10.736	7.864	14.152
80	0.7	19.718	6.121	10.602	3.584	3.772	10.996	5.746	19.461
80	0.3	21.402	4.771	11.564	2.722	2.579	9.639	4.421	22.902
80	0	21.828	4.673	10.346	2.315	2.282	10.951	5.369	22.236
120	1	12.804	13.233	15.164	18.799	12.840	13.197	15.128	18.835
120	0.9	24.069	11.577	15.615	8.346	8.431	16.210	10.653	25.099
120	0.7	34.098	6.872	14.337	4.194	3.977	15.040	7.265	34.217
120	0.3	37.799	5.791	15.596	2.742	2.752	13.698	5.864	35.758
120	0	36.854	7.053	15.007	2.629	2.635	15.271	6.218	34.333

Note: Entries are based on 1,000 simulations under 2-treatment 3-period RMDs with $\pi_1 = 0.5$, $\pi_2 = 0.6$, $\pi_3 = 0.7$, $\pi_4 = 0.5$, $\pi_5 = 0.4$, and $\pi_6 = 0.3$.

Table 6.14: Mean and Standard Deviation of the Parameters of Interest: $p = 3$

N	λ	$\hat{\pi_1}$	$\hat{\pi_2}$	$\hat{\pi_3}$	$\hat{\pi_4}$	$\hat{\pi_5}$	$\hat{\pi_6}$
40	1	0.497 (0.113)	0.606 (0.109)	0.696 (0.103)	0.497 (0.110)	0.397 (0.111)	0.305 (0.100)
40	0.9	0.498 (0.112)	0.592 (0.106)	0.704 (0.097)	0.489 (0.112)	0.393 (0.115)	0.296 (0.113)
40	0.7	0.485 (0.126)	0.590 (0.110)	0.693 (0.096)	0.482 (0.122)	0.383 (0.127)	0.280 (0.125)
40	0.3	0.471 (0.136)	0.583 (0.123)	0.691 (0.103)	0.472 (0.139)	0.371 (0.145)	0.279 (0.143)
40	0	0.464 (0.149)	0.571 (0.130)	0.692 (0.106)	0.455 (0.148)	0.360 (0.160)	0.264 (0.159)
80	1	0.502 (0.077)	0.599 (0.078)	0.700 (0.072)	0.499 (0.080)	0.397 (0.075)	0.294 (0.072)
80	0.9	0.500 (0.081)	0.598 (0.075)	0.696 (0.066)	0.488 (0.085)	0.395 (0.082)	0.292 (0.088)
80	0.7	0.475 (0.104)	0.589 (0.087)	0.698 (0.065)	0.481 (0.102)	0.384 (0.107)	0.283 (0.106)
80	0.3	0.468 (0.115)	0.578 (0.103)	0.692 (0.084)	0.464 (0.124)	0.376 (0.134)	0.276 (0.135)
80	0	0.462 (0.139)	0.570 (0.123)	0.687 (0.092)	0.458 (0.138)	0.354 (0.146)	0.263 (0.154)
120	1	0.499 (0.064)	0.598 (0.064)	0.699 (0.058)	0.498 (0.065)	0.399 (0.063)	0.302 (0.059)
120	0.9	0.491 (0.070)	0.598 (0.059)	0.698 (0.052)	0.494 (0.068)	0.392 (0.072)	0.293 (0.074)
120	0.7	0.481 (0.087)	0.592 (0.064)	0.700 (0.049)	0.477 (0.091)	0.372 (0.101)	0.279 (0.106)
120	0.3	0.470 (0.115)	0.574 (0.098)	0.692 (0.068)	0.461 (0.124)	0.356 (0.133)	0.278 (0.126)
120	0	0.461 (0.131)	0.565 (0.122)	0.685 (0.094)	0.453 (0.139)	0.349 (0.149)	0.265 (0.157)

Note: See notes for Table 6.13

Table 6.15: Median and Median Absolute Deviation of the Parameters of Interest: $p = 3$

N	λ	$\hat{\pi_1}$	$\hat{\pi_2}$	$\hat{\pi_3}$	$\hat{\pi_4}$	$\hat{\pi_5}$	$\hat{\pi_6}$
40	1	0.500 (0.148)	0.600 (0.074)	0.700 (0.074)	0.500 (0.074)	0.400 (0.148)	0.300 (0.074)
40	0.9	0.500 (0.117)	0.600 (0.109)	0.708 (0.101)	0.500 (0.117)	0.390 (0.115)	0.294 (0.118)
40	0.7	0.500 (0.124)	0.600 (0.099)	0.697 (0.093)	0.500 (0.124)	0.391 (0.134)	0.278 (0.142)
40	0.3	0.485 (0.126)	0.594 (0.108)	0.704 (0.089)	0.500 (0.124)	0.383 (0.144)	0.273 (0.152)
40	0	0.500 (0.126)	0.588 (0.105)	0.706 (0.087)	0.474 (0.126)	0.375 (0.164)	0.265 (0.146)
80	1	0.500 (0.074)	0.600 (0.074)	0.700 (0.074)	0.500 (0.074)	0.400 (0.074)	0.300 (0.074)
80	0.9	0.500 (0.082)	0.600 (0.076)	0.698 (0.066)	0.488 (0.083)	0.394 (0.079)	0.292 (0.091)
80	0.7	0.492 (0.089)	0.603 (0.071)	0.703 (0.058)	0.497 (0.089)	0.394 (0.095)	0.286 (0.112)
80	0.3	0.490 (0.091)	0.597 (0.069)	0.699 (0.062)	0.492 (0.094)	0.388 (0.116)	0.272 (0.136)
80	0	0.493 (0.096)	0.595 (0.076)	0.700 (0.059)	0.493 (0.096)	0.375 (0.116)	0.273 (0.156)
120	1	0.500 (0.049)	0.600 (0.049)	0.700 (0.049)	0.500 (0.074)	0.400 (0.074)	0.300 (0.049)
120	0.9	0.500 (0.067)	0.600 (0.059)	0.699 (0.048)	0.493 (0.069)	0.394 (0.075)	0.295 (0.076)
120	0.7	0.491 (0.073)	0.597 (0.056)	0.704 (0.044)	0.491 (0.069)	0.385 (0.083)	0.282 (0.112)
120	0.3	0.491 (0.081)	0.594 (0.060)	0.699 (0.044)	0.487 (0.083)	0.377 (0.115)	0.277 (0.114)
120	0	0.491 (0.078)	0.594 (0.060)	0.701 (0.047)	0.491 (0.084)	0.375 (0.120)	0.278 (0.144)

Note: See notes for Table 6.13

Table 6.16: Characteristics of Mean Squared Error (MSE) of $\boldsymbol{\theta}=(\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6)^T$: $p = 3$

N	λ	Trace(MSE)	Det(MSE)	Eigen(MSE)
40	1	0.1767575	3.58E-11	0.1136501
40	0.9	0.1769994	3.71E-11	0.1135412
40	0.7	0.1948637	7.72E-11	0.122277
40	0.3	0.21776	2.58E-10	0.1255553
40	0	0.2395586	5.89E-10	0.1322444
80	1	0.1385881	9.41E-13	0.1078347
80	0.9	0.1413767	1.21E-12	0.109019
80	0.7	0.1539715	6.14E-12	0.1090196
80	0.3	0.1858105	5.10E-11	0.1180341
80	0	0.2200243	2.35E-10	0.1302987
120	1	0.1228153	1.24E-13	0.1020549
120	0.9	0.1254579	1.67E-13	0.1034812
120	0.7	0.1418052	1.57E-12	0.1075898
120	0.3	0.1775834	2.50E-11	0.1187581
120	0	0.213967	1.60E-10	0.1301959

Note: See notes for Table 6.13

Chapter 7

Adaptive Repeated Measurement Design for Continuous Responses

In this chapter, we illustrate the application of the new allocation rule proposed in Chapter 5 for trials with continuous responses/outcomes. Based on the self and mixed carryover effects model proposed in Chapter 3, we construct an adaptive two-treatment two-period repeated measurement designs first, and then extend it to two-treatment three-period repeated measurement designs. In simulation studies, we demonstrate that the efficiency of the designs constructed under the new proposed allocation rule increases with sample size, and these adaptive designs are more efficient than fixed optimal designs in terms of the mean squared error. Finally, we discuss the challenges and difficulties in generalizing the results to arbitrary multi-treatment multi-period repeated measurement designs.

7.1 Adaptive Two-Treatment Repeated Measurement Design

7.1.1 Allocation Rule for Two-Period Repeated Measures Data

In an adaptive two-treatment two-period repeated measurement design, four different treatment sequences, AA, AB, BA and BB, are available for assignment. Suppose that N patients were randomly selected from a well-defined

population, and the first i patients were assigned using some optimal design suggested in the literature, a completely randomized design, for example. Let N_{1i}, N_{2i}, N_{3i} and N_{4i} be the number of patients who have received treatment sequence AA, AB, BA and BB, respectively, up to patient i.

The self and mixed carryover effect model for a two-treatment two-period repeated measurement design is defined as below.

$$\mathbf{y}_{jk} = \mathbf{X}_k \boldsymbol{\beta} + \boldsymbol{\xi}_j \mathbf{1}_{[2]} + \boldsymbol{\varepsilon}_{jk} \tag{7.1}$$

where $\mathbf{y}_{jk} = (y_{1jk}, y_{2jk})^T$ is the vector of observations from subject j in treatment sequence k, where k can be AA, AB, BA or BB, and $\mathbf{1}_{[2]}$ is a 2×1 vector of ones.

The parameter vector $\boldsymbol{\beta} = (\mu, \pi, \tau, \gamma, \varphi)^T$ consists of the overall mean effect μ, the period effect π (coefficient is 0 for the 1^{st} period, and coefficient is 1 for the 2^{nd} period), the direct treatment effect τ (coefficient is 1 if receiving treatment A, and coefficient is -1 if receiving treatment B), the mixed carryover effect γ (coefficient is 0 in the 1^{st} period, coefficient is 1 if receiving treatment B in the current period but receiving treatment A in the previous period, and coefficient is -1 if receiving treatment A in the current period but receiving treatment B in the previous period), and the self carryover effect φ (coefficient is 0 in the 1^{st} period, coefficient is 1 if receiving treatment A in both previous and current periods, and coefficient is -1 if receiving treatment B in both previous and current periods).

Therefore, the design matrix $\mathbf{X}_k$ for a given treatment sequence k is defined as follows.

$$\begin{aligned}
\mathbf{X}_{AA} &= \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 \end{pmatrix} \\
\mathbf{X}_{AB} &= \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & -1 & 1 & 0 \end{pmatrix} \\
\mathbf{X}_{BA} &= \begin{pmatrix} 1 & 0 & -1 & 0 & 0 \\ 1 & 1 & 1 & -1 & 0 \end{pmatrix} \\
\mathbf{X}_{BB} &= \begin{pmatrix} 1 & 0 & -1 & 0 & 0 \\ 1 & 1 & -1 & 0 & -1 \end{pmatrix}
\end{aligned} \tag{7.2}$$

$\boldsymbol{\xi}_j$ are random subject effects, assumed to have a multi-normal distribution with mean $\mathbf{0}$ and a variance-covariance matrix $\sigma_\xi^2 \mathbf{1}_{[2]} \mathbf{1}_{[2]}^T$. We assume

that $\boldsymbol{\xi}_j$ is independent of the random error $\boldsymbol{\varepsilon}_{jk} = (\varepsilon_{1jk}, \varepsilon_{2jk})^T$, which follows a multi-normal distribution with mean $\mathbf{0}$ and a variance-covariance matrix $\sigma_\varepsilon^2 \mathbf{I}_{[2]}$, where $\mathbf{I}_{[2]}$ is a 2×2 identity matrix.

The variance-covariance matrix of the vector $\mathbf{y}_{jk}$, $\mathbf{C}$, is then written as

$$Var(\mathbf{y}_{jk}) \triangleq \mathbf{C} = \sigma_\varepsilon^2 \mathbf{I}_{[2]} + \sigma_\xi^2 \mathbf{1}_{[2]} \mathbf{1}_{[2]}^T \tag{7.3}$$

Up to the i^{th} patient, based on the current observations, $\mathbb{H}_i$, the estimated information matrix up to the i^{th} patient is

$$\widehat{A}_i = \sum_{k \in \mathbb{H}_i} N_k \mathbf{X}_k^T \widehat{\mathbf{C}}^{-1} \mathbf{X}_k = \sum_{k \in \mathbb{H}_i} N_k \mathbf{X}_k^T (\hat{\sigma}_\varepsilon^2 \mathbf{I}_{[2]} + \hat{\sigma}_\xi^2 \mathbf{1}_{[2]} \mathbf{1}_{[2]}^T)^{-1} \mathbf{X}_k \tag{7.4}$$

where $\hat{\sigma}_\varepsilon^2$ and $\hat{\sigma}_\xi^2$ are restricted maximum likelihood estimates (REML) for σ_ε^2 and σ_ξ^2, respectively, using the EM algorithm proposed by Laird and Ware (1982).

Then the estimated information matrix, given the history $\mathbb{H}_i$ and the assumption that the $(i+1)^{th}$ patient receiving the treatment sequence AA, AB, BA or BB, will be defined as below, respectively.

$$\begin{aligned}
\widehat{A}_{i+1}^{AA}(\mathbb{H}_i) &= \widehat{A}_i + \mathbf{X}_{AA}^T \widehat{\mathbf{C}}^{-1} \mathbf{X}_{AA} \\
\widehat{A}_{i+1}^{AB}(\mathbb{H}_i) &= \widehat{A}_i + \mathbf{X}_{AB}^T \widehat{\mathbf{C}}^{-1} \mathbf{X}_{AB} \\
\widehat{A}_{i+1}^{BA}(\mathbb{H}_i) &= \widehat{A}_i + \mathbf{X}_{BA}^T \widehat{\mathbf{C}}^{-1} \mathbf{X}_{BA} \\
\widehat{A}_{i+1}^{BB}(\mathbb{H}_i) &= \widehat{A}_i + \mathbf{X}_{BB}^T \widehat{\mathbf{C}}^{-1} \mathbf{X}_{BB}
\end{aligned} \tag{7.5}$$

At the same time, an evaluation function, $g_{i,k}$, is defined to evaluate the quality of the treatment sequence k based on the information of the first i patients. To simplify, we assume a larger value of a response indicates a better treatment sequence and all the responses are nonnegative. Then the summation of all outcomes from a given treatment sequence is a straightforward way to define an evaluation function.

For the predetermined value of λ, the treatment sequence for the $(i+1)^{th}$ patient is determined by maximizing the criterion Λ as defined earlier in Chapter 5, Equation 5.5, i.e.,

$$\Lambda = \lambda \frac{\Theta(\widehat{A_{i+1}^k}(\mathbb{H}_i))}{\Theta(\widehat{A_{i+1}^{k^{(a)}}}(\mathbb{H}_i))} + (1-\lambda) \frac{g_{i,k}}{g_{i,k^{(b)}}}$$

where $\Theta(.)$ is the optimality criteria function such as the determinant (D-optimality), the trace (A-optimality) or the maximum eigenvalue (E-optimality) of the information matrix, $k^{(a)}$ is the treatment sequence which maximizes $\Theta(A_{i+1}^{k}(\mathbb{H}_i))$, and $k^{(b)}$ is the treatment sequence which maximizes $g_{k,i}$.

Repeat the same technique until all N patients have been allocated.

7.1.2 Allocation Rule for Three-Period Repeated Measures Data

The allocation rule for an adaptive two-treatment three-period repeated measurement design is a natural extension of that for a two-treatment two-period design. Eight different treatment sequences are available for assignment. We assume at the initial stage, first i patients are entered into the study. Let N_{ki} be the number of patients receiving treatment sequence k, where $k = AAA$, AAB, ABA, ABB, BBB, BBA, BAB and BAA.

The self and mixed carryover effects model for a two-treatment three-period repeated measurement design can be written as

$$y_{ij} = \mu + \pi_i + \tau_{d[i,j]} + (1 - \delta_{ij})\gamma_{d[i-1,j]} + \delta_{ij}\varphi_{d[i-1,j]} + \xi_j + \varepsilon_{ij} \qquad (7.6)$$

where y_{ij} denotes the response variable for subject j in period i, μ is an overall mean, π_i and ξ_j are the period and subject effects, respectively, and $d(i,j)$ denotes the treatment used for subject j in period i. Both $\gamma_{d[i-1,j]}$ and $\varphi_{d[i-1,j]}$ represent carryover effects, while δ_{ij} is an indicator variable, taking 1 if $d(i,j) = d(i-1,j)$ and 0 otherwise. Thus $\gamma_{d[i-1,j]}$ is the mixed carryover effect, while $\varphi_{d[i-1,j]}$ is the self carryover effect, with $\gamma_{d[0,j]} = \varphi_{d[0,j]} = 0$

A matrix format of the model (7.6) is

$$\mathbf{y}_{jk} = \mathbf{X}_k\, \boldsymbol{\beta} + \boldsymbol{\xi}_j \mathbf{1}_{[3]} + \boldsymbol{\varepsilon}_{jk} \qquad (7.7)$$

where $\mathbf{y}_{jk} = (y_{1jk}, y_{2jk}, y_{3jk})^T$ is the vector of observations from subject j in treatment sequence k, and $\mathbf{1}_{[3]}$ is a 3×1 vector of ones.

The parameter vector $\boldsymbol{\beta} = (\mu,\ \pi_2, \pi_3, \tau, \gamma, \varphi)^T$ consists of the overall mean effect μ, the second period effect π_2, the third period effect π_3, the

direct treatment effects τ, the mixed carryover effect γ and the self carryover effect φ.

$\boldsymbol{\xi}_j$ are random subject effects, assumed to have a multi-normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\sigma_\xi^2 \mathbf{1}_{[3]} \mathbf{1}_{[3]}^T$. We assume that $\boldsymbol{\xi}_j$ is independent of the random error $\boldsymbol{\varepsilon}_{jk} = (\varepsilon_{1jk}, \varepsilon_{2jk}, \varepsilon_{3jk})^T$, which follows a multi-normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\sigma_\varepsilon^2 \mathbf{I}_{[3]}$, where $\mathbf{I}_{[3]}$ is a 3×3 identity matrix.

Then we have the design matrix $\mathbf{X}_k$ for a given treatment sequence k, defined as follows.

$$\mathbf{X}_{AAA} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}$$

$$\mathbf{X}_{AAB} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & -1 & 1 & 0 \end{pmatrix}$$

$$\mathbf{X}_{ABA} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & -1 & 1 & 0 \\ 1 & 0 & 1 & 1 & -1 & 0 \end{pmatrix}$$

$$\mathbf{X}_{ABB} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & -1 & 1 & 0 \\ 1 & 0 & 1 & -1 & 0 & -1 \end{pmatrix}$$

$$\mathbf{X}_{BBB} = \begin{pmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 1 & 1 & 0 & -1 & 0 & -1 \\ 1 & 0 & 1 & -1 & 0 & -1 \end{pmatrix}$$

$$\mathbf{X}_{BBA} = \begin{pmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 1 & 1 & 0 & -1 & 0 & -1 \\ 1 & 0 & 1 & 1 & -1 & 0 \end{pmatrix}$$

$$\mathbf{X}_{BAB} = \begin{pmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 1 & 1 & 0 & 1 & -1 & 0 \\ 1 & 0 & 1 & -1 & 1 & 0 \end{pmatrix}$$

$$\mathbf{X}_{BAA} = \begin{pmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 1 & 1 & 0 & 1 & -1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \end{pmatrix} \tag{7.8}$$

Similar to the two-treatment two-period RMD, under the equicorrelated covariance assumption, the covariance matrix of the vector $\mathbf{y}_{jk}$, $\mathbf{C}$, is then

$$\mathbf{C} = \sigma_\varepsilon^2 \mathbf{I}_{[3]} + \sigma_\xi^2 \mathbf{1}_{[3]} \mathbf{1}_{[3]}^T \tag{7.9}$$

And the estimated information matrix up to the i^{th} patient is

$$\widehat{A}_i = \sum_{k \in \mathbb{H}_i} N_k \mathbf{X}_k^T \widehat{\mathbf{C}}^{-1} \mathbf{X}_k = \sum_{k \in \mathbb{H}_i} N_k \mathbf{X}_k^T (\hat{\sigma}_\varepsilon^2 \mathbf{I}_{[3]} + \hat{\sigma}_\xi^2 \mathbf{1}_{[3]} \mathbf{1}_{[3]}^T)^{-1} \mathbf{X}_k \qquad (7.10)$$

And then the estimated information matrix, given the $i+1$ patient receiving the treatment sequence k, will become

$$\widehat{A}_{i+1}^k(\mathbb{H}_i) = \widehat{A}_i + \mathbf{X}_k^T \widehat{\mathbf{C}}^{-1} \mathbf{X}_k$$

Let $g_{i,k}$ be an evaluation function for treatment sequence k up to the i^{th} patients. To simplify, we assume a larger value of $g_{i,k}$ indicates a better treatment sequence. For a predetermined value of λ, the treatment sequence for the $(i+1)^{th}$ patient is determined by maximizing the criterion Λ as defined earlier in Chapter 5, Equation 5.5, i.e.,

$$\Lambda = \lambda \frac{\Theta(\widehat{A_{i+1}^k}(\mathbb{H}_i))}{\Theta(\widehat{A_{i+1}^{k^{(a)}}}(\mathbb{H}_i))} + (1-\lambda) \frac{g_{i,k}}{g_{i,k^{(b)}}}$$

where $\Theta(.)$, $k^{(a)}$ and $k^{(b)}$ are defined the same as before.

Repeat the same technique until all N patients have been allocated.

7.1.3 Simulation Study

Two-Period Designs

We first study the properties of the allocation rule proposed in the previous section for two-period repeated measurement designs by simulations. Suppose at the initial stage, four patients have entered the study and each of them receives one of the four different treatment sequences (AA, AB, BA or BB). We then consider how to allocate the rest of the patients adaptively according to the observed data from these four patients.

In simulations, suppse $\sigma_\xi^2 = 2$, $\sigma_\varepsilon^2 = 1$, and $\mu = 100$. We choose $\lambda = 1, 0.9, 0.7, 0.3$ and 0, and $N = 10, 20, 40$ and 100, respectively, where N is the total number of patients in the study. $\lambda = 1$ indicates we only consider the objective of increasing the estimation precision, i.e., maximizing the information matrix, while, $\lambda = 0$ indicates the only objective of the design is to increase the proportion of patients assigned to a better treatment. When

$0 < \lambda < 1$, both objectives are taken into consideration, and the adaptive design will balance these two objectives according to the researchers' request. One thousand simulation data were generated for each situation. The R program code is available upon request.

For $\pi = \tau = \gamma = \varphi = 0$, Table 7.1 shows, for all combinations of N and λ values, the trial will assign an approximately equal number of subjects to each of the four treatment sequences. In addition, the estimation of each parameter with its standard error (reported in the bracket) is summarized in Table 7.2, which indicates when sample sizes increase, standard errors of the estimations of the parameters of interest decrease. Given a fixed number of patients, the estimations of the parameters and their standard errors are quite similar among different λ values, and they are all very close to the true values.

For $\pi = \tau = \varphi = 25$ and $\gamma = -25$, Table 7.3 shows when $\lambda = 1$, we will assign an equal number of subjects to each of the four treatment sequences. When $\lambda < 1$, more patients will be assigned to treatment sequence AA. The rest of the patients, in decreasing order, will receive treatments BA, AB or BB. Based on the values of the parameters of interest, treatment A is more effective than treatment B ($\tau > 0$); the treatment effect in the second period is stronger than that in the first period ($\pi > 0$); the self carryover effect for the treatment sequence AA is stronger than that for treatment sequence BB ($\varphi > 0$), and the mixed carryover effect for treatment sequence BA is stronger than that for treatment sequence AB ($\gamma < 0$). Therefore, the allocation results from the simulations are quite consistent with what one can expect (see Table 7.4).

The estimation of each parameter with its standard error in the bracket is summarized in Table 7.5. It indicates that in all cases, the estimated values are very close to the true values of the parameters of interest. For a fixed value of λ, when N increases, the precision of the estimation decreases. For a fixed value of N, the standard error slightly increases when the value of λ decreases. This happened because when λ decreased we gave more concern to the ethical criterion rather than the precision of the estimators. It is a

trade-off between benefit and cost. However, this result does not hold when $N = 10$. It seems that when sample sizes are small, the precision of the estimation does not vary, whether the ethical issues are taken into account or not.

Three-Period Designs

We then consider constructing three-period repeated measurement designs using the new adaptive allocation rule. Assume that 8 subjects were already entered in the study, one for each type of treatment sequence. Let $\sigma_\xi^2 = 2$, $\sigma_\varepsilon^2 = 1$, and $\mu = 100$. We choose $\lambda = 1, 0.9, 0.7, 0.3$ and 0, and $N = 40, 80$ and 120, respectively. One thousand simulation data were generated for each situation. To assess the efficiency of a design, we still use the mean squared error method described in Chapter 6.

Under the traditional model with an equi-correlated covariance structure, the design ABB/BAA is known to be the universally optimal design (Laska, Meisner and Kushner 1983, Kershner 1986), as shown in Chapter 3, while, under the self and mixed carryover effects model, as shown in Chapter 3, the design ABA/BAB is optimal for estimating the treatment difference. In this section, we also compare the adaptive designs constructed under the new allocation to the fixed designs, design ABB/BAA and design ABA/BAB.

No Treatment Difference

For $\pi_2 = \pi_3 = \tau = \gamma = \varphi = 0$, Table 7.6 shows when $\lambda < 1$, we will assign an approximately equal number of subjects to each of the 8 treatment sequences. When $\lambda = 1$, we will assign an approximately equal number of subjects to a treatment sequence and its dual treatment sequence. Most subjects were given ABB/BAA, which is, by the way, a very popular design in clinical trials for comparing three treatments.

Table 7.7 summarizes the estimation of each parameter with its standard error. It shows that when N increases, the standard error of the estimation of each parameter of interest decreases. For estimation of the treatment effect, τ, the designs with $\lambda = 1$ provide the smallest standard errors as expected. However, other designs also provide quite accurate estimates with slightly larger standard errors.

Table 7.8 shows the trace, the determinant and the maximum eigenvalue of the mean squared error (MSE) matrix for designs with various values of λ and for fixed designs ABA/BAB and ABB/BAA, where smaller entries indicate more efficient designs. Note that, for design ABA/BAB, the trace, the determinant and the maximum eigenvalue of MSE for the estimation of $\boldsymbol{\theta}=(\tau, \gamma, \varphi)^T$ are not applicable, because the self carryover effect φ is not estimable in this case.

Figures 7.1, 7.2, 7.3 and 7.4 plot the estimated relative efficiency (RE) with design ABB/BAA as the reference design for estimation of $\boldsymbol{\theta}=(\tau, \gamma, \varphi)^T$ under A-, D-, and E-optimality, and for estimation treatment effect τ, respectively, where RE > 1 indicates a more efficient design than the reference design. In all cases, adaptive designs increase the design efficiency 1.5 to 4 times when compared with the fixed design ABB/BAA. For estimation of the direct treatment contrast, design ABA/BAB has the highest efficiency. However, the new proposed adaptive designs can provide similar high efficiency while also taking the treatment performance into consideration, which is superior to fixed designs.

With Treatment Difference

For $\pi_2 = \pi_3 = \tau = \varphi = 25$ and $\gamma = -25$, Table 7.9 shows that, as before, when $\lambda = 1$, we will assign an approximately equal number of subjects to a treatment sequence and its dual treatment sequence, and most subjects were given ABB/BAA. However, when $\lambda < 1$, we will assign more subjects to treatment AAA and less subjects to treatment BBB, as λ decreases.

Table 7.10 summarizes the estimation of each parameter with its standard error in the bracket. It shows that when N increases, the standard error of the estimation of the parameter decreases. For estimation of the treatment contrast, τ, the design with $\lambda = 1$ provides the smallest standard error since we only focus on the estimation precision in this situation. However, other adaptive designs with $\lambda < 1$ also provide quite accurate estimations, with slightly larger standard errors.

Table 7.11 shows the trace, the determinant and the maximum eigenvalue of the mean squared error (MSE) matrix for designs with various values of

λ and for the fixed design ABA/BAB and design ABB/BAA. As before, smaller values indicate more efficient designs, and for design ABA/BAB, the trace, the determinant and the maximum eigenvalue of MSE for the estimation of $\boldsymbol{\theta}=(\tau, \gamma, \varphi)^T$ are not applicable.

Figures 7.5, 7.6, 7.7 and 7.8 plot the estimated relative efficiency (RE) with design ABB/BAA as the reference design for the estimation of $\boldsymbol{\theta}$ under A-, D-, and E-optimality, respectively, and for estimation treatment effect τ. As before, RE > 1 indicates a more efficient design than the reference design. In all cases, adaptive designs increase the design efficiency 1.5 to 4 times when compared with the fixed design ABB/BAA. For estimation of the direct treatment contrast, design ABA/BAB has the highest efficiency. However, the new proposed adaptive designs can provide similar high efficiency, especially when the total number of subjects is large. It means that those adaptive designs constructed under the new adaptive allocation rule not only take the treatment performance into consideration, but also have relatively high efficiency, which makes them more attractive.

7.2 Generalization

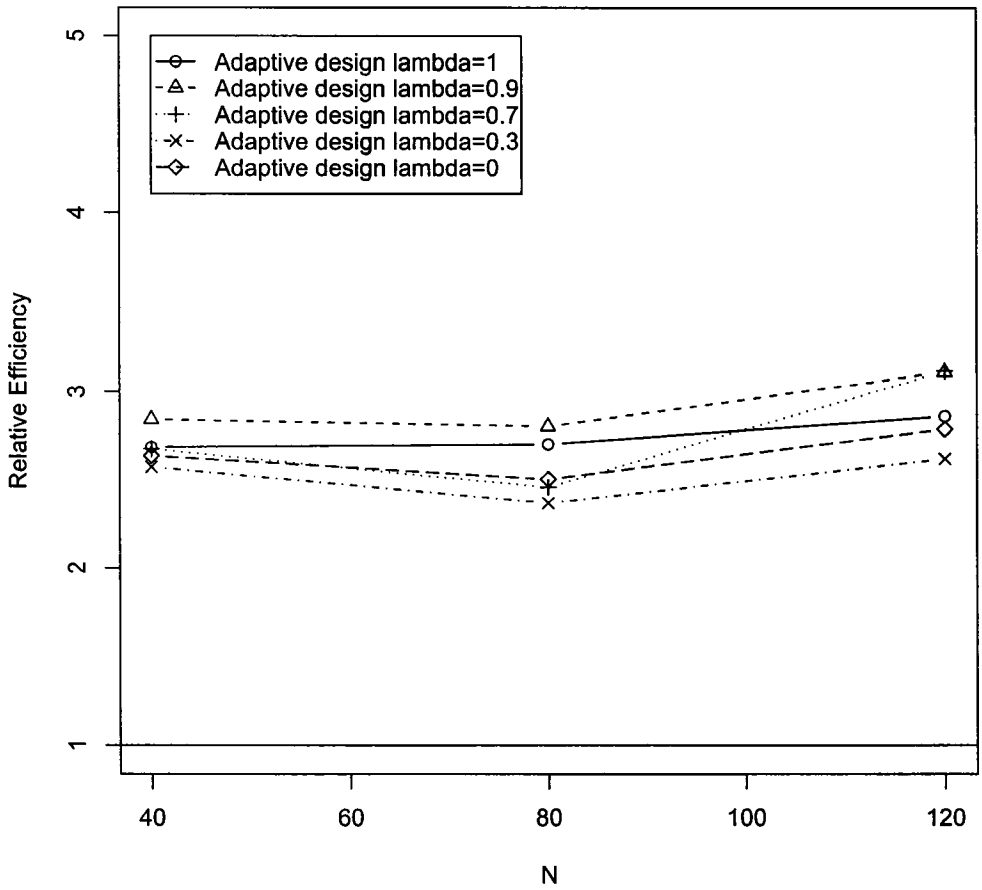
One can generalize the allocation rule to construct adaptive t-Treatment p-Period repeated measurement designs. However, similar to the case for dichotomous responses discussed in Chapter 6, the main challenge is to narrow down the number of treatment sequences out of t^p possibilities, which increase substantially as the number of treatments and periods increase. One can consider a particular subset of RMDs, for example uniform cross-over designs (Bate and Jones, 2003). Or one can refer to the fixed optimal design results available in the literature (Ebbutt 1984, Kershner 1986, Matthews 1987, Carriere and Reinsel 1992, 1993, Carriere 1994, Hedayat and Stufken 2003). For example, as shown in Chapter 3, the optimal two-treatment four-period design for estimating the treatment contrast based on the self and mixed carryover effects model is the design $ABBA/BAAB$ and $AABA/BBAB$; based on the traditional model is the design $ABBA/BAAB$ and $AABB/BBAA$, or

the design $ABBA/BAAB$, $ABAB/BABA$ and $AABB/BBAA$. Then the allocation rule can be applied to a smaller subset containing the treatment sequences of interest to construct the adaptive design.

7.3 Conclusion

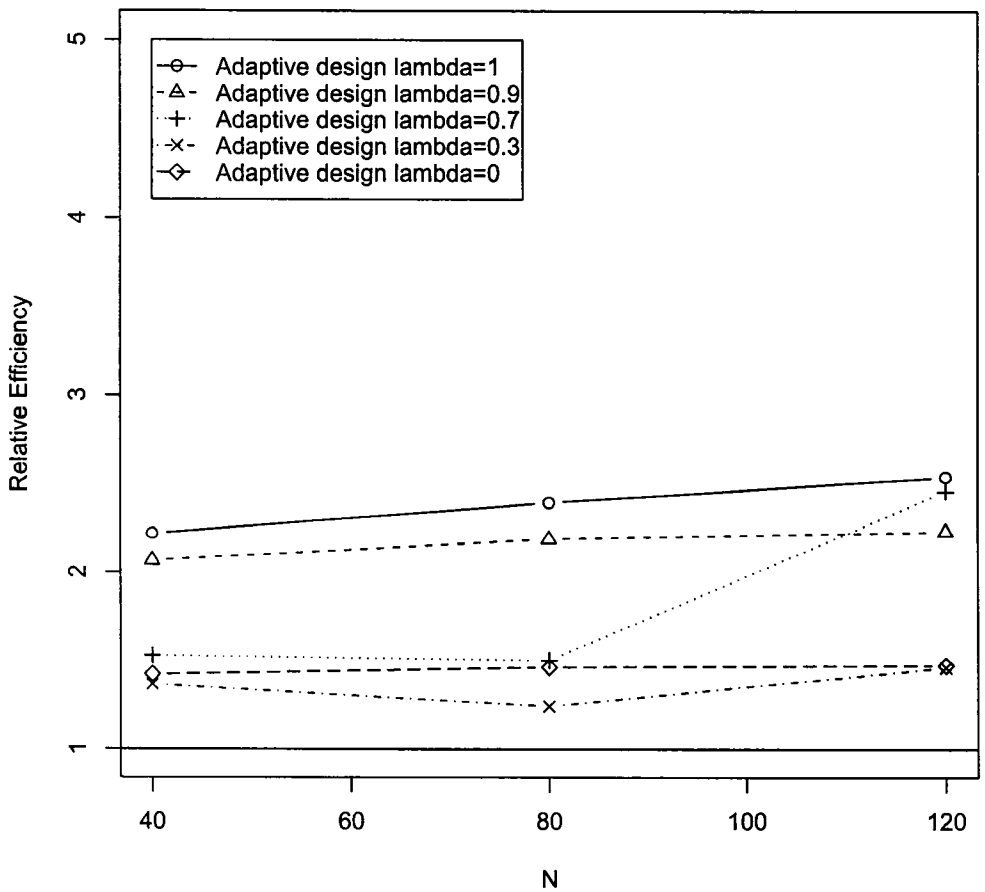
In this chapter, we use the allocation rule proposed in Chapter 5 to construct adaptive repeated measurement designs with continuous responses/outcomes, based on the self and mixed carryover effects model. We provide detailed allocation rule for constructing adaptive two-treatment two-period repeated measurement designs, and then extend it to two-treatment three-period repeated measurement designs. In simulation studies, we demonstrate that the efficiency of the designs constructed under the new proposed allocation rule increases with sample size. Moreover, those adaptive designs are more efficient than fixed design ABB/BAA in terms of the mean squared error. Similar to the dichotomous responses case in Chapter 6, the simulation study shows that designs with a high value of λ, say $\lambda = 0.9$, significantly skew the allocation results toward more effective treatment sequences without the loss of much estimation precision.

Figure 7.1: Relative Efficiency of $\boldsymbol{\theta}$ under A-optimality: no treatment difference



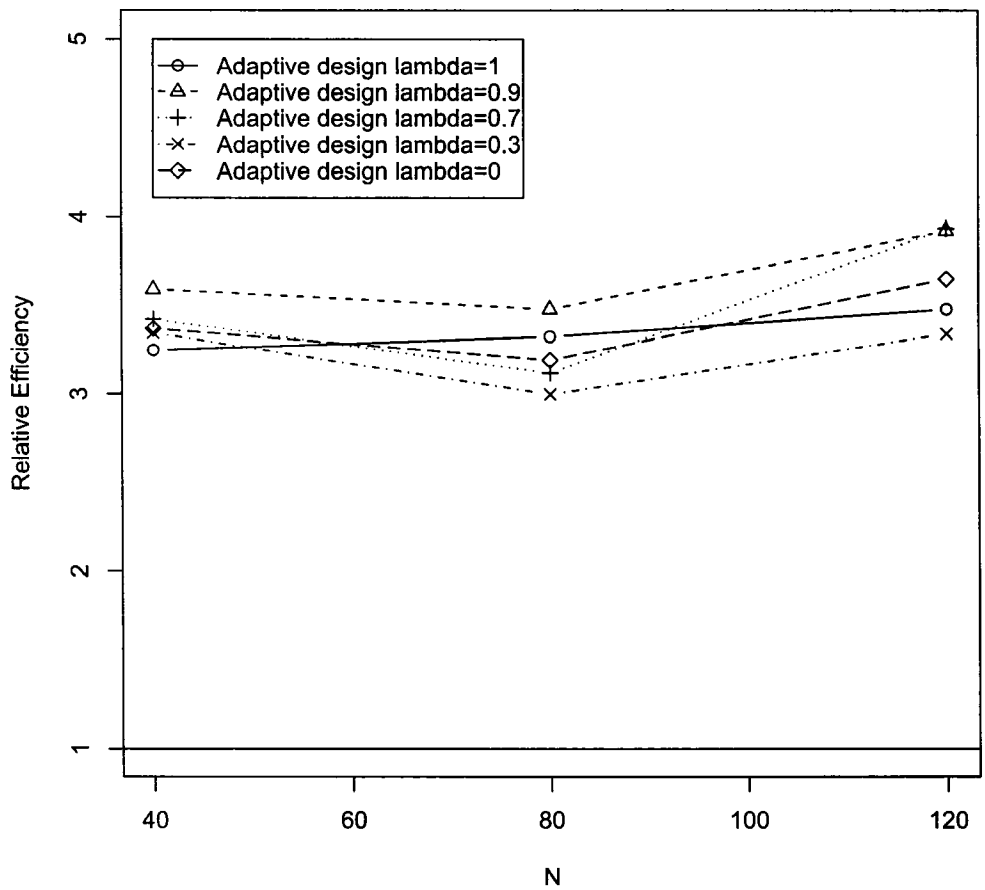
Note: $\boldsymbol{\theta}=(\tau,\gamma,\varphi)^T$. Design ABB/BAA is the reference design. Relative efficiencies are calculated based on 1,000 simulations under 2-treatment 3-period RMDs with $\pi_2=\pi_3=\tau=\gamma=\varphi=0$, $\sigma_\xi^2=2$, $\sigma_\epsilon^2=1$ and $\mu=100$

Figure 7.2: Relative Efficiency of $\boldsymbol{\theta}$ under D-optimality: no treatment difference



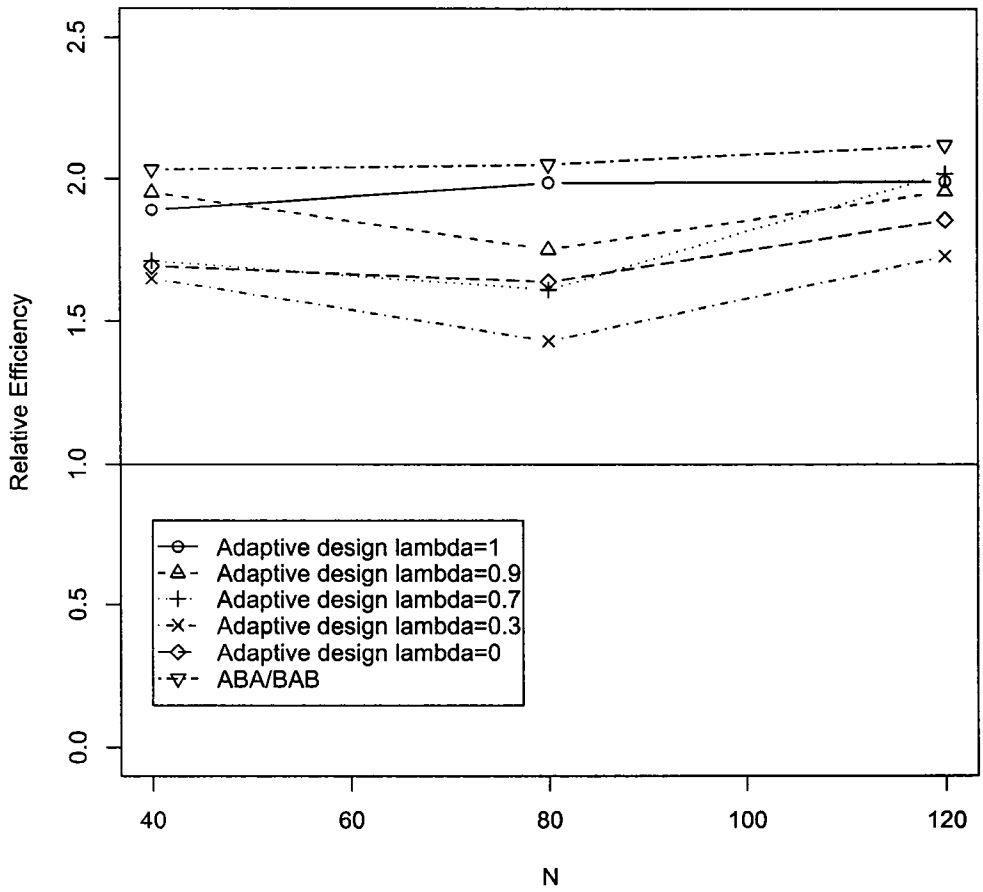
Note: see notes for Figure 7.1

Figure 7.3: Relative Efficiency of $\boldsymbol{\theta}$ under E-optimality: no treatment difference



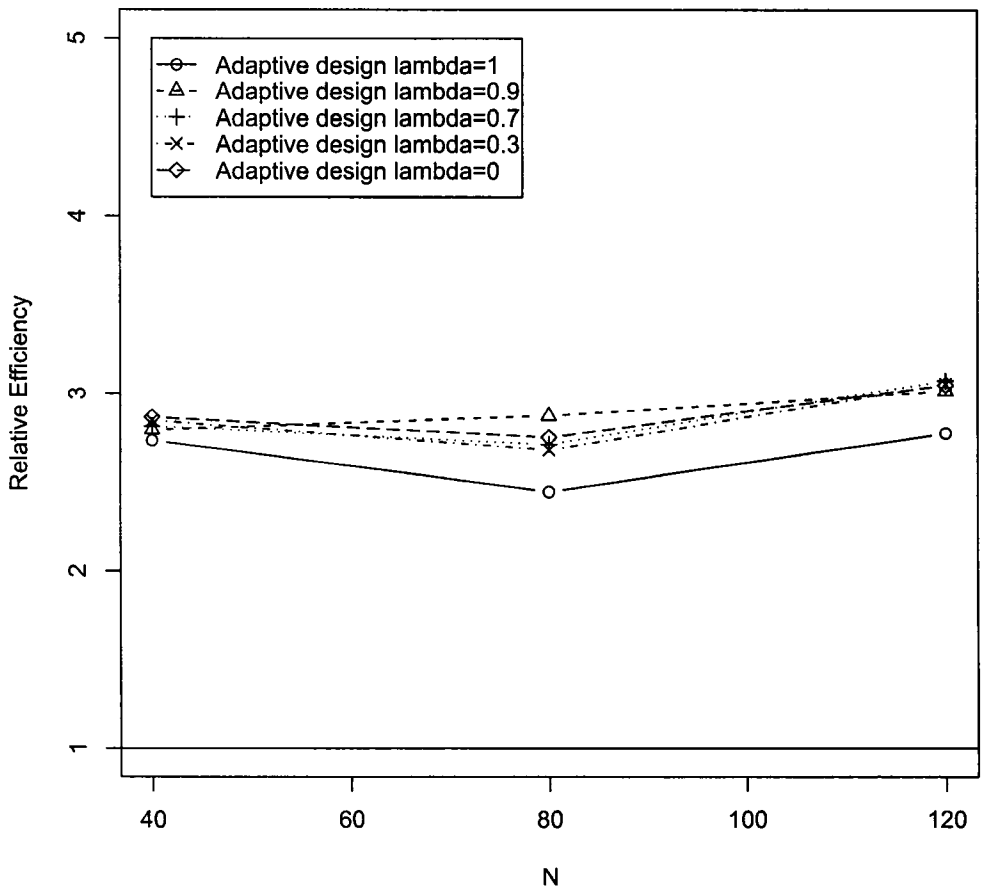
Note: see notes for Figure 7.1

Figure 7.4: Relative Efficiency of τ: no treatment difference



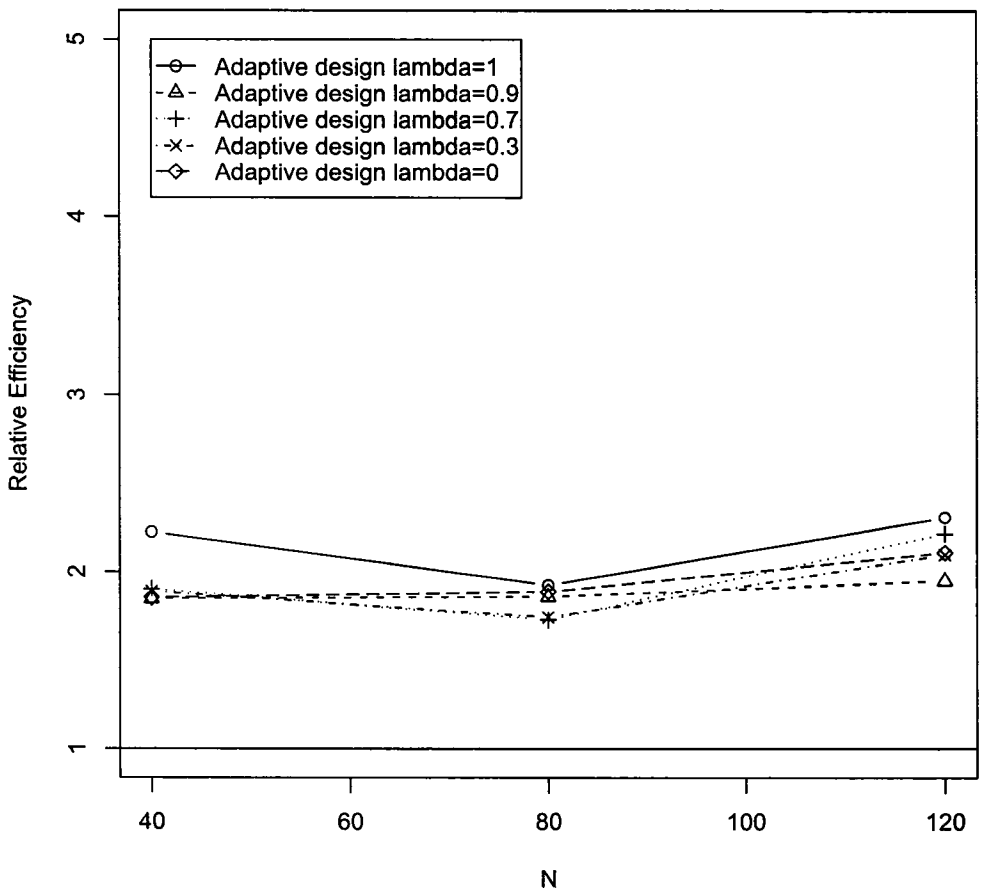
Note: Design ABB/BAA is the reference design. Relative efficiencies are calculated based on 1,000 simulations under 2-treatment 3-period RMDs with $\pi_2 = \pi_3 = \tau = \gamma = \varphi = 0$, $\sigma_\xi^2 = 2$, $\sigma_\epsilon^2 = 1$ and $\mu = 100$

Figure 7.5: Relative Efficiency of $\boldsymbol{\theta}$ under A-optimality



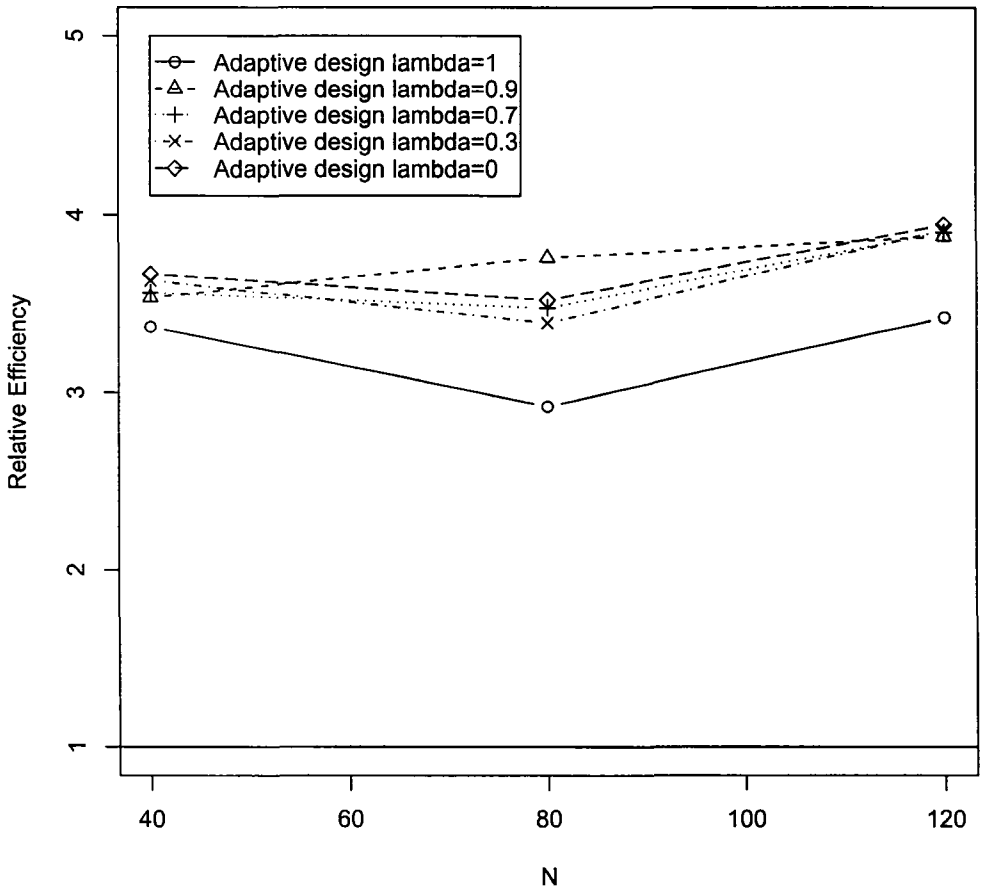
Note: $\boldsymbol{\theta}=(\tau,\gamma,\varphi)^T$. Design ABB/BAA is the reference design. Relative efficiencies are calculated based on 1,000 simulations under 2-treatment 3-period RMDs with $\pi_2 = \pi_3 = \tau = \varphi = 25$, $\gamma = -25$, $\sigma_\xi^2 = 2$, $\sigma_\epsilon^2 = 1$ and $\mu = 100$

Figure 7.6: Relative Efficiency of $\boldsymbol{\theta}$ under D-optimality



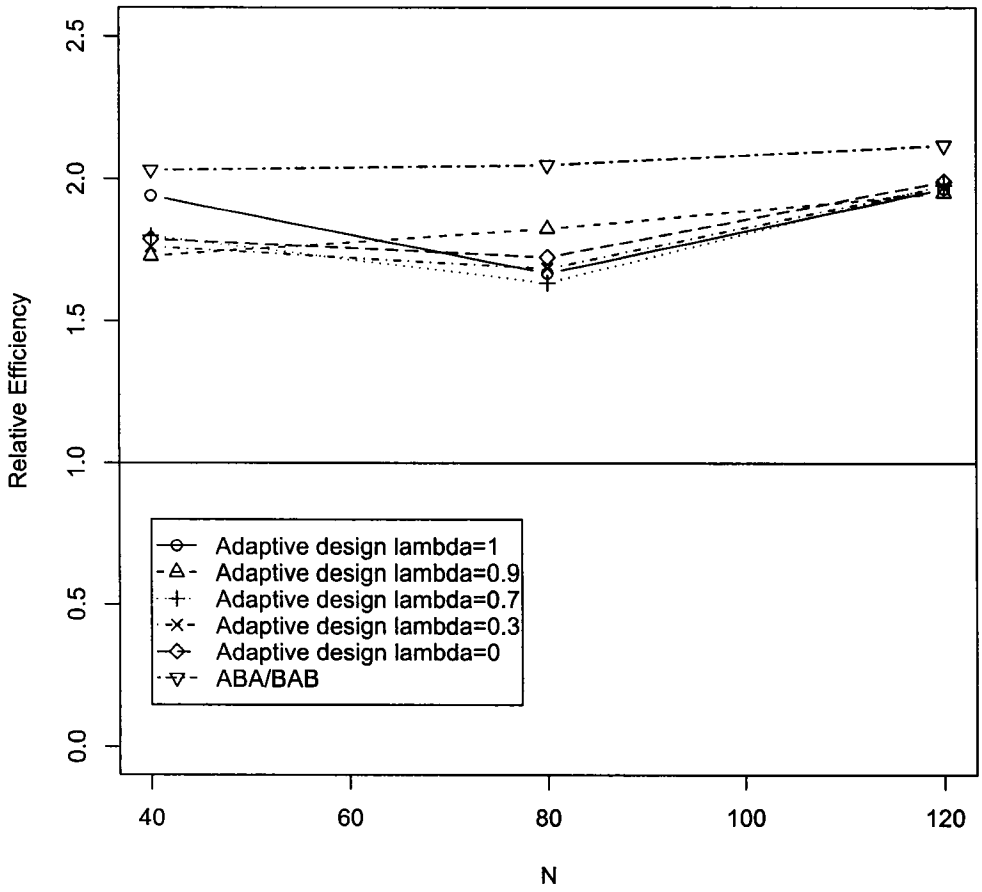
Note: see notes for Figure 7.5

Figure 7.7: Relative Efficiency of $\boldsymbol{\theta}$ under E-optimality



Note: see notes for Figure 7.5

Figure 7.8: Relative Efficiency of τ



Note: Design ABB/BAA is the reference design. Relative efficiencies are calculated based on 1,000 simulations under 2-treatment 3-period RMDs with $\pi_2 = \pi_3 = \tau = \varphi = 25$, $\gamma = -25$, $\sigma_\xi^2 = 2$, $\sigma_\epsilon^2 = 1$ and $\mu = 100$

Table 7.1: Estimated Numbers of Patients for Each Treatment Sequence: $p = 2$, no treatment difference

N	λ	N_{AA}	N_{AB}	N_{BA}	N_{BB}
10	1	2.502	2.498	2.466	2.534
10	0.9	2.497	2.503	2.510	2.490
10	0.7	2.501	2.499	2.488	2.512
10	0.3	2.479	2.518	2.491	2.512
10	0	2.527	2.506	2.484	2.483
40	1	10.000	10.000	10.000	10.000
40	0.9	10.000	10.000	10.000	10.000
40	0.7	10.000	10.000	10.000	10.000
40	0.3	10.000	10.000	10.000	10.000
40	0	10.000	10.000	10.000	10.000
80	1	20.000	20.000	20.000	20.000
80	0.9	19.998	20.007	20.000	19.995
80	0.7	19.997	20.009	19.999	19.995
80	0.3	20.001	19.995	20.002	20.002
80	0	19.995	20.000	19.998	20.007
100	1	25.000	25.000	25.000	25.000
100	0.9	25.008	24.991	25.001	25.000
100	0.7	24.999	24.999	24.997	25.005
100	0.3	25.004	24.986	24.997	25.013
100	0	24.971	25.014	25.006	25.009

Note: Entries are based on 1,000 simulations under 2-treatment 2-period RMDs with $\pi = \tau = \gamma = \varphi = 0$, $\sigma_\xi^2 = 2$, $\sigma_\epsilon^2 = 1$ and $\mu = 100$

Table 7.2: Estimated Parameters of Interest: $p = 2$, no treatment difference

N	λ	μ	π	τ	γ	φ	σ_ξ^2	σ_ε^2
10	1	99.980 (0.532)	0.009 (0.237)	0.042 (0.582)	0.096 (1.143)	-0.008 (0.627)	2.018	0.992
10	0.9	99.974 (0.534)	0.010 (0.236)	-0.017 (0.583)	-0.015 (1.145)	0.012 (0.631)	2.035	0.987
10	0.7	99.969 (0.534)	0.001 (0.238)	-0.015 (0.583)	-0.028 (1.146)	-0.003 (0.632)	2.027	0.998
10	0.3	99.973 (0.536)	-0.003 (0.238)	-0.011 (0.585)	-0.023 (1.149)	0.014 (0.634)	2.048	0.998
10	0	99.964 (0.525)	0.006 (0.236)	0.033 (0.575)	0.041 (1.128)	0.007 (0.625)	1.956	0.979
40	1	99.998 (0.255)	0.000 (0.113)	0.004 (0.279)	0.020 (0.550)	0.001 (0.306)	2.029	1.002
40	0.9	99.986 (0.253)	0.004 (0.113)	-0.003 (0.277)	0.000 (0.547)	-0.004 (0.305)	2.009	0.989
40	0.7	100.013 (0.255)	0.001 (0.114)	-0.005 (0.279)	-0.008 (0.550)	-0.001 (0.307)	2.026	1.008
40	0.3	99.996 (0.253)	-0.003 (0.114)	-0.007 (0.277)	-0.001 (0.546)	0.009 (0.306)	1.993	1.006
40	0	99.988 (0.252)	0.005 (0.114)	-0.012 (0.277)	-0.016 (0.545)	-0.013 (0.307)	1.971	1.009
80	1	99.993 (0.178)	0.000 (0.080)	-0.008 (0.195)	-0.010 (0.384)	0.003 (0.215)	2.000	1.001
80	0.9	100.001 (0.178)	-0.001 (0.080)	-0.006 (0.195)	-0.017 (0.385)	-0.001 (0.216)	2.004	1.007
80	0.7	100.000 (0.178)	-0.002 (0.079)	-0.001 (0.195)	0.001 (0.385)	-0.010 (0.214)	2.016	0.993
80	0.3	99.994 (0.178)	-0.005 (0.079)	-0.005 (0.195)	-0.003 (0.385)	0.007 (0.215)	2.018	0.995
80	0	99.993 (0.178)	-0.001 (0.080)	0.005 (0.195)	0.012 (0.384)	-0.004 (0.215)	2.000	1.002
100	1	99.998 (0.159)	-0.002 (0.071)	0.009 (0.174)	0.012 (0.344)	0.001 (0.192)	2.007	1.003
100	0.9	100.007 (0.159)	0.002 (0.071)	0.006 (0.174)	0.015 (0.343)	0.005 (0.193)	2.001	1.003
100	0.7	99.998 (0.159)	0.001 (0.071)	0.000 (0.174)	0.012 (0.343)	-0.001 (0.193)	1.995	1.006
100	0.3	100.007 (0.159)	-0.002 (0.071)	-0.007 (0.174)	-0.014 (0.344)	0.000 (0.192)	2.019	0.994
100	0	99.997 (0.159)	0.000 (0.071)	-0.008 (0.174)	-0.011 (0.343)	0.002 (0.192)	1.999	1.000

Note: Entries are estimated values (standard errors) based on 1,000 simulations under 2-treatment 2-period RMDs with $\pi = \tau = \gamma = \varphi = 0$, $\sigma_\xi^2 = 2$, $\sigma_\varepsilon^2 = 1$ and $\mu = 100$

Table 7.3: Estimated Numbers of Patients for Each Treatment Sequence: $p = 2$

N	λ	N_{AA}	N_{AB}	N_{BA}	N_{BB}
10	1	2.519	2.481	2.441	2.559
10	0.9	3.000	2.000	3.000	2.000
10	0.7	3.000	2.000	3.000	2.000
10	0.3	3.000	2.000	3.000	2.000
10	0	3.000	2.000	3.000	2.000
40	1	10.000	10.000	10.000	10.000
40	0.9	12.000	9.000	11.000	8.000
40	0.7	13.000	9.000	11.001	6.999
40	0.3	13.416	8.986	11.030	6.568
40	0	13.717	8.979	11.022	6.282
80	1	20.000	20.000	20.000	20.000
80	0.9	25.547	17.930	22.405	14.118
80	0.7	26.810	17.693	22.451	13.046
80	0.3	27.127	17.624	22.426	12.823
80	0	27.243	17.602	22.374	12.781
100	1	25.000	25.000	25.000	25.000
100	0.9	32.299	22.197	28.103	17.401
100	0.7	33.664	22.021	28.089	16.226
100	0.3	34.051	21.959	28.054	15.936
100	0	34.134	21.955	28.051	15.860

Note: Entries are based on 1,000 simulations under 2-treatment 2-period RMDs with $\pi = \tau = \varphi = 25$, $\gamma = -25$, $\sigma_\xi^2 = 2$, $\sigma_\epsilon^2 = 1$ and $\mu = 100$

Table 7.4: Expected Outcome for Each Treatment Sequence Based on the Values Used for Simulations

Treatment Sequence	Expected Outcomes
AA	$\begin{pmatrix} 125 \\ 175 \end{pmatrix}$
AB	$\begin{pmatrix} 125 \\ 75 \end{pmatrix}$
BA	$\begin{pmatrix} 75 \\ 175 \end{pmatrix}$
BB	$\begin{pmatrix} 75 \\ 75 \end{pmatrix}$

Table 7.5: Estimated Parameters of Interest: $p = 2$

N	λ	μ	π	τ	γ	φ	σ_ξ^2	σ_ε^2
10	1	99.986 (0.538)	25.013 (0.237)	25.009 (0.587)	-24.993 (1.156)	24.978 (0.626)	2.079	0.989
10	0.9	100.010 (0.538)	25.005 (0.238)	24.985 (0.587)	-25.030 (1.169)	25.014 (0.604)	2.069	0.998
10	0.7	100.007 (0.529)	24.992 (0.243)	24.973 (0.582)	-25.047 (1.156)	25.004 (0.615)	1.969	1.038
10	0.3	99.999 (0.532)	24.989 (0.236)	25.012 (0.582)	-24.988 (1.158)	25.027 (0.601)	2.023	0.986
10	0	99.992 (0.531)	25.003 (0.237)	25.007 (0.581)	-25.004 (1.156)	24.967 (0.604)	2.004	0.995
40	1	100.000 (0.253)	24.994 (0.113)	25.003 (0.277)	-24.998 (0.546)	25.001 (0.304)	2.006	0.988
40	0.9	100.008 (0.253)	24.996 (0.115)	25.002 (0.277)	-25.009 (0.544)	25.006 (0.313)	1.977	0.999
40	0.7	100.000 (0.257)	25.006 (0.117)	25.011 (0.280)	-24.971 (0.552)	24.999 (0.315)	2.025	1.012
40	0.3	99.991 (0.257)	25.001 (0.117)	24.997 (0.280)	-25.013 (0.550)	25.003 (0.318)	2.017	0.999
40	0	100.008 (0.258)	24.994 (0.117)	25.012 (0.281)	-24.958 (0.551)	25.003 (0.317)	2.028	0.992
80	1	100.001 (0.178)	25.000 (0.079)	25.003 (0.195)	-24.999 (0.384)	25.000 (0.214)	2.010	0.992
80	0.9	100.002 (0.179)	25.001 (0.081)	24.995 (0.195)	-25.011 (0.384)	24.998 (0.220)	1.987	1.001
80	0.7	99.999 (0.179)	25.002 (0.082)	25.000 (0.195)	-25.002 (0.385)	25.004 (0.221)	1.982	0.999
80	0.3	100.006 (0.181)	25.000 (0.082)	25.003 (0.197)	-24.986 (0.388)	24.987 (0.222)	2.023	1.002
80	0	99.990 (0.180)	25.003 (0.082)	25.002 (0.197)	-25.002 (0.387)	25.008 (0.222)	2.008	1.002
100	1	99.995 (0.159)	24.998 (0.071)	25.000 (0.174)	-25.008 (0.343)	24.999 (0.193)	1.987	1.006
100	0.9	100.005 (0.160)	25.000 (0.073)	24.999 (0.175)	-25.007 (0.344)	25.002 (0.197)	2.000	1.002
100	0.7	99.997 (0.161)	25.003 (0.073)	25.003 (0.175)	-24.998 (0.345)	25.006 (0.199)	1.992	1.005
100	0.3	100.001 (0.161)	25.001 (0.073)	25.021 (0.175)	-24.960 (0.346)	24.989 (0.198)	2.007	0.997
100	0	99.999 (0.161)	24.999 (0.073)	24.996 (0.175)	-25.011 (0.345)	25.002 (0.198)	1.994	0.997

Note: Entries are estimated values (standard errors) based on 1,000 simulations under 2-treatment 2-period RMDs with $\pi = \tau = \varphi = 25$, $\gamma = -25$, $\sigma_\xi^2 = 2$, $\sigma_\varepsilon^2 = 1$ and $\mu = 100$

Table 7.6: Estimated Numbers of Patients for Each Treatment Sequence: $p = 3$, no treatment difference

N	λ	N_{AAA}	N_{AAB}	N_{ABA}	N_{ABB}	N_{BBB}	N_{BBA}	N_{BAB}	N_{BAA}
40	1	1.016	5.984	5.977	7.023	1.018	5.982	5.974	7.026
40	0.9	4.144	5	5.591	5.263	4.149	5	5.609	5.244
40	0.7	5	5	5	5	5	5	5	5
40	0.3	5	5	5	5	5	5	5	5
40	0	5	5	5	5	5	5	5	5
80	1	1.009	12.998	11.848	14.145	1.007	13	11.842	14.151
80	0.9	9.065	9.996	10.647	10.297	9.062	10	10.628	10.305
80	0.7	10	10	10.001	10	9.998	10	10.001	10
80	0.3	10	10	10	10	10	10	10	10
80	0	10	10	10	10	10	10	10	10
120	1	1.01	20.016	17.649	21.325	1.007	20.019	17.63	21.344
120	0.9	14.098	14.972	15.629	15.324	14.089	14.964	15.592	15.332
120	0.7	14.95	14.994	15.039	15.019	14.953	14.999	15.028	15.018
120	0.3	14.997	14.999	15	15.004	14.999	14.996	15.002	15.003
120	0	15.001	14.996	15.003	15.003	14.999	15.004	14.998	14.996

Note: Entries are based on 1,000 simulations under 2-treatment 3-period RMDs with $\pi_2 = \pi_3 = \tau = \gamma = \varphi = 0$, $\sigma_\xi^2 = 2$, $\sigma_\varepsilon^2 = 1$ and $\mu = 100$

Table 7.7: Estimated Parameters of Interest: $p = 3$, no treatment difference

N	λ	μ	π_2	π_3	τ	γ	φ	σ_ξ^2	σ_ε^2
40	1	99.999 (0.278)	0.003 (0.227)	0.000 (0.227)	0.003 (0.194)	0.009 (0.275)	-0.002 (0.295)	2.005	1.005
40	0.9	100.000 (0.277)	0.010 (0.226)	0.004 (0.226)	-0.005 (0.195)	0.005 (0.281)	0.001 (0.264)	2.006	0.995
40	0.7	99.995 (0.277)	-0.007 (0.227)	0.003 (0.227)	-0.016 (0.197)	-0.010 (0.286)	0.009 (0.262)	1.994	1.006
40	0.3	100.002 (0.276)	-0.007 (0.225)	0.002 (0.225)	-0.003 (0.196)	-0.003 (0.285)	0.000 (0.258)	1.984	0.991
40	0	100.006 (0.278)	-0.010 (0.227)	-0.004 (0.227)	0.004 (0.198)	0.004 (0.288)	0.003 (0.259)	2.005	1.004
80	1	99.995 (0.195)	0.010 (0.159)	0.008 (0.159)	-0.003 (0.136)	-0.003 (0.194)	-0.007 (0.208)	2.000	1.000
80	0.9	99.994 (0.195)	0.002 (0.159)	-0.002 (0.159)	-0.007 (0.138)	-0.011 (0.200)	0.004 (0.184)	2.001	1.001
80	0.7	99.997 (0.195)	-0.004 (0.160)	-0.006 (0.160)	-0.001 (0.139)	0.002 (0.202)	0.008 (0.183)	1.986	1.005
80	0.3	100.004 (0.195)	-0.005 (0.159)	0.000 (0.159)	0.000 (0.139)	0.002 (0.203)	0.001 (0.182)	2.007	1.004
80	0	99.993 (0.194)	0.006 (0.159)	0.001 (0.159)	-0.005 (0.139)	-0.007 (0.202)	0.005 (0.181)	1.978	0.998
120	1	100.008 (0.159)	-0.002 (0.130)	-0.004 (0.130)	-0.004 (0.111)	0.000 (0.158)	0.005 (0.170)	1.991	0.999
120	0.9	99.996 (0.158)	0.009 (0.129)	0.012 (0.129)	0.001 (0.112)	0.001 (0.163)	0.003 (0.149)	1.985	0.994
120	0.7	99.999 (0.159)	-0.001 (0.130)	-0.003 (0.130)	-0.003 (0.113)	-0.007 (0.165)	0.002 (0.149)	2.013	1.005
120	0.3	100.004 (0.159)	-0.001 (0.130)	-0.003 (0.130)	-0.002 (0.113)	-0.008 (0.165)	0.001 (0.148)	1.994	1.003
120	0	99.999 (0.159)	0.005 (0.130)	0.002 (0.130)	-0.001 (0.113)	-0.004 (0.165)	0.003 (0.148)	1.995	1.002

Note: Entries are estimated values (standard errors) based on 1,000 simulations under 2-treatment 3-period RMDs with $\pi_2 = \pi_3 = \tau = \gamma = \varphi = 0$, $\sigma_\xi^2 = 2$, $\sigma_\varepsilon^2 = 1$ and $\mu = 100$

Table 7.8: Characteristics of Mean Squared Error (MSE): no treatment difference

Design	λ	N	Trace(MSE($\boldsymbol{\theta}$))	Det(MSE($\boldsymbol{\theta}$))	Eigen(MSE($\boldsymbol{\theta}$))	MSE(τ)
1	1	40	0.202	4.55E-05	0.157	0.038
1	1	80	0.096	5.06E-06	0.073	0.017
1	1	120	0.066	1.53E-06	0.051	0.012
2	0.9	40	0.191	4.89E-05	0.142	0.037
2	0.9	80	0.093	5.53E-06	0.070	0.019
2	0.9	120	0.060	1.74E-06	0.045	0.013
3	0.7	40	0.203	6.63E-05	0.149	0.042
3	0.7	80	0.106	8.08E-06	0.078	0.021
3	0.7	120	0.060	1.58E-06	0.045	0.012
4	0.3	40	0.210	7.41E-05	0.152	0.044
4	0.3	80	0.110	9.76E-06	0.081	0.023
4	0.3	120	0.072	2.66E-06	0.053	0.014
5	0	40	0.205	7.11E-05	0.151	0.043
5	0	80	0.104	8.29E-06	0.076	0.020
5	0	120	0.067	2.64E-06	0.049	0.013
6	ABA/BAB	40	NA	NA	NA	0.036
6	ABA/BAB	80	NA	NA	NA	0.016
6	ABA/BAB	120	NA	NA	NA	0.012
7	ABB/BAA	40	0.542	1.01E-04	0.509	0.072
7	ABB/BAA	80	0.261	1.21E-05	0.244	0.034
7	ABB/BAA	120	0.188	3.88E-06	0.177	0.025

Note: $\boldsymbol{\theta}=(\tau,\gamma,\varphi)^T$. MSE are calculated based on 1,000 simulations under 2-treatment 3-period RMDs with $\pi_2=\pi_3=\tau=\gamma=\varphi=0$, $\sigma_\xi^2=2$, $\sigma_\epsilon^2=1$ and $\mu=100$

Table 7.9: Estimated Numbers of Patients for Each Treatment Sequence: $p = 3$

N	λ	N_{AAA}	N_{AAB}	N_{ABA}	N_{ABB}	N_{BBB}	N_{BBA}	N_{BAB}	N_{BAA}
40	1	1.014	5.986	5.968	7.032	1.014	5.986	5.98	7.02
40	0.9	5.144	5	6	4.856	3	5	5	6
40	0.7	6.009	5.003	5.977	4	3.011	5	5	6
40	0.3	6.975	5.001	5.01	4	3.014	5	5	6
40	0	6.996	5	5	4	3.005	5	4.999	6
80	1	1.006	12.998	11.842	14.154	1.009	12.995	11.839	14.157
80	0.9	12.001	10.28	11.038	8.736	6.328	9.501	9.999	12.117
80	0.7	12.999	10.81	10.996	8.001	6.749	9.119	9.321	12.005
80	0.3	13.067	10.947	10.973	8	6.859	9.071	9.083	12
80	0	13.151	10.944	10.947	8	6.84	9.057	9.06	12.001
120	1	1.011	20.021	17.645	21.323	1.008	20.024	17.651	21.317
120	0.9	18.662	15.878	16.712	12.31	9.767	14.008	14.444	18.219
120	0.7	19.968	15.999	16.029	11.998	9.971	13.999	14.007	18.029
120	0.3	20.01	15.998	16.005	11.999	9.98	14	14	18.008
120	0	20.019	16	16	11.992	9.988	13.999	13.998	18.004

Note: Entries are based on 1,000 simulation under 2-treatment 3-period RMD with $\pi_2 = \pi_3 = \tau = \varphi = 25$, $\gamma = -25$, $\sigma_\xi^2 = 2$, $\sigma_\varepsilon^2 = 1$ and $\mu = 100$

Table 7.10: Estimated Parameters of Interest: $p = 3$

N	λ	μ	π_2	π_3	τ	γ	φ	σ_ξ^2	σ_ε^2
40	1	100.003 (0.278)	24.999 (0.227)	25.003 (0.227)	24.999 (0.193)	-25.006 (0.275)	24.999 (0.294)	2.011	1.000
40	0.9	99.989 (0.277)	25.003 (0.227)	25.007 (0.228)	25.002 (0.196)	-24.994 (0.281)	24.996 (0.267)	1.998	1.004
40	0.7	100.001 (0.278)	24.984 (0.228)	24.981 (0.229)	24.995 (0.197)	-24.997 (0.286)	25.001 (0.263)	2.008	1.006
40	0.3	100.005 (0.278)	25.004 (0.227)	25.004 (0.228)	25.006 (0.196)	-24.995 (0.285)	24.999 (0.262)	2.015	0.997
40	0	99.976 (0.278)	25.014 (0.227)	25.006 (0.228)	25.004 (0.196)	-25.004 (0.286)	25.006 (0.262)	2.019	0.996
80	1	100.005 (0.195)	25.013 (0.159)	25.004 (0.160)	25.001 (0.137)	-25.000 (0.194)	24.995 (0.208)	2.003	1.005
80	0.9	100.004 (0.195)	25.003 (0.160)	25.004 (0.160)	24.998 (0.138)	-25.001 (0.200)	25.004 (0.186)	2.001	1.001
80	0.7	100.004 (0.195)	24.999 (0.160)	25.005 (0.161)	25.009 (0.139)	-24.996 (0.202)	24.992 (0.184)	1.989	1.005
80	0.3	100.000 (0.196)	24.994 (0.160)	24.997 (0.161)	25.000 (0.139)	-24.998 (0.203)	25.005 (0.184)	2.008	1.004
80	0	100.000 (0.194)	25.004 (0.160)	24.996 (0.160)	24.997 (0.139)	-25.001 (0.202)	25.013 (0.183)	1.978	0.998
120	1	99.998 (0.159)	24.997 (0.130)	24.997 (0.130)	25.003 (0.111)	-24.992 (0.158)	24.991 (0.170)	2.003	1.000
120	0.9	99.997 (0.158)	25.009 (0.130)	25.012 (0.130)	24.996 (0.113)	-25.004 (0.163)	24.997 (0.150)	1.982	0.995
120	0.7	99.997 (0.159)	25.000 (0.131)	24.998 (0.131)	25.004 (0.113)	-24.992 (0.165)	24.992 (0.150)	2.009	1.005
120	0.3	100.005 (0.159)	24.999 (0.131)	24.997 (0.131)	24.998 (0.113)	-25.004 (0.165)	24.999 (0.150)	1.991	1.004
120	0	99.999 (0.159)	25.004 (0.131)	25.001 (0.131)	24.999 (0.113)	-25.003 (0.165)	25.003 (0.150)	1.994	1.002

Note: Entries are estimated values (standard errors) based on 1,000 simulations under 2-treatment 3-period RMDs with $\pi_2 = \pi_3 = \tau = \varphi = 25$, $\gamma = -25$, $\sigma_\xi^2 = 2$, $\sigma_\varepsilon^2 = 1$ and $\mu = 100$

Table 7.11: Characteristics of Mean Squared Error (MSE)

Design	λ	N	Trace(MSE($\boldsymbol{\theta}$))	Det(MSE($\boldsymbol{\theta}$))	Eigen(MSE($\boldsymbol{\theta}$))	MSE(τ)
1	1	40	0.198	4.54E-05	0.151	0.037
1	1	80	0.107	6.29E-06	0.084	0.020
1	1	120	0.068	1.68E-06	0.052	0.013
2	0.9	40	0.194	5.47E-05	0.144	0.042
2	0.9	80	0.091	6.51E-06	0.065	0.018
2	0.9	120	0.062	1.99E-06	0.046	0.013
3	0.7	40	0.192	5.31E-05	0.143	0.040
3	0.7	80	0.096	7.00E-06	0.070	0.021
3	0.7	120	0.061	1.75E-06	0.045	0.013
4	0.3	40	0.190	5.35E-05	0.140	0.041
4	0.3	80	0.097	6.94E-06	0.072	0.020
4	0.3	120	0.062	1.85E-06	0.045	0.013
5	0	40	0.189	5.45E-05	0.139	0.040
5	0	80	0.095	6.41E-06	0.069	0.019
5	0	120	0.062	1.84E-06	0.045	0.012
6	ABA/BAB	40	NA	NA	NA	0.036
6	ABA/BAB	80	NA	NA	NA	0.016
6	ABA/BAB	120	NA	NA	NA	0.012
7	ABB/BAA	40	0.542	1.01E-04	0.509	0.072
7	ABB/BAA	80	0.261	1.21E-05	0.244	0.034
7	ABB/BAA	120	0.188	3.88E-06	0.177	0.025

Note: $\boldsymbol{\theta}=(\tau,\gamma,\varphi)^T$. MSE are calculated based on 1,000 simulations under 2-treatment 3-period RMDs with $\pi_2=\pi_3=\tau=\varphi=25$, $\gamma=-25$, $\sigma_\xi^2=2$, $\sigma_\varepsilon^2=1$ and $\mu=100$

Chapter 8

Conclusion

This thesis is a study of design issues in clinical trials that use repeated measurement designs, especially those that are concerned with response-adaptive designs for comparing two treatments. We have constructed optimal repeated measurement designs under more general models than those constructed in previous studies, examined the influence of baseline measurements on constructing repeated measurement designs, and proposed new design strategies to construct response-adaptive repeated measurement designs. Further, the R codes developed for this study can be used as design software to provide an optimal treatment sequence for the next patient entering the study.

8.1 Main Contribution

8.1.1 Optimal Designs and Baseline Measurement Study Under the Self and Mixed Carryover Effects Model

We apply the Lagrange multiplier method to solve the optimal design problems under both the traditional and a more general model, the self and mixed carryover effects model with random subject effects where the direct treatment effect will manifest itself no matter where and when the treatment is applied. The model with both self and mixed carryover effects was proposed by Afsarinejad and Hedayat (2002). They studied two-period optimal designs under this model with fixed subject effects. In this thesis, we consider random subject effects in the self and mixed carryover effects model and construct p-period optimal designs comparing two treatments. We also consider the

effect of the baseline measurements on constructing optimal designs, under both the traditional model and the self and mixed carryover effects model, and we find that

1. For two-treatment repeated measurement designs, the duality *Lemma* 3.3.2.1 proves that optimal designs allocate equal number of subjects to a treatment sequence and its dual.

2. Optimal designs, in terms of minimizing the variance of the parameters of interest, are strongly model dependent and are not unique under the same model. Under the traditional model, the results here are identical to what is available in the literature. When $p = 2$, for estimation of the treatment difference, the design AA/BB and AB/BA with an equal number subjects per sequence is optimal. For estimation of the carryover effect, the optimal design depends on the value of ρ: when $\rho = 0$, the optimal design is AA/BB and AB/BA with an equal number of subjects per sequence; when $\rho \rightarrow 1$ the optimal design is AA/BB with an equal number of subjects per sequence. When $p = 3$, the design ABB/BAA, with an equal number of subjects per sequence, is optimal for estimation of direct treatment and carryover effects. When $p = 4$, the design $ABBA/BAAB$ and $AABB/BBAA$ with an equal number of subjects per sequence is optimal. We also find for the first time that the design $ABBA/BAAB$, $ABAB/BABA$ and $AABB/BBAA$ with 1/6, 1/24 and 7/24 of the total subjects per sequence, respectively, is also optimal. The former design is more popular in practice because it utilizes less treatment sequences, uses an equal number of subjects per treatment, and requires the total number of subjects to be a multiplier of 4 instead of 24, as in the latter design. Under the self and mixed carryover effects model with random subject effects, we present the optimal results for the first time in the literature. When $p = 2$, without baseline measurements, the estimation of the treatment difference uses only the data in the first period, therefore, it is not efficient. For estimation of the mixed carryover effect, the design AB/BA is op-

timal. For estimation of the self carryover effect, the design AA/BB is optimal. When $p = 3$, the optimal design for estimation of the treatment difference is ABA/BAB, however, there are no self carryover effects with this design. Other designs are recommended such as design AAB/BBA and ABA/BAB, design ABB/BAA and ABA/BAB, and design AAB/BBA, ABA/BAB and ABB/BAA. Those designs are almost as efficient as the design ABA/BAB, especially when ρ is small. Therefore, there is a price to be paid for allowing different types of carryover effects in the model. When $p = 4$, the simplest optimal design is the design $ABBA/BAAB$ and $AABA/BBAB$ with an equal number of subjects per sequence.

3. The use of baseline measurements should be discussed in each specific situation. Under the assumptions about baseline measurements in Section 3.4.1, we can conclude that under the traditional model, when $p = 2$, the efficiency of the design with baseline measurements is 1 to 2.5 times that of the design without baseline measurements. Therefore, it is recommended to use the baseline measurements. When $p = 3$, the baseline measurements improve the efficiency only slightly. The relative efficiency between the design with baseline measurements vs. the design without is equal to 1 to 1.0625 when ρ increases from 0 to 1. Therefore, use of the baseline measurement does not appear to be helpful in improving the design efficiency. When $p = 4$, the baseline measurements do not improve the efficiency at all. While under the self and mixed carryover effects model, we also gain something from the baseline measurements. When $p = 2$, the baseline measurements improve the efficacy of the design measurements by at least 1.5 times. Therefore, it is strongly recommended to use the baseline measurements. When $p = 3$, the baseline measurements improve the efficiency significantly: the relative efficiency between the design with baseline measurements vs. the design without baseline measurements is equal to 2 to 3 when ρ increases from 0 to 1. Therefore, it is recommended to

use the baseline measurements. When $p = 4$, the efficiency of the design with baseline measurements is 2.5 to 3 times that of the efficiency of the design without. Therefore, it is worthwhile to add the baseline measurements in the study.

4. There is a dramatic reduction in variability for estimating the direct treatment effect contrast when extending two-period designs to three-period or four-period designs. In particular, under the traditional model, no matter using baseline measurements or not, the three-period designs achieve at least a 33% reduction in variance compared to the two-period designs, and the four-period designs achieve at least a 25% reduction in variance compared to the three-period designs. Similar patterns are found under the self and mixed carryover effects model. For the designs utilizing baseline measurements in each period, there is at least a 25% reduction in variance in three-period designs compared to two-period designs, and at least a 20% reduction in four-period designs compared to three-period designs. Without baseline measurements, when within subject correlation is 0.5 or more, compared to the two-period designs, the three-period designs achieve a 27% or more reduction in variability. In addition, a 14% or more reduction in variability is achieved when add one more period after the third period.

8.1.2 Stratified and Randomized Play-the-Winner Rule (SRPWR)

Since Zelen (1969) proposed the PWR for controlled clinical trials, various researchers proposed and investigated allocation rules for better patient treatment (Wei and Durham 1978, Wei 1978 and 1979, Smith 1984, Durham and Yu 1990, Andersen et al. 1994, Eisele 1994, Smythe 1996, Durham et al. 1998, Bai and Hu 1999, Biswas 1999, Bandyopadhyay and Biswas 2000, Ivanova et al. 2000, Ivanova and Durham 2000, Ivanova and Flournoy 2001, Bai et al. 2002, Hu and Zhang 2004). The main contribution of this chapter is in proposing a simpler allocation rule, which considers the heterogeneity

of subjects. Simulation studies show that, on average, the SRPWR tends to assign more patients to the better treatment. SRPWR is superior to existing allocation rules, and it successfully adjusts the allocation results while accommodating the heterogeneity of the patients, leading to a better treatment strategy.

8.1.3 Multiple-Objective Approach for Constructing Response-Adaptive Repeated Measurement Designs

One of the main contributions of this thesis is to extend single objective designs to multiple objective designs. In this thesis we develop a new adaptive allocation rule, which can provide good estimates of the treatment differences and assign more patients to a better treatment. The basic idea is to modify the allocation rule based on the observed data from previous patients. We assume patients enter the study sequentially. The first few patients are assigned using the optimal design suggested in the literature or a completely randomized design. Then the information matrix can be calculated based on the observed data. We also propose an evaluation function to evaluate the performance of each treatment sequence. For the next patient, we consider all possible treatment sequences and choose the treatment sequence to maximize the criteria, which has two components: the first component is to choose a treatment sequence to maximize the information matrix; the second is to choose a treatment sequence which gives the best performance based on the observed data. A weight parameter λ is used to balance the two objectives, and can be chosen by the investigator prior to the experiment. A large value of λ will place more emphasis on the estimation precision. When $\lambda = 1$ the allocation rule becomes the usual response adaptive designs as considered by Kushner (2003). A small value of λ will emphasize the performance/benefit of the treatment. When $\lambda = 0$, the allocation rule becomes a typical play-the-winner rule (Zelen, 1969). Note that Kushner's adaptive allocation rule is for trails with continuous outcomes, and Zelen's play-the-winner rule is for trials with dichotomous outcomes. However, our new adaptive allocation

rule is applicable to trials with both continuous and dichotomous outcomes.

In Chapter 6, we utilize this allocation strategy to construct adaptive repeated measurement designs with dichotomous responses/outcomes. We provide the detailed allocation rule for constructing adaptive two-treatment two-period repeated measurement designs, and then extend it to two-treatment p-period repeated measurement designs. In simulation studies, we demonstrate that the designs with $\lambda < 1$ constructed under the new proposed allocation rule are not as efficient as the design with $\lambda = 1$ in terms of the mean squared error, but it successfully put more patients to the better treatment group. In addition, simulation studies show that the design with a high value of λ, say $\lambda = 0.9$, significantly favors the allocation results to more effective treatment sequences without loss of much estimation precision.

In Chapter 7, we utilize this allocation rule to construct adaptive repeated measurement designs with continuous responses/outcomes, based on the self and mixed carryover effects model. We provide a detailed allocation rule for constructing adaptive two-treatment two-period repeated measurement designs, and then extend it to two-treatment three-period repeated measurement designs. In simulation studies, we demonstrate that the designs constructed under the new proposed allocation rule are more efficient than the fixed design ABB/BAA in terms of the mean squared error. The value of λ is pre-determined by researchers, which is used to balance the two objectives between increasing the estimation precision and decreasing the proportion of patients receiving inferior treatments. In the simulation study, we notice that the design with $\lambda = 0.9$ substantially skews the allocation results to more effective treatment sequences without loss of much estimation precision. Therefore, choosing a high value of λ is recommended in practice.

8.2 Future Research

There are still some issues that need to be further studied. When Afsarinejed and Hedayat (2002) introduced for the first time the topic of simple and mixed carryover effects, they presented some two-period optimal design re-

sults under the model with fixed subject effects. We considered the optimal designs under the self and mixed carryover effects model with random subject effects for designs comparing two treatments in p periods ($p = 2, 3, 4$), under the equicorrelated variance-covariance matrix structure. It is well known that no design is optimal under all models, and under a specific model there could be many optimal designs. Further study should explore optimal designs comparing more than two treatments under the model with various types of carryover effects, variance-covariance matrix structures and random subject effects. In addition, instead of constructing a design that is optimal under a particular model, it is needed to build a design that is reasonably simple and has relatively high efficiency under different models.

We are working on extending the stratified and randomized play-the-winner rule we proposed for clinical trials with dichotomous outcomes, to experiments with ordinal responses. One possibility is to give different weights for each response, with the number of balls added to the urn being proportional to the weights.

Finally, one of the main contributions of this thesis is in proposing an adaptive allocation rule to construct multiple-objective repeated measurement designs. We studied the performance of the design strategy for dichotomous and continuous responses, respectively, through simulation studies. In clinical trials, it is more often that several outcomes, continuous or discrete or both, will be measured to examine many aspects of the interventions. Therefore, further research is needed to apply the allocation rule to mixed outcome data.

Bibliography

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

[2] Andersen, J.; Faries, D. and Tamura, R.N. A randomized play-the-winner design for multi-arm clinical trials. *Communications in Statistics – Theory and Methods*, 23, 309-323.

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

[4] Armitage, P. (1985). The search for optimality in clinical trials. *International Statistical Review*, Vol 53, No.1, 15-24.

[5] Atkinson, A.C.(1982). Optimal bised coin desings for sequential clinical trials with prognostic factors. *Biometrika*, 69, 61-67.

[6] Atkinson, A.C.(1999). Optimal bised-coin desings for sequential treatment allocation with covariate information. *Stiatistics in Medicine*, 18, 1741-1752.

[7] Bai, Z.D. and Hu F. (1999). Asymptotic theorems for urn models with non-homogeneous generating matrices. *Stochastic Processes and Their Applications*, 80, 87-101.

[8] Bai, Z.D.; Hu F. and Shen, L. (2002). An adaptive design for multi-arm clinical trials. *Journal of Multivariate Analysis*, 81, 1-18.

[9] Bandyopadhyay, U. and Biswas A. (2000) Comparison of two treatments with heterogeneous experimental units. *Journal of Statistical Planning and Inference* 83, 441-448.

[10] Bandyopadhyay, U. and Biswas A. (1999)

[11] Bartlett, R.H.; Roloff, D.W.; Cornell, R.G.; Andrews, A.F.; Dillon, P.W., and Zwischenberger, J.B. (1985) Extracorporeal circulation in neonatal respiratory failure: a prospective randomized study. *Pediatrics* 76, 479-487.

[12] Bate, S.T. and Jones, B. The construction of universally optimal uniform cross-over designs. Unpublished manuscprit.

[13] Begg, C.B., Iglewicz, B. (1980). A treatment allocation procedure for sequential clinical trials. *Biometrics*, 36, 81-90.

[14] Bellavance, F. and Tardif, S. (1995). A nonparametric approach to the analysis of three-treatment three-period crossover designs. *Biometrika*, Vol. 82, No. 4 (Dec., 1995), 865-875.

[15] Biswas, A. (1999). Allocation by randomized play-the-winner rule in the presence of prognostic factors. *The Indian Journal of Statistics*, Vol. 61, Series B, Pt. 3, 397-412.

[16] Biswas, A. (2001). Adaptive designs for binary treatment responses in phase III clinical trials: controversies and progress. *Statistical Methods in Medical Rsearch*, 10, 353-364.

[17] Biswas, A. (2004). A randomized longitudinal play-the-winner design for repeated binary data. Unpublished manuscprit.

[18] Boos, D.D. (1992). On Generalized Score Tests *The American Statistician*, Vol. 46, No. 4, 327-333.

[19] Breslow, N.E. and Clayton, D.G. (1993). Approximate Inference in Generalized Linear Mixed Models *Journal of the American Statistical Association*, Vol. 88, No. 421, 9-25.

[20] Carrière, K.C. (1994a) Incomplete repeated measures data analysis in the presence of treatment effects. *Journal of the American Statistical Association.* Vol 89, 680-686.

[21] Carrière, K.C. (1994b) Cross-over designs for clinical trials. *Statistics in Medicine.* 13, 1063-1069.

[22] Carrière, K.C. (1998) Regulatory affairs in biotechnology: optimal statistical designs for biomedical experiments. *Biotechnology Annual Review.* Vol. 4, 215-238.

[23] Carrière, K.C. (1999). Methods for repeated measures data analysis with missing values. *Journal of Statistical Planning and Inference.* Vol. 77, 221-236.

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

[25] Carrière, K.C. and Reinsel, G.C. (1993). Optimal two-period repeated measurement designs with two or more treatments. *Biometrics.* 80, 924-929.

[26] Carrière, K.C., Roos, L.L, and Dover, D.C. (2000). Across time and space: variations in hospital use during Canadian health reform. *Health Services Research* 35 (2): 467-487.

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

[28] Cook, R.J. (1995). Interim Analyses in 2×2 crossover trials. *Biometrics*, Vol 51, No. 3, 932-945.

[29] Cook, R.J. and Willan A.R. (1996). Design considerations in crossover trials with a single interim analysis and serial patient entry. *Biometrics*, Vol 52, No. 2, 732-739.

[30] Durham, S.D.; Flournoy, N. and Li, W. Sequential designs for maximizing the probability of a favorable response. *Canadian Journal of Statistics*, 3, 479-495.

[31] Durham, S.D. and Yu, C.F. Randomized play-the-leader rules for sequential sampling from two populations. *Probability in the Engineering and Informational Sciences*, 4, 355-367.

[32] Diggle, P.J., Liang, K.Y., and Zeger, S.L. (1994). Analysis of Longitudinal Data, Oxford: Clarendon Press.

[33] Ebbutt, A.F. (1984). Three-period crossover designs for two treatments. *Biometrics*, 40, 219-224.

[34] Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika*, 58, 403-417.

[35] Eisele, J. (1994). The Doubly Adaptive Biased Coin Design for Sequential Clinical Trials. *Journal of Statistical Planning and Inference*, 38, 249-262.

[36] Eisele, J. and Woodroofe, M. (1995). Central Limit Theorems for Doubly Adaptive Biased Coin Designs. *The Annals of Statistics*, 23, 234-254.

[37] Fleiss, J.L. Wallenstein, S. and Rosenfeld, R. (1985). Adjusting for baseline measurements in the two-period crossover study: a cautionary note. *Controlled Clinical Trials*, 6, 192-197.

[38] Freedman, David A. (1965). Bernard Friedman's Urn. *The Annals of Mathematrical Statistics*, Vol 36, No. 3, 956-970.

[39] Friedman, L.M., Furberg, C.D. and DeMets, D.L. (1998). Fundamentals of Clinical Trials. Thrid Edition, Springer, New York.

[40] Grizzle, J.E. (1965). The Two-Period Change-Over Design and Its Use in Clinical Trials. *Biometrics*. 21, 467-480.

[41] Grizzle, J.E. (1974). Corrections: The Two-Period Change-Over Design and Its Use in Clinical Trials. *Biometrics*. 30, 727.

[42] Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and Its Application*, Academic, san Diego.

[43] Hedayat, A. and Afsarinejad, K. (1978). Repeated measurements designs, II. *The Annals of Statistics*, 6(3), 619-628.

[44] Hedayat, A.S. and Stufken, J. (2003). Optimal and efficient crossover designs under different assumptions about the carryover effects. *Journal of Biopharmaceutical Statistics*, Vol 13, No. 3, 519-528.

[45] Hedayat, A.S. and Zhao, W. (1990). Optimal two-period repeated measurements designs. *The Annals of Statistics*. 18, 1805-1816.

[46] Hills, M. and Armitage, P. (1979). The two-period cross-over clinical trial. *British Journal of Clinical Pharmacology*. 8, 7-20.

[47] Hoel, D. and Sobel M. (1971). Comparisons of sequential procedures for selecting the best binomial population. *Proceeding of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*. IV, 53-69.

[48] Hu, F. and Rosenberger, W.F. (2003), Optimality, Variability, Power: Evaluating Responses-Adaptive Randomizeation Procedures for Treatment Comparison. *Journal of the American Statistical Association*. 98, 671-678.

[49] Hu, F. and Zhang L.X. (2004). Asymptotic Properties of Doubly Adaptive Biased Coin Designs for Multi-Treatment Clinical Trials. *The Annals of Statistics*. 32, 268-301.

[50] Huang, R. (2001). Issues with Repeated measures data. Unpublished Ph.D. Thesis. University of Alberta.

[51] Ivanova, A. (2003). A Play-the-Winner-Type Urn Design with Reduced Variabiltiy. *Metrika*, 58, 1-13.

[52] Ivanova, A. and Durham, S.D. (2000). Drop the loser rule. *Technical Report* $TR00-01$, Universtiy of North Carolina, Chapel Hill.

[53] Ivanova, A. and Flournoy, N. (2001). A birth and death urn for ternary outcomes: stochastic processes applied to urn models. Probabiltiy and

Statistical Models with Applications; Charalambides, C.A., Koutras, M.V., Balakrishnan, N., Eds.; Boca Raton: Chapman and Hall, 2001; 583-600.

[54] Ivanova, A.; Rosenberger, W.F.; Durham, S.D. and Flournoy, N. (2000). A birth and death urn for randomized clinical trials: asymptotic methods. *Sankhya B*, 62, 104-118.

[55] Kershner, R.P. (1986). Optimal 3-period 2-treatment crossover designs with and without baseline measurements. *Proceedings of the Biopharmacerutical Section of the American Statistical Association*, 152-156.

[56] Kiefer, J. (1958). On the nonrandomized optimality and randomized nonoptimality of symmetrical designs. *The Annals of Mathematical Statistics*. 29, 675-699.

[57] Kiefer. J. (1975). Construction and optimality of generalized Youden designs. In *A Survey of Statistical Designs and Linear Models* (J. N. Srivastava, Ed.) 333-353. North-Holland, Amsterdam.

[58] Kiefer, J. and Wolfowitz J. (1959). Optimum designs in regression problems. *The Annals of Mathematical Statistics*. 30, 271-294.

[59] Kunert, J. (1983). Optimal design and refinement of the linear model with applications to repeated measurements designs. *The Annnals Of Statistics*. 11, 247-257.

[60] Kunert, J. (1984). Optimality of balanced uniform repeated measurements designs. *The Annnals Of Statistics*. 12, 1006-1017.

[61] 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.

[62] Kushner, H.B. (1997b). optimal repeated measurement desings: the linear optimality equations. *The Annals of Statistics*, 25, 2328-2344.

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

[64] Laird, N. M. and Ware J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38, 963-974.

[65] 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.

[66] Laska, E.M., Meisner, M. and Kushner, H.B. (1983). Optimal crossover designs in the presence of carryover effects. *Biometrics*, 39, 1087-1091.

[67] Lehmann, E.L. (1975). *Nonparametrics: Statistical Methods Based on Ranks.* Holden-Day, San Francisco.

[68] Liang, K.Y. and Zeger, S.L. (1986). Longitudianl data analysis using generalized models. *Biometrika* 73, 13–22.

[69] Liang, Y.Y. (2003). Analysis of familial data on alcoholism. Unpublished Master project. University of Alberta.

[70] Lipsitz, S.H., Fitzmaurice, G.M., Orav, E.J., and Laird, N.M. (1994a). Performance of generalized estimating equations in practical situations. *Biometrics*, 50, 270 -278.

[71] Lipsitz, S.H., Kim, K., and Zhao, L. (1994b). Analysis of repeated categorical data using generalized estimating equations. *Statistics in Medicine*, 13, 1149 -1163.

[72] Lipsitz, S.H., Laird, N.M., and Harrington, D.P. (1991). Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. *Biometrika*, 78, 153 -160.

[73] Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979). Multivariate Analysis. Academic Press, New York.

[74] Martin, R.J. and Eccleston, J.A. (1998). Variance-balanced change-over designs for dependent observations. *Biometrika* 85, 883–892.

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

[76] Melfi, V. and Page, C. Variability in Adaptive Desings for Estimation of Success Probabilities. In New Developments and Applications in Experimental Design (Flournoy, N. , Rosenberger, W.F. and Wong, W.K., eds) Hayward, CA: Insititute of Mathematical Statistics, 106-114.

[77] Moerbeek, M. and Wong, W.K. (2002). Multiple-objective optimal designs for the hierarchiecal linear model. *Journal of Official Statistics* Vol. 18., No. 2, 291–303.

[78] Pocock, S.J. (1979). Allocation of patients to treatments in clinical trials. *Biometrics*, 35, 183-197.

[79] Pocock, S.J. and Simon R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trials. *Biometrics*, 35, 183-197.

[80] Putt, M.E. (2006). Power to detect clinically relevant carry-over in a series of cross-over studies. *Statistics in Medicine*, in press.

[81] Rosenberger, W.F. (1992). Asymptotic inference proglems arising from clinical trials using response-adaptive treatment allocation. Ph.D. dissertation, Graduate School of Arts and Sciences, George Washington Univ.

[82] Rosenberger, W.F. (1993). Asympototic inference with response-adaptive treatment allocation desings. *The Annals of Statistics* Vol. 21, No. 4, 2098-2107.

[83] Rosenberger, W.F. (1995). Adaptive designs in clinical trials: some issurs with emphasis on asymptotic inference. In Flournoy, N. and Rosenberger W.F. (eds.), *Adaptive Designs*, Hayward, CA:Institute of Mathematical Statistics, pp. 1-12.

[84] Rosenberger, W.F. (2002). Randomized urn models and sequential design (with discussion), *Sequential Analysis*, 21, 1-41.

[85] Rosenberger, W.F. and Lachin, J.M. (1993). The use of response-adaptive designs in clinical trials. *Controlled Clin. Trials* 14, 471–484.

[86] Rotnitzky, A. and Jewell, N.P. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika* Vol.77, No.3, 485–497.

[87] Schall, R. (1991). Estimation in Generalized Liear Models with Random Effects. *Biometrika* Vol.78, No.4, 719–727.

[88] Schwabe, R. (1987). On an adaptive design in regression. *Statistics* 18, 521–525.

[89] Silvey, S.D. (1980). Optimal Design. Monographs on Applied Probability and Statistics. Chapman & Hall, London.

[90] Simon, Richard; Weiss, George H. and Hoel, David G. (1975). Sequential analysis of binomial clinical trials. *Biometrika*, Vol. 62, No. 1, 195-200.

[91] Simon, R. (1977). Adaptive treatment assignment methods and clinical trials. *Biometrics*, 33, 743-749

[92] Simon, R. (1991) A decade of progress in statistical methodology for clinical trials. *Statistics in Medicine*, 10, 1789-1817.

[93] Smith, R.L. Properties of biased coin designs in sequential clinical trials. *The Annals of Statistics*, 12, 1018-1034.

[94] Smythe, R.T. Central limit theorems for urn models. *Stochastic Processes and Their Applicaitons*, 65, 115-137.

[95] Stufken, J. (1991). Some families of optimal and efficient repeated measurements designs. *Journal of Statistical Pleanning and Inference*, 27, 75-83.

[96] Tamura, R.N.; Faries, D.E.; Andersen, J.S. and Heiligenstein, J.H. (1994) A case study of an adaptive clinical trials in the treatment of out-patients with depreeive disorder. *Journal of the American Statistical Association*, 89, 768-776.

[97] Wallenstein, S. (1979). Inclusion of Baseline Values in the Analysis of Crossover Designs (abstract). *Biometrics* 35, 894.

[98] Wei, L.J. (1977). A class of designs for sequential clinical trials. *Journal of the American Statistical Association* 72,382-386.

[99] Wei, L.J. (1978). The adaptive biased coin design for sequential experiments. *The Annals of Statistics* 6, 92-100.

[100] Wei, L.J. (1979). The generalized polya's urn design for sequential medical trials. *The Annals of Statistics* Vol. 7, No. 2, 291-296.

[101] Wei, L.J.(1988). Exact two-sample permutation tests based ont the randomized play-the-winner rule. *Biometrika* 75, 603–606.

[102] 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.

[103] Wei, L.J.; Smythe, R.T.; Lin, D.Y. and Park T.S. (1990). Statistical inference with data-dependent treatment allocation rules. *Journal of the American Statistical Association* 85, 156–162.

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