# Models with Subject by Treatment and Subject by Carryover Interactions and Use of Baseline Measurements in Crossover Trials

By

Lingling Han

(Under the direction of John Stufken)

## Abstract

Because treatments can be compared on the same subject and fewer subjects are needed to obtain the same number of observations as in a parallel trial, crossover designs have been applied extensively in various fields. With the development and application of crossover designs, researchers realized that one serious potential problem involved in their use is the presence of carryover effects.

In this dissertation, in order to capture the variabilities due to different direct and carryover treatments, we propose a model that includes interactions of subject by treatment and subject by carryover effects. We assume subject effects to be random, and therefore take these interaction terms to be random too. We study the identifiability properties of the corresponding variance components and the conditions under which the parameters of the model are identifiable. The REML estimation method for those variance components and other model parameters is also considered. Some special cases and practical applications of this model are studied as well.

A second objective of this dissertation is to investigate appropriate ways to handle baseline measurements in crossover studies in different situations. Four different methods are considered to incorporate the baseline measurements in the $2 \times 2$ crossover design for both single measurements and repeated measurements. Analytical expressions of variances of the estimators of the treatment contrast from those methods are derived and compared under different scenarios. Simulation studies are conducted to evaluate the performance of each method when the variance components are unknown. For the case of repeated measurements, graphical methods are discussed to study the change in treatment effect over time; different types of baselines and different assumptions for the random error terms are also considered. The methods are applied to real data analysis. Designs with more than two treatments are discussed briefly.

Index words: Crossover design; carryover effect; identifiability; REML estimate; baseline measurements; single measurements; repeated measurements.

# Models with Subject by Treatment and Subject by Carryover Interactions and Use of Baseline Measurements in Crossover Trials

by

Lingling Han

M.S., University of Georgia, Athens, GA, 2002

A Dissertation Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

DOCTOR OF PHILOSOPHY

Athens, Georgia

2007

# Models with Subject by Treatment and Subject by Carryover Interactions and Use of Baseline Measurements in Crossover Trials

by

Lingling Han

Approved:

Major Professor: John Stufken

Committee: Daniel Hall
Abhyuday Mandal
Jaxk Reeves
Lynne Seymour

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2007

## Acknowledgements

While the work for this dissertation has been challenging and at times trying, I am greatly indebted to many people; without their help, support, trust and encouragement, this dissertation would not have been possible.

First of all, I would like to express my sincere appreciation to my advisor, Dr. John Stufken, for his insight, wise direction, constant encouragement, and for his support in so many aspects. I would never have finished my dissertation without his help. He is such a wonderful advisor: not only knowledgeable, enthusiastic and confident, but he also truly cares for his students. He is a true mentor and it has been such a rewarding time to work with him.

I would like to thank my committee members: Dr. Daniel Hall, Dr. Abhyuday Mandal, Dr. Jaxk Reeves and Dr. Lynne Seymour for their assistance and valuable suggestions for my dissertation and also for all the help they have offered to me for the past five years. I also want to take this opportunity to thank all the professors, staff, and graduate students in this department, who made my life enjoyable and memorable. Special thanks go to Dr. Hui Xu and Dr. Haitao Zheng for their help and encouragement on both my course work and dissertation; special thanks also go to XiaoA Zhen for helping me run my programs.

I thank my family, in particular, my parents Rensheng Han and Juqiu Shi, my brothers Shiwei Han and Jianxin Han, and my boyfriend Ke Xu for their unconditional love and endless support to pursue my dream, and for always believing that I am the best one in the world.

I truly thank my high school teacher Yongjian Lin; it is your words that made me persist in working hard to obtain this degree. I am sorry that I am still like “iron” and not like “steel” after so many years. Special thanks also go to my best friends: Weiyi He, Saiping Liu, Linhuo Lai, Qiuyan Ying and Aiping Zhang, for their love and support for so many years; I cherish our friendship forever.

Table of Contents

# Chapter 1

# Introduction

A crossover design, also called change-over design, is a repeated measurements design such that each subject receives different or identical treatments during different time periods (Hedayat and Afsarinejad, 1975; Cheng and Wu, 1980; Bishop and Jones, 1984; Matthews, 1988).

Because treatments can be compared on the same subject and fewer subjects are required to obtain the same number of observations as in a parallel trial (Grieve, 1985; Stufken, 1996), crossover designs have been applied extensively in various fields, such as animal husbandry, food science, biological research, sensory testing, marketing, psychological experiments, social engineering, educational studies as well as medical sciences. Among these, the most prevalent use is no doubt in pharmaceutical studies and clinical trials in medical research (Brown, 1980), especially in small clinical trials (Matthews, 1994).

With the development and application of crossover designs, researchers realized that one serious potential problem involved in their use is the presence of carryover effects, which are the lingering effects of treatments given in one of the previous periods. Many different models have been introduced to incorporate these undesired effects.

The simplest model with carryover effects is the traditional model (Cheng and Wu, 1980; Hedayat and Afsarinejad, 1978), which assumes, starting from the second

period, that a carryover effect from the treatment in the previous period (a first-order carryover effect) always exists, and that it is additive and constant for that treatment (Jones and Kenward, 2003).

The validity of this assumption has been questioned by many authors, for example, Fleiss (1989) and Senn (2002). Perhaps in reaction to these criticisms, Afsarinejad and Hedayat (2002) introduced self and simple mixed carryover effects, allowing for two different carryover effects from each treatment depending on whether the treatment is followed by itself or by another treatment. Bose and Mukherjee (2000) studied the model with higher-order carryover effects and all interactions between direct treatments and carryover treatments. Kempton, Ferris and David (2001) considered a model with carryover effects that are proportional to the direct treatment effects.

All of those papers emphasized theoretical derivations or identification and construction of optimal designs under the model being focused on. But there is little evidence based on data to support any of these models. In medical studies, some treatments have obvious effects on some patients, but less on others; moreover, the lingering effects also affect different patients differently. We would like to have a model that captures this variability.

Jones, Kunert and Wynn (1992) tried a model with random carryover effects, which were assumed to be randomly distributed with mean 0 and variance $\sigma_\epsilon^2$. However, in practice, since there is no infinite population of carryover effects from which one is sampling, but just one carryover effect for each treatment in the design, this formulation is rather questionable.

Instead, we propose a model that includes subject by treatment and subject by carryover interactions. We assume subject effects to be random, and therefore take these interaction terms to be random too.

Thus the model has three random terms in addition to the random error. First, the identifiability properties of these variance components are investigated, and the conditions under which the parameters of the model are identifiable are also determined. Next, some special cases and practical applications of this model are studied as well. Finally, the estimation methods for those variance components and other model parameters are considered. Thus, studying the new model which incorporates the subject by carryover interaction is one of the objectives of this dissertation.

A second objective of this dissertation is to investigate appropriate ways to handle baseline measurements in crossover studies in different situations.

By far the most prevalent crossover design in clinical trials is the $2 \times 2$ design (a definition will be given in Chapter 2). However, it is also regarded by many statisticians as particularly problematic because of the carryover effects and aliasing of several effects (Senn, 2002; Jones and Kenward, 2003). Using baseline measurements at the beginning of each period is introduced as a technique to rescue the *AB/BA* design from its deficiencies and to provide additional information to eliminate nuisance effects from the treatment effects (Patel, 1983; Kenward and Jones, 1987).

But, contrary to some authors' intuition and initial purposes, using baseline measurements before each period to eliminate the carryover effects receives serious criticism in many papers (Fleiss, Wallenstein and Rosenfeld, 1985; Willan and Pater, 1986; Fleiss,1989). Of course, the conclusions depend on different models, methods and assumptions considered and all of those papers considered carryover effects on the wash-out period. In a strict sense, the measurements from the wash-out period are not true baselines, if they were affected by the treatments from the previous period. Therefore, some authors, for example, Senn (2002), considered measurements obtained prior to the second treatment period as baselines only when wash-out periods are long enough to eliminate any carryover effects, and study how to use the information provided by the baseline measurements effectively.

In clinical trials, the $2 \times 2$ design is often used with a long enough wash-out period to ensure no carryover effects. If baseline measurements are made, they may be obtained at one time point or multiple time points (single measurements or repeated measurements) before each period, which leads to the question of how to use this information effectively. Four different methods are considered to incorporate the baseline measurements for both single measurements and repeated measurements. Analytical expressions of variances of the estimators of the treatment contrast from those methods are derived and compared under different scenarios. For the case of repeated measurements, different types of baselines and different assumptions for the random error terms are considered.

Thus, this dissertation is organized as follows. Chapter 2 gives a literature review of crossover designs: the traditional model and its limitations are presented; definitions, terminology and special design issues for crossover designs are also discussed. In particular, various models that incorporate carryover effects are investigated in detail; different methods to handle baseline measurements in crossover studies are reviewed. In Chapter 3, we propose a new model for carryover effects; the identifiability problem for its variance components is investigated, flexibility of the model is pointed out, and parameter estimation of the model is considered. A simulation study is conducted to estimate the parameters for particular designs and to check whether the new model performs better. At the end of Chapter 3, a real data example is used to illustrate the results. We study the use of baseline measurements for the $2 \times 2$ crossover design for both single measurements and repeated measurements in Chapter 4 and Chapter 5, respectively. Potential methods are proposed, variances for estimators of a treatment contrast for each method are derived and compared. A simulation study is conducted to evaluate the performance of each method, and real data sets are used to illustrate our results. In addition, Chapter 5 gives a graphical method to study the change in treatment effect over time if the treatment by time

interaction is significant. We present studies for selected designs with more than two treatments in Chapter 6.

## 1.1 References


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

2. Bishop, S. H. and Jones, B. (1984). A review of higher order crossover designs. *Journal of Applied Statistics* 11, 29-50.

3. Bose, M. and Mukherjee, B. (2000). Cross-over designs in the presence of higher order carry-overs. *Australian and New Zealand Journal of Statistics* 42, 235-244.

4. Brown, B. W. (1980). The crossover experiment for clinical trials. *Biometrics* 36, 69-79.

5. Cheng, C. S. and Wu, C. F. (1980) Balanced repeated measurements designs. *The Annals of Statistics* 8, 1272-1283.

6. Fleiss, J. L. (1989) A critique of recent research on the two-treatment crossover design. *Controlled Clinical Trials* 10, 237-243.

7. 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.

8. Grieve, A. P. (1985). A bayesian analysis of the two-period crossover design in clinical trials. *Biometrics* 41, 979-990.

9. Hedayat, A. and Afsarinejad, K. (1975). Repeated measurements designs, I. *In a survey of Statistical Design and Linear Models* (J.N. Srivastava, Ed.) 229-242. North-Holland, Amsterdam.

10. Hedayat, A. and Afsarinejad, K. (1978). Repeated measurements designs, II. *The Annals of Statistics* 6, 619-628.

11. Jones, B. and Kenward, M. G. (2003). *Design and Analysis of Cross-Over Trials.* Second Edition, CRC press.

12. Jones, B., Kunert, J. and Wynn, H. P. (1992). Information matrices for mixed effects models with applications to the optimality of repeated measurements designs. *Journal of Statistical Planning and Inference* 33, 261-274.

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

14. Kenward, M. G. and Jones, B. (1987) The analysis of data from $2 \times 2$ cross-over trials with baseline measurements. *Statistics in Medicine* 6, 911-926.

15. Matthews, J. N. S. (1988). Recent developments in crossover designs. *International Statistical Review* 56, 117-127.

16. Matthews, J. N. S. (1994). Modeling and optimality in the design of crossover studies for medical applications. *Journal of Statistical Planning and Inference* 42, 89-108.

17. Patel, H. I. (1983). The use of baseline measurements in the two-period cross-over design. *Communications in Statistics* A12, 2693-2721.

18. Senn, S. (2002). *Crossover Trials in Clinical Research.* Second Edition, U.K. Wiley.

19. Stufken, J. (1996). Optimal crossover designs. *Handbook of Statistics.* Vol.13, Ghosh, S. and Rao, C. R. (Eds), 63-90. Amsterdam: North Holland.

20. Willan, A. R., and Pater, J. L. (1986). Using baseline measurements in the two-period crossover clinical trial. *Controlled Clinical Trials* 7, 282-289.

# Chapter 2

# Literature Review

## 2.1 Introduction

As mentioned in Chapter 1, in a crossover design, subjects are exposed to a sequence of different or identical treatments. This design is different from a parallel design under which each subject receives only one treatment during the entire experiment (Grieve, 1990; Vonesh and Chinchilli, 1997; Jones and Kenward, 2003).

The subjects in a crossover design could be patients in clinical trials, animals in animal husbandry experiments, plots of land in agricultural science, and so on. The order of treatments assigned to a subject in a crossover experiment is called a treatment sequence, with the first treatment to be used in the first period, the second treatment in the second period, and so on. Normally, the treatments are denoted by capital Latin letters, such as $A$, $B$, $C$, ect.

Typically, the main purpose of an experiment that uses a crossover design is the comparison of treatment effects.

The simplest and most popular design in the crossover family is for two treatments in two periods, which is also called a $2 \times 2$ crossover design and is sometimes denoted by $AB/BA$. In that design, subjects randomized to the $AB$ sequence receive treatment $A$ in the first period and treatment $B$ in the second period, whereas subjects randomized to the $BA$ sequence receive the treatments $A$ and $B$ in reverse order. This design has been applied in clinical trials extensively (Grizzle, 1965; Hills and Armitage, 1979; Brown, 1980).

Crossover designs were first applied in agricultural science, and can be traced back to 1853 when Laws and Liebig applied the crossover idea on nutrition of crop plants (Jones and Kenward, 2003). Since then, crossover designs have been widely used in all kinds of fields. Nowadays the most common application of crossover designs is in pharmaceutical studies and clinical trials in medical research. Detailed description, explanations, discussion and examples can be found in several books, such as Ratkowsky et al. (1992), Senn (2002), Jones and Kenward (2003); for models and analysis, the reader may refer to the papers by Grizzle (1965) and Brown (1980); Grieve (1985, 1994, 1995) used Bayesian methods to study crossover designs in clinical trials. Good review papers for crossover designs include those by Hedayat and Afsarinejad (1975); Hedayat (1981); Bishop and Jones (1984); Matthew (1988, 1994); Afsarinejad (1990); Stufken (1996); Jones and Deppe (2001).

A compelling reason for using crossover designs is that different treatments can be compared on the same subject, which, in general, facilitates more precise comparisons of the treatments. This property makes crossover designs especially important when there is large variability between subjects (Kershner and Federer, 1981).

Secondly, fewer subjects might be required in a crossover design in order to obtain the same level of statistical power, precision and efficiency as in a parallel design. So crossover designs are appealing for practical reasons when subjects are scarce or costly. For example, a crossover design could be a better choice in a clinical trial, when it is difficult and expensive to recruit more patients.

Thirdly, if the intent is to find the effect of different sequences of treatment applications as in drug, nutrition or learning experiments, or to discover whether or not a trend can be seen by successive applications of several treatments on the same subject, crossover design is a natural choice (Hedayat and Afsarinejad, 1978).

Although crossover designs possess attractive advantages, they also have some potential shortcomings.

First of all, the application of crossover designs in clinical trials has some limitations: it can be applied only if the treatments do not cure a disease but merely alleviate its condition (Vonesh and Chinchilli, 1997).

Secondly, crossover designs tend to take longer to complete compared to parallel designs (Stufken, 1996). The longer the duration of the experiment, the larger the chance of a subject dropping out, which complicates the analysis and adversely affects the information obtained.

In addition, the most serious disadvantage of a crossover design is the possibility that the effect of a treatment given in one period is still present in a subsequent period. This effect is called a carryover effect or residual effect. For example, in the $2 \times 2$ design, the observation in the second period for a subject that receives the $AB$ sequence could be affected by treatment $B$ assigned in the second period, but also by a carryover effect from treatment $A$ from the first period. Sufficiently long wash-out periods between treatment periods may be used to eliminate the carryover effects. But it is not always possible to know what "sufficiently long" means or to have long enough wash-out periods due to practical or ethical reasons.

Assuming subject effects to be random effects, Grizzle (1965) suggested a preliminary test for the equality of carryover effects in the $2 \times 2$ design, and to use observations from both periods only when there is no evidence of significant carryover effects. But Freeman (1989) found that the power for this preliminary test is low; also we would only use observations from the first period and lose the advantage of the crossover design when carryover effects exist. Under certain model assumptions, Willan and Pater (1986) observed that baseline measurements at the beginning of each period can be used to eliminate the carryover effects, but it can also raise other concerns.

Many authors even advise that a crossover design should not be used if carryover effects may be present (Brown, 1980; Fleiss, 1989; Senn, 2002). Therefore, inves-

tigators should make every effort to avoid these undesirable effects in the design stage, but it is not always possible. Furthermore, carryover effects should not prohibit experimenters from ever using a crossover design since the design's advantages may outweigh its flaws. Therefore, it is necessary and meaningful to use appropriate models to study the carryover effects, and also study optimal and efficient designs for models including carryover effects.

Among the papers that try to identify optimal or efficient crossover designs in the presence of carryover effects we mention Cheng and Wu (1980), Magda (1980), Kunert (1984, 1987), Sen and Mukerjee (1987), Matthews (1988, 1994), Afsarinejad (1989, 1990), Hedayat and Zhao (1990), Carrière and Reinsel (1993), Stufken (1996), Kushner (1998), Bose and Mukerjee (2000, 2003), Kempton, Ferris and David (2001), Kunert and Stufken (2002), Afsarinejad and Hedayat (2002), Hedayat and Stufken (2003) and Hedayat, Stufken and Yang (2006).

On the other hand, baseline measurements before each period are often available for a simple $2 \times 2$ crossover trial. Different methods are used to incorporate the baseline measurements in the analysis, including methods discussed by Patel (1983), Fleiss, Wallenstein and Rosenfeld (1985), Willan and Pater (1986), Kenward and Jones (1987), Grieve (1994), Senn (2002), Jones and Kenward (2003).

In this chapter, we present different ideas in the literature for modeling carryover effects and different ways to handle baseline measurements in crossover studies. We also provide a discussion about the traditional model and definitions, terminology and special design issues of crossover trials. The response variable here is assumed to be continuous. For the binary response case, the reader may refer to Koch (1972); Armitage and Hill (1982); Farewell (1985); Jones and Kenward (2003).

## 2.2 Definitions, terminology and design issues in crossover designs

Similar to other areas, crossover studies have their own special definitions and terminology. Following are some basic concepts and terminology.

The effect that a treatment has on the response of a subject during the period in which it is applied is called the direct treatment effect, also referred to as the direct effect or treatment effect. An effect of a treatment that lasts beyond the period of application is called a carryover or residual effect (Kershner and Federer, 1981). First-order carryover effect refers to the effect from the treatment in the previous period, second-order carryover effect to the effect from the treatment given two periods ago, and so on.

A wash-out period is an intermittent, inactive period between two active periods, which may be used to eliminate or reduce carryover effects.

A crossover design is said to be uniform on subjects, if each treatment appears the same number of times for any subject; and it is said to be uniform on periods if each treatment is assigned to the same number of subjects within each period. If a crossover design is uniform on both subjects and periods then it is said to be uniform.

A crossover design is said to be balanced with respect to first-order carryover effects (or balanced for short), if each treatment immediately precedes every other treatment the same number of times, and no treatment is immediately preceded by itself. A design is said to be strongly balanced with respect to first-order carryover effects (or just strongly balanced) if each treatment immediately precedes every treatment the same number of times, including itself. For a uniform balanced design with an equal number of periods as treatments, a strongly balanced design can be constructed by repeating the last period. For example, $AB/BA$ is a uniform, bal-

anced design, and $ABB/BAA$ is the strongly balanced design obtained by repeating the last period in the balanced design.

Optimality criteria used for crossover designs are usually a function of the information matrix for the treatment effects or of the variance-covariance matrix V of $(t-1)$ orthogonal and normalized contrasts between the $t$ treatments (Jones and Kenward, 2003). A small value of this function of V is an indicator of a good design. Some of the functions commonly used include:

D-optimality : A D-optimal design minimizes the determinant of V;

A-optimality : An A-optimal design minimizes the average variance of the best linear unbiased estimator of all pairwise treatment comparisons;

E-optimality : An E-optimal design minimizes the maximum among the variances of the best linear unbiased estimators for all nomalized treatment contrasts.

If we use $C_d$ and $\tilde{C}_d$ to denote the information matrices for the direct and carryover effects when a design $d$ is used, then $d$ is called $\Phi$-optimal for direct or carryover effects if it minimizes $\Phi(C_d)$ or $\Phi(\tilde{C}_d)$ respectively, where $\Phi$ is a function: $\beta_{t,0} \rightarrow (-\infty, \infty)$, and $\beta_{t,0}$ is the collection of $t \times t$ nonnegative definite matrices with zero row and column sums (Cheng and Wu, 1980).

Kiefer (1975) introduced the concept of universal optimality, which is an optimality criterion that includes many other criteria as special cases. He also provided sufficient conditions for a design $d$ to be universally optimal. It suffices that $d$ maximizes trace $(C_d)$ (trace $(\tilde{C}_d)$) in addition to $C_d$ ( $\tilde{C}_d$) being completely symmetric, where a completely symmetric matrix is a matrix for which all the diagonal elements are equal and all off diagonal elements are equal.

By the definition of universal optimality, it can be shown that a universally optimal design is also D-, A-, and E-optimal.

## 2.3 Traditional model and notation

Let the collection of all crossover designs, which are based on $t$ treatments, $n$ subjects and $p$ periods, be denoted by $\Omega_{t,n,p}$. If $d$ is a design in $\Omega_{t,n,p}$, let $d(i,j)$ denote the treatment assigned by $d$ to the $j^{th}$ subject in the $i^{th}$ period.

For the traditional model for crossover designs, it is assumed that the $np$ observations from $p$ periods and $n$ subjects results from a continues random variable $Y_{ij}$, which yields observed values $y_{ij}$ for the $j^{th}$ subject in the $i^{th}$ period; furthermore, the data are assumed to be uncorrelated with common variance $\sigma_\epsilon^2$. The model has been introduced many years ago, the idea underlying this model was used by Cochran, Autrey and Cannon (1941) to analyze the data from a crossover trial of different feeds for dairy cattle, and the model can be written as

$$\begin{aligned} &Y_{ij} = \mu + \alpha_i + \beta_j + \tau_{d(i,j)} + \gamma_{d(i-1,j)} + \epsilon_{ij}, \qquad (2.1)\\ &i = 1,\ldots,p\,,\ j = 1,\ldots,n,\ \gamma_{d(0,j)} \equiv 0 \quad \forall j. \end{aligned}$$

Here $\mu$ is an overall mean, $\alpha_i$ is an effect due to the $i^{th}$ period, $\beta_j$ is an effect due to the $j^{th}$ subject, $\tau_{d(i,j)}$ is a treatment effect due to treatment $d(i,j)$, $\gamma_{d(i-1,j)}$ is a first-order carryover effect due to treatment $d(i-1,j)$, and the $\epsilon_{ij}$ are random error terms, which are assumed to be $i.i.d.N(0,\sigma_\epsilon^2)$.

Let $\underline{1}_a$ and $\underline{0}_a$ indicate a $\times$ 1 vectors of 1's and 0's respectively, and let $I_a$ indicate the a $\times$ a identity matrix. We write Model 2.1 in matrix notation as:

$$Y = \mu 1_{pn} + X_1\alpha + X_2\beta + X_{d3}\tau + X_{d4}\gamma + \epsilon.$$

Here $Y = (Y_{11}, Y_{21}, ..., Y_{pn})'$, $\alpha = (\alpha_1, ..., \alpha_p)'$, $\beta = (\beta_1, ..., \beta_n)'$, $\tau = (\tau_1, ..., \tau_t)'$, $\gamma = (\gamma_1, ..., \gamma_t)'$, $\epsilon = (\epsilon_{11}, \epsilon_{21}, ..., \epsilon_{pn})'$, the matrices $X_1(pn \times p)$ and $X_2(pn \times n)$ are given by

$$X_1 = \begin{bmatrix} I_p \\ \vdots \\ I_p \end{bmatrix} = \underline{1}_n \otimes I_p, \quad X_2 = \begin{bmatrix} \underline{1}_p & \underline{0}_p & \cdots & \underline{0}_p \\ \underline{0}_p & \underline{1}_p & \cdots & \underline{0}_p \\ \vdots & \vdots & \ddots & \vdots \\ \underline{0}_p & \underline{0}_p & \cdots & \underline{1}_p \end{bmatrix} = I_n \otimes \underline{1}_p$$

($\otimes$ denotes the Kronecker product) and the matrices $X_{d3}$ and $X_{d4}$, both $pn \times t$, are of the form

$$X_{d3} = \begin{bmatrix} X_{d31} \\ X_{d32} \\ \vdots \\ X_{d3n} \end{bmatrix}, \qquad X_{d4} = \begin{bmatrix} X_{d41} \\ X_{d42} \\ \vdots \\ X_{d4n} \end{bmatrix},$$

where $X_{d3j}$ is the $p \times t$ period-treatment incidence matrix for subject $j$ under design $d$ with $X_{d4j} = LX_{d3j}$, and L is the $p \times p$ matrix defined as

$$L = \begin{bmatrix} 0 & 0 & \ldots & 0 & 0 \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}.$$

Optimal designs for this traditional model have been studied extensively, such as in Hedayat and Afsarinejad (1975, 1978), Cheng and Wu (1980), Kunert (1983, 1984), Stufken (1991, 1996), Kushner (1998) and Hedayat and Yang (2003, 2004).

However, the traditional model has been criticized for being too simplistic in assumptions about the carryover effects and the random error terms. Moreover, there are many other concerns about the model. For example, the model assumes the subject effects are fixed, which is inappropriate if these subjects are to be thought of as representatives of a larger population; the model does not include any interaction

terms, whose effects may be significant in some situations. Furthermore, if baseline measurements are available at the beginning of each period, which is not uncommon, then we would like to use a model and analysis that facilitate the incorporation of such additional information.

## 2.4 Different models to handle carryover effects and corresponding optimal designs

For reasons mentioned in the last section, many different models have been proposed to relax the implicit or explicit assumptions in the traditional model for different situations. For example, models with random subject effects have been studied widely (Brown, 1980; Carrière and Reinsel, 1993; Jones and Kenward, 2003; Hedayat, Stufken and Yang, 2006); models in which the errors are assumed to follow an autoregressive structure have been investigated extensively (Bora, 1984, 1985; Gill, 1992; Bellavance and Stephens, 1996; Kunert and Martin, 2000) since first proposed by Williams (1952).

Besides those alternatives, particular attention has been paid to the simplistic assumptions about the carryover effects, with a number of different models with alternative assumptions having been proposed in the literature. We will review those models and corresponding optimal designs in the next subsections. However, for detailed results on optimal designs, the reader should refer to the original papers.

The reader should be aware that no design is good for all models and for each model there could be many good designs (Afsarinejad and Hedayat, 2002).

All of the terms in the models in the following subsections have the same interpretation as they do in the traditional model, unless explicitly stated otherwise.

### 2.4.1 Circular and non-circular models

If carryover effects exist in the initial period and these effects are thought to come from the treatments in the last period, then a model that includes such kind of carryover effects is called a circular model; models discussed so far are non-circular.

Carryover effects in the first period may sound unreasonable intuitively, but Magda (1980) argued that there are practical situations where the carryover effects in the first period come from the treatments in the last period. For example, in agricultural experiments, if the same experiment is conducted repeatedly on the same land, then the carryover effects in the first period can be assumed to come from the treatments in the last period of the previous experiment (Hedayat, 1981). Moreover, if no such effects exist in the first period, a pre-period (or period 0) could be introduced in which treatments are given to the subjects only for the carryover effects for the first period. If the treatments in this pre-period are taken to be the same as in the last period, the carryover effects from the treatments in the last period will be observed during the first period.

Magda (1980) studied circular models in the presence or absence of period and subject effects, i.e. he considered the following four different models:

$$Y_{ij} = \alpha_i + \beta_j + \tau_{d(i,j)} + \gamma_{d(i-1,j)} + \epsilon_{ij}, \tag{2.2}$$

$$Y_{ij} = \alpha_i + \tau_{d(i,j)} + \gamma_{d(i-1,j)} + \epsilon_{ij}, \tag{2.3}$$

$$Y_{ij} = \beta_j + \tau_{d(i,j)} + \gamma_{d(i-1,j)} + \epsilon_{ij}, \tag{2.4}$$

$$Y_{ij} = \tau_{d(i,j)} + \gamma_{d(i-1,j)} + \epsilon_{ij}, \tag{2.5}$$

$$\gamma_{d(0,j)} = \gamma_{d(p,j)} \quad \forall j.$$

Thus, instead of assuming $\gamma_{d(0,j)} \equiv 0$ in the traditional model, we now have $\gamma_{d(0,j)} = \gamma_{d(p,j)}$.

Let $d_1^*$, $d_2^*$, $d_3^*$ and $d_4^*$ ($\delta_1^*$, $\delta_2^*$, $\delta_3^*$ and $\delta_4^*$) denote a circular strongly balanced (balanced) uniform crossover design, a circular strongly balanced (balanced) crossover design that is uniform on subjects, a circular strongly balanced (balanced) crossover design that is uniform on periods, and a circular strongly balanced (balanced) crossover design, respectively. Here circular strongly balanced and circular balanced are defined similarly as strongly balanced and balanced in Section 2.2, except that a treatment in the last period has now carryover effect on a measurement in the first period. Magda (1980) proved that whenever designs $d_1^*$, $d_2^*$, $d_3^*$ and $d_4^*$ exist, they are universally optimal for the estimation of direct as well as carryover effects over the collection of designs with the same parameters under Models 2.2, 2.3, 2.4 and 2.5, respectively; whenever designs $\delta_1^*$, $\delta_2^*$, $\delta_3^*$ and $\delta_4^*$ exist, they are universally optimal for the estimation of direct as well as carryover effects over the collection of designs with the same parameters and the restriction that no treatment precedes itself under Models 2.2, 2.3, 2.4 and 2.5 respectively.

Hedayat (1981) restated Magda's results in a review paper, and discussed the practical applicability of the four models in detail. Afsarinejad (1989) provided constructions of the circular balanced uniform designs.

The circular models avoid lack of carryover effects in the first period by assuming a seemingly artificial carryover effect incurred from the last period, which brought some technical advantages in proving results. But introducing a "pre-period" before the first treatment period just for the carryover effects and without collecting or using the data generated seems wasteful in practice. "In the medical context, it would not only appear bizarre to clinical colleagues but would be ethically dubious" (Matthews, 1994).

### 2.4.2 Models with interaction between the direct treatment and the carryover treatment

In the traditional model, a carryover effect depends only on the treatment assigned in the previous period, no matter which treatment provides the direct effect. Conceivably, in some situations, there could be an interaction between the treatment in current period and the treatment from the previous period (i.e., an interaction between the direct treatment and the carryover treatment).

Sen and Mukerjee (1987) considered a model with interaction between direct and carryover treatment, such that each treatment is allowed to have a different carryover effect depending on the treatment in the next period. They looked at both circular and non-circular models:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \xi_{d(i,j)d(i-1,j)} + \epsilon_{ij}, \tag{2.6}$$

$$Y_{ij} = \begin{cases} \mu + \alpha_i + \beta_j + \tau_{d(i,j)} + \epsilon_{ij}, & i = 1, \\ \mu + \alpha_i + \beta_j + \xi_{d(i,j)d(i-1,j)} + \epsilon_{ij}, & i = 2, \ldots, p, \end{cases} \tag{2.7}$$

where $\xi_{d(i,j)d(i-1,j)}$ is the sum of a direct treatment effect, a carryover effect and an interaction.

They investigated whether the optimality results under the traditional model (Cheng and Wu, 1980) and the circular model (Magda, 1980) remain valid when the interaction of the direct and carryover treatment is taken into account. They showed that a strongly balanced uniform crossover design under the non-additive circular model is still universally optimal over $\Omega_{t,n,p}$ for the estimation of direct treatments as well as for carryover effects. For the traditional model, this conclusion holds also for the estimation of direct treatments, but for the estimation of carryover effects, additional conditions on the design are needed. Furthermore, it is shown that the optimality results for balanced uniform crossover designs are no longer valid under this non-additive model.

This model has however been criticized for containing too many parameters to be practically useful, especially when the number of treatments is large.

### 2.4.3 Model with self and simple mixed carryover effects

Instead of allowing each treatment to have a different carryover effect depending on the treatment in the next period, Afsarinejad and Hedayat (2002) allowed each treatment to have two different types of carryover effects, the self and simple mixed carryover effect. Here a self-carryover effect refers to a carryover effect that applies if the treatment providing the direct effect is the same as that providing the carryover effect, while a simple mixed carryover effect refers to a carryover effect that applies if the two treatments are different.

The model used for the self and simple mixed carryover effects can be formulated as

$$Y_{ij} = \begin{cases} \alpha_i + \beta_j + \tau_{d(i,j)} + \chi_{d(i-1,j)} + \epsilon_{ij}, & \text{if } d(i,j) = d(i-1,j), \\ \alpha_i + \beta_j + \tau_{d(i,j)} + \rho_{d(i-1,j)} + \epsilon_{ij}, & \text{if } d(i,j) \neq d(i-1,j), \end{cases} \tag{2.8}$$

where $\chi_l$ and $\rho_l$ are the self-carryover effect and simple mixed carryover effect of treatment $l$, respectively.

As Afsarinejad and Hedayat (2002) pointed out, studying and estimating the self and simple mixed carryover effects can be very important in many fields. For example, it will be very helpful to know the impact of the simple mixed-carryover effects to arrange the best crop rotation schedule in agricultural science, and it will be very meaningful to know the self-carryover effect of the drug for a patient in a single drug study in medical science.

Afsarinejad and Hedayat (2002) were the first to investigate this model for two-period crossover designs with two or more treatments. By using a statistical tool developed by Hedayat and Zhao (1990), they connected the problem of optimal two-period crossover designs with optimal block designs. They showed that if a study is

designed properly, then unbiased and efficient estimators of all contrasts in direct treatment effects can be obtained if $t \geq 3$. However, the contrasts in both the direct treatment effects and the simple mixed carryover effects cannot be unbiasedly estimated if $t = 2$. Furthermore, if all treatments happened at least once for subjects with the same treatment in the two periods, then all contrasts in self-carryover effects are estimable.

Kunert and Stufken (2002) studied the same model for designs with more than two treatments, and they showed that this model leads for $t \geqslant 3$ to optimal designs for direct treatment effects that have the attractive feature that they avoid pairs of consecutive identical treatments.

So far the results we discussed for the self and simple mixed carryover effects are for designs with more than two treatments. Results for two treatments are studied by Kunert and Stufken (2007). They pointed out that the model with self and simple mixed carryover effects for only two treatments is equivalent to the model with interaction of the direct and carryover treatments. They identified the optimal designs for the case of two treatments under this model. Meanwhile, they are very strict about the model for the design with only two treatments and two periods, since there is no unbiased estimator for the treatment contrast.

### 2.4.4 Model with carryover effects proportional to direct treatment effects

During crossover experiments, it seems plausible that a treatment with a large direct effect should generally have a larger carryover effect. Based on this phenomenon, Patterson and Lucas (1962) considered a model where carryover effects are proportional to the direct treatment effects. They provided a test for common values of the proportionality parameter for all the treatments. Sen and Sinha (1986) applied this model for data analysis based on situations whether the proportionality parameter

$\lambda_t$ is known or unknown, and equal or unequal for all treatments $t$. Kempton, Ferris and David (2001) studies the optimality of this model, where they assumed that the proportionality parameter $\lambda$ is unknown. The model may be given by

$$Y_{ij} = \mu + \alpha_i + \beta_j + \tau_{d(i,j)} + \lambda\tau_{d(i-1,j)} + \epsilon_{ij}, \tag{2.9}$$

where $\lambda\tau$ is the carryover effect and $\lambda$ is the constant of proportionality. It would normally be expected that $|\lambda| \leq 1$.

The complication of this model is that it is nonlinear for the direct treatment effects and the constant of proportionality. Kempton, Ferris and David (2001) considered least squares estimation for the parameters and used a linear approximation of the model at the true values of the parameters. By using a combination of analytical results and computer search, they identified optimal designs for the estimation of direct effects based on some extensions of the A-optimality criterion.

Bose and Stufken (2007) considered the same model but assumed that the constant of proportionality $\lambda$ is a known constant. By doing so, they avoided the problem raised by the nonlinearity of the direct treatment effects and the constant of proportionality. They used the sufficient conditions introduced by Kiefer (1975) to search for universally optimal designs, and gave a list of optimal designs for different combinations of the number of treatments and periods for $2 \leq p \leq 4$.

However, in practice, it is impossible that the constant of proportionality $\lambda$ is known. Considering the average performance over a distribution of unknown parameter $\tau_0$ and small absolute value for $\lambda_0$, Bailey and Kunert (2006) studied this model in determination of an $\bar{A}$-optimal design, where a $\bar{A}$-optimal design is a design that minimizes the A-criterion averaged on the distribution of the unknown parameter $\tau_0$.

### 2.4.5 Models with higher-order carryover effects

The models we have considered so far include only first-order carryover effects; some authors have considered models that also include higher-order, often second-order carryover effects (Kershner and Federer, 1981).

Carryover effects of each treatment on all succeeding periods were studied by Lakatos and Raghavarao (1987), while Bose and Mukherjee (2000) introduced a general model for possible carryover effects up to the $k^{th}$ order and interactions among the successive treatments applied on a subject. Their model may be written as

$$Y_{ij} = \begin{cases} \mu + \alpha_i + \beta_j + \xi_{d(i,j),d(i-1,j),\ldots,d(1,j)} + \epsilon_{ij}, & \text{if } 1 \leq i \leq k-1, \\ \mu + \alpha_i + \beta_j + \xi_{d(i,j),d(i-1,j),\ldots,d(i-k+1,j)} + \epsilon_{ij}, & \text{if } k \leq i \leq p, \end{cases} \tag{2.10}$$

where $\xi_{h_1,h_2,\ldots,h_m} (1 \leq m \leq k)$ stands for the effect due to the treatment $h_1$ being applied in the current period, $h_2$ in the previous period, ..., $h_m$ in the $(m-1)^{th}$ preceding period $(1 \leq h_1, h_2, \ldots, h_m \leq t)$. Thus, the term $\xi_{h_1,h_2,\ldots,h_m}$ is then modeled as the direct treatment effect of $h_1$, the first-order carryover effect of $h_2$, ..., the $(m-1)^{th}$ order carryover effect of $h_m$, together with interactions of these $m$ factors.

This model can be viewed as a generalization of the traditional model by incorporating higher-order carryover effects and interaction between direct and carryover treatment. It is more flexible in the sense that an experimenter can choose an appropriate value for $k$ to allow carryover effects of different order and their interactions in the model. By applying the calculus for factorial arrangements, Bose and Mukherjee (2000) obtained a class of optimal designs under this model; they also provided one method of constructing these optimal designs.

Bose and Mukherjee (2003) extended this model by considering random subject effects, thereby relaxing a major assumption of the traditional model. They identified some universally optimal designs under this model.

Even if higher-order carryover effects exist, the modeling of such effects as in Model 2.10 may be too simplistic. If the models are good approximations, we may still wind up with rather inefficient estimators of the direct treatment effects (which are typically of most interest) due to the presence of so many nuisance parameters. Therefore, we should do everything possible at the design stage to avoid having to deal with higher-order carryover effects.

The above is a brief review of the literature concerning various ways to model carryover effects. There is also a considerable literature on the use of baseline measurements in crossover trials; we consider this topic in the next section.

## 2.5 Use of baseline measurements in crossover studies

Baselines are measurements made on the subjects to give general or background information during non-treatment periods. These can be used to improve inferences for direct treatment effects (Senn, 2002). Baseline measurements can often be obtained before the beginning of each treatment period. Most of the research done on baseline measurements in crossover trials is for two-period crossover trials. Different models and methods have been introduced to use the information provided by the baseline measurements. We review several of them in this section.

First, let us look at the different models that incorporate the baseline measurements in the literature. The differences between those models can be characterized by the different assumptions about fixed effects and random effects. Therefore, they can be described by cell means and variance-covariance structures. In a two-period crossover design, if baseline measurements are available at the beginning of each period, each subject yields four measurements, and the corresponding time periods may be referred to as the run-in period, the first treatment period, the wash-out period and the second treatment period, respectively.

Willan and Pater (1986) assumed that period effects of the run-in period are the same as for the first treatment period, the wash-out period is the same as the second treatment period, and carryover effects in wash-out period are identical to those in the second treatment period. In addition to assuming random subject effects, they also allowed for an interaction between subjects and periods. Fleiss, Wallenstein and Rosenfeld (1985) had the same assumption for the period effects, but they allowed for different carryover effects in the wash-out period and the second treatment period and assumed the variance-covariance structure for observations on the same subject to be compound symmetric. Different from those models, Chi (1993) and Kenward and Jones (1987) assumed that the period effects in the periods of run-in, first treatment, wash-out and second treatment are all different, and Chi (1993) assumed that the carryover effects are the same in both wash-out period and second treatment period, while Kenward and Jones (1987) made no assumption about the equality of the carryover effects from those two periods. In addition to the random subject effects, Chi (1993) considered an error structure following an autoregressive process for each subject, while Kenward and Jones (1987) used an unstructured variance-covariance matrix for the measurements from each subject. Kenward and Jones (1987) also included a sequence effect in their model.

To summarize, we illustrate the cell means and variance-covariance structures for these models in Tables 2.1 and 2.2, where we denote models proposed by Willan and Pater (1986), Fleiss, Wallenstein and Rosenfeld (1985), Chi (1993) and Kenward and Jones (1987) by models I, II, III and IV, respectively.

Table 2.1 Cell Means for Models that Incorporate Baseline Measurements

| | | Period 1 | | Period 2 | |
|---|---|---|---|---|---|
| Model | Sequence | Run-in | First Treatment | Wash-out | Second Treatment |
| I | $AB$ | $\mu+\pi_1$ | $\mu+\pi_1+\tau$ | $\mu+\pi_2+\lambda$ | $\mu+\pi_2-\tau+\lambda$ |
| | $BA$ | $\mu+\pi_1$ | $\mu+\pi_1-\tau$ | $\mu+\pi_2-\lambda$ | $\mu+\pi_2+\tau-\lambda$ |
| II | $AB$ | $\mu+\pi_1$ | $\mu+\pi_1+\tau$ | $\mu+\pi_2+\theta$ | $\mu+\pi_2-\tau+\lambda$ |
| | $BA$ | $\mu+\pi_1$ | $\mu+\pi_1-\tau$ | $\mu+\pi_2-\theta$ | $\mu+\pi_2+\tau-\lambda$ |
| III | $AB$ | $\mu+\pi_1$ | $\mu+\pi_2+\tau$ | $\mu+\pi_3+\lambda$ | $\mu+\pi_4-\tau+\lambda$ |
| | $BA$ | $\mu+\pi_1$ | $\mu+\pi_2-\tau$ | $\mu+\pi_3-\lambda$ | $\mu+\pi_4+\tau-\lambda$ |
| IV | $AB$ | $\mu+\gamma+\pi_1$ | $\mu+\gamma+\pi_2+\tau$ | $\mu+\gamma+\pi_3+\theta$ | $\mu+\gamma+\pi_4-\tau+\lambda$ |
| | $BA$ | $\mu-\gamma+\pi_1$ | $\mu-\gamma+\pi_2-\tau$ | $\mu-\gamma+\pi_3-\theta$ | $\mu-\gamma+\pi_4+\tau-\lambda$ |

$\mu$: overall mean; $\pi_i$: $i^{th}$ period effect; $\theta$: carryover effect in wash-out period
$\gamma$: sequence effect; $\tau$: treatment effect; $\lambda$: carryover effect in the second treatment effect

Table 2.2 Covariance Structures for Models that Incorporate Baseline Measurements

| Model | Covariance Structure |
|---|---|
| I | $\sigma_s^2\begin{pmatrix}1&1&1&1\\1&1&1&1\\1&1&1&1\\1&1&1&1\end{pmatrix}+\sigma_{sp}^2\begin{pmatrix}1&1&0&0\\1&1&0&0\\0&0&1&1\\0&0&1&1\end{pmatrix}+\sigma_\epsilon^2\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}$ |
| II | $\sigma_s^2\begin{pmatrix}1&1&1&1\\1&1&1&1\\1&1&1&1\\1&1&1&1\end{pmatrix}+\sigma_\epsilon^2\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}$ |
| III | $\sigma_s^2\begin{pmatrix}1&1&1&1\\1&1&1&1\\1&1&1&1\\1&1&1&1\end{pmatrix}+\sigma_\epsilon^2\begin{pmatrix}1&\rho&\rho^{1+w}&\rho^{2+w}\\\rho&1&\rho^{w}&\rho^{1+w}\\\rho^{1+w}&\rho^{w}&1&\rho\\\rho^{2+w}&\rho^{1+w}&\rho&1\end{pmatrix}$ |
| IV | no assumptions about the covariance structure |

$\sigma_s^2$: variance of subject; $\sigma_{sp}^2$: variance of subject by period interaction; $\sigma_\epsilon^2$ : random error
$\rho$: autocorrelation parameter; $w$: time units for the wash-out period

Next, let us look at the different methods which use the information provided by the baseline measurements.

Based on their assumptions for the period effects, carryover effects as well as random effects, Willan and Pater (1986) considered the analyses based on the outcomes only as well as on the change from baselines. They obtained estimators for the treatment contrast and their variances from both methods, and showed that the analysis of change from baselines can eliminate the carryover effects for the treatment comparison under their model and assumptions. However, after further investigation of the power for the test of a difference in treatment effects and the relative precision of the estimators of the treatment contrast in terms of mean squared error (MSE) under both methods, they concluded that the analysis of change from baselines reduces the power and precision under many situations. They also derived a condition to decide whether or not to include the baseline measurements in the analysis.

Assuming that carryover effects are not equal in the wash-out period in two sequences, but the total time duration until the end of the second treatment period is long enough to eliminate any carryover effects, Fleiss, Wallenstein and Rosenfeld (1985) showed that the analysis of change from baselines may produce different carryover effects, when none actually exist if one analyzes the data from outcomes only. Therefore, they also concluded that use of baseline measurements from both periods should be undertaken cautiously.

Patel (1983) suggested that baseline measurements be used for a number of preliminary tests to determine the validity of test of treatment contrast. If all of the null hypotheses are not rejected, then one may use the baseline measurements as a covariate in making inferences about the treatment effects. He also compared the variances for the estimators of the treatment contrast from different methods. However, his results for the method of using baseline measurements as a covariate

were based on the assumption that the random terms from the two periods are independent. When, actually, they are correlated under the model assumptions.

However, all the analyses discussed considered carryover effects on the wash-out period. Therefore, in a strict sense, the measurements from the wash-out period are not true baselines; they were affected by the treatments from the previous period. Fleiss (1989) pointed out some problems with including baseline measurements made at the start of the second period in the analysis. Kenward and Jones (1987) also mentioned that there is no satisfactory statistical analysis for the $2 \times 2$ crossover trial if there is any possible presence of carryover effects. Therefore, some authors, for example Senn (2002), consider measurements obtained prior to the second treatment period as baselines only when wash-out periods are long enough to eliminate any carryover effects. We will also consider this case in Chapters 4-6.

## 2.6 References


1. Afsarinejad, K. (1989). Circular balanced uniform repeated measurements designs. *Statistics and Probability Letters* 7, 187-189.

2. Afsarinejad, K. (1990). Repeated measurement designs - A review. *Communications in Statistics* 19, 3985-4028.

3. Afsarinejad, K. and Hedayat, A. S. (2002). Repeated measurements designs for a model with self and simple mixed carryover effects. *Journal of Statistical Planning and Inference* 106, 449-459.

4. Armitage, P. and Hills, M. (1982). The two-period crossover trial. *The Statistician* 32, 119-131.

5. Bailey, R. A. and Kunert, J. (2006). On optimal cross-over designs when carryover effects are proportional to direct effects. *Biometrika* 93, 613-625.

6. Bellavance, F., Tardif, S. and Stephens, M. A. (1996). Tests for the analysis of variance of crossover designs with correlated errors. *Biometrics* 52, 608-612.

7. Bishop, S. H. and Jones, B. (1984). A review of higher order crossover designs. *Journal of Applied Statistics* 11, 29-50.

8. Bora, A. C. (1984). Change over design with errors following a first order autoregressive process. *Australian Journal of Statistics* 26, 179-188.

9. Bora, A. C. (1985). Change over design with first order residual effects and errors following a first order autoregressive process. *The Statistician* 34, 161-173.

10. Bose, M. and Mukherjee, B. (2000). Cross-over designs in the presence of higher order carry-overs. *Australian and New Zealand Journal of Statistics* 42, 235-244.

11. Bose, M. and Mukherjee, B. (2003). Optimal cross-over designs under a general model. *Statistics and Probability Letters* 62, 413-418.

12. Bose, M. and Stufken, J. (2007). Optimal crossover designs when carryover effects are proportional to direct effects. *Journal of Statistical Planning and Inference*, to appear.

13. Brown, B. W. (1980). The crossover experiment for clinical trials. *Biometrics* 36, 69-79.

14. Carrière, K. C. and Reinsel, G. C. (1993). Optimal two-period repeated measurement designs with two or more treatments. *Biometrika* 80, 924-929.

15. Cheng, C. S. and Wu, C. F. (1980) Balanced repeated measurements designs. *The Annals of Statistics* 8, 1272-1283.

16. Chi, E. M. (1993) Analysis of cross-over trials when within-subject errors follow an AR(1) process. *Biometrical Journal* 34, 359-365.

17. Cochran, W. G., Autrey, K. M. and Cannon, C. Y. (1941) A double change-over design for dairy cattle feeding experiments. *Journal of Dairy Sciences* 24, 937-951.

18. Farewell, V. T. (1985) Some remarks on the analysis of cross-over trials with binary response. *Applied Statistics* 34, 121-128.

19. Fleiss, J. L. (1989) A critique of recent research on the two-treatment crossover design. *Controlled Clinical Trials* 10, 237-243.

20. 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.

21. Freeman, P. (1989) The performance of the two-stage analysis of two-treatment, two-period crossover trials. *Statistics in Medicine* 8, 1421-1432.

22. Gill, P. S. (1992). Balanced change-over designs for autocorrelated observations. *The Australian Journal of Statistics* 34, 415-420.

23. Grizzle, J. E. (1965). The two-period change-over design and its use in clinical trials. *Biometrics* 21, 467-480.

24. Grieve, A. P. (1985). A Bayesian analysis of the two-period crossover design for clinical trials. *Biometrics* 41, 979-990.

25. Grieve, A. P. (1990). *Crossover versus parallel designs. In: Statistical Methodology in the Pharmaceutical Sciences (Berry DA, ED).* Marcel Dekker, New York.

26. Grieve, A. P. (1994). Extending a Bayesian analysis of the two-period crossover to allow for baseline measurements. *Statistics in Medicine* 13, 905-929.

27. Grieve, A. P. (1995). Extending a Bayesian analysis of the two-period crossover to accommodate missing data. *Biometrika* 82, 227-286.

28. Hedayat, A. and Afsarinejad, K. (1975). Repeated measurements designs, I. *In a survey of Statistical Design and Linear Models* (J.N. Srivastava, Ed.) 229-242. North-Holland, Amsterdam.

29. Hedayat, A. and Afsarinejad, K. (1978). Repeated measurements designs, II. *The Annals of Statistics* 6, 619-628.

30. Hedayat, A. S. (1981). Repeated measurements designs, IV: Recent advances *Invited paper 43rd Session of the International Statistical Institute* November 30th - December 11th, 1981, 1-14. Buenos Aires - Argentina.

31. Hedayat, A. S. and Zhao, W. (1990) Optimal two-period repeated measurements designs. *The Annals of Statistics* 18, 1805-1816.

32. Hedayat, A. S. and Stufken, J. (2003) Optimal and efficient crossover designs under different assumptions about the carryover effects. *Journal of Biopharmaceutical Statistics* 13, 519-528.

33. Hedayat, A, Stufken, J. and Yang, M. (2006). Optimal and efficient crossover designs when subject effects are random. *Journal of the American Statistical Association* 101, 1031-1038.

34. Hedayat, A. S. and Yang, M. (2003) Universal optimality of balanced uniform crossover designs. *The Annals of Statistics* 31, 978-983.

35. Hedayat, A. S. and Yang, M. (2004) Universal optimality for selected crossover designs. *Journal of the American Statistical Association* 99, 461-466.

36. Hills, M. and Armitage, P. (1979) The two-period cross-over clinical trial. *British Journal of Clinical Pharmacology* 8, 7-20.

37. Jones, B. and Deppe, C. (2001) Recent developments in the design of cross-over trials: a brief review and bibliography. *Recent Advances in Experimental Designs and Related topics* 153-173, New York.

38. Jones, B. and Kenward, M. G. (2003). *Design and Analysis of Cross-Over Trials.* Second Edition, CRC press.

39. Kiefer, J. (1975) Construction and optimality of generalized Youden design. *In a Survey of Statistical Designs and Linear Models* Amsterdam: North Holland, 333-353.

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

41. Kenward, M. G. and Jones, B. (1987) The analysis of data from 2 $\times$ 2 cross-over trials with baseline measurements. *Statistics in Medicine* 6, 911-926.

42. Kershner, R. P., Federer, W. T. (1981) Two-treatment crossover designs for estimating a variety of effects. *Journal of the American Statistical Association* 76, 612-619.

43. Koch, G. G. (1972) The use of non-parametric methods in the statistical analysis of the two-period change-over design. *Biometrics* 28, 577-584.

44. Kushner, H. B. (1998) Optimal and efficient repeated-measurement designs for uncorrelated observations. *Journal of the American Statistical Association* 93, 1176-1187.

45. Kunert, J. (1983). Optimal design and refinement of the linear model with applications to repeated measurements designs. *The Annals of Statistics* 11, 247-257.

46. Kunert, J. (1984). Optimality of balanced uniform repeated measurements designs. *The Annals of Statistics* 12, 1006-1017.

47. Kunert, J. (1987). On variance estimation in crossover designs. *Biometrics* 43, 833-845.

48. Kunert, J. and Martin, R. J. (2000). Optimality of type I orthogonal arrays for cross-over models with correlated errors. *Journal of Statistical Planning and Inference* 87, 119-124.

49. 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.

50. Kunert, J. and Stufken, J. (2007). Optimal crossover designs for two treatments in the presence of mixed and self carryover effects. Under revision.

51. Lakatos, E. and Raghavarao, D. (1987). Undiminished residual effects designs and their suggested applications. *Communications in Statistics* A16, 1345-1350.

52. Magda, C. G. (1980). Circular balanced repeated measurements designs. *Communications in Statistics* A9, 1901-1918.

53. Matthews, J. N. S. (1988). Recent developments in crossover designs. *International Statistical Review* 56, 117-127.

54. Matthews, J. N. S. (1994). Modeling and optimality in the design of crossover studies for medical applications. *Journal of Statistical Planning and Inference* 42, 89-108.

55. Patel, H. I. (1983). The use of baseline measurements in the two-period cross-over design. *Communications in Statistics* 12, 2693-2721.

56. Patterson, H. D. and Lucas, H. L. (1962) Change-over designs. *North Carolina Agriculture Experimental Station, Technical Bulletin* 147, 3-53.

57. Ratkowsky, D. A., Evans, M. A., and Alldredge, J. R. (1992). *Cross-over Experiments: Design, Analysis and Application.* Marcel Dekker, New York.

58. Sen, M. and Mukerjee, R. (1987). Optimal repeated measurements designs under interaction. *Journal of Statistical Planning and Inference* 17, 81-91.

59. Sen, M. and Sinha, B. K. (1986). A statistical analysis of serially balanced sequences: first order residual proportional to direct effects. *Calcutta Statistical Associate Bulletin* 35, 31-36.

60. Senn, S. (1994) The *AB/BA* crossover: past, present, and future? *Statistical Methods in Medical Research* 3, 303-324.

61. Senn, S. (2002). *Crossover Trials in Clinical Research.* Second Edition, U.K. Wiley.

62. Stufken, J. (1991). Some families of optimal and efficient cross-over designs. *Journal of Statistical Planning and Inference* 27, 75-83.

63. Stufken, J. (1996). Optimal crossover designs. *Handbook of Statistics.* Vol.13, Ghosh, S. and Rao, C. R. (Eds), 63-90. Amsterdam: North Holland.

64. Vonesh, E. F. and Chinchilli, V. M. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*, New York.

65. Willan, A. R., and Pater, J. L. (1986). Using baseline measurements in the two-period crossover clinical trial. *Controlled Clinical Trials* 7, 282-289.

66. Williams, R. M. (1952). Experimental designs for serially correlated observations. *Biometrika* 39, 151-167.

# Chapter 3

# Models with Subject by Carryover Interaction

## 3.1 Introduction

Crossover designs, in which various treatments are applied to each subject over different time periods of the study, have been widely used in different fields. The primary goal of an experiment that uses crossover designs is typically to compare the different treatment effects, which can be done on the same subject with these designs. This characteristic makes crossover designs attractive especially when there is substantial variability among the subjects.

The traditional model for crossover designs, which includes period, subject, direct treatment and carryover effects, has been introduced in the early 40's. In this model carryover effects are assumed to exist only from the treatment in the previous period, and it is assumed that they do not depend on the treatment applied in the current period. This model has been adopted by many authors during the last four or five decades, and optimal designs for this model have also been studied extensively by a number of authors after the initial work of Hedayat and Afsarinejad (1978), Cheng and Wu (1980), and Magda (1980).

However, more recently, this model has been criticized by many authors, especially for its simplistic and unrealistic assumptions about the carryover effects (Fleiss, 1989; Senn, 2002). Partly in reaction to these criticisms, several models with more complicated assumptions about carryover effects have been investigated recently, for

example, Kempton, Ferris and David (2001), Afsarinejad and Hedayat (2002) and Bose and Mukherjee (2000, 2003).

But their contributions focused on theoretical aspects of the models or on identification and construction of optimal designs under these models. There is little evidence based on data to support any of these models. In medical studies, some treatments have obvious effects on the response in the current period or subsequent periods for some patients, but less for others. It may be of importance to know that treatment and carryover effects are different for different subjects.

Matthews (1988, 1994) mentioned that the interaction of subject by treatment could be of interest, however he presented no further discussion about any research on this topic. Chinchilli and Esinhart (1996), Ghosh and Fairchild (2000) and Ghosh and Crosby (2005) studied the model with treatment by subject interactions, but consideration of different carryover effects in different subjects does not seem to have received any attention in the literature.

Jones, Kunert and Wynn (1992) considered a model with random carryover effects, which were assumed normally distributed with mean 0 and variance $\sigma^2$, and independent of the random errors. In their argument, crossover designs should only be applied when carryover effects are small; treating carryover effects as random can provide more efficient estimates in the sense of avoiding over-correcting for them. However, their assumption seems to be rather ad hoc, since, conceptually, there is no large or infinite population of carryover effects from which those in the experiment were randomly selected. This paper, too, focused mostly on theoretical derivations.

In order to capture the mentioned variabilities, we formulate a model which assumes random subject effects, and includes subject by treatment and subject by carryover interactions. Since subject effects are assumed to be random, so are these interactions. This yields a more general variance-covariance structure for the responses from each subject.

By introducing several additional random effects, we must deal with model identifiability. Estimation of the model parameters is another problem that we face, as is a study of the flexibility and practical applicability of the model.

Our main goal in this chapter is to investigate basic statistical properties of this model, demonstrate its flexibility, show its practical usefulness, and discuss methods estimating the relevant parameters associated with the model.

This chapter is organized as follows. In Section 3.2, we propose models to incorporate the subject by carryover interaction and study identifiability of the random components. The flexibility of the models is established in Section 3.3. Maximum likelihood based estimations of the model parameters are discussed in Section 3.4. In Section 3.5, a simulation study is conducted for particular designs to evaluate the performance of REML estimation, and the *AIC* model selection criterion is used to compare the models with and without the random interaction terms. A data set available in the literature is used to illustrate the methods in Section 3.6. Finally, we provide a brief discussion in Section 3.7.

## 3.2 Statistical models and their properties

Consider a crossover experiment in which $t$ treatments are allocated to $n$ subjects in $p$ time periods. The response from the $j^{th}$ subject in the $i^{th}$ period will be denoted by $Y_{ij}$. A possible model to incorporate subject by carryover and subject by treatment interactions may be written as:

$$\begin{aligned} Y_{ij} &= \mu + \pi_i + \tau_{d(i,j)} + \gamma_{d(i-1,j)} + s_j + (s\tau)_{jd(i,j)} + (s\gamma)_{jd(i-1,j)} + \epsilon_{ij}, \qquad (3.1) \\ & \qquad i = 1, \ldots, p\,;\; j = 1, \ldots, n;\; \gamma_{d(0,j)} \equiv 0 \quad \forall\, j;\; (s\gamma)_{jd(0,j)} \equiv 0 \quad \forall\, j, \end{aligned}$$

where $\mu$, $\pi_i$, $\tau_{d(i,j)}$, $\gamma_{d(i-1,j)}$, $s_j$ and $\epsilon_{ij}$ are, respectively, an overall mean, the $i^{th}$ period effect, a treatment effect due to the treatment assigned to the $j^{th}$ subject in the $i^{th}$ period (i.e., treatment $d(i,j)$), a first-order carryover effect due to treatment

$d(i-1,j)$, the $j^{th}$ subject effect, and a random error term. Terms $(s\tau)_{jd(i,j)}$ and $(s\gamma)_{jd(i-1,j)}$ are the subject by treatment interaction and the subject by carryover interaction, respectively.

In the most general form of the model, we assume that $s_j$, $(s\tau)_{jd(i,j)}$, $(s\gamma)_{jd(i-1,j)}$ and $\epsilon_{ij}$ are distributed as $N(0,\sigma_s^2)$, $N(0,\sigma_{s\tau}^2)$, $N(0,\sigma_{s\gamma}^2)$ and $N(0,\sigma_\epsilon^2)$, respectively, and that the only non-zero covariances are those shown in the following variance-covariance matrix:

$$\text{var}\begin{pmatrix} s_j \\ (s\tau)_{j1} \\ \vdots \\ (s\tau)_{jt} \\ (s\gamma)_{j1} \\ \vdots \\ (s\gamma)_{jt} \end{pmatrix} = \begin{pmatrix} \sigma_s^2 & 0 & \dots & \dots & 0 & 0 & \dots & \dots & 0 \\ & \sigma_{s\tau}^2 & \theta_{s\tau} & \dots & \theta_{s\tau} & 0 & \dots & \dots & 0 \\ & & \ddots & \ddots & \vdots & \vdots & \ddots & & \vdots \\ & & & \sigma_{s\tau}^2 & \theta_{s\tau} & \vdots & & \ddots & \vdots \\ & & & & \sigma_{s\tau}^2 & 0 & \dots & \dots & 0 \\ & & & & & \sigma_{s\gamma}^2 & \theta_{s\gamma} & \dots & \theta_{s\gamma} \\ & & & & & & \ddots & \ddots & \vdots \\ & & & & & & & \sigma_{s\gamma}^2 & \theta_{s\gamma} \\ & & & & & & & & \sigma_{s\gamma}^2 \end{pmatrix}.$$

Thus, $\theta_{s\tau}$ and $\theta_{s\gamma}$ are the covariances for different subject by treatment interactions and subject by carryover interactions, respectively, belonging to the same subject.

In matrix notation, Model 3.1 may be written as follows:

$$Y = \mu 1_{pn} + X_1\pi + X_{d1}\tau + X_{d2}\gamma + Z_1 s + Z_{d1}(s\tau) + Z_{d2}(s\gamma) + \epsilon,$$

where

$$X_1 = \begin{pmatrix} I_p \\ \vdots \\ I_p \end{pmatrix} = 1_n \otimes I_p, \qquad X_{d1} = \begin{pmatrix} X_{d11} \\ X_{d12} \\ \vdots \\ X_{d1n} \end{pmatrix}, \qquad X_{d2} = \begin{pmatrix} X_{d21} \\ X_{d22} \\ \vdots \\ X_{d2n} \end{pmatrix},$$

$$Z_1 = \begin{pmatrix} 1_p & 0_p & \dots & 0_p \\ 0_p & 1_p & \dots & 0_p \\ \vdots & \vdots & \ddots & \vdots \\ 0_p & 0_p & \dots & 1_p \end{pmatrix} = I_n \otimes 1_p, \qquad Z_{d1} = \begin{pmatrix} X_{d11} & 0 & \dots & 0 & 0 \\ 0 & X_{d12} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & X_{d1(n-1)} & 0 \\ 0 & 0 & \dots & 0 & X_{d1n} \end{pmatrix},$$

$$Z_{d2} = \begin{pmatrix} X_{d21} & 0 & \dots & 0 & 0 \\ 0 & X_{d22} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & X_{d2(n-1)} & 0 \\ 0 & 0 & \dots & 0 & X_{d2n} \end{pmatrix}.$$

Here $X_{d1j}$ is the period-treatment incidence matrix for subject $j$ under design $d$, and $X_{d2j} = LX_{d1j}$, where L is the matrix defined as

$$L = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}.$$

Based on the assumptions, we obtain that $E(Y) = \mu 1_{pn} + X_1\pi + X_{d1}\tau + X_{d2}\gamma$ and

$$\text{var}(Y) = \begin{pmatrix} \Sigma_1 & 0 & \dots & 0 & 0 \\ 0 & \Sigma_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \Sigma_{n-1} & 0 \\ 0 & 0 & \dots & 0 & \Sigma_n \end{pmatrix},$$

where

$$\Sigma_j = \sigma_s^2 J_p + \sigma_{s\tau}^2 X_{d1j} X_{d1j}^T + \sigma_{s\gamma}^2 L X_{d1j} X_{d1j}^T L^T + \sigma_\epsilon^2 I_p$$
$$+\theta_{s\tau}(J_p - X_{d1j} X_{d1j}^T) + \theta_{s\gamma}(L J_p L^T - L X_{d1j} X_{d1j}^T L^T),$$
$$j = 1, 2, \ldots, n.$$

Since the model contains more than one random component, we will study the identifiability problem for those variance components. If any two different sets of values for the variance components result in two different variance-covariance matrices, then the variance components of the model are identifiable. If we consider the variance components of Model 3.1 as $\sigma_s^2$, $\sigma_{s\tau}^2$, $\sigma_{s\gamma}^2$, $\sigma_\epsilon^2$, $\theta_{s\tau}$ and $\theta_{s\gamma}$, then it is easy to see from the variance-covariance structure that the sum of the coefficients for $\sigma_{s\tau}^2$ and $\theta_{s\tau}$ is equal to the coefficient for $\sigma_s^2$ for any design. This implies that two different sets of values for the variance components could result in the same variance-covariance matrix, which indicates that the variance components in Model 3.1 are not identifiable for any design.

To alleviate this problem, we add constraints to the model:

$$Y_{ij} = \mu + \pi_i + \tau_{d(i,j)} + \gamma_{d(i-1,j)} + s_j + (s\tau)_{jd(i,j)} + (s\gamma)_{jd(i-1,j)} + \epsilon_{ij}, \tag{3.2}$$

where $\sum_{i=1}^p \pi_i = 0$, $\sum_{l=1}^t \tau_l = 0$, $\sum_{l=1}^t \gamma_l = 0$, $\sum_{l=1}^t (s\tau)_{jl} = 0 \quad \forall\, j$, $\sum_{l=1}^t (s\gamma)_{jl} = 0 \quad \forall\, j$.

As before, we assume that $\mathrm{Var}((s\tau)_{jl})$ is constant and that the covariance of any two of these terms is constant if they belong to the same subject. A similar assumption is used for the subject by carryover interaction. Random effects pertaining to different subjects are assumed to be uncorrelated. With these assumptions, we obtain:

$$\mathrm{cov}(s_j, s_{j'}) = \begin{cases} \sigma_s^2 & \text{if } j = j' \\ 0 & \text{otherwise} \end{cases},$$

$$\text{cov}((s\tau)_{jl},(s\tau)_{j'l'}) = \begin{cases} \sigma^2_{s\tau} & \text{if } j=j', l=l' \\ -\sigma^2_{s\tau}/(t-1) & \text{if } j=j', l\neq l', \\ 0 & \text{if } j\neq j' \end{cases}$$

$$\text{cov}((s\gamma)_{jl},(s\gamma)_{j'l'}) = \begin{cases} \sigma^2_{s\gamma} & \text{if } j=j', l=l' \\ -\sigma^2_{s\gamma}/(t-1) & \text{if } j=j', l\neq l'\cdot \\ 0 & \text{if } j\neq j' \end{cases}$$

Thus, the variance-covariance structure for the random effects for any subject is:

$$\text{var}\begin{pmatrix} s_j \\ (s\tau)_{j1} \\ \vdots \\ (s\tau)_{jt} \\ (s\gamma)_{j1} \\ \vdots \\ (s\gamma)_{jt} \end{pmatrix} = \begin{pmatrix} \sigma^2_s & 0 & \dots & \dots & 0 & 0 & \dots & \dots & 0 \\ & \sigma^2_{s\tau} & -\frac{\sigma^2_{s\tau}}{(t-1)} & \dots & -\frac{\sigma^2_{s\tau}}{(t-1)} & 0 & \dots & \dots & 0 \\ & & \ddots & \ddots & \vdots & \vdots & \ddots & & \vdots \\ & & & \sigma^2_{s\tau} & -\frac{\sigma^2_{s\tau}}{(t-1)} & \vdots & & \ddots & \vdots \\ & & & & \sigma^2_{s\tau} & 0 & \dots & \dots & 0 \\ & & & & & \sigma^2_{s\gamma} & -\frac{\sigma^2_{s\gamma}}{(t-1)} & \dots & -\frac{\sigma^2_{s\gamma}}{(t-1)} \\ & & & & & & \ddots & \ddots & \vdots \\ & & & & & & & \sigma^2_{s\gamma} & -\frac{\sigma^2_{s\gamma}}{(t-1)} \\ & & & & & & & & \sigma^2_{s\gamma} \end{pmatrix},$$

and the variance for $Y$ is:

$$\text{var}(Y) = \begin{pmatrix} \Sigma_1 & 0 & \dots & 0 & 0 \\ 0 & \Sigma_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \Sigma_{n-1} & 0 \\ 0 & 0 & \dots & 0 & \Sigma_n \end{pmatrix}, \tag{3.3}$$

where

$$\Sigma_j \;=\; \sigma^2_s J_p + \sigma^2_{s\tau}(\tfrac{t}{t-1}X_{d1j}X^T_{d1j} - \tfrac{1}{t-1}J_p) + \sigma^2_{s\gamma}(\tfrac{t}{t-1}LX_{d1j}X^T_{d1j}L^T - \tfrac{1}{t}LJ_pL^T) + \sigma^2_\epsilon I_p.$$

**Theorem 3.2.1** All the variance components in Model 3.2 are identifiable if and only if the design contains a sequence with repeated treatments.

**Proof**: Since the variance-covariance structure is a block diagonal matrix with each block corresponding to one subject, we focus on an arbitrary subject $j$. Consider two sets of variance components, $\sigma_s^2$, $\sigma_{s\tau}^2$, $\sigma_{s\gamma}^2$, $\sigma_\epsilon^2$ and $\tilde{\sigma}_s^2$, $\tilde{\sigma}_{s\tau}^2$, $\tilde{\sigma}_{s\gamma}^2$ and $\tilde{\sigma}_\epsilon^2$. If $\sigma_s^2 J_p + \sigma_{s\tau}^2(\frac{t}{t-1}X_{d1j}X_{d1j}^T - \frac{1}{t-1}J_p) + \sigma_{s\gamma}^2(\frac{t}{t-1}LX_{d1j}X_{d1j}^T L^T - \frac{1}{t}LJ_pL^T) + \sigma_\epsilon^2 I_p = \tilde{\sigma}_s^2 J_p + \tilde{\sigma}_{s\tau}^2(\frac{t}{t-1}X_{d1j}X_{d1j}^T - \frac{1}{t-1}J_p) + \tilde{\sigma}_{s\gamma}^2(\frac{t}{t-1}LX_{d1j}X_{d1j}^T L^T - \frac{1}{t}LJ_pL^T) + \tilde{\sigma}_\epsilon^2 I_p$ implies that $\sigma_s^2 = \tilde{\sigma}_s^2$, $\sigma_{s\tau}^2 = \tilde{\sigma}_{s\tau}^2$, $\sigma_{s\gamma}^2 = \tilde{\sigma}_{s\gamma}^2$ and $\sigma_\epsilon^2 = \tilde{\sigma}_\epsilon^2$, then all the variance components are identifiable, and vice verse.

Therefore, suppose that $\sigma_s^2 J_p+\sigma_{s\tau}^2(\frac{t}{t-1}X_{d1j}X_{d1j}^T-\frac{1}{t-1}J_p)+\sigma_{s\gamma}^2(\frac{t}{t-1}LX_{d1j}X_{d1j}^T L^T-\frac{1}{t}LJ_pL^T)+\sigma_\epsilon^2 I_p = \tilde{\sigma}_s^2 J_p+\tilde{\sigma}_{s\tau}^2(\frac{t}{t-1}X_{d1j}X_{d1j}^T-\frac{1}{t-1}J_p)+\tilde{\sigma}_{s\gamma}^2(\frac{t}{t-1}LX_{d1j}X_{d1j}^T L^T-\frac{1}{t}LJ_pL^T)+\tilde{\sigma}_\epsilon^2 I_p$, from the first and second diagonal element of the variance-covariance structure, we have $(a)$ $\sigma_s^2+\sigma_{s\tau}^2+\sigma_\epsilon^2 = \tilde{\sigma}_s^2+\tilde{\sigma}_{s\tau}^2+\tilde{\sigma}_\epsilon^2$ and $(b)$ $\sigma_s^2+\sigma_{s\tau}^2+\sigma_{s\gamma}^2+\sigma_\epsilon^2 = \tilde{\sigma}_s^2+\tilde{\sigma}_{s\tau}^2+\tilde{\sigma}_{s\gamma}^2+\tilde{\sigma}_\epsilon^2$. Subtracting $(a)$ from $(b)$, we obtain $\sigma_{s\gamma}^2 = \tilde{\sigma}_{s\gamma}^2$. Then we can ignore the terms $\sigma_{s\gamma}^2$ and $\tilde{\sigma}_{s\gamma}^2$ from now on. If the design contains a sequence with repeated treatments, from the off diagonal elements, once considering two observations for the same treatment and once two observations for distinct treatments, we have $(c)$ $\sigma_s^2 + \sigma_{s\tau}^2 = \tilde{\sigma}_s^2 + \tilde{\sigma}_{s\tau}^2$ and $(d)$ $\sigma_s^2 - \frac{1}{t-1}\sigma_{s\tau}^2 = \tilde{\sigma}_s^2 - \frac{1}{t-1}\tilde{\sigma}_{s\tau}^2$, respectively. Subtracting $(d)$ from $(c)$, we obtain $\sigma_{s\tau}^2 = \tilde{\sigma}_{s\tau}^2$, and substituting that into $(d)$, we also have $\sigma_s^2 = \tilde{\sigma}_s^2$. From $(a)$ we now also obtain $\sigma_\epsilon^2 = \tilde{\sigma}_\epsilon^2$.

Thus, we showed that all the variance components are identifiable if the design contains a sequence with repeated treatments. However, if there is no such sequence in the design, then we do not have $(c)$, and we cannot obtain $\sigma_s^2 = \tilde{\sigma}_s^2$, $\sigma_{s\tau}^2 = \tilde{\sigma}_{s\tau}^2$, and $\sigma_\epsilon^2 = \tilde{\sigma}_\epsilon^2$, so that not all of the variance components are identifiable in that case.

Therefore, we emphasize that investigators should know objectives of the experiment at the design stage. If one of the objectives of the crossover experiment is to

estimate and compare all the variance components, then investigators should use a crossover design with repeated treatments.

## 3.3 Some special cases of Model 3.2

Model 3.2 contains several other models used in the literature as special cases. For example, if $\sigma^2_{s\gamma} = 0$, then Model 3.2 reduces to:

$$Y_{ij} = \mu + \pi_i + \tau_{d(i,j)} + \gamma_{d(i-1,j)} + s_j + (s\tau)_{jd(i,j)} + \epsilon_{ij}, \tag{3.4}$$

which is the model with only an interaction between subjects and treatments. If, in addition, it is assumed that $\sigma^2_{s\tau} = 0$, then Model 3.2 becomes:

$$Y_{ij} = \mu + \pi_i + \tau_{d(i,j)} + \gamma_{d(i-1,j)} + s_j + \epsilon_{ij}. \tag{3.5}$$

For Model 3.5, we consider two extreme cases. If it is assumed that $\sigma^2_s = 0$, then the model corresponds to the situation of no subject effects. Conceptually, this case may be thought of as the subjects being carbon copies of each other (Hedayat, Stufken and Yang, 2006). The second case, $\sigma^2_s \to \infty$, is considered in the next result.

**Theorem 3.3.1** The traditional model, i.e., the model with fixed subject effects, is obtained from Model 3.5 for $\sigma^2_s \to \infty$.

**Proof**: Model 3.5 can be written in matrix notation as $Y = \mu 1_{pn} + X_1\pi + X_{d1}\tau + X_{d2}\gamma + Z_0 s + \epsilon$. Based on the model assumptions, we have $\text{var}(Y) = \sigma^2_\epsilon I_{np} + \sigma^2_s Z_0 Z_0' = V$. We rewrite the above model as $Y = D\eta + X_{d1}\tau + Z_0 s + \epsilon$, where $\eta$ combines the overall mean, and the period and carryover effects, while $D$ is the corresponding design matrix for those parameters. Then the information matrix $C_d$ for the direct

treatment effects $\tau$ under Model 3.5 can be expressed as

$$\begin{aligned} C_d &= (V^{-\frac{1}{2}}X_{d1})' pr^{\perp}(V^{-\frac{1}{2}}D)V^{-\frac{1}{2}}X_{d1} \\ &= X'_{d1}\{V^{-1} - V^{-1}D(D'V^{-1}D)^{-}D'V^{-1}\}X_{d1} \\ &= X'_{d1}\{(\sigma^2_\epsilon I_{np} + \sigma^2_s Z_0 Z'_0)^{-1} - (\sigma^2_\epsilon I_{np} + \sigma^2_s Z_0 Z'_0)^{-1} \\ &\quad D(D'(\sigma^2_\epsilon I_{np} + \sigma^2_s Z_0 Z'_0)^{-1}D)^{-}D'(\sigma^2_\epsilon I_{np} + \sigma^2_s Z_0 Z'_0)^{-1}\}X_{d1}, \end{aligned}$$

where $pr^{\perp}(X) = I - pr(X)$ and $pr(X) = X(X'X)^{-}X'$.

The information matrix for $\tau$ under the model with fixed subject effect is

$$\begin{aligned} C_{d_{\text{fix}}} &= X'_{d1} pr^{\perp}([D, Z_0])X_{d1} \\ &= X'_{d1} pr^{\perp}(D)X_{d1} - X'_{d1} pr^{\perp}(D)Z_0(Z'_0 pr^{\perp}(D)Z_0)^{-}Z'_0 pr^{\perp}(D)X_{d1}. \end{aligned}$$

Based on Proposition 1 of Jones, Kunert and Wynn (1992),
$(\sigma^2_\epsilon I_{np} + \sigma^2_s Z_0 Z'_0)^{-1} - (\sigma^2_\epsilon I_{np} + \sigma^2_s Z_0 Z'_0)^{-1}D(D'(\sigma^2_\epsilon I_{np} + \sigma^2_s Z_0 Z'_0)^{-1}D)^{-}D'(\sigma^2_\epsilon I_{np} + \sigma^2_s Z_0 Z'_0)^{-1} = pr^{\perp}(D) - pr^{\perp}(D)Z_0(\frac{1}{\sigma^2_s}I_n + Z'_0 pr^{\perp}(D)Z_0)^{-1}Z'_0 pr^{\perp}(D)$, which goes to $pr^{\perp}(D) - pr^{\perp}(D)Z_0(Z'_0 pr^{\perp}(D)Z_0)^{-}Z'_0 pr^{\perp}(D)$ if $\sigma^2_s \to \infty$. The result follows now immediately.

## 3.4 Estimation of the parameters

In this section, we discuss the estimation procedures for Model 3.2. Our interest lies in estimating the variance components and the fixed effects for the direct treatment and the carryover effects.

Since Model 3.2 is a linear mixed model, and the random terms are assumed to be normally distributed, the commonly used efficient estimation methods are the likelihood-based methods, such as, Maximum Likelihood (ML) and Restricted Maximum Likelihood (REML) estimation. The validation of the ML and REML estimation methods in the setting of a crossover design is discussed by Vonesh and Chinchilli (1997). We also consider these methods to estimate the parameters for Model 3.2.

To distinguish the subjects in different sequences, in this and the next section, instead of using $j$ as the only index for the subjects, we use both $j$ and $k$ to denote the $k^{th}$ subject in the $j^{th}$ sequence. So the outcome of the response variable for the $k^{th}$ subject in the $i^{th}$ period in the $j^{th}$ sequence is denoted by $Y_{ijk}$, where $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, s$, and $k = 1, 2, \ldots, n_j$. Here $s$ denotes the number of distinct sequences in the design, and $n_j$ is the number of subjects in the $j^{th}$ sequence.

Let $\mathbf{Y}_{jk} = [Y_{1jk}\ Y_{2jk}\ \ldots\ Y_{pjk}]'$ denote the p-vector of responses for the $k^{th}$ subject in the $j^{th}$ sequence. Based on the normality assumptions, we obtain that

$$\mathbf{Y}_{jk} \sim N_p(\mu_j, \Sigma_j), \quad j = 1, 2, \ldots, s; \quad k = 1, 2, \ldots, n_j,$$

where $\mu_j = (E(\bar{Y}_{1j.}), E(\bar{Y}_{2j.}), \ldots, E(\bar{Y}_{pj.}))'$ and $\Sigma_j$ is as in Equation 3.3.

Observations from different subjects are independent, so the joint density for $Y$ is

$$\prod_{j=1}^{s}\prod_{k=1}^{n_j} \frac{1}{(2\pi)^{(p/2)}|\Sigma_j|^{(1/2)}} e^{-\frac{1}{2}(\mathbf{Y}_{jk}-\mu_j)'\Sigma_j^{-1}(\mathbf{Y}_{jk}-\mu_j)},$$

and, thus, the log-likelihood function is

$$\sum_{j=1}^{s}\sum_{k=1}^{n_j}\left[-\frac{p}{2}\log(2\pi) - \frac{1}{2}\log(|\Sigma_j|) - \frac{1}{2}(\mathbf{Y}_{jk}-\mu_j)'\Sigma_j^{-1}(\mathbf{Y}_{jk}-\mu_j)\right]. \tag{3.6}$$

The ML estimates for the parameters can be obtained by maximizing 3.6. But the ML estimation takes no account for the degrees of freedom lost in estimating the fixed effects when estimating the variance components, and we will instead use Restricted Maximum Likelihood (REML) estimation. REML estimation was introduced by Patterson and Thompson (1971), and is also referred to as a modified maximum likelihood method. It improves ML estimators of the variance components by taking into account the loss of degrees of freedom associated with estimating the fixed effects parameters. Their ideas about using the error contrast instead of all the data had been adapted by Harville (1977). Following Harville (1977), the restricted

log-likelihood function can be written as

$$\sum_{j=1}^{s}\sum_{k=1}^{n_j}\left[-\frac{p}{2}\log(2\pi)-\frac{1}{2}\log(|\Sigma_j|)-\frac{1}{2}(\mathbf{Y}_{jk}-\mu_j)'\Sigma_j^{-1}(\mathbf{Y}_{jk}-\mu_j)\right] \\ -\frac{1}{2}\log(|\tilde{X}'\Sigma_j^{-1}\tilde{X}|),$$

where $\tilde{X}$ is any n × rank(X) matrix such that $C(\tilde{X}) = C(X)$ with $X$ as the design matrix $[1_{pn}|X_1|X_{d1}|X_{d2}]$ for the fixed effects and $C(X)$ as the column space of $X$.

Unfortunately, it is difficult to derive a closed form for the estimators of the variance components for our model. Thus, the Newton-Raphson iteration algorithm is applied to obtain the estimates.

## 3.5 Simulation study

In this section, we consider designs *AB/BA* and *ABB/BAA*. *AB/BA* is a balanced uniform design without repeated treatments; *ABB/BAA* is a strongly balanced design with repeated treatments, which is formed by repeating the treatments in the last period of the popular design *AB/BA*. For design *AB/BA*, if the carryover effects from the two treatments are not equal, then we cannot obtain unbiased estimators for the treatment contrast, so we consider the situation that the carryover effects from the two treatments are equal in the simulation study. In addition, as we proved in Section 3.2, since design *AB/BA* has no repeated treatment, there is an identifiability problem for the variance components. We will therefore focus on estimating functions of these variance components that are identifiable. Design *ABB/BAA* is advocated by several authors to ameliorate the deficiencies of design *AB/BA*.

The outcome of the response variable for the $k^{th}$ subject in the $i^{th}$ period for the $j^{th}$ sequence is denoted by $Y_{ijk}$, where $i = 1, 2$ for design *AB/BA* and $i = 1, 2, 3$ for

design ABB/BAA, $j = 1, 2$ and $k = 1, 2, \ldots, n_j$. The notation is illustrated in Table 3.1.

Table 3.1 Layout of Design *ABB/BAA*

| | Period 1 | Period 2 | Period 3 |
|---|---|---|---|
| | A | B | B |
| Sequence 1 | $Y_{111} \ldots Y_{11n_1}$ | $Y_{211} \ldots Y_{21n_1}$ | $Y_{311} \ldots Y_{31n_1}$ |
| | B | A | A |
| Sequence 2 | $Y_{121} \ldots Y_{12n_2}$ | $Y_{221} \ldots Y_{22n_2}$ | $Y_{321} \ldots Y_{32n_2}$ |

the layout of design *AB/BA* can be obtained from Table 3.1 by discarding Period 3.

A dual design refers to a design which contains a sequence and its dual, which is formed by interchanging the treatment labels, equally often. It is clear that both *AB/BA* and *ABB/BAA* are dual designs. According to assumptions of Model 3.2, the variance-covariance matrix for the response variable from both designs is a block diagonal matrix with the same block along the diagonal. The block under Model 3.2 with design *AB/BA* and *ABB/BAA* can be expressed as

$$\Sigma = \Sigma_{jk} = \begin{pmatrix} \sigma_s^2 + \sigma_{s\tau}^2 + \sigma_\epsilon^2 & \sigma_s^2 - \sigma_{s\tau}^2 \\ \sigma_s^2 - \sigma_{s\tau}^2 & \sigma_s^2 + \sigma_{s\tau}^2 + \sigma_{s\gamma}^2 + \sigma_\epsilon^2 \end{pmatrix}$$

and

$$\Sigma = \Sigma_{jk} = \begin{pmatrix} \sigma_s^2 + \sigma_{s\tau}^2 + \sigma_\epsilon^2 & \sigma_s^2 - \sigma_{s\tau}^2 & \sigma_s^2 - \sigma_{s\tau}^2 \\ \sigma_s^2 - \sigma_{s\tau}^2 & \sigma_s^2 + \sigma_{s\tau}^2 + \sigma_{s\gamma}^2 + \sigma_\epsilon^2 & \sigma_s^2 + \sigma_{s\tau}^2 - \sigma_{s\gamma}^2 \\ \sigma_s^2 - \sigma_{s\tau}^2 & \sigma_s^2 + \sigma_{s\tau}^2 - \sigma_{s\gamma}^2 & \sigma_s^2 + \sigma_{s\tau}^2 + \sigma_{s\gamma}^2 + \sigma_\epsilon^2 \end{pmatrix},$$

respectively.

Due to the identifiability problems for the estimation of variance components for design *AB/BA*, we estimate only the terms $\sigma_s^2+\sigma_{s\tau}^2+\sigma_\epsilon^2$, $\sigma_s^2-\sigma_{s\tau}^2$ and $\sigma_s^2+\sigma_{s\tau}^2+\sigma_{s\gamma}^2+\sigma_\epsilon^2$ for that design, which can be interpreted as the variance for the observation for a subject in the first period, the covariance between observations for a subject in the first and second period, and the variance for an observation for a subject in the

second period, respectively. For design *ABB/BAA*, we estimate $\sigma_s^2$, $\sigma_{s\tau}^2$, $\sigma_{s\gamma}^2$ and $\sigma_\epsilon^2$ individually.

We use the REML estimation method discussed in Section 3.4 to estimate those variance components, as well as the fixed effect contrasts, $\tau_A - \tau_B$ and $\gamma_A - \gamma_B$. Tables 3.2 - 3.4 present the results of the mean and variance of the estimates for 500 simulations for different numbers of subjects. The numbers in parentheses are the true values, and $n$ denotes the common number of subjects for each sequence.

Table 3.2 Simulation Results under Model 3.2 for Design *AB/BA*

| | $\sigma_s^2+\sigma_{s\tau}^2+\sigma_\epsilon^2$ (4.61) | | $\sigma_s^2-\sigma_{s\tau}^2$ (-0.23) | | $\sigma_s^2+\sigma_{s\tau}^2+\sigma_{s\gamma}^2+\sigma_\epsilon^2$ (6.3) | | $\tau_A-\tau_B$ (-1.0) | | $\gamma_A-\gamma_B$ (0.0) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | mean | var | mean | var | mean | var | mean | var | mean | var |
| n=20 | 4.4385 | 0.8715 | -0.2516 | 0.8098 | 6.5660 | 1.8362 | -1.0174 | 0.4256 | -0.0047 | 0.9851 |
| n=30 | 4.4742 | 0.5850 | -0.2369 | 0.4927 | 6.5183 | 1.1698 | -0.9981 | 0.3282 | 0.0055 | 0.6847 |
| n=50 | 4.5259 | 0.4110 | -0.2110 | 0.2992 | 6.3661 | 0.6796 | -1.0068 | 0.1787 | 0.0002 | 0.4430 |
| n=100 | 4.5887 | 0.2289 | -0.2168 | 0.1497 | 6.2997 | 0.3640 | -0.9909 | 0.0930 | -0.0092 | 0.2014 |

Table 3.3 Simulation Results for Random Effects for Design *ABB/BAA*

| | $\sigma_s^2$ (1.21) | | $\sigma_{s\tau}^2$ (1.44) | | $\sigma_{s\gamma}^2$ (1.69) | | $\sigma_\epsilon^2$ (1.96) | |
|---|---|---|---|---|---|---|---|---|
| | mean | var | mean | var | mean | var | mean | var |
| n=20 | 1.1480 | 0.2650 | 1.4243 | 0.3125 | 1.6743 | 0.4945 | 2.0328 | 0.9919 |
| n=30 | 1.1806 | 0.2091 | 1.4240 | 0.2325 | 1.6943 | 0.3653 | 1.9991 | 0.8041 |
| n=50 | 1.1909 | 0.1531 | 1.4367 | 0.1766 | 1.6836 | 0.2928 | 2.0109 | 0.7147 |
| n=100 | 1.2266 | 0.0873 | 1.4354 | 0.0949 | 1.6954 | 0.1685 | 1.9868 | 0.4835 |

Table 3.4 Simulation Results for Fixed Effects for Design *ABB/BAA*

| | $\tau_A - \tau_B$ (-1.0) | | $\gamma_A - \gamma_B$ (-0.5) | |
|---|---|---|---|---|
| | mean | var | mean | var |
| n=20 | -0.9999 | 0.2170 | -0.5003 | 0.2949 |
| n=30 | -1.0065 | 0.1452 | -0.4937 | 0.1776 |
| n=50 | -1.0021 | 0.0788 | -0.5035 | 0.1087 |
| n=100 | -0.9976 | 0.0404 | -0.4923 | 0.0538 |

From the results of Tables 3.2 - 3.4, the estimates of both the random and fixed terms for design *ABB/BAA* are fairly close to the true values, but the estimates of the functions of the variance components for design *AB/BA* is not very close to the true values. On average, the larger the sample size, the more precise the estimates.

To show that the model with subject by treatment and subject by carryover interactions is better than the model without those terms, we should test the hypothesis that $\sigma^2_{s\tau} = \sigma^2_{s\gamma} = 0$. However, this hypothesis is on the boundary of the parameter space and distribution under the null for the likelihood ratio test statistic is complicated. Therefore, it is difficult to use standard statistical methods to test whether one of the variance components is equal to 0. Instead, we use the model selection criterion *AIC* to compare Model 3.2 with Models 3.4 and 3.5 for design *ABB/BAA*. The results are presented in Table 3.5.

Table 3.5 Model Selection Results under Design *ABB/BAA*

| $\sigma_s$ | $\sigma_{s\tau}$ | $\sigma_{s\gamma}$ | $\sigma_\epsilon$ | $AIC_{diff1} < 0$ | $AIC_{diff2} < 0$ |
|---|---|---|---|---|---|
| 1.1 | 0.8 | 0.9 | 1.4 | 129/500 | 100/500 |
| 1.1 | 1.2 | 1.3 | 1.4 | 226/500 | 245/500 |
| 1.1 | 2.2 | 2.3 | 1.4 | 420/500 | 475/500 |

Note, $AIC_{diff1}$ is the difference in *AIC* values between Model 3.2 and Model 3.4, and $AIC_{diff2}$ is the difference in *AIC* values between Model 3.2 and Model 3.5.

The number below "/" is the total number of tries, and the number above "/" is the number of times that the difference of $AIC$ is less than zero, which indicates Model 3.2 is better than Model 3.4 or Model 3.5.

From the results of Table 3.5, we can see that when the variances of the interaction terms are smaller, there is no evidence that Model 3.2 is better based on the $AIC$ criterion. However, when the variance of the interaction terms become larger, then there is evidence to show that the models with the interaction terms perform better. These results also tell us that in order for the model with the interaction terms to be selected as the best model by $AIC$ criterion, the variance of the interaction terms should be relatively large.

## 3.6 Numerical example

The data set in this example consists of systolic blood pressure measurements from a trial on hypertension using a two-treatment three-period crossover design. In the original study, subjects were randomly assigned to the four sequence groups $ABB/BAA/ABA/BAB$. Each treatment period lasted for six weeks, and there were no wash-out periods for ethical reasons. These data have been used by different authors as an example for the design $ABB/BAA$ by only considering the observations from the first two sequences (Ebbutt, 1984; Matthews, 1989; Jones and Kenward, 2003). The complete data set can be found in Jones and Kenward (2003, pages 232-233).

We also only use the observations from the first two sequences of this data set, and the number of subjects in each sequence is 22. We obtain the following results by fitting the data to Models 3.2, 3.4 and 3.5:

Table 3.6 Summary of the Results for A Numerical Example

| | Estimates of Variance Components | | | | Model Selection Criteria | |
|---|---|---|---|---|---|---|
| | $\hat{\sigma}_s^2$ | $\hat{\sigma}_{s\tau}^2$ | $\hat{\sigma}_{s\gamma}^2$ | $\hat{\sigma}_\epsilon^2$ | -2LogL | $AIC$ |
| Model 3.2 | 185.5166 | 35.6118 | 41.8076 | 88.1132 | 858.5978 | 876.5978 |
| Model 3.4 | 162.0344 | 7.7089 | - | 167.4737 | 858.9144 | 874.9143 |
| Model 3.5 | 160.7905 | - | - | 177.1573 | 859.0850 | 873.0850 |

From the results in Table 3.6, we can see that Model 3.2 has a slightly smaller $-2LogL$ value than Models 3.4 and 3.5, but that all are very close. Since Model 3.2 has one more parameter than Model 3.4. and two more parameters than Model 3.5, after taking account of the parameter penalty, Model 3.2 has a larger $AIC$ value. But that does not necessarily mean that Model 3.2 is not appropriate, since the estimates of $\sigma_{s\tau}^2$ and $\sigma_{s\gamma}^2$ are not even close to zero. Also, notice that the sum of $\hat{\sigma}_{s\tau}^2$, $\hat{\sigma}_{s\gamma}^2$ and $\hat{\sigma}_\epsilon^2$ for Model 3.2 is close to the the sum of $\hat{\sigma}_{s\tau}^2$ and $\hat{\sigma}_\epsilon^2$ of Model 3.4 and $\hat{\sigma}_\epsilon^2$ for Model 3.5, but $\hat{\sigma}_{s\tau}^2$ and $\hat{\sigma}_{s\gamma}^2$ are relatively small compared to $\hat{\sigma}_s^2$ and $\hat{\sigma}_\epsilon^2$. Because the variances of the interaction terms are relatively small, it is not surprising that we cannot select Model 3.2 is the best one with the model selection criteria, a result which is supported by our simulation results.

## 3.7 Discussion

We introduced a new model which incorporates the interactions of subject by treatment and subject by carryover to capture the variabilities of direct treatment and carryover effects on different subjects. We showed that the new model can be generalized to different models including the traditional model, and all of the variance components for that model are identifiable if there are repeated treatments in the design. We use model selection criteria $AIC$ to compare the model with and without the random interaction terms, the reason we do not use $BIC$ is that $BIC$ tends to

select too simple or parsimonious model for the data (Fitzmaurice, Laird and Ware, 2004). "In general, we do not recommend the use of *BIC* for covariance model selection as it entails a high risk of selecting a model that is too simple or parsimonious for the data at hand." However, if based only on the model selection criteria, the model with interactions can not demonstrate that it is superior to the model without interactions, if the variability due to the interaction is not significantly great.

## 3.8 References


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

2. Bose, M. and Mukherjee, B. (2000). Cross-over designs in the presence of higher order carry-overs. *Australian and New Zealand Journal of Statistics* 42, 235-244.

3. Bose, M. and Mukherjee, B. (2003). Optimal cross-over designs under a general model. *Statistics and Probability Letters* 62, 413-418.

4. Cheng, C. S. and Wu, C. F. (1980) Balanced repeated measurements designs. *The Annals of Statistics* 8, 1272-1283.

5. Chinchilli, V. M. and Esinhart, J. D. (1996) Designs and analysis of intra-subject variability in cross-over experiments. *Statistics in Medicine* 15, 1619-1634.

6. Ebbutt, A. F. (1984). Three-period crossover designs for two treatments. *Biometrics* 40, 219-224.

7. Fitzmaurice, G. M., Larid, N.M. and Ware, J. H. (2004) *Applied Longitudinal Analysis*. Wiley.

8. Fleiss, J. L. (1989) A critique of recent research on the two-treatment crossover design. *Controlled Clinical Trials* 10, 237-243.

9. Ghosh, S. and Crosby, H. R. (2005) Subject-treatment interactions in crossover trials: performance evaluation of subgroup methods. *Journal of Statistical Planning and Inference* 132, 63-73.

10. Ghosh, S. and Fairchild, L. D. (2000) Inference on interactions between treatments and subjects within groups in a two-period crossover trial. *Journal of Statistical Planning and Inference* 88, 301-317.

11. Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 72, 320-338.

12. Hedayat, A. and Afsarinejad, K. (1978). Repeated measurements designs, II. *The Annals of Statistics* 6, 619-628.

13. Hedayat, A, Stufken, J. and Yang, M. (2006). Optimal and efficient crossover designs when subject effects are random. *Journal of the American Statistical Association* 101, 1031-1038.

14. Jones, B. and Kenward, M. J. (2003) *Design and Analysis of Cross-over Trials*. Second Edition, CRC Press.

15. Jones, B., Kunert, J. and Wynn, H. P. (1992). Information matrices for mixed effects models with applications to the optimality of repeated measurements designs. *Journal of Statistical Planning and Inference* 33, 261-274.

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

17. Kershner, R. P., Federer, W. T. (1981) Two-treatment crossover designs for estimating a variety of effects. *Journal of American Statistical Association* 76, 612-619.

18. Magda, C. G. (1980). Circular balanced repeated measurements designs. *Communications in Statistics* A9, 1901-1918.

19. Matthews, J. N. S (1988). Recent developments in crossover designs. *International Statistical Review* 56, 117-127.

20. Matthews, J. N. S (1989). Estimating dispersion parameters in the analysis of data from crossover trials. *Biometrika* 76, 239-244.

21. Matthews, J. N. S (1994). Modeling and optimality in the design of crossover studies for medical applications. *Journal of Statistical Planning and Inference* 42, 89-108.

22. Patterson, H. D. and Thompson, R (1971) Recovery of interblock information when block sizes are unequal. *Biometrika* 58, 545C554.

23. Senn, S. (2002). *Crossover Trials in Clinical Research.* Second Edition, U.K. Wiley.

24. Vonesh, E. F. and Chinchilli, V. M. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*, New York.

# Chapter 4

# Use of Baseline Measurements in the $2 \times 2$ Crossover Trial for the Case of Single Measurements

## 4.1 Introduction

The $2 \times 2$ crossover design (two treatments in two periods, using the sequences $AB$ and $BA$) is the simplest crossover design and is extremely popular in clinical trials (Grizzle, 1965; Brown, 1980; Armitage and Hill, 1982). However, it is also regarded by many statisticians as particularly problematic because of the carryover effects and aliasing of several effects (Senn, 1994; Jones and Kenward, 2003). Using baseline measurements at the beginning of each period is introduced as a technique to rescue the $AB/BA$ design from its deficiencies and to provide additional information to eliminate nuisance effects from the treatment effects (Patel, 1983; Kenward and Jones, 1987).

However, contrary to some authors' intuition and initial purposes, using baseline measurements at the beginning of each period to eliminate the carryover effects receives serious criticism in many papers (Fleiss, Wallenstein and Rosenfeld, 1985; Willan and Pater, 1986; Fleiss,1989). Of course, the conclusions depend on different models and assumptions considered. For example, Willan and Pater (1986) considered a model with fixed period, treatment and carryover effects, and random subject and subject by period interaction effects as well as random error. Assuming that the carryover effects in the wash-out period and the second treatment period are the same, they found that analyzing change from baselines can eliminate the carryover

effects but at the cost of reducing the power and precision for inferences about the treatment contrast under many situations. Assuming that there is a carryover effect in the wash-out period but not in the second treatment period and that carryover effect is proportional to the direct treatment effect, Fleiss, Wallenstein and Rosenfeld (1985) showed that analysis of change from baselines created a carryover effect where none would have existed if only observations from treatment period (we use term outcomes later) had been used. In addition, the contrast of carryover effects is opposite in sign to the treatment contrast.

According to Ratkowsky *et al.* (1993) and Kenward and Jones (2003), the measurements taken at the beginning of the second treatment period as discussed in the previous paragraph are not true baselines, since they are affected by the treatments in the previous period. Kenward and Jones (1987) also argued that there is no satisfactory statistical analysis for the $AB/BA$ crossover trial if carryover effects from treatments $A$ and $B$ are different. So, in the pharmaceutical industry, an adequate wash-out period is usually conducted between two active periods to eliminate any carryover effects, but investigators often also collect measurements before the active treatment periods. In that situation, measurements taken at the beginning of both periods can be viewed as baselines (Senn, 2002). How to incorporate those baseline measurements in the analysis is the problem we intend to investigate in this chapter.

In a clinical trial in the pharmaceutical industry, data can be collected at a single time or at different time points (repeated measurements, which will be discussed in Chapters 5 and 6) before the treatments are assigned (off-drug day) and during the treatment periods (on-drug day). The data from the off-drug day can be viewed as baseline measurements, while the data from the on-drug day are outcomes of the response variable. We assume that the wash-out periods are sufficiently long so that there is no carryover effect and the baseline measurements are not affected by the treatments.

Unlike in parallel design, where it is recommended to collect baseline measurements and to use them as a covariate, there is no general consensus on how to properly handle baseline measurements in crossover trials. Different models and assumptions can lead to different conclusions. In this chapter, we will investigate various potential methods to handle baseline measurements for the case of measurements at a single time point in each period, and compare the variances for the estimators of the treatment contrast under various scenarios by using the relative efficiency (RE). The RE for a method is defined as the ratio of the minimum standard error (SE) of the estimators of the treatment contrast from all considered methods and the SE for the particular method. Since the primary goal of an experiment using a crossover design is to compare the treatment effects, we will also discuss testing the hypothesis of equal treatment effects.

This chapter is organized as follows. In Section 4.2, potential methods and models to handle baseline measurements in crossover studies are discussed, and analytical expressions of variances of the estimators of the treatment contrast from different methods are derived and compared. In Section 4.3, a thorough simulation study is conducted for the *AB/BA* crossover design to evaluate the performance of different methods under different scenarios. A real data example is analyzed for illustration purposes in Section 4.4. We conclude this chapter with a discussion, conclusion and recommendation in Section 4.5.

We focus on the $2 \times 2$ crossover design in Chapters 4 and 5, and will extend our study to more general designs in Chapter 6.

## 4.2 Methods and models

Suppose that we have a single observation for each subject from both the off-drug day and the on-drug day in both periods in a *AB/BA* crossover trial. Denote these

by $X_{ijk}$ and $Y_{ijk}$ respectively, where $i = 1, 2$, $j = 1, 2$ and $k = 1, 2, \ldots, n_j$ are the indices for period, sequence and subject within the $j^{th}$ sequence, respectively. The data layout is presented in Table 4.1.

Table 4.1 A Two-period Crossover Trial with Baseline Measurements

| | | Sequence AB | | | Sequence BA | | |
|---|---|---|---|---|---|---|---|
| Period 1 | off-drug day | $X_{111}$ | ... | $X_{11n_1}$ | $X_{121}$ | ... | $X_{12n_2}$ |
| | on-drug day | $Y_{111}$ | ... | $Y_{11n_1}$ | $Y_{121}$ | ... | $Y_{12n_2}$ |
| Period 2 | off-drug day | $X_{211}$ | ... | $X_{21n_1}$ | $X_{221}$ | ... | $X_{22n_2}$ |
| | on-drug day | $Y_{211}$ | ... | $Y_{21n_1}$ | $Y_{221}$ | ... | $Y_{22n_2}$ |

From Table 4.1, each subject has four observations, which are from four different days (two from each period). Part of the variation can be attributable to differences in subjects (between subjects), while other parts can result from differences between periods and days within periods. In order to capture all of these sources of variability, we include random subject and subject by period interaction effects as well as random error in the model. As explained in Section 4.1, we assume that there is no carryover effect, but include other commonly considered fixed effects in crossover trials, such as an overall mean, period effects and treatment effects, as well as day effects. A similar model, which includes the carryover effects as well, was considered by Willan and Pater (1986).

Following the notations from the traditional model, the model for the observations can be written as

$$Y_{hijk} = \mu + \pi_i + D_{hi} + h\tau_{t(i,j)} + s_{jk} + \zeta_{ijk} + \epsilon_{hijk}, \tag{4.1}$$
$$i = 1, 2,\ j = 1, 2,\ k = 1, 2, \ldots, n_j,\ h = 0, 1,$$

where $Y_{hijk}$ corresponds to $X_{ijk}$ when $h = 0$ and to $Y_{ijk}$ when $h = 1$. The terms $\mu$, $\pi_i$, $D_{hi}$ and $\tau_{t(i,j)}$ represent an overall mean, the $i^{th}$ period effect, the $h^{th}$ day effect

in the $i^{th}$ period and the direct treatment effect due to the treatment assigned to the $i^{th}$ period in the $j^{th}$ sequence, respectively. The remaining three terms $s_{jk}$, $\zeta_{ijk}$ and $\epsilon_{ijkh}$ are random subject effects, subject by period interaction and random error, and they are assumed to be mutually independently distributed as $s_{jk} \sim N(0, \sigma_s^2)$, $\zeta_{ijk} \sim N(0, \sigma_{sp}^2)$ and $\epsilon_{hijk} \sim N(0, \sigma_\epsilon^2)$, respectively.

From the model assumptions, we obtain

$$\text{Corr}(X_{ijk}, Y_{i'jk}) = \begin{cases} (\sigma_s^2 + \sigma_{sp}^2)/(\sigma_s^2 + \sigma_{sp}^2 + \sigma_\epsilon^2) = \rho_{w/p} & \text{if } i = i', \\ \sigma_s^2/(\sigma_s^2 + \sigma_{sp}^2 + \sigma_\epsilon^2) = \rho_{b/p} & \text{if } i \neq i', \end{cases}$$

where $\rho_{w/p}$ is a within period correlation, while $\rho_{b/p}$ is a between period correlation. Notice that $\sigma_{sp}^2 > 0$ indicates $\rho_{w/p}$ is greater than $\rho_{b/p}$ , which implies that the correlation between the baseline measurements and the outcomes is greater when they are from the same period. From the assumptions, the correlation between the first baseline measurements and the second outcomes is the same as the correlation between the first outcomes and the second baseline measurements, which is not unreasonable if the wash-out periods are relatively long compared to the treatment periods.

For a given subject, the within period correlation can be denoted as $\rho = \text{Corr}(X_{ijk}, Y_{ijk}|s_{jk}) = \sigma_{sp}^2/(\sigma_{sp}^2 + \sigma_\epsilon^2)$, and is referred to as partial correlation (Senn, 2002; Jones and Kenward, 2003). Senn (2002) suggested that methods of incorporating baseline measurements can be chosen based on the value of $\rho$. We consider Model 4.1 to retain the baseline measurements as part of the response vector, and compare it with the other three commonly used methods at different values for $\rho$. Therefore, we consider four ways for handling the baseline measurements:

1) Retain the baseline measurements as part of the response vector as described in Model 4.1.

2) Ignore the baseline measurements, and use the observations for outcomes only:

$$Y_{ijk} = \mu + \pi_i + \tau_{t(i,j)} + s_{jk} + \epsilon_{ijk}. \tag{4.2}$$

3) Use change from baseline measurements as the response variable:

$$Y_{ijk} - X_{ijk} = \mu + \pi_i + \tau_{t(i,j)} + \epsilon_{ijk}. \tag{4.3}$$

4) Use baseline measurements as a covariate and model the outcomes conditional on the covariate:

$$Y_{ijk} = \mu + \pi_i + \tau_{t(i,j)} + \beta x_{ijk} + s_{jk} + \epsilon_{ijk}. \tag{4.4}$$

The assumptions for the random effects in Models 4.2-4.4 are identical to those for Model 4.1. It is clear that Models 4.2 and 4.3 are consistent with the assumptions for baseline measurements and outcomes postulated by Model 4.1. However, this is less obvious for Model 4.4. It is however easy to show that $Y_{ijk}$ and $X_{ijk}$ in Model 4.1 follow a bivariate normal distribution, and the form of Model 4.4 emerges by considering the conditional distribution of $Y_{ijk}$ given $X_{ijk} = x_{ijk}$, with the value of $\beta$ being the ratio of the covariance between $Y_{ijk}$ and $X_{ijk}$ and the variance of $X_{ijk}$.

Assuming that the variance components are known, we can obtain the Best Linear Unbiased Estimators (BLUE) of the treatment contrast and the corresponding variances for Models 4.1 - 4.4. We summarize the results in Table 4.2.

Table 4.2 BLUEs and their Variances for Estimating the Treatment Contrast

| Model | Estimator | Variance |
|---|---|---|
| 4.1 | $\frac{1}{2}[\bar{Y}_{11.} + \bar{Y}_{22.} - \bar{Y}_{12.} - \bar{Y}_{21.} - \rho(\bar{X}_{11.} + \bar{X}_{22.} - \bar{X}_{12.} - \bar{X}_{21.})]$ | $\frac{n_1+n_2}{2n_1n_2}(1+\rho)\sigma_\epsilon^2$ |
| 4.2 | $\frac{1}{2}(\bar{Y}_{11.} + \bar{Y}_{22.} - \bar{Y}_{12.} - \bar{Y}_{21.})$ | $\frac{n_1+n_2}{2n_1n_2}\frac{1}{1-\rho}\sigma_\epsilon^2$ |
| 4.3 | $\frac{1}{2}[\bar{Y}_{11.} + \bar{Y}_{22.} - \bar{Y}_{12.} - \bar{Y}_{21.} - (\bar{X}_{11.} + \bar{X}_{22.} - \bar{X}_{12.} - \bar{X}_{21.})]$ | $\frac{n_1+n_2}{2n_1n_2}2\sigma_\epsilon^2$ |
| 4.4 | $\frac{1}{2}[\bar{Y}_{11.} + \bar{Y}_{22.} - \bar{Y}_{12.} - \bar{Y}_{21.} - \beta(\bar{x}_{11.} + \bar{x}_{22.} - \bar{x}_{12.} - \bar{x}_{21.})]$ | $\frac{n_1+n_2}{2n_1n_2}(1+a)\sigma_\epsilon^2$ |
| | $\rho = \frac{\sigma_{sp}^2}{\sigma_{sp}^2+\sigma_\epsilon^2}, \quad \beta = \frac{\sigma_s^2+\sigma_{sp}^2}{\sigma_s^2+\sigma_{sp}^2+\sigma_\epsilon^2}, \quad a = \frac{(\sigma_s^2+\sigma_{sp}^2)^2+\sigma_{sp}^2\sigma_\epsilon^2}{(\sigma_s^2+\sigma_{sp}^2+\sigma_\epsilon^2)^2}$ | |

It can be shown that $\rho \leq a \leq 1$, and the minimum and maximum value of $a$ can be obtained when $\sigma_s^2 = 0$ and $\sigma_\epsilon^2 = 0$, respectively. So the differences between the variances for the estimators of the treatment contrast from different models can be roughly expressed as a function of the partial correlation $\rho$.

From Table 4.2, we can see that only the variance from the method of using baseline measurements as a covariate (Model 4.4) depends on the subject variability $\sigma_s^2$, and the value is between the variances for the method of retaining the baseline measurements as part of the response vector (Model 4.1) and the method of analyzing change from baselines (Model 4.3). Furthermore, the variance for the method of retaining the baseline measurements as part of the response vector is the smallest as long as $\rho$ is not equal to zero. If $\rho = 0$, then this variance is the same as that from the method of ignoring baseline measurements (Model 4.2), which is only 50% of that for the method of analyzing change from baselines. When $\rho$ is equal to 0.5, the variance for the method of ignoring baseline measurements is the same as that for the method of analyzing change from baselines, but is almost 25% greater than that for the methods of retaining the baseline measurements as part of the response vector and using baseline measurements as a covariate. When $\rho$ is around 1, the variances for the three methods of using baseline measurements are close to each other, but ignoring baseline measurements could in that case result in a very large variance.

## 4.3 Simulation studies

In the last section, the theoretical derivations and comparisons of the variances for the different methods under different scenarios for the variance components were made under the assumption that the variance components are known. However,

in practice, variance components are typically unknown and need to be estimated from the data. Thus, in this section, we conduct a simulation study for the *AB/BA* crossover design to evaluate the performance of the discussed methods.

As discussed in Section 4.2, we obtain the measurements from both the off-drug day and the on-drug day within each period, and we analyze the data by the various methods discussed. The primary purpose of the simulation study is to compare the relative efficiencies (RE) of the methods for estimating of the treatment contrast at different values of the partial correlation $\rho$. Since the main purpose of an experiment which uses a crossover design is to compare the treatment effects, we may also be interested in testing the hypothesis of equal treatment effects. Thus, we will compute the type I error rate for testing $\tau_A = \tau_B$ in the simulation study.

In the simulation study, we generate the data based on Model 4.1. We set $\tau_A - \tau_B = 0$ and $\tau_A - \tau_B = 0.5$ for the evaluation of type I error rate and power, respectively, for testing $H0 : \tau_A = \tau_B$. We set $\mu = 100$ and set all other fixed effects to be zero. The relative efficiencies depend on the relative value of $\sigma_s^2$, $\sigma_{sp}^2$ to $\sigma_\epsilon^2$, and we can set $\sigma_\epsilon^2 = 1$. We let the subject variance $\sigma_s^2$ take the values 0.2, 0.5, 1, 2, 5 and 10, which then correspond to the ratio of $\sigma_s^2$ and $\sigma_\epsilon^2$. For $\sigma_{sp}^2$, we select each time a different value so that the partial correlations $\rho$ between the data from the off-drug day and on-drug day are 0, 0.2, 0.4, 0.5, 0.6, 0.8. We consider 6, 12 and 24 subjects in each sequence, which are common numbers in phase I clinical trials. Each scenario is simulated 5000 times.

We analyze the simulated data by using PROC MIXED in SAS with "KR" degree of freedom adjustment for all the methods mentioned in Section 4.2. We estimate the treatment contrast $\tau_A - \tau_B$, compute the standard error (SE) for the estimators of the treatment contrasts, and convert SE to relative efficiency (RE). We also investigate the type I error rate and power for testing the hypothesis of equal treatment effects.

The results of REs for different methods does not depend on the value of $\tau_A - \tau_B$, therefore, we do not present the results from the case of $\tau_A - \tau_B$ here. Table 4.3 presents the results for 24 subjects per sequence when $\sigma_s^2/\sigma_\epsilon^2 = 1$. In Table 4.3, $\hat{\tau}_A - \hat{\tau}_B$ is the mean of the treatment contrast estimates, SE_est is the mean of the estimated standard errors for $\hat{\tau}_A - \hat{\tau}_B$, and the numbers in parentheses under column called SE_true are theoretical results of the standard errors calculated from Table 4.2 based on the values of the variance components used in the simulation. With SE_MC as the Monte Carlo standard deviation, i.e. the standard deviation of the 5000 estimates of $\tau_A - \tau_B$, SE_rediff is the relative difference between SE_est and SE_MC and is computed as $\frac{\text{SE\_est - SE\_MC}}{\text{SE\_MC}} \times 100\%$. RE is the relative efficiency, which is the ratio of the minimum SE_est of all methods for that scenario and SE_est for that particular method. Type I error rate is calculated by recording the percentage of times out of 5000 that the test of no treatment difference is rejected at the level of 0.05.

Table 4.3 Results for 24 Subjects when $\sigma_s^2/\sigma_\epsilon^2 = 1$

| $\rho$ | Methods | $\hat{\tau}_A - \hat{\tau}_B$ | SE_est(SE_true) | SE_rediff | RE | Type I error |
|---|---|---|---|---|---|---|
| 0.0 | Baselines in response vector | -0.0014 | 0.2064(0.2041) | 1.10 | 98.27 | 4.96 |
| | Ignore baselines | -0.0015 | 0.2028(0.2041) | -0.29 | 100 | 4.72 |
| | Change from baselines | -0.0025 | 0.2880(0.2887) | -0.01 | 70.44 | 5.24 |
| | Baselines as a covariate | -0.0022 | 0.2214(0.2282) | -0.99 | 91.59 | 5.06 |
| 0.2 | Baselines in response vector | 0.0032 | 0.2250(0.2236) | -0.94 | 100 | 5.10 |
| | Ignore baselines | 0.0039 | 0.2273(0.2282) | -1.44 | 98.99 | 5.24 |
| | Change from baselines | 0.0025 | 0.2879(0.2887) | -0.92 | 78.15 | 5.06 |
| | Baselines as a covariate | 0.0030 | 0.2334(0.2379) | -1.88 | 96.37 | 5.12 |
| 0.4 | Baselines in response vector | 0.0042 | 0.2430(0.2415) | -1.20 | 100 | 5.46 |
| | Ignore baselines | 0.0040 | 0.2625(0.2635) | -1.39 | 92.59 | 5.86 |
| | Change from baselines | 0.0052 | 0.2884(0.2887) | -0.41 | 84.26 | 5.38 |
| | Baselines as a covariate | 0.0050 | 0.2447(0.2487) | -2.55 | 99.30 | 5.78 |
| 0.5 | Baselines in response vector | 0.0015 | 0.2509(0.2500) | 1.09 | 99.57 | 4.98 |
| | Ignore baselines | -0.0019 | 0.2868(0.2887) | 1.22 | 87.1 | 4.56 |
| | Change from baselines | 0.0047 | 0.2880(0.2887) | 1.84 | 86.74 | 4.64 |
| | Baselines as a covariate | 0.0026 | 0.2498(0.2546) | -0.56 | 100 | 4.52 |
| 0.6 | Baselines in response vector | -0.0006 | 0.2590(0.2582) | -0.31 | 98.66 | 5.02 |
| | Ignore baselines | -0.0041 | 0.3204(0.3227) | -0.46 | 79.75 | 4.94 |
| | Change from baselines | 0.0009 | 0.2879(0.2887) | -0.92 | 88.75 | 5.00 |
| | Baselines as a covariate | 0.0002 | 0.2555(0.2608) | -2.61 | 100 | 5.28 |
| 0.8 | Baselines in response vector | -0.0037 | 0.2742(0.2739) | 1.42 | 97.59 | 4.88 |
| | Ignore baselines | -0.0054 | 0.4520(0.4564) | -2.34 | 59.19 | 5.62 |
| | Change from baselines | -0.0033 | 0.2881(0.2887) | 1.41 | 92.86 | 4.56 |
| | Baselines as a covariate | -0.0041 | 0.2676(0.2743) | -1.38 | 100 | 5.24 |

From Table 4.3, we can see that the estimates of the treatment contrast are close to the true values, and the estimates of the standard errors are close to the theoretical values under all scenarios. The small value of SE_rediff indicates that SE_est is close to SE_MC. However, the SE_rediff when using the baseline measurements as a covariate is relatively larger than that for other methods. In addition, SE_rediff for smaller sample sizes is relatively larger (results are not shown here). Furthermore, we can see that the type I error rate maintains the nominal level of 5%.

We present the pattern of relative efficiencies (RE) at different values of the partial correlation $\rho$ for different numbers of subjects when $\sigma_s^2/\sigma_\epsilon^2 = 1$ in Figure 4.1, where REs in the top, middle and lower panel correspond to the REs based on SE_est, SE_true and SE_MC, respectively.

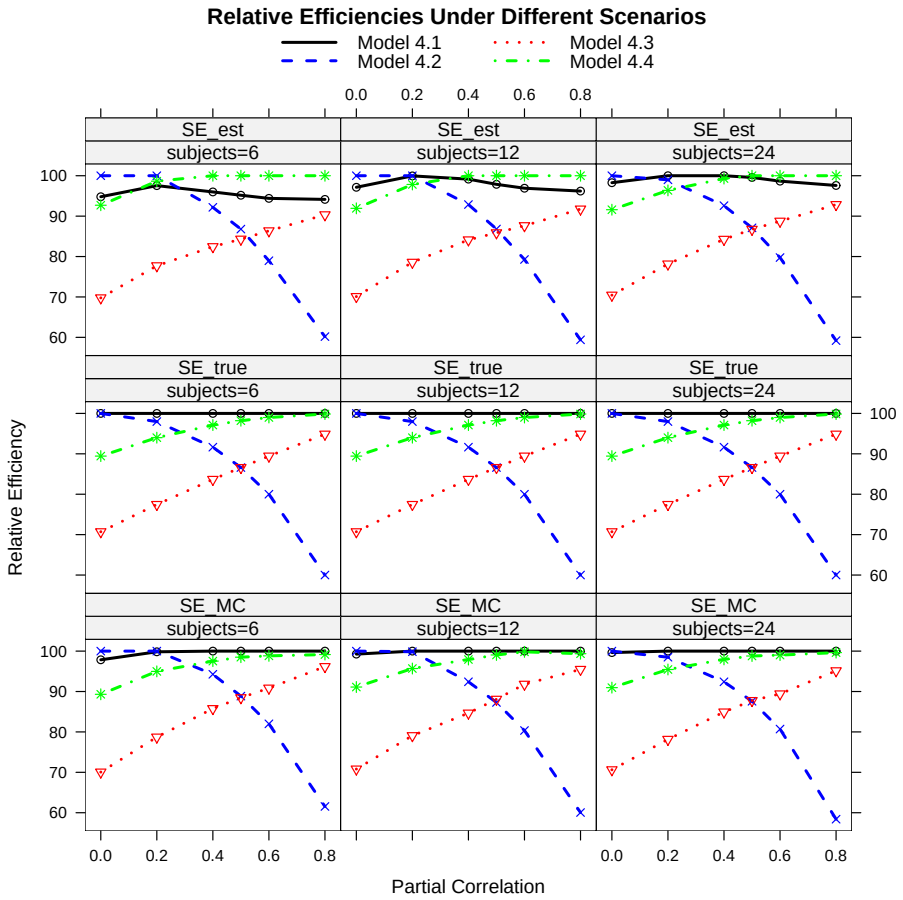


Figure 4.1: Relative Efficiencies for Each Method at Different Values of the Partial Correlation for Different Sample Sizes when $\sigma_s^2/\sigma_\epsilon^2 = 1$ (REs in the top, middle and lower panel correspond to REs based on SE_est, SE_true and SE_MC, respectively)

For the middle panels (based on the theoretical SE), the REs do not depend on the numbers of subjects, as is also immediately clear from Table 4.2. The pattern for RE based on SE_MC from the simulation study is closer to the theoretical result than that based on SE_est. This is especially clear at the smaller sample size. This reflects that it is more difficult to obtain precise estimates for the variance components at smaller sample sizes. The larger the sample size, the smaller the values for SE_rediff, and the closer SE_est is to the theoretical results.

If we look at the RE for one particular sample size in Figure 4.1 (say 24 subjects per sequence), then the results confirm the theoretical results we derived in Section

4.2. For example, we can see that retaining baselines as part of the response vector has 100% relative efficiency (RE) almost all the time, the RE for baselines as a covariate and change from baselines increases as the partial correlation $\rho$ increases, and the RE for ignoring baseline measurements decreases. The RE for baselines as a covariate is greater than that for change from baselines for all values of $\rho$.

When the partial correlation $\rho$ is around zero, which indicates that $\sigma^2_{sp}$ is small compared to $\sigma^2_\epsilon$, then the observations on the same subject in the same period are no more correlated than the observations on the same subject in different periods, and ignoring baseline measurements has the largest RE, i.e. the smallest standard error. Retaining baseline measurements in the response vector has a similar RE, while incorporating baseline measurements as a covariate results in almost 10% reduction of RE and change from baselines reduces the RE dramatically. The RE for ignoring baseline measurements is close to that of using change from baselines as the response variable when $\rho$ is around 0.5, but is smaller than that of using baseline measurements as a covariate and retaining baseline measurements as part of the response vector. When $\rho$ is greater than 0.5, all of the methods, which include baseline measurements in the analysis, has a higher RE than ignoring baselines. This is especially true when $\rho$ is close to 1, where using baseline measurements has a significantly larger RE than ignoring baseline measurements. In addition, using baseline measurements as a covariate and retaining baseline measurements as part of the response vector are comparable, and both methods have more efficiency than using change from baselines as the response variable.

Since the variance for baseline measurements as a covariate also depends on the subject variance, we display the relative efficiencies of the different methods for different values of $\sigma^2_s/\sigma^2_\epsilon$ in Figure 4.2.

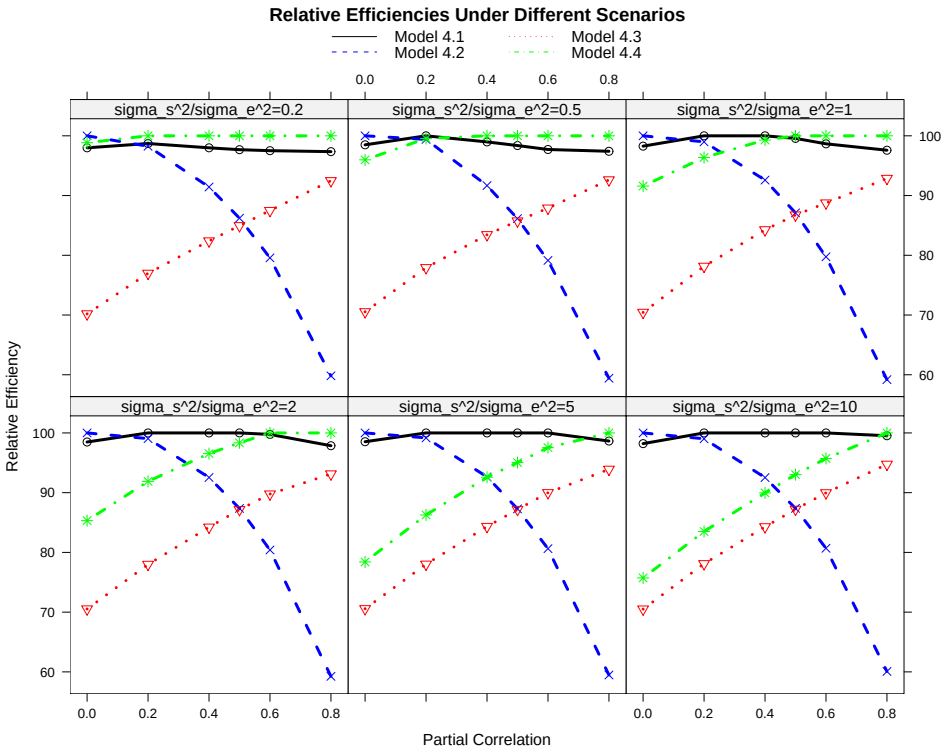


Figure 4.2: Relative Efficiencies at Different Values for the Partial Correlation for Six Values of the Ratio of Subject Variance to Random Error

From Figure 4.2, we see confirmation that the general pattern of relative efficiencies (RE) for three of the methods is similar for different values of $\sigma_s^2/\sigma_\epsilon^2$. However, the RE for baseline measurements as a covariate depends dramatically on the ratio of $\sigma_s^2/\sigma_\epsilon^2$. The RE for baseline measurements as a covariate decreases as the ratio increases, especially when the partial correlation $\rho$ is small. When the ratio of $\sigma_s^2/\sigma_\epsilon^2$ is larger, we can see that using baseline measurements as a covariate has a similar efficiency as using change from baselines. That is because $\beta$ is close to 1 at larger ratio of $\sigma_s^2/\sigma_\epsilon^2$, then these two methods are similar to each other.

Figure 4.3 displays the standard errors of the estimators of the treatment contrast for different methods at different values of $\sigma_s^2/\sigma_\epsilon^2$.

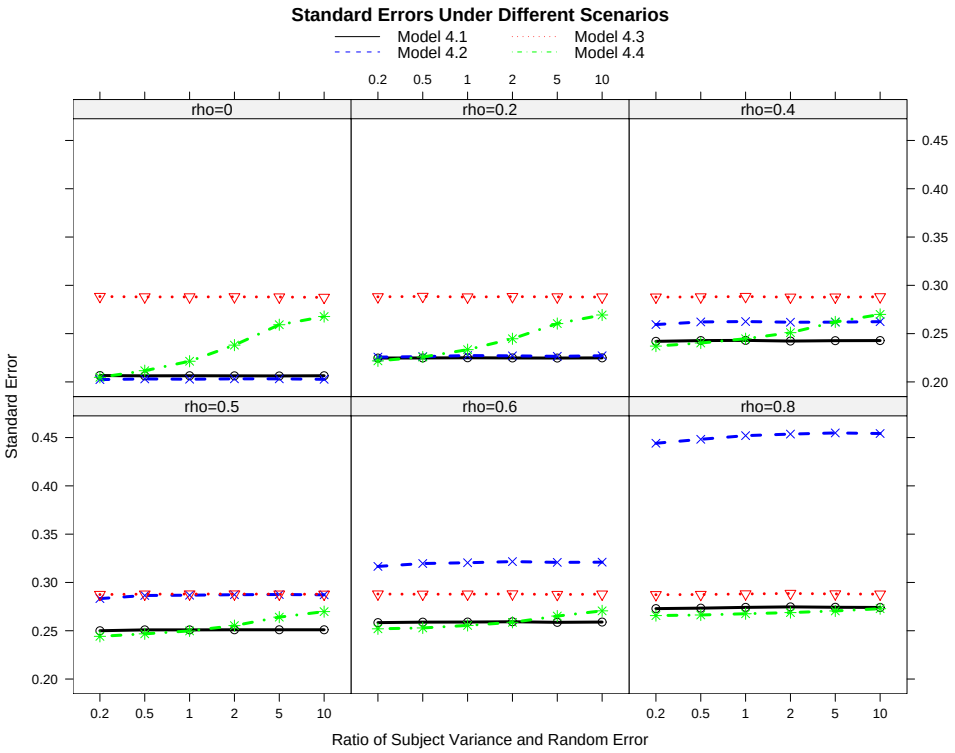


Figure 4.3: Standard Errors for Estimating $\tau_A - \tau_B$ for Different Values of the Ratio of Subject Variance and Random Error as a function of the Partial Correlation

As expected, the standard errors (SE) from the methods of retaining the baseline measurements as part of the response vector, ignoring baseline measurements and analysis of change from baselines remain the same as the ratio of $\sigma_s^2/\sigma_\epsilon^2$ changes. However, for the method of using baseline measurements as a covariate, the SE increases as the ratio of $\sigma_s^2/\sigma_\epsilon^2$ increases, and the change is more pronounced when $\rho$ is small, with almost no change when $\rho$ is large. That is why the method of using baseline measurements as a covariate is not as efficient for large values of the ratio of $\sigma_s^2/\sigma_\epsilon^2$ as it is for small values for small $\rho$.

## 4.4 Real data analysis

We now present a real data example for the case of single measurements to illustrate our methods.

### 4.4.1 Comparison of transdermal nitrate and isosorbide dinitrate in chronic stable angina

Nicholls *et al.* (1986) carried out a study to compare the treatment effects of transdermal nitrate (TN) with isosorbide dinitrate (ISDN) for angina in a group of 20 patients with chronic stable angina pectoris. Half of the patients were randomly allocated to the sequence consisting of treatment with TN for 4 weeks followed by treatment with ISDN for 4 weeks, whereas the other half of the patients received the treatments in reversed order. Before each treatment period, there was a 2-week period in which no treatment was administrated to obtain the baseline measurements, so that the trial lasted 12 weeks for each patient. The measurements are the weekly anginal attack rates and the data are displayed in Table 4.4.

Table 4.4 Weekly Anginal Attack Rates

| | | Period 1 | | Period 2 | |
|---|---|---|---|---|---|
| Sequence | Patient | Baseline | Outcomes | Baseline | Outcomes |
| TN → ISDN | 1 | 1.00 | 2.00 | 2.00 | 1.25 |
| | 4 | 41.50 | 30.00 | 31.50 | 27.00 |
| | 10 | 20.50 | 20.50 | 21.00 | 25.50 |
| | 12 | 15.50 | 14.50 | 14.50 | 13.25 |
| | 14 | 16.00 | 18.00 | 12.50 | 9.00 |
| | 15 | 2.00 | 3.50 | 3.00 | 2.25 |
| | 17 | 10.00 | 9.00 | 7.50 | 5.50 |
| | 20 | 10.00 | 8.50 | 6.00 | 4.25 |
| | 22 | 14.00 | 2.00 | 2.00 | 1.25 |
| | 24 | 5.50 | 2.50 | 1.50 | 2.50 |
| ISDN → TN | 3 | 17.50 | 19.25 | 19.00 | 21.25 |
| | 5 | 11.00 | 6.50 | 7.50 | 6.50 |
| | 7 | 4.00 | 2.00 | 1.50 | 3.00 |
| | 9 | 11.00 | 16.50 | 10.00 | 18.25 |
| | 13 | 6.50 | 4.25 | 0.50 | 1.25 |
| | 16 | 6.00 | 3.25 | 2.00 | 4.00 |
| | 18 | 1.00 | 0.00 | 0.00 | 0.00 |
| | 21 | 3.00 | 0.75 | 3.00 | 5.25 |
| | 23 | 9.50 | 1.00 | 0.50 | 8.50 |
| | 25 | 10.50 | 14.00 | 11.00 | 17.25 |

From Table 4.4, it is easy to notice that there is substantial variability among subjects. Crossover designs are especially important in such situations, since each subject receives both treatments so that the between-subject variability can be eliminated when comparing the treatment effects.

From Nicholls *et al.* (1986), there was no evidence of any carry-over effects in the second active treatment period. We analyzed the data by the various methods discussed in Section 4.2, and we obtained $\hat{\sigma}_s^2 = 65.3032$, $\hat{\sigma}_{sp}^2 = 0$ and $\hat{\sigma}^2 = 8.9099$ based on Model 4.1, so the partial correlation $\rho = 0/(0 + 8.9099) = 0$. Since the partial correlation equals zero, we would expect that ignoring baseline measurements will have the smallest variance for the estimator of the treatment comparison.

We present the results from the different methods in Table 4.5, where we also include a 95% confidence interval and a p-value for the treatment difference between TN and ISDN.

Table 4.5 Analysis Results for Anginal Attack Rates

| | | | | 95% CI | | |
|---|---|---|---|---|---|---|
| Method | Estimate | StdErr | RE | Lower | Upper | P-value |
| 1) Baselines in response vector | 2.3200 | 0.9439 | 90.97 | 0.4291 | 4.2109 | 0.0171 |
| 2) Ignoring baselines | 2.3200 | 0.8587 | 100 | 0.5159 | 4.1241 | 0.0146 |
| 3) Change from baselines | 1.8450 | 1.3982 | 61.41 | -0.9881 | 4.6781 | 0.1951 |
| 4) Baselines as a covariate | 1.9387 | 1.1078 | 77.52 | -0.4013 | 4.2787 | 0.0984 |

From Table 4.5, in terms of the relative efficiency (RE), for ignoring baseline measurements, retaining baseline measurements in the response vector, using baseline measurements as a covariate and using change from baseline measurements presents the methods in decreasing order. Using change from baselines as the response variable has only 61% efficiency compared to ignoring baseline measurements, and the

RE of using baseline measurements as a covariate is also low. This is because the ratio of $\sigma_s^2/\sigma_\epsilon^2$ is relatively large for these data, and this observation agrees with our simulation results. Notice that treatment effect is significant under only two of the methods. By using Grizzle (1965) and Koch (1972)'s method, Nicholls *et al.* (1986) concluded that there are significant differences between TN and ISDN.

In addition, for the method of using baseline measurements as a covariate, based on the estimated variance components and the formula for $\beta$ in Table 4.2, $\hat{\beta}$ is equal to 0.8799. Estimating this coefficient directly from the data, we obtain an estimate of 0.8027, with a standard error 0.08997. This indicates that these two estimates are close.

## 4.5 Discussion, conclusion and recommendation

It is common that baseline measurements are obtained prior to each treatment period in $2 \times 2$ crossover trials in clinical studies. Different methods under various models have been suggested to incorporate the baseline measurements in the analysis (Hills and Armitage, 1979; Patel, 1983; Fleiss, Wallenstein and Rosenfeld, 1985; Willan and Pater, 1986; Kenward and Jones, 1987; Chi, 1993; Grieve, 1994; Grieve and Senn, 1998).

However, all of these papers allowed for carryover effects in the wash-out period. Fleiss (1989) pointed out some issues associated with this topic. Moreover, there is a debate about whether, in the case of carryover effects, such measurements should be used as baselines at all. That is why Senn (2002) suggested the the second baseline measurements should be used only when there is no carryover effect from the previous period. In addition, he suggested to use whether $\rho$ is greater or less than 0.5 to decide whether to use change from baselines as the response variable or ignore baseline measurements.

In clinical trials, the FDA accepts the results of experiments using crossover designs only when there is no carryover effect. Therefore, long wash-out periods are usually applied, and data are often obtained on both the off-drug and on-drug days, because pharmacologists believe that there maybe some diurnal variation. In this chapter, we studied the problem of how to incorporate the baseline measurements for this situation. We proposed a method, retaining the baseline measurements as part of the response variable, which is new in the crossover setting. We compared it with three other commonly used methods. The variances of the treatment contrast estimators for the different methods are derived and compared at different values of $\rho$ in the range from 0 to 1. We also conducted a simulation study to evaluate the performance of four methods.

Our theoretical results, which were obtained under the assumption of known variances, are confirmed by our simulation results, which do not make this assumption. From both results, we can conclude that different methods can be compared based on the partial correlation $\rho$. Among the four methods, only the variance of the contrast estimator from the method that uses baseline measurements as a covariate depends on the subject variance. However, the general pattern of the RE for the different methods at different partial correlations $\rho$ is consistent for a broad range of true values of the subject variance. In general, the RE of ignoring baseline measurements decreases as $\rho$ increases, whereas the RE of using baseline measurements, no matter which method is used, increases as $\rho$ increases. This is not surprising, since larger values of $\rho$ imply a stronger correlation between the baseline measurements and outcomes. On average, our new method, retaining baseline measurements as part of the response vector has the largest RE, as long as $\rho$ is not near 0. If $\rho$ is very small, then ignoring baseline measurements has a relatively large RE. The RE of using baseline measurements as a covariate is also high as long as the subject variance is not large. However, for the analysis of change from baselines, the RE can

be considerably smaller and it is best to avoid this method. That agrees with the recommendation in earlier work (Fleiss, 1989; Senn, 2002).

There are some implicit assumptions in our study, and changing those could lead to different conclusions. For example, the fact that the same effects $\mu$ and $s_{jk}$ are assumed for both baselines and outcomes implies that these measurements are comparable and on the same scale, which is reasonable for most cases. Moreover, we assume that the period effect for baselines and the corresponding outcomes are the same, but we also include the day effects to capture the difference between data from off-drug and on-drug day. We also assume that the variance of the baseline measurements is equal to that for the outcomes. For a detailed discussion of the validation of the assumptions made in the baseline analysis, the reader may refer to Grieve and Senn (1998).

Based on our results, we formulate the following recommendations:

First, the experimenter must use whatever information is available from earlier studies with the product (drug) to design the experiment appropriately. For example, appropriate lengths for treatment wash-out periods should be selected. If significant unequal carryover effects are suspected, then a crossover design should not be used.

Second, if there is no prior knowledge about the partial correlation from previous data analysis in past experiences, collect the baseline measurements and retain the baseline measurements as part of the response vector in the analysis.

Third, if there is evidence from previous data analysis in past experiences that partial correlation $\rho$ is very small, there is no benefit to collect baselines, including baseline measurements as a covariate can result in substantial SE inflation especially when subject variance is large.

Fourth, if baseline measurements are obtained, and the partial correlation is relatively large and subject variance is relatively small then we should use them

as a covariate; however, analysis of change from baseline measurements should be avoided at all times.

## 4.6 References


1. Armitage, P. and Hills, M. (1982). The two-period crossover trial. *The Statistician* 32, 119-131.

2. Brown, B. W. (1980). The crossover experiment for clinical trials. *Biometrics* 36, 69-79.

3. Chi, E. M. (1993) Analysis of cross-over trials when within-subject errors follow an AR(1) process. *Biometrical Journal* 34, 359-365.

4. 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.

5. Fleiss, J. L. (1989) A critique of recent research on the two-treatment crossover design. *Controlled Clinical Trials* 10, 237-243.

6. Grieve, A. P. (1994). Extending a Bayesian analysis of the two-period crossover to allow for baseline measurements. *Statistics in Medicine* 13, 905-929.

7. Grieve, A. and Senn, S. (1998). Estimating treatment effects in clinical crossover trials. *Journal of Biopharmaceutical Statistics* 8, 191-233.

8. Grizzle, J. E. (1965). The two-period change-over design and its use in clinical trials. *Biometrics* 21, 467-480.

9. Hills, M. and Armitage, P. (1979) The two-period cross-over clinical trial. *British Journal of Clinical Pharmacology* 8, 7-20.

10. Jones, B. and Kenward, M. J. (2003) *Design and Analysis of Cross-over Trials.* Second Edition, CRC Press.

11. Kenward, M. J. and Jones, B. (1987) The analysis of data from $2 \times 2$ cross-over trials with baseline measurements. *Statistics in Medicine* 6, 911-926.

12. Koch, G. G. (1972) The use of non-parametric methods in the statistical analysis of the two-period change-over design. *Biometrics* 28, 577-584.

13. Nicholls, D. P., Moles, K., Gleadhill, D. N. S., Booth, K., Rowan, J. and Morton, P. (1986). Comparison of transdermal nitrate and isosorbide dinitrate in chronic stable angina. *British Journal of Clinical Pharmacology* 22, 15-20.

14. Patel, H. I. (1983). The use of baseline measurements in the two-period cross-over design. *Communications in Statistics* 12, 2693-2721.

15. Ratkowsky, D. A., Evans, M. A., and Alldredge, J. R. (1992). *Cross-over Experiments, Design, Analysis and Application.* Marcel Dekker, New York.

16. Senn, S. (1994) The *AB/BA* crossover: past, present, and future? *Statistical Methods in Medical Research* 3, 303-324.

17. Senn, S. (2002). *Crossover Trials in Clinical Research.* Second Edition, U.K. Wiley.

18. Willan, A. R., and Pater, J. L. (1986). Using baseline measurements in the two-period crossover clinical trial. *Controlled Clinical Trials* 7, 282-289.

# Chapter 5

# Use of Baseline Measurements in the $2 \times 2$ Crossover Trial for the Case of Repeated Measurements

## 5.1 Introduction

The discussion of baseline measurements in Chapter 4 deals with the case of a single measurement on each day within each period. However, very often, in practice, the response in a clinical trial is measured repeatedly over time for each day within each period. In this case, each subject produces a set of profiles of repeated measurements in each period. Similar to Putt and Chinchilli (1999), we refer to this as a repeated measurements crossover design to emphasize that repeated measures are collected at different time points under the same treatment within each period. A crossover design with one observation per treatment period has been called a basic crossover design by Wallenstein and Fisher (1977).

There can be many reasons to collect measurements over time, for example, to obtain as much information about the effects of the treatments as possible; to investigate dynamic changes in treatment effects over time, and so on. How to approach the statistical analysis may depend on the reasons for collecting such data. Jones and Kenward (2003) observed that summary statistics, such as particular end point, the area under the profile and the average slope of the profile, may be used to reduce the repeated measurements crossover design to the basic crossover design. This will simplify the problem, and summary statistics may provide useful information in some situations. However, it is neither always possible, nor always desirable, depending on

the objective, to base the analysis on a summary statistic. Direct methods of analysis for repeated measurements with no baseline measurements in crossover design have been discussed by many authors. For example, by using the sum and difference of the observations from two periods, both Wallenstein and Fisher (1977) and Jones and Kenward (2003) applied the conventional split-plot approach for this design, while both Patel and Hearne (1980) and Grender and Johnson (1993) discussed a multivariate linear model approach. Dunsmore (1981) compared two approaches to study the repeated measurements crossover design: one applied Wallenstein and Fisher's (1977) model, the other adapted Fearn (1975)'s Bayesian analysis of growth curves for a quadratic curve over time. Putt and Chinchilli (1999) presented a mixed effects model to analyze repeated measures crossover studies.

However, to our knowledge, the analysis of the repeated measurements crossover design including the baseline measurements has not been formally described. To motivate our work, we introduce studies in which data were collected at different time points on both the off-drug and on-drug day in both periods. We consider both time-matched baselines and averaged baselines. Time-matched baselines refer to the situation where each measurement on an off-drug day is the baseline for the measurement on the on-drug day at the same time point, whereas averaged baselines refer to the average of the measurements from all time points on an off-drug day as the baseline for each on-drug day measurement in that period.

The goal of our study in this chapter is twofold. As a preliminary step in the analysis, we suggest a test for the treatment by time interaction by retaining the baseline measurements as part of the response vector. If this interaction is statistically significant, we wish to develop graphical methods to detect the trend of treatment changes over time. Alternatively, if this interaction is not significant, we wish to determine appropriate methods to incorporate the information from the baseline measurements to compare the average treatment contrast for different methods and

different types of baselines considered. In repeated measurements crossover designs, random errors are commonly assumed to be independent, such as in Wallenstein and Fisher (1977). However, the measurements taken closer together in time could be more highly correlated than measurements further apart in time, such as for an autocorrelation structure, or maybe even more complicated. We would like to consider different assumptions for the random errors to study the treatment contrast. For the theoretical results, we will focus on independent random errors and random errors that follow an AR(1) structure.

Therefore, this chapter is organized as follows. In Section 5.2, examples are introduced to illustrate the motivation of the study. In Section 5.3, potential models for the repeated measurements crossover design are discussed, and the effect of treatment by time interaction is tested. If the interaction effect is significant, graphical methods are discussed in Section 5.4 to visualize the treatment effect changes over time. Otherwise, different methods and models to handle baseline measurements are discussed in Section 5.5, and analytical expressions of variances of the estimators of the treatment contrast from different methods are derived and compared. A thorough simulation study is conducted to evaluate the performance of different methods. A numerical example discussed in Section 5.2 is analyzed for illustration purposes. Section 5.6 provides some discussion for this chapter. The proofs for the results in Section 5.5 are deferred to the Appendix.

## 5.2 Motivating examples

Numerical examples of repeated measurements crossover experiments from the pharmaceutical industry are used to motivate our study. Because these data are from real studies, due to confidentiality issues, we can describe the studies only abstractly, and we can not report the original data here either.

The first example (Study I) is from a single-center, randomized, double-blind 2 $\times$ 2 crossover trial. In that experiment, 12 healthy males were randomly assigned to one of two sequences, 6 in each sequence. The subjects in sequence $AB$ received treatment $A$ first then followed by treatment $B$; the subjects in sequence $BA$ received the treatments in the reverse order. Between two periods, there were 14 days of wash-out period. The data were collected repeatedly at 3 different time points (equally spaced) before the treatments were assigned (off-drug day) and during the treatment periods (on-drug day) for both periods.

The second example (Study II) is also from a single-center, randomized, double-blind, but 4 $\times$ 4 crossover trial. There are 4 subjects in each sequences. Since period effects are not significant, we will extract a 2 $\times$ 2 design by using two treatments only for illustration purpose. So it is realized similarly as Study I except that the data were collected repeatedly at 10 different time points (not equally spaced) on both days in each period.

The goal of Study I is to compare the treatment contrast averaged over time, and the goal of Study II is to detect trend of treatment changes over time. The question is how to use the information provided by the baseline measurements more efficiently in such studies.

## 5.3 Statistical models

For designs described in Section 5.2, in general, we assume that measurements are obtained at $q$ ($m = 1, 2, \ldots, q$) time points on both off-drug and on-drug days in both periods in a $AB/BA$ crossover trial, and we denote them by $X_{ijkm}$ and $Y_{ijkm}$ respectively, where $i = 1, 2$, $j = 1, 2$, $k = 1, 2, \ldots, n_j$ and $m = 1, 2, \ldots, q$ are the indices for period, sequence, subject within the $j^{th}$ sequence and time point, respectively. Schematically, the data can be displayed as in Table 5.1.

Table 5.1 A Two-Period Crossover Trial with Repeated Measurements

| | | | Sequence $AB$ | | | Sequence $BA$ | | |
|---|---|---|---|---|---|---|---|---|
| Period 1 | off-drug day | time 1 | $X_{1111}$ | ... | $X_{11n_11}$ | $X_{1211}$ | ... | $X_{12n_21}$ |
| | | time 2 | $X_{1112}$ | ... | $X_{11n_12}$ | $X_{1212}$ | ... | $X_{12n_22}$ |
| | | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| | | time $q$ | $X_{111q}$ | ... | $X_{11n_1q}$ | $X_{121q}$ | ... | $X_{12n_2q}$ |
| | on-drug day | time 1 | $Y_{1111}$ | ... | $Y_{11n_11}$ | $Y_{1211}$ | ... | $Y_{12n_21}$ |
| | | time 2 | $Y_{1112}$ | ... | $Y_{11n_12}$ | $Y_{1212}$ | ... | $Y_{12n_22}$ |
| | | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| | | time $q$ | $Y_{111q}$ | ... | $Y_{11n_1q}$ | $Y_{121q}$ | ... | $Y_{12n_2q}$ |
| Period 2 | off-drug day | time 1 | $X_{2111}$ | ... | $X_{21n_11}$ | $X_{2211}$ | ... | $X_{22n_21}$ |
| | | time 2 | $X_{2112}$ | ... | $X_{21n_12}$ | $X_{2212}$ | ... | $X_{22n_22}$ |
| | | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| | | time $q$ | $X_{211q}$ | ... | $X_{21n_1q}$ | $X_{221q}$ | ... | $X_{22n_2q}$ |
| | on-drug day | time 1 | $Y_{2111}$ | ... | $Y_{21n_11}$ | $Y_{2211}$ | ... | $Y_{22n_21}$ |
| | | time 2 | $Y_{2112}$ | ... | $Y_{21n_12}$ | $Y_{2212}$ | ... | $Y_{22n_22}$ |
| | | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| | | time $q$ | $Y_{211q}$ | ... | $Y_{21n_1q}$ | $Y_{221q}$ | ... | $Y_{22n_2q}$ |

From Table 5.1, each subject has $4q$ observations, which are from $q$ different time points on four different days within two different periods. Here we assume that days are nested in periods, but that times are crossed with days and periods. So, besides the possible variations discussed in the case of single measurements, additional sources of random variation may be due to different time points. Possible additional fixed effects could be time effects and a treatment by time interaction. Although the primary purpose of an experiment which uses crossover design is to compare the treatment effects, if the treatment by time interaction is significant, it is usually not meaningful to compare the average treatment effects. So we consider a model with a treatment by time interaction to examine the change in treatment

difference over time. If this interaction is not significant, then we ignore it to study the average treatment effects. Thus, one possible model, with treatment by time interaction, for the observations in the case of repeated measurements may be written as

$$\begin{aligned} Y_{hijkm} = \mu + \pi_i + D_{hi} + T_m + h\tau_{t(i,j)} + h(\tau T)_{t(i,j)m} \\ + s_{jk} + \zeta_{ijk} + \omega_{jkm} + \xi_{hijk} + \epsilon_{hijkm}, \\ i = 1,2,\ j = 1,2,\ k = 1,2,\ldots,n_j,\ m = 1,2,\ldots,q,\ h = 0,1, \end{aligned} \tag{5.1}$$

where $Y_{hijkm}$ corresponds to $X_{ijkm}$ when $h = 0$ and to $Y_{ijkm}$ when $h = 1$, respectively. Model 5.1 without treatment by time interaction can be written as

$$Y_{hijkm} = \mu + \pi_i + D_{hi} + T_m + h\tau_{t(i,j)} + s_{jk} + \zeta_{ijk} + \omega_{jkm} + \xi_{hijk} + \epsilon_{hijkm}. \tag{5.2}$$

The terms $T_m$ and $(\tau T)_{t(i,j)m}$ stand for a fixed time effect and treatment by time interaction, respectively. The subject by time interaction $\omega_{jkm}$ is the random effect due to the $k^{th}$ subject in the $j^{th}$ sequence at the $m^{th}$ time point, and we assume that $\omega_{jkm} \sim N(0, \sigma^2_{st})$. The term $\xi_{hijk}$ is the random effect of the $k^{th}$ subject in the $j^{th}$ sequence at the $i^{th}$ period on the $h^{th}$ day, and we assume that $\xi_{hijk} \sim N(0, \sigma^2_{sd})$. The random terms $s_{jk}$ and $\zeta_{ijk}$ have same distribution as in Chapter 4. For the random error terms $\epsilon_{hijkm}$'s, we consider two different assumptions for theoretical derivations. First, as usual, we assume that the error terms are independently distributed as $\epsilon_{hijkm}$ $\sim N(0, \sigma^2_\epsilon)$. Second, as discussed in Section 5.1, we assume that random error terms follow the first-order autocorrelated structure (AR(1)) for the different time points $m$ (equally spaced) for the same subject on the same day in the same period with variance $\sigma^2_\epsilon$ and autocorrelation coefficient $r$. It is assumed that the random errors and all other random terms are independent to each other. Other model terms have the same interpretations as in Chapter 4.

From the model assumptions, we obtain

$$\text{Corr}(X_{ijkm}, Y_{i'jkm}) = \begin{cases} (\sigma_s^2 + \sigma_{sp}^2 + \sigma_{st}^2)/(\sigma_s^2 + \sigma_{sp}^2 + \sigma_{st}^2 + \sigma_{sd}^2 + \sigma_\epsilon^2) & \text{if } i = i', \\ \sigma_s^2 + \sigma_{st}^2)/(\sigma_s^2 + \sigma_{sp}^2 + \sigma_{st}^2 + \sigma_{sd}^2 + \sigma_\epsilon^2) & \text{if } i \neq i'. \end{cases}$$

### 5.3.1 Testing for the treatment by time interaction

First, we would like to use Model 5.1 to test for a possible treatment by time interaction. We use the motivating examples, and fit the data with different variance-covariance structures for the random errors. Tables 5.2 and 5.3 present the results.

Table 5.2 Results of Test for Treatment by Time Interaction for Study I

| Var-cov | F-value | ProbF | *AIC* | *AICC* | *BIC* |
|---|---|---|---|---|---|
| IND | 1.68 | 0.1656 | -142.5* | -142.1* | -140.1* |
| AR(1) | 1.63 | 0.1804 | -140.6 | -139.9 | -137.6 |
| CS | 1.68 | 0.1656 | -140.5 | -139.9 | -137.6 |
| TOEP | 1.68 | 0.1684 | -140.6 | -139.9 | -137.6 |
| ANTE(1) | 1.57 | 0.2016 | -135.1 | -133.7 | -130.8 |
| SP(POW)(t) | 1.81 | 0.1395 | -140.6 | -139.9 | -137.6 |
| SP(EXP)(t) | 1.28 | 0.2883 | -140.6 | -139.9 | -137.6 |

Table 5.3 Results of Test for Treatment by Time Interaction for Study II

| Var-cov | F-value | ProbF | *AIC* | *AICC* | *BIC* |
|---|---|---|---|---|---|
| IND | 3.77 | $< .0001$ | 4159.5 | 4159.6 | 4162.6 |
| AR(1) | 3.21 | $< .0001$ | 4156.5 | 4156.6 | 4160.3 |
| CS | 3.77 | $< .0001$ | 4161.5 | 4161.6 | 4165.4 |
| TOEP | 3.37 | $< .0001$ | 4164.7 | 4165.3 | 4174.0 |
| ANTE(1) | 2.76 | 0.0004 | 4170.9 | 4172.6 | 4187.9 |
| SP(POW)(t) | 3.24 | $< .0001$ | 4152.1 * | 4152.2 * | 4156.0 * |
| SP(EXP)(t) | 3.25 | $< .0001$ | 4152.1 * | 4152.2 * | 4156.0 * |

In Tables 5.2 and 5.3, the abbreviated names of the variance-covariance structure "IND", "AR(1)", "CS", "TOEP", "ANTE(1)", "SP(POW)(t)" and "SP(EXP)(t)"

stand for "Independent", "Autoregressive (1)", "Compound Symmetry", "Toeplitz", "Ante-dependence (1)", "Spatial Power (t)" and "Spatial Exponential (t)", respectively. The examples for the structures "ANTE(1)", "TOEP" and "SP(POW)" can be displayed as

$$\begin{bmatrix} \sigma^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_1\rho_2 \\ \sigma_1\sigma_2\rho_1 & \sigma^2 & \sigma_2\sigma_3\rho_2 \\ \sigma_3\sigma_1\rho_2\rho_1 & \sigma_3\sigma_2\rho_2 & \sigma_3^2 \end{bmatrix}, \begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix} \text{ and } \sigma^2 \begin{bmatrix} 1 & \rho^{d_{12}} & \rho^{d_{13}} & \rho^{d_{14}} \\ \rho^{d_{21}} & 1 & \rho^{d_{23}} & \rho^{d_{24}} \\ \rho^{d_{31}} & \rho^{d_{32}} & 1 & \rho^{d_{34}} \\ \rho^{d_{41}} & \rho^{d_{42}} & \rho^{d_{43}} & 1 \end{bmatrix},$$

respectively, where $d_{ij}$ is the absolute distance between the $i^{th}$ and $j^{th}$ observations.

"F-value" and "ProbF" are the values of the test statistic and p-values for tests for a treatment by time interaction. "*AIC*", "*AICC*" and "*BIC*" are model selection criteria. From the tables, models with independent variance-covariance structure for Study I and Spatial structure for Study II have the smallest *AIC*, *AICC* and *BIC*. However, different variance-covariance structures lead to the same conclusion for the significance of treatment by time interaction. Clearly, the treatment by time interaction is not significant for Study I, but it is for Study II. So we will use Study II as an example to demonstrate the graphical methods to study the treatment effect changes over time, and use Study I to study treatment effects averaged over time by using different methods to incorporate the baseline measurements.

## 5.4 Graphical methods for treatment effects over time

As seen in Section 5.3, the treatment by time interaction from Study II is significant. We will use graphical methods to display how the treatment effects change over time. Most often in repeated measurements crossover designs with no baselines, profile plots are used of the response versus time points for each period for each individual (Putt and Chinchilli, 1999; Jones and Kenward, 2003). The latter also mentioned the possibility of plotting treatment differences from the original data, against the

different time points for each treatment group, either individually or averaged over each sequence.

In the presence of the baseline measurements, every subject has four observations at each time point. By focusing on a fixed time point, we are back to a situation similar to that in Chapter 4 and estimators obtained there suggest now how to estimate the treatment difference at that time points. As an illustration, we consider the method of retaining baseline measurements as part of the response vector. Based on the results from Chapter 4, we use coefficients $(\rho_m, -1, -\rho_m, 1)$ and $(-\rho_m, 1, \rho_m, -1)$ for the first and second sequence, respectively, for estimating $\tau_B - \tau_A$ at a particular time point, where $\rho_m$ can be obtained by estimating $\sigma^2_{sp}/(\sigma^2_{sp} + \sigma^2_\epsilon)$, $\sigma^2_{sp}$ and $\sigma^2_\epsilon$ are estimated based on the measurements on time point $m$. These linear combinations are calculated for each subject at each time point, and the resulting values can be plotted against time for each subject or averaged over each sequence.

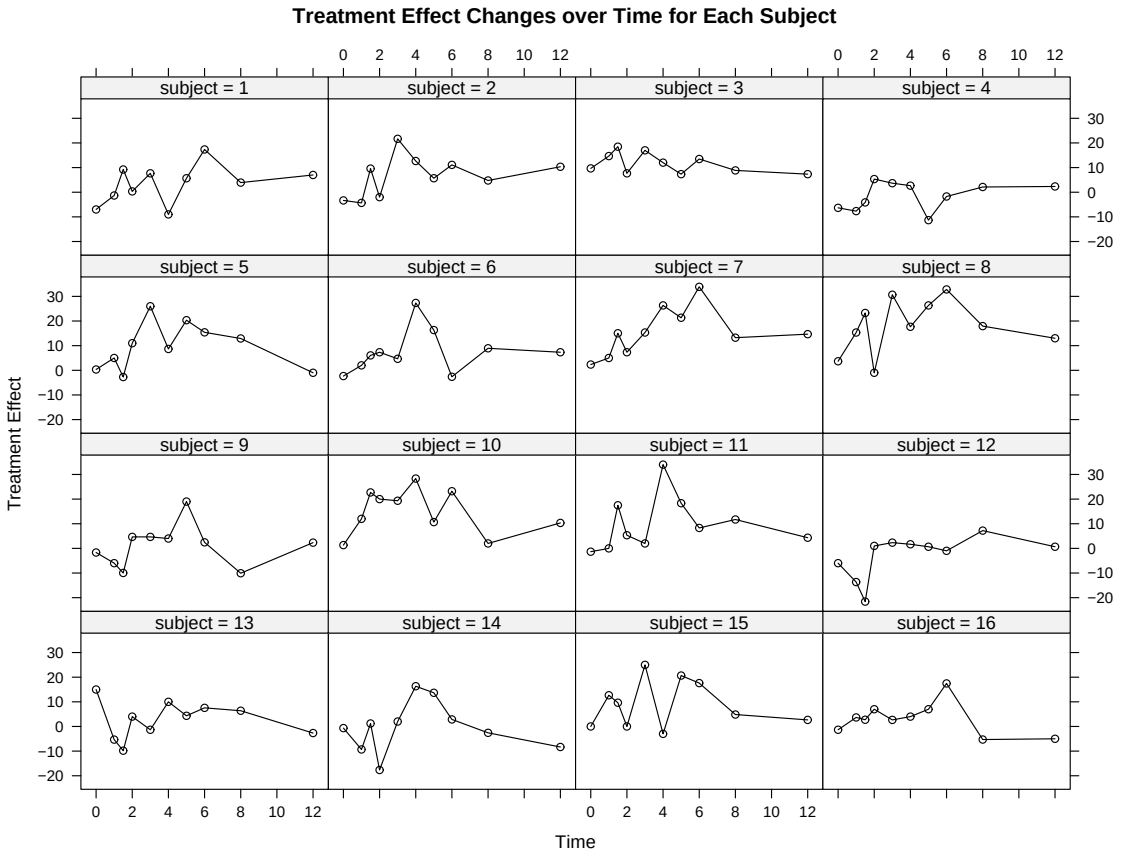


Figure 5.1: Treatment Effect Changes over Time for Each Subject

Figure 5.1 shows the plot for each subject, where subjects 1-8 belong to sequence $AB$ and subjects 9-16 to sequence $BA$. However, from Figure 5.1, it is difficult to recognize any trend of change in treatment differences over time. Figure 5.2 displays averages over each sequence.

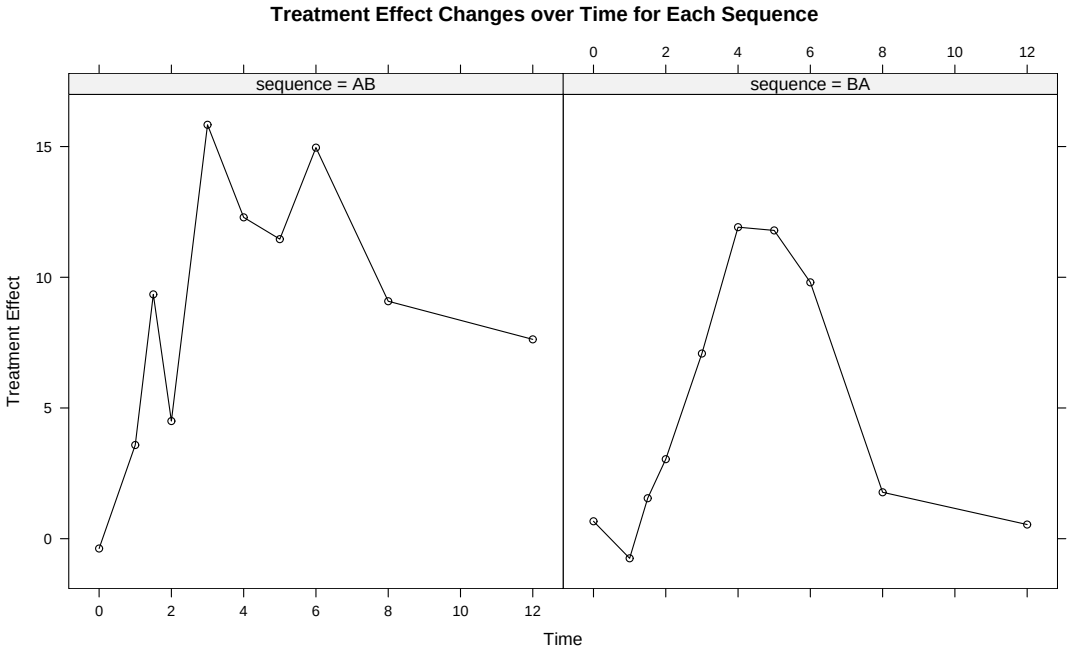


Figure 5.2: Treatment Effect Changes over Time for Each Sequence

From Figure 5.2, there is a clearer pattern of treatment effect change over time. Most of the plotted points are greater than zero, which indicates that the effect of treatment $B$ is larger than the effect of treatment $A$. In addition, in general, the treatment effect first increases with time, and then decreases after reaching a peak at the middle time points. A polynomial growth curve might be appropriate to study the overall trend of treatment effect changes over time, and the estimates of the parameters of the growth curves can be obtained by the same approach used in the analysis of variance for basic designs.

## 5.5 Methods to incorporate baseline measurements to study average treatment effects

### 5.5.1 Methods and models

If the treatment by time interaction is not significant, we consider Model 5.2 to study the average treatment effects. Similar to the case of single measurements, we consider four different methods to incorporate the baseline measurements in the analysis. We start from Model 5.2, and obtain the corresponding models for the other methods. In addition, we consider both time-matched baselines and averaged baselines when using either change from baseline measurements or using baseline measurements as a covariate in this section.

Thus, the various models are:

1) Retain the baseline measurements as part of the response vector as in Model 5.2;

2) Ignore the baseline measurements, and use the observations for outcomes only:

$$Y_{ijkm} = \mu + \pi_i + T_m + \tau_{t(i,j)} + s_{jk} + \zeta_{ijk} + \omega_{jkm} + \epsilon_{ijkm} \tag{5.3}$$

3) Use change from baseline measurements as the response variable:

3a) Time-matched baselines:

$$Y_{ijkm} - X_{ijkm} = \mu + \pi_i + \tau_{t(i,j)} + \zeta_{ijk} + \epsilon_{ijkm} \tag{5.4}$$

3b) Averaged baselines:

$$Y_{ijkm} - \bar{X}_{ijk.} = \mu + \pi_i + T_m + \tau_{t(i,j)} + \zeta_{ijk} + \omega_{jkm} + \epsilon_{ijkm} \tag{5.5}$$

4) Use baseline measurements as a covariate and model the outcomes conditional on the covariate:

4a) Time-matched baselines:

$$Y_{ijkm} = \mu + \pi_i + T_m + \tau_{t(i,j)} + \beta x_{ijkm} + s_{jk} + \zeta_{ijk} + \omega_{jkm} + \epsilon_{ijkm} \tag{5.6}$$

4b) Averaged baselines:

$$Y_{ijkm} = \mu + \pi_i + T_m + \tau_{t(i,j)} + \beta\bar{x}_{ijk.} + s_{jk} + \zeta_{ijk} + \omega_{jkm} + \epsilon_{ijkm} \tag{5.7}$$

In Models 5.5 and 5.7, the $\omega_{jkm}$'s follow a compound symmetric correlation structure, while the $\epsilon_{ijkm}$'s follow a compound symmetric structure or Toeplitz correlation structure for an independent or AR(1) structure in Model 5.2, respectively. All other random terms in Models 5.3-5.7 have the same distribution as they do in Model 5.2.

Assuming that variance components are known, we can obtain the Best Linear Unbiased Estimators (BLUE) of the treatment contrast for Models 5.2-5.7. The results are presented in Tables 5.4 and 5.5 for independent and autocorrelated random errors respectively. The proofs of the results are deferred to the Appendix.

Table 5.4 BLUEs for Estimating the Treatment Contrast

(random error terms are independent)

| Model | Estimator |
|---|---|
| 5.2 | $\frac{1}{2}[\bar{Y}_{11..} + \bar{Y}_{22..} - \bar{Y}_{12..} - \bar{Y}_{21..} - \rho(\bar{X}_{11..} + \bar{X}_{22..} - \bar{X}_{12..} - \bar{X}_{21..})]$ |
| 5.3 | $\frac{1}{2}(\bar{Y}_{11..} + \bar{Y}_{22..} - \bar{Y}_{12..} - \bar{Y}_{21..})$ |
| 5.4 & 5.5 | $\frac{1}{2}[\bar{Y}_{11..} + \bar{Y}_{22..} - \bar{Y}_{12..} - \bar{Y}_{21..} - (\bar{X}_{11..} + \bar{X}_{22..} - \bar{X}_{12..} - \bar{X}_{21..})]$ |
| 5.6 | $\frac{1}{2}[\bar{Y}_{11..} + \bar{Y}_{22..} - \bar{Y}_{12..} - \bar{Y}_{21..} - \beta_1(\bar{x}_{11..} + \bar{x}_{22..} - \bar{x}_{12..} - \bar{x}_{21..})]$ |
| 5.7 | $\frac{1}{2}[\bar{Y}_{11..} + \bar{Y}_{22..} - \bar{Y}_{12..} - \bar{Y}_{21..} - \beta_2(\bar{x}_{11..} + \bar{x}_{22..} - \bar{x}_{12..} - \bar{x}_{21..})]$ |
| | $\rho = \frac{\sigma^2_{sp}}{\sigma^2_{sp}+\sigma^2_{sd}+\sigma^2_\epsilon/q}$, $\beta_1 = \frac{\sigma^2_s+\sigma^2_{sp}+\sigma^2_{st}}{\sigma^2_s+\sigma^2_{sp}+\sigma^2_{st}+\sigma^2_{sd}+\sigma^2_\epsilon}$, $\beta_2 = \frac{\sigma^2_s+\sigma^2_{sp}+\sigma^2_{st}/q}{\sigma^2_s+\sigma^2_{sp}+\sigma^2_{sd}+(\sigma^2_{st}+\sigma^2_\epsilon)/q}$ |

Table 5.5 BLUEs for Estimating the Treatment Contrast

(random error terms follow AR(1))

| Model | Estimator |
|---|---|
| 5.2 | $\frac{1}{2}\{\sum_{m=1}^{q}\phi_m[\bar{Y}_{11.m}+\bar{Y}_{22.m}-\bar{Y}_{12.m}-\bar{Y}_{21.m}-\rho(\bar{X}_{11.m}+\bar{X}_{22.m}-\bar{X}_{12.m}-\bar{X}_{21.m})]\}$ |
| 5.3 | $\frac{1}{2}\{\sum_{m=1}^{q}\phi_m(\bar{Y}_{11.m}+\bar{Y}_{22.m}-\bar{Y}_{12.m}-\bar{Y}_{21.m})\}$ |
| 5.4 | $\frac{1}{2}\{\sum_{m=1}^{q}\phi_m[\bar{Y}_{11.m}+\bar{Y}_{22.m}-\bar{Y}_{12.m}-\bar{Y}_{21.m}-(\bar{X}_{11.m}+\bar{X}_{22.m}-\bar{X}_{12.m}-\bar{X}_{21.m})]\}$ |
| 5.5 | $\frac{1}{2}\{\sum_{m=1}^{q}[\phi_m(\bar{Y}_{11.m}+\bar{Y}_{22.m}-\bar{Y}_{12.m}-\bar{Y}_{21.m})]-(\bar{X}_{11..}+\bar{X}_{22..}-\bar{X}_{12..}-\bar{X}_{21..})\}$ |
| 5.6 | $\frac{1}{2}\{\sum_{m=1}^{q}\phi_m[\bar{Y}_{11.m}+\bar{Y}_{22.m}-\bar{Y}_{12.m}-\bar{Y}_{21.m}-\beta_1(\bar{x}_{11.m}+\bar{x}_{22.m}-\bar{x}_{12.m}-\bar{x}_{21.m})]\}$ |
| 5.7 | $\frac{1}{2}\{\sum_{m=1}^{q}[\phi_m(\bar{Y}_{11.m}+\bar{Y}_{22.m}-\bar{Y}_{12.m}-\bar{Y}_{21.m})]-\beta_2(\bar{x}_{11..}+\bar{x}_{22..}-\bar{x}_{12..}-\bar{x}_{21..})\}$ |

$$\rho=\frac{\sigma_{sp}^2}{\sigma_{sp}^2+\sigma_{sd}^2+\frac{1+r}{q-(q-2)r}\sigma_\epsilon^2},\quad \phi_m=\begin{cases}\frac{1}{q-(q-2)r} & \text{if } m=1,q\\ \frac{1-r}{q-(q-2)r} & \text{if } m=2,\cdots,q-1\end{cases},$$

$$\beta_1=\frac{\sigma_s^2+\sigma_{sp}^2+\sigma_{st}^2}{\sigma_s^2+\sigma_{sp}^2+\sigma_{st}^2+\sigma_{sd}^2+\sigma_\epsilon^2},\quad \beta_2=\frac{\sigma_s^2+\sigma_{sp}^2+\sigma_{st}^2/q}{\sigma_s^2+\sigma_{sp}^2+\sigma_{sd}^2+\frac{\sigma_{st}^2}{q}+\frac{(1-r^2)q-2r(1-r^q)}{q^2(1-r)^2}\sigma_\epsilon^2}$$

When random errors are independent, we can see from Table 5.4 that the BLUE for the treatment contrast is comparable to the results in Table 4.2, and it is a linear combination of the averages over subjects and different time points for the two periods and two sequences. However, when random errors follow the AR(1) structure, as we can see from Table 5.5, the BLUE of the treatment contrast is no longer a linear combination of the $\bar{Y}_{ij.}$'s and $\bar{X}_{ij.}$'s, but becomes a weighted average of the $\bar{Y}_{ij.m}$'s and $\bar{X}_{ij.m}$'s. In addition, the weight is $1/[q-(q-2)r]$ for the first and last time points, and is $(1-r)/[q-(q-2)r]$ for all other points.

Tables 5.6 and 5.7 present the results for the corresponding variances of the BLUEs for the treatment contrast for independent and AR(1) random errors, respectively.

Table 5.6 Variances of BLUEs for the Treatment Contrast

(random error terms are independent)

| Model | Variance |
|---|---|
| 5.2 | $\frac{n_1+n_2}{2n_1n_2}\frac{(2\sigma^2_{sp}+\sigma^2_{sd}+\sigma^2_\epsilon/q)(\sigma^2_{sd}+\sigma^2_\epsilon/q)}{\sigma^2_{sp}+\sigma^2_{sd}+\sigma^2_\epsilon/q}$ |
| 5.3 | $\frac{n_1+n_2}{2n_1n_2}(\sigma^2_{sp}+\sigma^2_{sd}+\sigma^2_\epsilon/q)$ |
| 5.4 & 5.5 | $\frac{n_1+n_2}{2n_1n_2}[2(\sigma^2_{sd}+\sigma^2_\epsilon/q)]$ |
| 5.6 | $\frac{n_1+n_2}{2n_1n_2}[(1-\beta_1)^2\sigma^2_{sp}+(1+\beta_1^2)(\sigma^2_{sd}+\sigma^2_\epsilon/q)]$ |
| 5.7 | $\frac{n_1+n_2}{2n_1n_2}[(1-\beta_2)^2\sigma^2_{sp}+(1+\beta_2^2)(\sigma^2_{sd}+\sigma^2_\epsilon/q)]$ |

Table 5.7 Variances of BLUEs for the Treatment Contrast

(random error terms follow AR(1))

| Model | Variance |
|---|---|
| 5.2 | $\frac{n_1+n_2}{2n_1n_2}\frac{(2\sigma^2_{sp}+\sigma^2_{sd}+\frac{1+r}{q-(q-2)r}\sigma^2_\epsilon)(\sigma^2_{sd}+\frac{1+r}{q-(q-2)r}\sigma^2_\epsilon)}{\sigma^2_{sp}+\sigma^2_{sd}+\frac{1+r}{q-(q-2)r}\sigma^2_\epsilon}$ |
| 5.3 | $\frac{n_1+n_2}{2n_1n_2}(\sigma^2_{sp}+\sigma^2_{sd}+\frac{1+r}{q-(q-2)r}\sigma^2_\epsilon)$ |
| 5.4 | $\frac{n_1+n_2}{2n_1n_2}[2(\sigma^2_{sd}+\frac{1+r}{q-(q-2)r}\sigma^2_\epsilon)]$ |
| 5.5 | $\frac{n_1+n_2}{2n_1n_2}[2(\sigma^2_{sd}+\frac{1+r}{q-(q-2)r}\sigma^2_\epsilon)]+\frac{n_1+n_2}{n_1n_2}\frac{\sum_{m=1}^{q-1}(q-2m)r^{(m+1)}\sigma^2_\epsilon}{q^2[q-(q-2)r]}$ |
| 5.6 | $\frac{n_1+n_2}{2n_1n_2}[(1-\beta_1)^2\sigma^2_{sp}+(1+\beta_1^2)(\sigma^2_{sd}+\frac{1+r}{q-(q-2)r}\sigma^2_\epsilon)]$ |
| 5.7 | $\frac{n_1+n_2}{2n_1n_2}[(1-\beta_2)^2\sigma^2_{sp}+(1+\beta_2^2)(\sigma^2_{sd}+\frac{1+r}{q-(q-2)r}\sigma^2_\epsilon)]+\frac{n_1+n_2}{n_1n_2}\frac{\sum_{m=1}^{q-1}(q-2m)r^{(m+1)}\sigma^2_\epsilon}{q^2[q-(q-2)r]}$ |

To compare the variances for different models, similar to the case of single measurements, we can write some of variances as a function of $\rho$ in Table 5.8. We also rewrite the variances that are more complicated and are not easily written as a function of $\rho$ in the similar forms in Table 4.2 by using $a_1 - a_4$.

Table 5.8 Variances of BLUEs for the Treatment Contrast

| Model | Independent | AR(1) |
|---|---|---|
| 5.2 | $\frac{n_1+n_2}{2n_1n_2}(1+\rho)(\sigma^2_{sd}+\sigma^2_\epsilon/q)$ | $\frac{n_1+n_2}{2n_1n_2}(1+\rho)(\sigma^2_{sd}+\frac{1+r}{q-(q-2)r}\sigma^2_\epsilon)$ |
| 5.3 | $\frac{n_1+n_2}{2n_1n_2}\frac{1}{1-\rho}(\sigma^2_{sd}+\sigma^2_\epsilon/q)$ | $\frac{n_1+n_2}{2n_1n_2}\frac{1}{1-\rho}(\sigma^2_{sd}+\frac{1+r}{q-(q-2)r}\sigma^2_\epsilon)$ |
| 5.4 | $\frac{n_1+n_2}{2n_1n_2}2(\sigma^2_{sd}+\sigma^2_\epsilon/q)$ | $\frac{n_1+n_2}{2n_1n_2}[2(\sigma^2_{sd}+\frac{1+r}{q-(q-2)r}\sigma^2_\epsilon)]$ |
| 5.5 | $\frac{n_1+n_2}{2n_1n_2}2(\sigma^2_{sd}+\sigma^2_\epsilon/q)$ | $\frac{n_1+n_2}{2n_1n_2}[2(\sigma^2_{sd}+\frac{1+r}{q-(q-2)r}\sigma^2_\epsilon)]+c_1$ |
| 5.6 | $\frac{n_1+n_2}{2n_1n_2}(1+a_1)(\sigma^2_{sd}+\sigma^2_\epsilon/q)$ | $\frac{n_1+n_2}{2n_1n_2}[(1+a_3)(\sigma^2_{sd}+\frac{1+r}{q-(q-2)r}\sigma^2_\epsilon)]$ |
| 5.7 | $\frac{n_1+n_2}{2n_1n_2}(1+a_2)(\sigma^2_{sd}+\sigma^2_\epsilon/q)$ | $\frac{n_1+n_2}{2n_1n_2}[(1+a_4)(\sigma^2_{sd}+\frac{1+r}{q-(q-2)r}\sigma^2_\epsilon)]+c_1$ |
| | $c_1=\frac{n_1+n_2}{n_1n_2}\frac{\sum_{m=1}^{q-1}(q-2m)r^{(m+1)}\sigma^2_\epsilon}{q^2[q-(q-2)r]}$ | |

In Table 5.8, $a_1$-$a_4$ are very complicated and depend on the values of the variance components. As seen from Tables 5.4 and 5.5, for using change from baseline measurements, the BLUEs of the treatment contrast for time-matched baselines and averaged baselines are the same when random errors are independently distributed; however, they are different when random errors follow an AR(1) structure. This is also true for the corresponding variances for these estimators. In addition, for the case of AR(1) random errors, compared to the variance from time-matched baselines, the variance from averaged baselines has the additional term $\frac{n_1+n_2}{n_1n_2}\frac{\sum_{m=1}^{q-1}(q-2m)r^{(m+1)}\sigma^2_\epsilon}{q^2[q-(q-2)r]}$. This additional term can be shown to be nonnegative if $r$ is nonnegative. This indicates that the variance using averaged baselines is larger than the variance using time-matched baselines. However, for using baseline measurements as a covariate, the results are more complicated because of different $\beta_1$ and $\beta_2$ in Models 5.6 and 5.7.

From Table 5.8, we see that the general patten of the variances is similar to the case of single measurements in Table 4.2. However, when random errors are independent, the parameter $\rho$ is $\sigma^2_{sp}/(\sigma^2_{sp}+\sigma^2_{sd}+\sigma^2_\epsilon/q)$ and $\sigma^2_\epsilon$ in Table 4.2 is now replaced by $\sigma^2_{sd}+\sigma^2_\epsilon/q$. When random errors follow an AR(1) structure, the parameter

$\rho$ is $\sigma^2_{sp}/(\sigma^2_{sp}+\sigma^2_{sd}+\frac{1+r}{q-(q-2)r}\sigma^2_\epsilon)$ and $\sigma^2_\epsilon$ in Table 4.2 is now replaced by $\sigma^2_{sd}+\frac{1+r}{q-(q-2)r}\sigma^2_\epsilon$. Furthermore, when random errors follows an AR(1) structure, the comparison is rather complicated because of the additional term for the averaged baselines, though this is relatively small. Moreover, the $a_i$'s in Table 5.8 are different and complicated. Therefore, in a repeated measurements crossover design, it is not possible to compare the variances of different methods simply based on the values of the parameter $\rho$.

### 5.5.2 Simulation study

In this section, we conduct a simulation study for an *AB/BA* crossover design with repeated measurements to evaluate the performance of the discussed methods for incorporating the baseline measurements.

As discussed in Section 5.2, we consider two different assumptions for the random error terms to generate the data. That is, we generate data based on Model 5.2 both for independent and autocorrelated random errors. As in the case of single measurements, we generate the data by setting $\mu = 100$ and taking other fixed effects equal to zero. We also set $\tau_A - \tau_B = 0$ and $\tau_A - \tau_B = 0.5$ for the evaluation of type I error rate and power, respectively, for testing $H0 : \tau_A = \tau_B$. The relative efficiencies depend on the relative value of $\sigma^2_s$, $\sigma^2_{sp}$, $\sigma^2_{st}$ and $\sigma^2_{sd}$ to $\sigma^2_\epsilon$, and we can set $\sigma^2_\epsilon = 1$. We let $\sigma^2_s$, $\sigma^2_{st}$ and $\sigma^2_{sd}$ all take the values 0.1, 1 and 10. Therefore, we have 27 combinations (see Table 5.9) for the variance components $\sigma^2_s$, $\sigma^2_{st}$, $\sigma^2_{sd}$ and $\sigma^2_\epsilon$. For $\sigma^2_{sp}$, we select each time a value so that the parameter $\rho$ is 0, 0.2, 0.4, 0.5, 0.6 and 0.8. If the random error follows an AR(1) structure, then we consider values for the autocorrelation coefficient $r$ of 0.2, 0.5 and 0.8. We also consider different time points $q$ of 3, 6 and 10. For each sequence we use 24 subjects.

Table 5.9 Values of $\sigma^2_\epsilon$, $\sigma^2_s$, $\sigma^2_{st}$ and $\sigma^2_{sd}$ for the Simulation Study

| Case | $\sigma^2_\epsilon$ | $\sigma^2_s$ | $\sigma^2_{st}$ | $\sigma^2_{sd}$ |
|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 10 | 1 | 1 |
| 3 | 1 | 0.1 | 1 | 1 |
| 4 | 1 | 1 | 10 | 1 |
| 5 | 1 | 10 | 10 | 1 |
| 6 | 1 | 0.1 | 10 | 1 |
| 7 | 1 | 1 | 0.1 | 1 |
| 8 | 1 | 10 | 0.1 | 1 |
| 9 | 1 | 0.1 | 0.1 | 1 |
| 10 | 1 | 1 | 1 | 10 |
| 11 | 1 | 10 | 1 | 10 |
| 12 | 1 | 0.1 | 1 | 10 |
| 13 | 1 | 1 | 10 | 10 |
| 14 | 1 | 10 | 10 | 10 |
| 15 | 1 | 0.1 | 10 | 10 |
| 16 | 1 | 1 | 0.1 | 10 |
| 17 | 1 | 10 | 0.1 | 10 |
| 18 | 1 | 0.1 | 0.1 | 10 |
| 19 | 1 | 1 | 1 | 0.1 |
| 20 | 1 | 10 | 1 | 0.1 |
| 21 | 1 | 0.1 | 1 | 0.1 |
| 22 | 1 | 1 | 10 | 0.1 |
| 23 | 1 | 10 | 10 | 0.1 |
| 24 | 1 | 0.1 | 10 | 0.1 |
| 25 | 1 | 1 | 0.1 | 0.1 |
| 26 | 1 | 10 | 0.1 | 0.1 |
| 27 | 1 | 0.1 | 0.1 | 0.1 |

Each time the simulated data are analyzed by using PROC MIXED in SAS for each of the methods discussed in Section 5.2. With 24 subjects in each sequence, we do not use any degree freedom adjustment. Due to the complication of the variance-covariance structure, there is a convergence problem. More than 5000 times are

simulated for each scenario. The results for the first 5000 times in which convergence occurs for all methods are used as the simulation results.

The results of REs for different methods does not depend on the value of $\tau_A - \tau_B$, therefore, we do not present the results from the case of $\tau_A - \tau_B$ here. Table 5.10 presents the results of Case 1 in Table 5.9 for independent random errors when $q = 6$. The variables in each column have the same interpretation as they do in the case of single measurements. Recall from Chapter 4, SE_est is the mean of the estimated standard errors for $\hat{\tau}_A - \hat{\tau}_B$, SE_rediff is the relative difference between SE_est and SE_MC and is computed as $\frac{\text{SE\_est - SE\_MC}}{\text{SE\_MC}} \times 100\%$.

Table 5.10 Results of Case 1 for Independent Random Errors when $q = 6$

| $\rho$ | Methods | $\hat{\tau}_A - \hat{\tau}_B$ | SE_est | SE_rediff | RE | Type I error |
|---|---|---|---|---|---|---|
| 0 | Baselines in response vector | -0.0010 | 0.2211 | -0.25 | 99.34 | 4.78 |
| | Ignore baselines | -0.0005 | 0.2196 | -0.40 | 100 | 5.34 |
| | Change from baselines (T-match) | 0.0018 | 0.3107 | -0.75 | 70.67 | 5.52 |
| | Change from baselines (Average) | 0.0018 | 0.3104 | -0.84 | 70.74 | 5.78 |
| | Baselines as a covariate (T-match) | 0.0005 | 0.2388 | -1.17 | 91.97 | 5.62 |
| | Baselines as a covariate (Average) | 0.0002 | 0.2412 | -1.00 | 91.05 | 5.42 |
| 0.2 | Baselines in response vector | -0.0016 | 0.2393 | -2.33 | 100 | 6.10 |
| | Ignore baselines | -0.0017 | 0.2456 | -0.73 | 97.42 | 5.68 |
| | Change from baselines (T-match) | -0.0032 | 0.3108 | -1.73 | 76.99 | 5.68 |
| | Change from baselines (Average) | -0.0032 | 0.3100 | -2.00 | 77.2 | 5.88 |
| | Baselines as a covariate (T-match) | -0.0025 | 0.2473 | -2.33 | 96.76 | 6.12 |
| | Baselines as a covariate (Average) | -0.0024 | 0.2544 | -2.04 | 94.07 | 6.20 |
| 0.4 | Baselines in response vector | 0.0018 | 0.2584 | -1.60 | 100 | 5.78 |
| | Ignore baselines | 0.0047 | 0.2830 | -1.28 | 91.33 | 5.76 |
| | Change from baselines (T-match) | -0.0006 | 0.3106 | 0.69 | 83.19 | 5.04 |
| | Change from baselines (Average) | -0.0006 | 0.3095 | 0.34 | 83.49 | 5.36 |
| | Baselines as a covariate (T-match) | 0.0020 | 0.2585 | -1.11 | 99.97 | 5.66 |
| | Baselines as a covariate (Average) | 0.0009 | 0.2670 | -0.26 | 96.80 | 5.38 |
| 0.5 | Baselines in response vector | -0.001 | 0.2678 | -0.74 | 99.47 | 5.12 |
| | Ignore baselines | -0.0041 | 0.3091 | 0.04 | 86.16 | 5.70 |
| | Change from baselines (T-match) | 0.0024 | 0.3110 | 0.18 | 85.64 | 5.00 |
| | Change from baselines (Average) | 0.0024 | 0.3103 | -0.06 | 85.84 | 5.12 |
| | Baselines as a covariate (T-match) | -0.0012 | 0.2663 | -0.93 | 100 | 5.78 |
| | Baselines as a covariate (Average) | -0.0002 | 0.2740 | 0.10 | 97.20 | 5.34 |
| 0.6 | Baselines in response vector | 0.0030 | 0.2766 | -0.49 | 99.90 | 5.36 |
| | Ignore baselines | 0.0033 | 0.3457 | -0.25 | 79.92 | 5.62 |
| | Change from baselines (T-match) | 0.0024 | 0.3107 | -0.29 | 88.93 | 5.20 |
| | Change from baselines (Average) | 0.0024 | 0.3097 | -0.62 | 89.22 | 5.54 |
| | Baselines as a covariate (T-match) | 0.0029 | 0.2763 | -0.90 | 100 | 5.38 |
| | Baselines as a covariate (Average) | 0.0029 | 0.2806 | -0.22 | 98.45 | 5.44 |
| 0.8 | Baselines in response vector | -0.0012 | 0.2939 | -2.01 | 100 | 6.02 |
| | Ignore baselines | -0.0030 | 0.4866 | -2.15 | 60.40 | 5.64 |
| | Change from baselines (T-match) | -0.0005 | 0.3107 | -1.15 | 94.58 | 5.48 |
| | Change from baselines (Average) | -0.0005 | 0.3097 | -1.48 | 94.91 | 5.72 |
| | Baselines as a covariate (T-match) | -0.0019 | 0.3100 | -3.65 | 94.82 | 6.48 |
| | Baselines as a covariate (Average) | -0.0007 | 0.2956 | -1.61 | 99.44 | 6.02 |

From Table 5.10, we can see that the averages of the estimates of the treatment contrast are close to the true value of 0. Small values of SE_rediff indicates that SE_est is close to SE_MC, and the type I error rate maintains the nominal level of 5%. The estimates of the treatment contrast are the same for time-matched and averaged baselines when using change from baseline measurements as the response variable, as anticipated from Table 5.4.

We display the relative efficiencies for the different methods at each combination of the variance components in Figures 5.3-5.5.

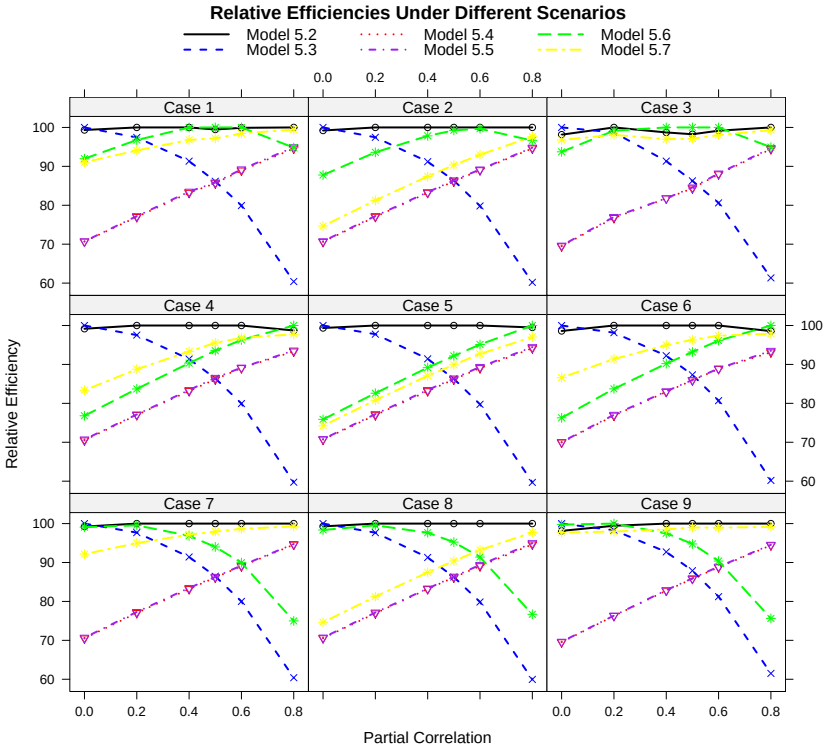


Figure 5.3: Relative Efficiencies for Each Method at Different Values of the Parameter $\rho$ for the First 9 Cases ($\sigma^2_{sd}/\sigma^2_{\epsilon} = 1$) in Table 5.9 when $q = 6$ and Independent Random Errors

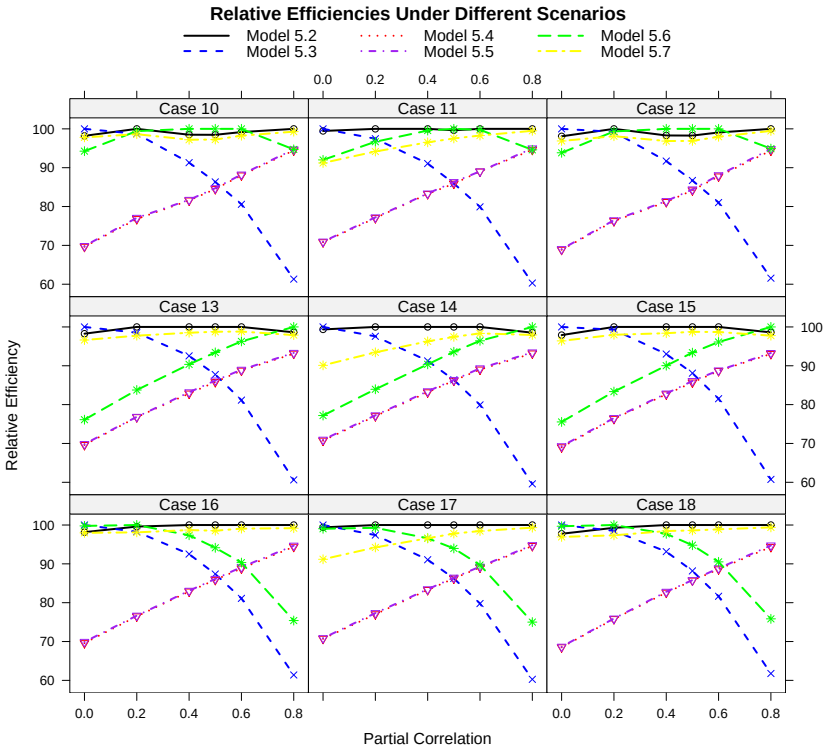


Figure 5.4: Relative Efficiencies for Each Method at Different Values of the Parameter $\rho$ for the Second 9 Cases ($\sigma^2_{sd}/\sigma^2_{\epsilon} = 10$) in Table 5.9 when $q = 6$ and Independent Random Errors

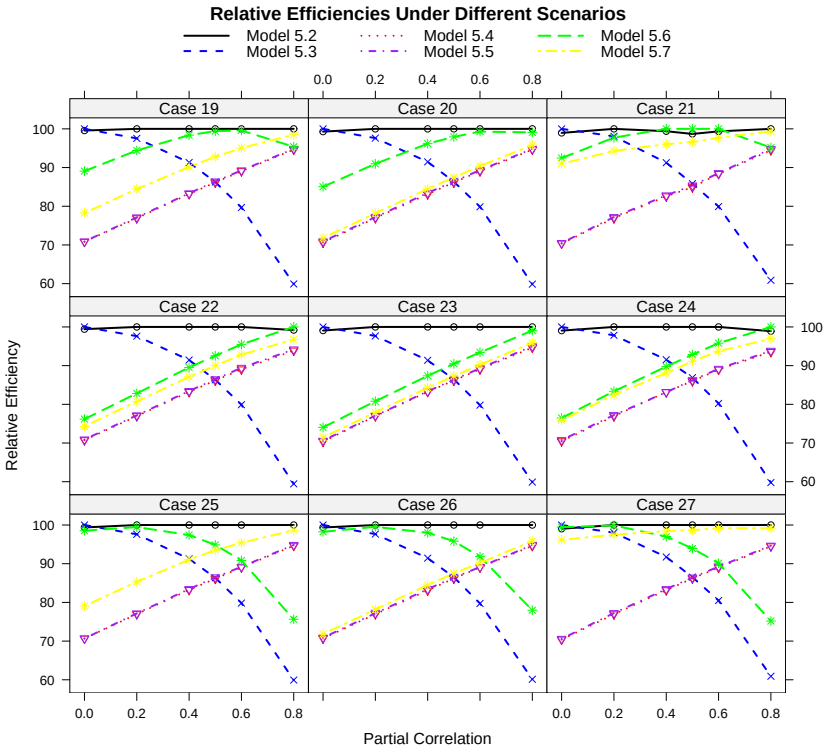


Figure 5.5: Relative Efficiencies for Each Method at Different Values of the Parameter $\rho$ for the Last 9 Cases ($\sigma^2_{sd}/\sigma^2_{\epsilon} = 0.1$) in Table 5.9 when $q = 6$ and Independent Random Errors

From Figures 5.3-5.5, we can see that the relative efficiencies of Models 5.2-5.5 are similar for all 27 choices of the variance components, reflecting that they are only a function of $\rho$. The general pattern for these efficiencies is also similar to what for the case of single measurements, regardless of the different values for the variance components. It is virtually impossible to recognize the plot for Model 5.4, since it is coincides with that for Model 5.5. The same performance of time-matched (Model 5.4) and averaged baselines (Model 5.5) for using change from baselines as the response variable is as expected. However, the change of REs is complicated for using baseline measurements as a covariate: time-matched baselines and averaged baselines demonstrated differently not only at different values for the parameter $\rho$, but also for different ratios of the variance components. For example, the RE of using time-matched baselines as a covariate (Model 5.6) performs similarly for different values of $\sigma^2_{sd}/\sigma^2_\epsilon$ and $\sigma^2_{st}/\sigma^2_\epsilon$, but demonstrates differently for different values of $\sigma^2_s/\sigma^2_\epsilon$. The RE does not change very much at different values of the parameter $\rho$ when $\sigma^2_s/\sigma^2_\epsilon = 1$, but increases significantly as $\rho$ increases when $\sigma^2_s/\sigma^2_\epsilon = 10$, and decreases significantly as $\rho$ increases when $\sigma^2_s/\sigma^2_\epsilon = 0.1$. In addition, the RE at small and larger value of $\rho$ could be relatively small for $\sigma^2_s/\sigma^2_\epsilon = 10$ and $\sigma^2_s/\sigma^2_\epsilon = 0.1$, respectively. However, the RE of using averaged baselines as a covariate (Model 5.7) increases as $\rho$ increases, and it is relatively larger when $\sigma^2_{sd}/\sigma^2_\epsilon = 10$. When $\sigma^2_{sd}/\sigma^2_\epsilon = 1$ and $\sigma^2_{sd}/\sigma^2_\epsilon = 0.1$, the RE of using averaged baselines increases dramatically as $\rho$ increases except for Cases 3, 9, 21 and 27.

To see the effect of a different number of time points $q$, we display the relative efficiencies for Cases 1, 15 and 27 in Table 5.9 at $q$ =3, 6 and 10 in Figure 5.6.

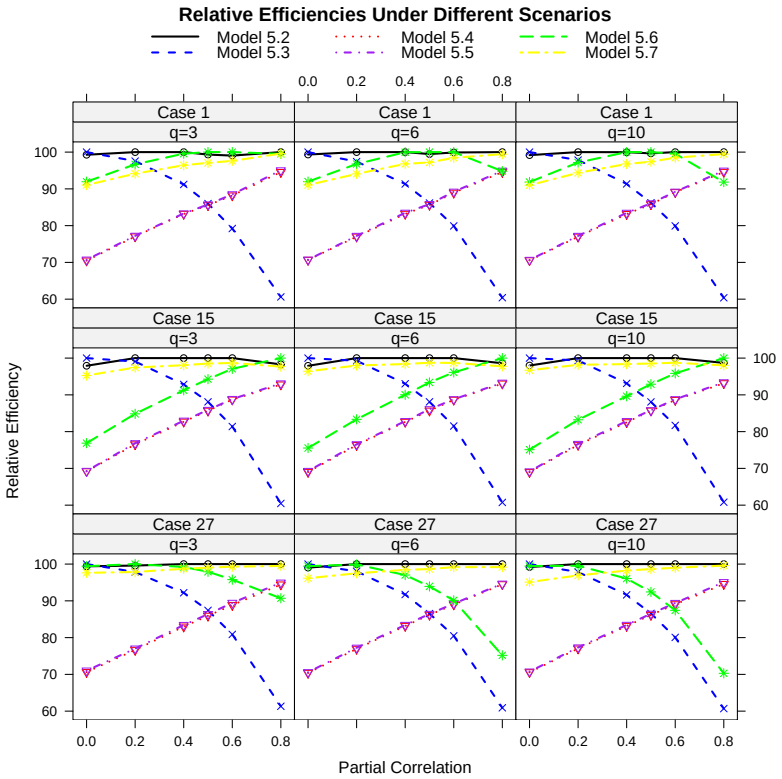


Figure 5.6: Relative Efficiencies for Each Method at Different Values for the Parameter $\rho$ and with Different Number of Time Points for Cases 1, 15 and 27 in Table 5.9

From Figure 5.6, for these particular choices of the random components, we can see that the relative efficiencies are similar for various methods at different time points except when using time-matched baselines as a covariate. The RE for using time-matched baseline measurements as a covariate decreases as the number of time points increases when $\rho$ is larger for Cases 1 and 27. There is almost no difference for Case 15.

Next, we present the results of a simulation study for the case that random errors follow an AR(1) structure. First, we focus on $q = 6$ and $r = 0.8$ for the 27 combinations in Table 5.9. Then we study at different values for $q$ and $r$ for one particular choice. Table 5.11 presents the results for Case 1 in Table 5.9 when $q = 6$ and $r = 0.8$.

Table 5.11 Results for Case 1 in Table 5.9 for AR(1) Random Errors ($q = 6$, $r = 0.8$ )

| $\rho$ | Methods | $\hat{\tau}_A - \hat{\tau}_B$ | SE_est | SE_rediff | RE | Type I error |
|---|---|---|---|---|---|---|
| 0.0 | Baselines in response vector | -0.0012 | 0.2625 | -0.03 | 99.42 | 5.02 |
| | Ignore baselines | -0.0009 | 0.2610 | -0.09 | 100 | 5.86 |
| | Change from baselines (T-match) | -0.0007 | 0.3692 | 0.11 | 70.68 | 4.80 |
| | Change from baselines (Average) | 0.0000 | 0.3719 | 0.17 | 70.17 | 5.08 |
| | Baselines as a covariate (T-match) | -0.0008 | 0.3193 | -2.45 | 81.72 | 5.50 |
| | Baselines as a covariate (Average) | -0.0006 | 0.2785 | -0.38 | 93.71 | 5.34 |
| 0.2 | Baselines in response vector | -0.0004 | 0.2838 | -0.90 | 100 | 4.86 |
| | Ignore baselines | -0.0009 | 0.2907 | 0.40 | 97.64 | 4.94 |
| | Change from baselines (T-match) | -0.0040 | 0.3689 | -1.11 | 76.94 | 5.26 |
| | Change from baselines (Average) | -0.0049 | 0.3713 | -1.01 | 76.44 | 5.38 |
| | Baselines as a covariate (T-match) | -0.0032 | 0.3208 | -3.31 | 88.47 | 5.54 |
| | Baselines as a covariate (Average) | -0.0027 | 0.2968 | -0.59 | 95.62 | 5.40 |
| 0.4 | Baselines in response vector | 0.0045 | 0.3069 | -3.30 | 100 | 6.14 |
| | Ignore baselines | 0.0044 | 0.3355 | -2.23 | 91.46 | 6.10 |
| | Change from baselines (T-match) | 0.0043 | 0.3691 | -2.27 | 83.13 | 5.64 |
| | Change from baselines (Average) | 0.0047 | 0.3712 | -2.38 | 82.67 | 5.72 |
| | Baselines as a covariate (T-match) | 0.0039 | 0.3241 | -4.70 | 94.70 | 6.42 |
| | Baselines as a covariate (Average) | 0.0043 | 0.3155 | -2.46 | 97.26 | 5.94 |
| 0.5 | Baselines in response vector | 0.0002 | 0.3183 | -1.45 | 100 | 5.52 |
| | Ignore baselines | 0.0029 | 0.3684 | -0.87 | 86.41 | 5.72 |
| | Change from baselines (T-match) | -0.0013 | 0.3695 | 0.09 | 86.14 | 5.12 |
| | Change from baselines (Average) | -0.0014 | 0.3723 | 0.13 | 85.5 | 5.56 |
| | Baselines as a covariate (T-match) | -0.0004 | 0.3273 | -2.65 | 97.25 | 6.12 |
| | Baselines as a covariate (Average) | -0.0004 | 0.3255 | -0.49 | 97.78 | 5.74 |
| 0.6 | Baselines in response vector | 0.0022 | 0.3284 | -0.23 | 100 | 5.52 |
| | Ignore baselines | -0.0045 | 0.4105 | -1.25 | 80.00 | 5.68 |
| | Change from baselines (T-match) | 0.0063 | 0.3689 | 0.99 | 89.03 | 4.94 |
| | Change from baselines (Average) | 0.0061 | 0.3706 | 0.41 | 88.61 | 5.50 |
| | Baselines as a covariate (T-match) | 0.0036 | 0.3292 | -1.69 | 99.76 | 5.92 |
| | Baselines as a covariate (Average) | 0.0031 | 0.3338 | 0.04 | 98.37 | 5.98 |
| 0.8 | Baselines in response vector | -0.0034 | 0.3492 | -0.19 | 97.76 | 4.98 |
| | Ignore baselines | 0.0019 | 0.5756 | -1.29 | 59.31 | 5.58 |
| | Change from baselines (T-match) | -0.0052 | 0.3692 | 0.36 | 92.45 | 5.02 |
| | Change from baselines (Average) | -0.0043 | 0.3717 | 0.24 | 91.83 | 4.94 |
| | Baselines as a covariate (T-match) | -0.0034 | 0.3414 | -2.25 | 100 | 5.74 |
| | Baselines as a covariate (Average) | -0.0028 | 0.3537 | 0.41 | 96.51 | 5.00 |

As anticipated from Table 5.5, the estimates of the treatment contrast from time-matched and averaged baselines are different both for using change from baselines as the response variable and for using baseline measurements as a covariate. When using change from baselines as the response variable, SE_est of time-matched baselines is slightly smaller than averaged baselines.

We present the relative efficiencies of the different methods at each combination of the variance components when $q = 6$ and $r = 0.8$ in Figures 5.7-5.9.

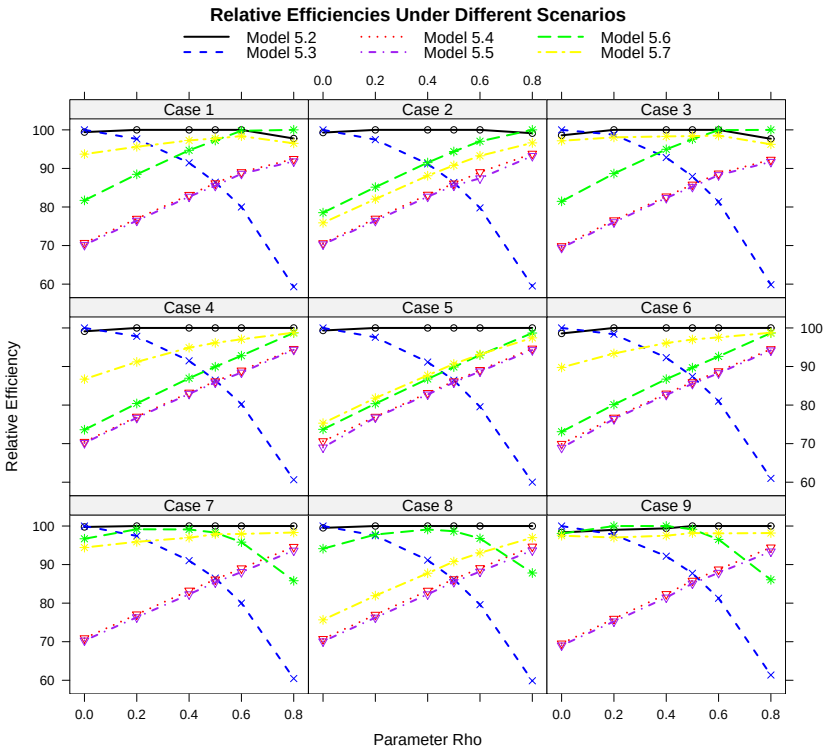


Figure 5.7: Relative Efficiencies at Different Values of the Parameter $\rho$ for the First 9 Cases ($\sigma^2_{sd}/\sigma^2_{\epsilon} = 1$) in Table 5.9 of the Variance Components when $q = 6$ and AR(1) Random Errors with $r = 0.8$

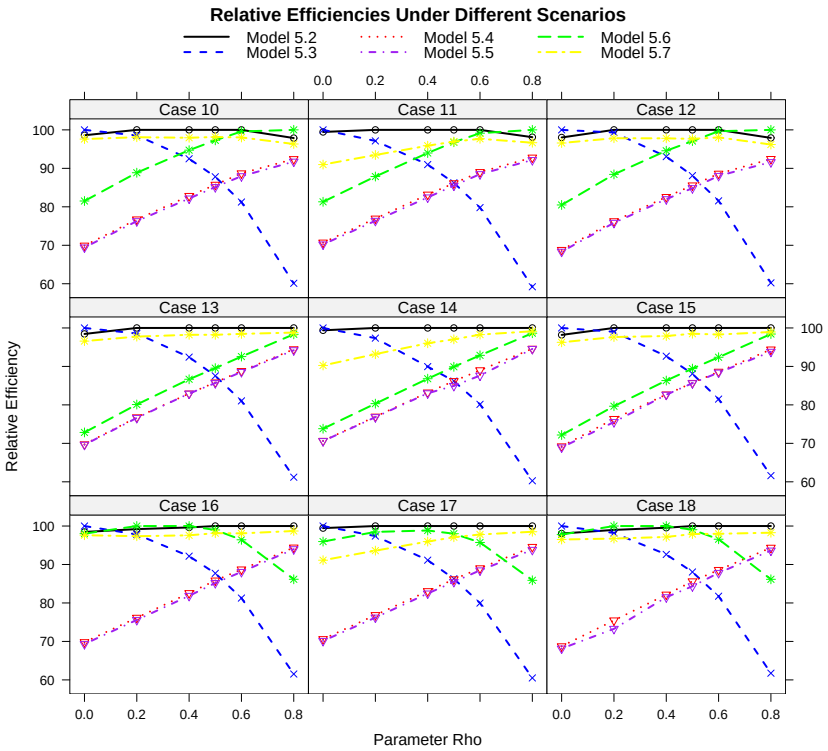


Figure 5.8: Relative Efficiencies at Different Values of the Parameter $\rho$ for the Second 9 Cases ($\sigma^2_{sd}/\sigma^2_{\epsilon} = 10$) in Table 5.9 of the Variance Components when $q = 6$ and AR(1) Random Errors with $r = 0.8$

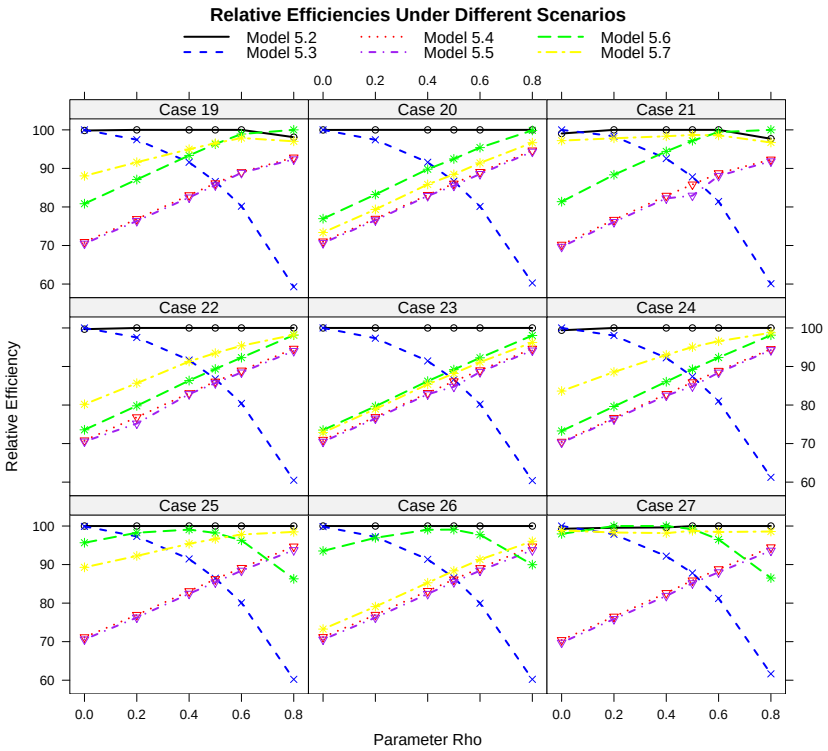


Figure 5.9: Relative Efficiencies at Different Values of the Parameter $\rho$ for the Last 9 Cases ($\sigma^2_{sd}/\sigma^2_{\epsilon} = 0.1$) in Table 5.9 of the Variance Components when $q = 6$ and AR(1) Random Errors with $r = 0.8$

From the simulation results, when using change from baseline measurements as a response variable, SE_est of time-matched baselines is slightly smaller than averaged baselines. However, from Figures 5.7-5.9, we can only recognize the difference for some cases. That is because that the additional term $c_1$ is too small to be detected in the figures. This results are confirmed by our theoretical results. For using baseline measurements as a covariate, the comparison is complicated as already seen for the case of independent random errors, and the pattern of the REs at different values of variance components is similar to the pattern in the case of independent random errors except when $\sigma^2_{st}/\sigma^2_{\epsilon} = 1$, where the RE of using time-matched baselines also increases as $\rho$ increases.

Next, we present the relative efficiencies at different values for the number of time points $q$ for Cases 1 and 27 in Table 5.9 when $r = 0.8$ in Figure 5.10.

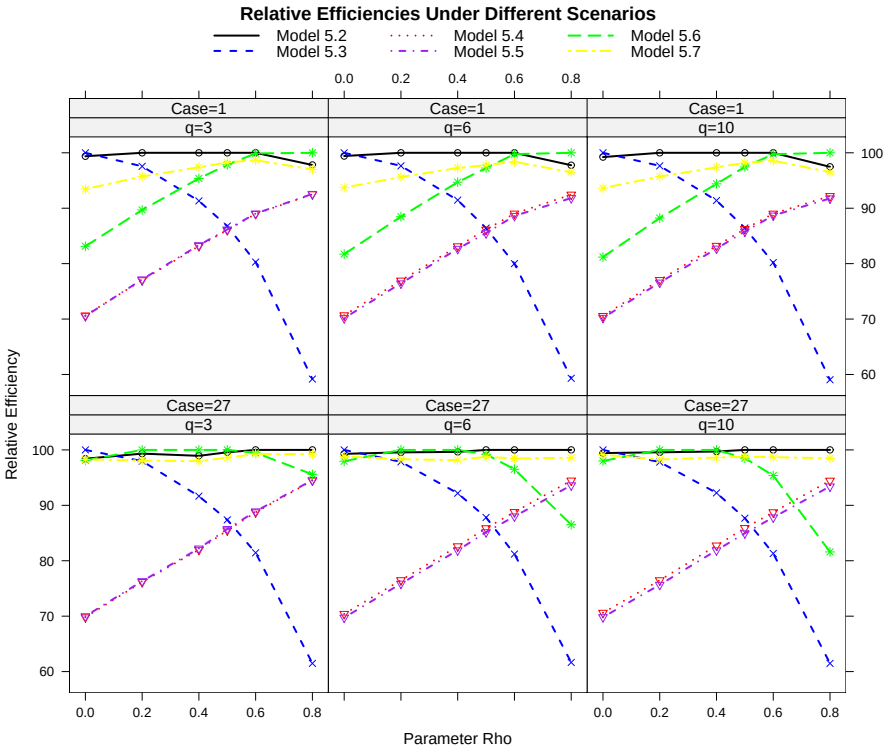


Figure 5.10: Relative Efficiencies for Each Method at Different Values of the Parameter $\rho$ at Different Values for the Number of Time Points

From Figure 5.10, we cannot see much difference for the general pattern of RE at different time points for Case 1. However, for Case 27, the RE of using averaged baselines as a covariate decreases as time points increases at larger values of $\rho$. For the method of using change from baseline measurements as a response variable, time-matched baselines performance slightly better than averaged baselines, especially when the number of the time points is larger.

Figure 5.11 shows the effect of different autocorrelation coefficients on the RE. The number of time points and the values of the variance components were fixed at $q = 6$ and the combination for Cases 1 and 27 in Table 5.9.

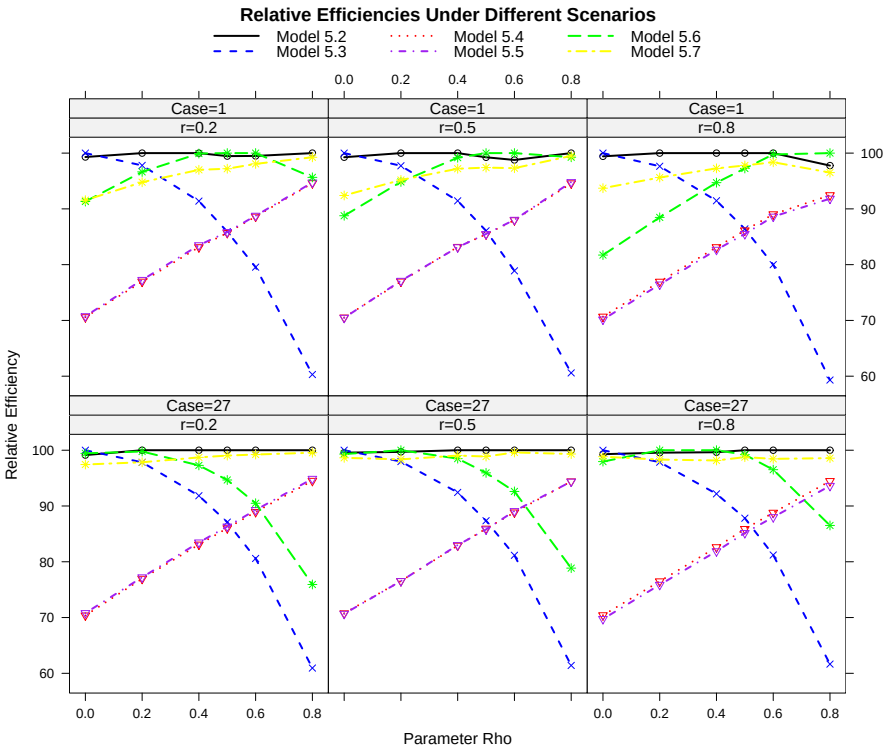


Figure 5.11: Relative Efficiencies for Each Method at Different Values of the Parameter $\rho$ for Different Autocorrelation Coefficients

From Figure 5.11, there is no much difference for the general pattern of RE at different autocorrelation coefficients $r$ except for using time-matched baselines as a

covariate. The RE of time-matched baselines increases as $r$ increases at large value of $\rho$, and decreases as $r$ increases at small value of $\rho$ for Cases 1 and 27, respectively.

### 5.5.3 Numerical examples

We return to the numerical example from Study I that we introduced in Section 5.3. From the results in Section 5.3, the treatment by time interaction is not significant, so we will compare the discussed methods for incorporating the baseline measurements.

The results in Table 5.2 suggested that independent variance-covariance structure is the best for the random errors. From the data analysis based on Model 5.2 with independent variance-covariance structure, we obtained $\hat{\sigma}_s^2 = 0.02235$, $\hat{\sigma}_{sp}^2 = 0.00031$, $\hat{\sigma}_{st}^2 = 0.00033$, $\hat{\sigma}_{sd}^2 = 0.00085$ and $\hat{\sigma_\epsilon}^2 = 0.01060$. So the parameter $\rho = 0.00031/(0.00031 + 0.00086 + 0.01060/3) = 0.0659$, which is close to zero. We would thus expect that retaining the baseline measurements as part of the response vector and ignoring baseline measurements would have high relative efficiencies. We present the results for the different methods in Table 5.12.

Table 5.12 Analysis Results for the Example with Repeated Measurements

| | | | | 95%CI | | |
|---|---|---|---|---|---|---|
| Methods | Estimate | SE | RE | Lower | Upper | P-value |
| Baselines in response vector | 0.0629 | 0.0298 | 93.13 | 0.0020 | 0.1238 | 0.0433 |
| Ignore baselines | 0.0618 | 0.0277 | 100.00 | -0.0000 | 0.1236 | 0.0500 |
| Change from baselines (T-match) | 0.0787 | 0.0385 | 72.04 | -0.0014 | 0.1588 | 0.0537 |
| Change from baselines (Average) | 0.0787 | 0.0359 | 77.19 | -0.0014 | 0.1588 | 0.0533 |
| Baselines as a covariate (T-match) | 0.0678 | 0.0284 | 97.76 | 0.0045 | 0.1312 | 0.0383 |
| Baselines as a covariate (Average) | 0.0758 | 0.0334 | 82.96 | 0.0005 | 0.1511 | 0.0487 |

From Table 5.12, the RE when ignoring baseline measurements is 100 %, while the REs for retaining baseline measurements as part of the response vector and for time-matched baselines as a covariate are also high. However, the RE for using

change from baseline measurements is relatively low for either type of baselines. The results indicate that we should either ignore the baseline measurements or use time-matched baselines as a covariate, or retain baseline measurements as part of the response vector. For this example, the number of time points is 3, which is relative small; the parameter $\rho$ is also relative small. For the variance components, the estimates of the ratio of $\sigma_s^2$, $\sigma_{st}^2$ and $\sigma_{sd}^2$ to $\sigma_\epsilon^2$ are around 2, 0.03, and 0.08 respectively. This is close to Case 25 in Table 5.9, and the RE for Case 25 in Figure 5.5 indicates that the time-matched baselines as a covariate has high RE at small value of $\rho$. So the results of RE in this study are confirmed from our simulation results.

## 5.6 Discussion, conclusion and recommendation

In this chapter, we extended our study in Chapter 4 to accommodate repeated measurements within each period both for the baseline measurements and the response variable. We introduced a preliminary test for the treatment by time interaction, and considered different analyses based on whether this interaction term is statistically significant. When the interaction term is significant, we proposed a graphical method to study the change in treatment differences over time. Alternatively, if the interaction term is not significant, we studied different methods to incorporate baseline measurements for estimating the treatment contrast. We considered two assumptions for the random error terms.

In general, similar to the case of single measurements, we find that retaining baseline measurements as part of the response vector has the largest relative efficiency among all the methods as long as the parameter $\rho$ is not vary small. When the parameter $\rho$ is close to zero, ignoring baseline measurements has a larger relative efficiency. However, when the parameter $\rho$ increases, the relative efficiency for ignoring

baseline measurements decreases dramatically. When using change from baseline measurements as the response variable, time-matched baselines perform better than averaged baselines for an AR(1) random error structure, especially when the number of time points is large and the autocorrelation coefficient is large. When using baseline measurements as a covariate, the comparison is complicated, since it depends on the estimator of $\beta$, which in turn depends on $\sigma_s^2$, $\sigma_{st}^2$ and $\sigma_{sd}^2$. Using baselines as a covariate can be more or less efficient than other methods depending on the variance components. Regardless of whether time-matched or averaged baselines are used, the RE for using change from baselines can be clearly less than for retaining baselines in the response vector, and is never much more.

In general, if possible, we would suggest to collect the baseline measurements, and retain them as part of the response vector to do the analysis. Using change from baseline measurements should be avoided.

## Appendix

We now present details for a proof of the results in Table 5.5.

**Theorem 5.1** Under the Aitken model $E(Y) = X\beta, Var(Y) = \sigma^2 V$, where $V$ is p.d., $a^T Y$ is the BLUE of its expected value (i.e., of $a^T X\beta$) if and only if $Cov(a^T Y, l^T Y) = 0$ for all vectors $l$ for which $l^T Y$ is an unbiased estimator of 0 (i.e. $E(l^T Y) = 0$).

**Theorem 5.2** For Model 5.2, under the assumptions that the random error terms follow an AR(1) process, variance components and autocorrelation coefficient $r$ are known and all the time points are equally spaced, the BLUE of $\tau_A - \tau_B$ is $\frac{1}{2}\{\sum_{m=1}^{q} \phi_m[\bar{Y}_{11.m} - \bar{Y}_{21.m} - \bar{Y}_{12.m} + \bar{Y}_{22.m} - \rho(\bar{X}_{11.m} - \bar{X}_{21.m} - \bar{X}_{12.m} + \bar{X}_{22.m})]\}$, where

$$\phi_m = \begin{cases} \frac{1}{q-(q-2)r} & \text{if } m = 1,\ q \\ \frac{1-r}{q-(q-2)r} & \text{if } m = 2, \cdots, q-1 \end{cases} \qquad \text{and } \rho = \frac{\sigma_{sp}^2}{\sigma_{sp}^2 + \sigma_{sd}^2 + \frac{1+r}{q-(q-2)r}\sigma_\epsilon^2}.$$

**Proof**: When random error terms follow an AR(1) process, Model 5.2 can be written in matrix notation as follows:

$$Y = \mu 1_{4nq} + X_1\pi + X_2D + X_3T + X_d\tau + Z_0s + Z_1\zeta + Z_2\xi + Z_3\omega + \epsilon,$$

where

$$X_1 = 1_n \otimes I_2 \otimes 1_{2q}, \qquad X_2 = 1_n \otimes I_4 \otimes 1_q, \qquad X_3 = 1_n \otimes 1_4 \otimes I_q,$$

$$X_d = (X_{d11}^T, X_{d12}^T, \cdots X_{d2n_2}^T)^T \otimes 1_q,\ X_{d1k} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}^T \text{ and } X_{d2k} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}^T;$$

$$Z_0 = I_n \otimes 1_{4q}, \quad Z_1 = I_n \otimes I_2 \otimes 1_{2q}, \quad Z_2 = I_n \otimes I_4 \otimes 1_q, \quad Z_3 = I_n \otimes 1_4 \otimes I_q$$

and $s$, $\zeta$, $\xi$ and $\omega$ are independently normally distributed with their elements being independently normally distributed with mean 0 and variances $\sigma_s^2$, $\sigma_{sp}^2$, $\sigma_{sd}^2$ and $\sigma_{st}^2$ respectively, $Var(\epsilon) = \sigma_\epsilon^2(I_n \otimes I_4 \otimes A)$. Here the $q \times q$ matrix $A$ is given by $A = \begin{pmatrix} 1 & r & \cdots & r^{q-1} \\ r & 1 & \cdots & r^{q-2} \\ \vdots & \vdots & \ddots & \vdots \\ r^{q-1} & r^{q-2} & \cdots & 1 \end{pmatrix}$.

Let the proposed estimator of $\tau_A - \tau_B$ in the statement of Theorem 5.2 be $a^TY$ with $a^T = (a_{11}^T, \cdots, a_{1n_1}^T, a_{21}^T, \cdots, a_{2n_2}^T)$, where $a_{jk}^T = (a_{01jk}^T, a_{11jk}^T, a_{02jk}^T, a_{12jk}^T)$. Thus $a_{hijk}$ is the $q \times 1$ vector $(a_{hijk1}, \cdots, a_{hijkq})^T$ with $a_{hijkm} = \frac{1}{2}\lambda(-1)^{(i+j)}\frac{1}{n_j}\phi_m$ where $\lambda = 1$ for $h = 1$ and $\lambda = -\rho$ for $h = 0$. It is easy to verify that $a^TY$ is a linear unbiased estimator for $\tau_A - \tau_B$. To prove that $a^TY$ is also the BLUE of $\tau_A - \tau_B$, by Theorem 5.1 we need to show that $Cov(a^TY, l^TY) = 0$ for all $l$ such that $E(l^TY) = 0$.

Let $l^T = (l_{11}^T, \cdots, l_{1n_1}^T, l_{21}^T, \cdots, l_{2n_2}^T)$, where $l_{jk}^T = (l_{01jk}^T, l_{11jk}^T, l_{02jk}^T, l_{12jk}^T)$ and $l_{hijk}$ is a $q \times 1$ vector. Based on Model 5.2, we have that $E(l^TY) = 0$ if and only if $l^TX_2 = 0^T$, $l^TX_3 = 0^T$ and $l^TX_d = 0^T$, i.e., we have

$$
\begin{aligned}
l^T X_2 &= l^T(1_n \otimes I_4 \otimes 1_q) \\
&= \sum_{j=1}^{2}\sum_{k=1}^{n_j} l_{jk}^T (I_4 \otimes 1_q) \\
&= \sum_{j=1}^{2}\sum_{k=1}^{n_j} (l_{01jk}^T 1_q, l_{11jk}^T 1_q, l_{02jk}^T 1_q, l_{12jk}^T 1_q) \\
&= 0^T, \qquad (5.8)
\end{aligned}
$$

$$
\begin{aligned}
l^T X_3 &= l^T(1_n \otimes 1_4 \otimes I_q) \\
&= \sum_{j=1}^{2}\sum_{k=1}^{n_j} l_{jk}^T (1_4 \otimes I_q) \\
&= \sum_{j=1}^{2}\sum_{k=1}^{n_j}\sum_{h=0}^{1}\sum_{i=1}^{2} l_{hijk}^T \\
&= 0^T \qquad (5.9)
\end{aligned}
$$

$$
\begin{aligned}
l^T X_d &= l^T(X_{d11}^T, X_{d12}^T, \cdots X_{d2n_2}^T)^T \otimes 1_q \\
&= \sum_{j=1}^{2}\sum_{k=1}^{n_j} l_{jk}^T (X_{djk} \otimes 1_q) \\
&= \sum_{k=1}^{n_1} l_{1k}^T \left( \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}^T \otimes 1_q \right) + \sum_{k=1}^{n_2} l_{2k}^T \left( \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}^T \otimes 1_q \right) \\
&= \left(\sum_{k=1}^{n_1} l_{111k}^T 1_q + \sum_{k=1}^{n_2} l_{122k}^T 1_q, \sum_{k=1}^{n_1} l_{121k}^T 1_q + \sum_{k=1}^{n_2} l_{112k}^T 1_q\right) \\
&= 0^T. \qquad (5.10)
\end{aligned}
$$

Based on the model assumptions for the random components and using that subjects are independent to each other, we have

$$
\begin{aligned}
&Cov(a^TY, l^TY)\\
&= a^TVar(Y)l\\
&= \sigma_s^2(a^TZ_0Z_0^Tl) + \sigma_{sp}^2(a^TZ_1Z_1^Tl) + \sigma_{sd}^2(a^TZ_2Z_2^Tl) + \sigma_{st}^2(a^TZ_3Z_3^Tl) + \sigma_\epsilon^2[a^T(I_n\otimes I_4\otimes A)l]\\
&= \sigma_s^2[a^T(I_n\otimes J_{4q})l] + \sigma_{sp}^2[a^T(I_n\otimes I_2\otimes J_{2q})l] + \sigma_{sd}^2[a^T(I_n\otimes I_4\otimes J_q)l]\\
&\quad +\sigma_{st}^2[a^T(I_n\otimes J_4\otimes I_q)l] + \sigma_\epsilon^2[a^T(I_n\otimes I_4\otimes A)l]\\
&= \textstyle\sum_{j=1}^2\sum_{k=1}^{n_j}\{\sigma_s^2(a_{jk}^TJ_{4q}l_{jk}) + \sigma_{sp}^2[a_{jk}^T(I_2\otimes J_{2q})l_{jk}] + \sigma_{sd}^2[a_{jk}^T(I_4\otimes J_q)l_{jk}]\\
&\quad +\sigma_{st}^2[a_{jk}^T(J_4\otimes I_q)l_{jk}] + \sigma_\epsilon^2[a_{jk}^T(I_4\otimes A)l_{jk}]\}.
\end{aligned}
$$

Because $a_{01jkm} = -a_{02jkm}$ and $a_{11jkm} = -a_{12jkm}$, it is easy to obtain that $\sum_{j=1}^2\sum_{k=1}^{n_j}\sigma_s^2(a_{jk}^TJ_{4q}l_{jk}) = 0$ and $\sum_{j=1}^2\sum_{k=1}^{n_j}\sigma_{st}^2[a_{jk}^T(J_4\otimes I_q)l_{jk}] = 0$. Furthermore, we have

$$
\begin{aligned}
&\textstyle\sum_{j=1}^2\sum_{k=1}^{n_j}[a_{jk}^T(I_2\otimes J_{2q})l_{jk}]\\
=\ &\textstyle\sum_{j=1}^2\sum_{k=1}^{n_j}[(a_{01jk}^T + a_{11jk}^T)J_q(l_{01jk}+l_{11jk}) + (a_{02jk}^T + a_{12jk}^T)J_q(l_{02jk}+l_{12jk})]\\
=\ &\textstyle\sum_{j=1}^2\sum_{k=1}^{n_j}[(a_{01jk}^T + a_{11jk}^T)J_q(l_{01jk}+l_{11jk}-l_{02jk}-l_{12jk})]\\
=\ &\textstyle\sum_{k=1}^{n_1}[(a_{011k}^T + a_{111k}^T)J_q(l_{011k}+l_{111k}-l_{021k}-l_{121k})]\\
&\textstyle+\sum_{k=1}^{n_2}[(a_{012k}^T + a_{112k}^T)J_q(l_{012k}+l_{112k}-l_{022k}-l_{122k})]\\
=\ &\textstyle\frac{1}{2n_1}\sum_{k=1}^{n_1}\sum_{m=1}^{q}[(1-\rho)(l_{011km}-l_{021km}+l_{111km}-l_{121km})]\\
&\textstyle-\frac{1}{2n_2}\sum_{k=1}^{n_2}\sum_{m=1}^{q}[(1-\rho)(l_{012km}-l_{022km}+l_{112km}-l_{122km})],
\end{aligned}
$$

$$
\begin{aligned}
&\textstyle\sum_{j=1}^2\sum_{k=1}^{n_j}[a_{jk}^T(I_4\otimes J_q)l_{jk}]\\
=\ &\textstyle\sum_{j=1}^2\sum_{k=1}^{n_j}[a_{01jk}^TJ_ql_{01jk} + a_{11jk}^TJ_ql_{11jk} + a_{02jk}^TJ_ql_{02jk} + a_{12jk}^TJ_ql_{12jk}]\\
=\ &\textstyle\sum_{j=1}^2\sum_{k=1}^{n_j}[a_{01jk}^TJ_q(l_{01jk}-l_{02jk}) + a_{11jk}^TJ_q(l_{11jk}-l_{12jk})]\\
=\ &\textstyle\sum_{k=1}^{n_1}[a_{011k}^TJ_q(l_{011k}-l_{021k}) + a_{111k}^TJ_q(l_{111k}-l_{121k})]\\
&\textstyle+\sum_{k=1}^{n_2}[a_{012k}^TJ_q(l_{012k}-l_{022k}) + a_{112k}^TJ_q(l_{112k}-l_{122k})]\\
=\ &\textstyle\frac{1}{2n_1}\sum_{k=1}^{n_1}\sum_{m=1}^{q}[-\rho(l_{011km}-l_{021km}) + (l_{111km}-l_{121km})]\\
&\textstyle-\frac{1}{2n_2}\sum_{k=1}^{n_2}\sum_{m=1}^{q}[-\rho(l_{012km}+l_{022km}) + (l_{112km}-l_{122km})],
\end{aligned}
$$

$$
\begin{aligned}
& \textstyle\sum_{j=1}^{2}\sum_{k=1}^{n_j}[a_{jk}^T(I_4\otimes A)l_{jk}] \\
= \; & \textstyle\sum_{j=1}^{2}\sum_{k=1}^{n_j}[a_{01jk}^T Al_{01jk}+a_{11jk}^T Al_{11jk}+a_{02jk}^T Al_{02jk}+a_{12jk}^T Al_{12jk}] \\
= \; & \textstyle\sum_{j=1}^{2}\sum_{k=1}^{n_j}[a_{01jk}^T A(l_{01jk}-l_{02jk})+a_{11jk}^T A(l_{11jk}-l_{12jk})] \\
= \; & \textstyle\sum_{k=1}^{n_1}[a_{011k}^T A(l_{011k}-l_{021k})+a_{111k}^T A(l_{111k}-l_{121k})] \\
& \textstyle+\sum_{k=1}^{n_2}[a_{012k}^T A(l_{012k}-l_{022k})+a_{112k}^T A(l_{112k}-l_{122k})] \\
= \; & \tfrac{1+r}{q-(q-2)r}\{\tfrac{1}{2n_1}\textstyle\sum_{k=1}^{n_1}\sum_{m=1}^{q}[-\rho(l_{011km}-l_{021km})+(l_{111km}-l_{121km})] \\
& -\tfrac{1}{2n_2}\textstyle\sum_{k=1}^{n_2}\sum_{m=1}^{q}[-\rho(l_{012km}+l_{022km})+(l_{112km}-l_{122km})]\}.
\end{aligned}
$$

The last equality holds because $\frac{1}{q-(q-2)r}(1,1-r,\cdots,1-r,1)\times A$ is equal to $\frac{1}{q-(q-2)r}\{r^{m-1}+r^{q-m}+(1-r)[\sum_{\psi=1}^{q-1-m}r^{\psi}+\sum_{\psi=0}^{m-2}r^{\psi}]\}$. By some algebra, we can show that this is equal to $\frac{1}{q-(q-2)r}(1+r,\cdots,1+r)$.

Thus,

$$
\begin{aligned}
& Cov(a^TY,l^TY) \\
= \; & \sigma_{sp}^2(a^TZ_1Z_1^Tl)+\sigma_{sd}^2(a^TZ_2Z_2^Tl)+\sigma_{\epsilon}^2[a^T(I_4\otimes A)l] \\
= \; & (\sigma_{sp}^2+\sigma_{sd}^2+\tfrac{1+r}{q-(q-2)r}\sigma_{\epsilon}^2)\{\tfrac{1}{2n_1}\textstyle\sum_{k=1}^{n_1}\sum_{m=1}^{q}[l_{111km}-l_{121km}-\rho(l_{011km}-l_{021km})] \\
& \qquad -\tfrac{1}{2n_2}\textstyle\sum_{k=1}^{n_2}\sum_{m=1}^{q}[(l_{112km}-l_{122km})-\rho(l_{012km}-l_{022km})]\} \\
& -\sigma_{sp}^2\{\tfrac{1}{2n_1}\textstyle\sum_{k=1}^{n_1}\sum_{m=1}^{q}[\rho(l_{111km}-l_{121km})-(l_{011km}-l_{021km})] \\
& \qquad -\tfrac{1}{2n_2}\textstyle\sum_{k=1}^{n_2}\sum_{m=1}^{q}\rho[(l_{112km}-l_{122km})-(l_{012km}-l_{022km})]\} \\
= \; & [\sigma_{sp}^2-\rho(\sigma_{sp}^2+\sigma_{sd}^2+\tfrac{1+r}{q-(q-2)r}\sigma_{\epsilon}^2)]\{\tfrac{1}{2n_1}\textstyle\sum_{k=1}^{n_1}\sum_{m=1}^{q}(l_{011km}-l_{021km}) \\
& \qquad -\tfrac{1}{2n_2}\textstyle\sum_{k=1}^{n_2}\sum_{m=1}^{q}(l_{012km}-l_{022km})\} \\
& +(\sigma_{sp}^2+\sigma_{sd}^2+\tfrac{1+r}{q-(q-2)r}\sigma_{\epsilon}^2)\{\tfrac{1}{2n_1}\textstyle\sum_{k=1}^{n_1}\sum_{m=1}^{q}(l_{111km}-l_{121km}) \\
& \qquad -\tfrac{1}{2n_2}\textstyle\sum_{k=1}^{n_2}\sum_{m=1}^{q}(l_{112km}-l_{122km})\} \\
& -\sigma_{sp}^2\{\tfrac{1}{2n_1}\textstyle\sum_{k=1}^{n_1}\sum_{m=1}^{q}[\rho(l_{111km}-l_{121km})]-\tfrac{1}{2n_2}\textstyle\sum_{k=1}^{n_2}\sum_{m=1}^{q}[\rho(l_{112km}-l_{122km})]\} \\
= \; & [(1-\rho)\sigma_{sp}^2+\sigma_{sd}^2+\tfrac{1+r}{q-(q-2)r}\sigma_{\epsilon}^2]\{\tfrac{1}{2n_1}\textstyle\sum_{k=1}^{n_1}[(l_{111k}^T1_q-l_{121k}^T1_q)]+\tfrac{1}{2n_2}\textstyle\sum_{k=1}^{n_2}[(l_{112k}^T1_q-l_{122k}^T1_q)]\} \\
= \; & 0.
\end{aligned}
$$

The second to last equality holds because $\sigma_{sp}^2-\rho(\sigma_{sp}^2+\sigma_{sd}^2+\frac{1+r}{q-(q-2)r}\sigma_{\epsilon}^2)=0$ since $\rho=\sigma_{sp}^2/(\sigma_{sp}^2+\sigma_{sd}^2+\frac{1+r}{q-(q-2)r}\sigma_{\epsilon}^2)$. The last equality holds because from Equation 5.8, we

have $\sum_{j=1}^{2}\sum_{k=1}^{n_j} l_{11jk}^T 1_q = 0$, which indicates that $\sum_{k=1}^{n_1} l_{111k}^T 1_q + \sum_{k=1}^{n_2} l_{112k}^T 1_q = 0$; from Equation 5.10, we have $\sum_{k=1}^{n_1} l_{111k}^T 1_q + \sum_{k=1}^{n_2} l_{122k}^T 1_q = 0$ and $\sum_{k=1}^{n_1} l_{121k}^T 1_q + \sum_{k=1}^{n_2} l_{112k}^T 1_q = 0$. By separately subtracting each of the latter two equations from the former, we obtain that $\sum_{k=1}^{n_1}(l_{111k}^T 1_q - l_{121k}^T 1_q) = 0$ and $\sum_{k=1}^{n_2}(l_{112k}^T 1_q - l_{122k}^T 1_q) = 0$.

Therefore, we show that $Cov(a^T Y, l^T Y) = 0$ for all $l$ such that $E(l^T Y) = 0$. Thus, $a^T Y$ is the BLUE for $\tau_A - \tau_B$.

Similar arguments can be used to prove the results for the other models.

## 5.7 References


1. Dunsmore, I. R. (1981). Growth curves in two-period change over models. *Applied Statistics* 30, 223-229.

2. Fearn, T. (1975). A Bayesian approach to growth curves. *Biometrika* 62, 89-100.

3. Grender, J. M. and Johnson, W. D. (1993). Analysis of crossover designs with multivariate response. *Statistics in Medicine* 12, 69-89.

4. Grender, J. M. and Johnson, W. D. (1993). Fitting multivariate polynomial growth curves in two-period crossover designs. *Statistics in Medicine* 13, 931-943.

5. Jones, B. and Kenward, M. J. (2003) *Design and analysis of cross-over trials.* Second Edition, CRC Press.

6. Patel, H. I. and Hearne III, E. M. (1980). Multivariate analysis for the two-period repeated measurements crossover design with application to clinical trials. *Communications in Statistics* A9, 1919-1929.

7. Putt, M. and Chinchilli V. M. (1999). A mixed effect model for the analysis of repeated measures cross-over studies. *Statistics in Medicine* 18, 3037-3058.

8. Wallenstein, S. and Fisher, A. C. (1977). The analysis of the two-period repeated measurements crossover design with application to clinical trials. *Biometrics* 33, 261-269.

# Chapter 6

# Designs with more than Two Treatments

## 6.1 Introduction

In Chapters 4 and 5, we focused on the $2 \times 2$ crossover design. However, in some applications, it is of interest to compare three or more treatments in a crossover design with one of them possibly being a placebo. The primary interest of such a design could be in all pairwise comparisons, in comparing each treatment to the placebo (Pigeon and Raghavarao, 1987), or in comparing the active treatments first, and then test whether the active treatments are better than the placebo (Koch et al., 1989). The optimality and efficiency of designs with purpose of comparing more than two active treatments and a placebo treatment have been studied by Hedayat and Yang (2005, 2006), Yang and Park (2007) and Yang and Stufken (2007). Often, a uniform design with equal numbers of treatments and periods is adopted. In this chapter, we discuss some of those designs.

As explained above, the analysis of data from a crossover design with more than two treatments is typically still focused on comparisons between two treatments. Since we will only consider designs for which the comparison of any two treatments has the same precision, we will focus on the comparison between treatments $A$ and $B$. Furthermore, since the results for single measurements can be obtained as a special case from the results for repeated measurements, and the results for the independent random error can be obtained as a special case from the results for an AR(1) random error, we only consider the repeated measurements with an AR(1)

random error structure in this chapter. In particular, we present the results of BLUE of the treatment contrast when retaining the baseline measurements as part of the response vector. Sections 6.2 and 6.3 provide the results for selected $3 \times 3$ and $4 \times 4$ designs, respectively. We provide some discussion in Section 6.4.

## 6.2 Three treatments in three periods

A uniform design with three periods, three treatments and three sequences must be one of the following two (rows are sequences):

$$\begin{array}{ccc} A & B & C \\ B & C & A \\ C & A & B \end{array} \quad \text{or} \quad \begin{array}{ccc} A & C & B \\ B & A & C \\ C & B & A \end{array}.$$

Following the discussion for the $2 \times 2$ design in Chapter 5, a model for retaining the baseline measurements as part of the response vector for $3 \times 3$ designs can be formulated as:

$$Y_{hijkm} = \mu + \pi_i + D_{hi} + T_m + h\tau_{t(i,j)} + s_{jk} + \zeta_{ijk} + \omega_{jkm} + \xi_{hijk} + \epsilon_{hijkm}, \qquad (6.1)$$
$$i = 1, 2, 3,\ j = 1, 2, 3,\ k = 1, 2, \ldots, n_j,\ m = 1, 2, \ldots, q,\ h = 0, 1,$$

where $Y_{hijkm}$ corresponds to $X_{ijkm}$ when $h = 0$ and to $Y_{ijkm}$ when $h = 1$, respectively. All the random terms have the same distribution as in Model 5.2, and all the fixed effects have the same interpretation as in Model 5.2 too.

We focus on the design $ABC/BCA/CAB$ to study the BLUEs of $\tau_A - \tau_B$. The results for the design $ACB/BAC/CBA$ can be obtained similarly.

**Theorem 6.1** For Model 6.1, we assume that the random error terms follow an AR(1) process, variance components and autocorrelation coefficient $r$ are known and all the time points are equally spaced. We denote $n_1$, $n_2$ and $n_3$ as the numbers of subjects for sequences 1, 2 and 3 respectively and $n_1 + n_2 + n_3 = n$.

Then, we can obtain the BLUE of $\tau_A - \tau_B$ for design $ABC/BCA/CAB$ as $\frac{1}{3(n_1n_2+n_1n_3+n_2n_3)} \sum_{m=1}^{q} \phi_m \sum_{i=1}^{3} \sum_{j=1}^{3} \psi_{ij}(\bar{Y}_{ij.m} - \rho \bar{X}_{ij.m})$, where $\rho = \frac{\sigma_{sp}^2}{\sigma_{sp}^2+\sigma_{sd}^2+\frac{1+r}{q-(q-2)r}\sigma_\epsilon^2}$,

$$\phi_m = \begin{cases} \frac{1}{q-(q-2)r} & \text{if } m = 1,\ q \\ \frac{1-r}{q-(q-2)r} & \text{if } m = 2, \cdots, q-1 \end{cases}$$

and $\psi_{ij}$ is shown in Table 6.1.

Table 6.1 Coefficients $\psi_{ij}$ of the BLUE of $\tau_A - \tau_B$ for Design $ABC/BCA/CAB$

| | Period i | | |
|---|---|---|---|
| Sequence j | 1 | 2 | 3 |
| 1 $(ABC)$ | $n_1(n_2 - n_1 + n)$ | $n_1(n_1 - n_3 - n)$ | $n_1(n_3 - n_2)$ |
| 2 $(BCA)$ | $n_2(n_2 - n_1 - n)$ | $n_2(n_1 - n_3)$ | $n_2(n_3 - n_2 + n)$ |
| 3 $(CAB)$ | $n_3(n_2 - n_1)$ | $n_3(n_1 - n_3 + n)$ | $n_3(n_3 - n_2 - n)$ |

Notice here, if $n_1 = n_2 = n_3$, the BLUE of $\tau_A - \tau_B$ will reduce to $\frac{1}{3}\{\sum_{m=1}^{q} \phi_m[\bar{Y}_{11.m} - \bar{Y}_{21.m} - \bar{Y}_{12.m} + \bar{Y}_{32.m} + \bar{Y}_{23.m} - \bar{Y}_{33.m} - \rho(\bar{X}_{11.m} - \bar{X}_{21.m} - \bar{X}_{12.m} + \bar{X}_{32.m} + \bar{X}_{23.m} - \bar{X}_{33.m})]\}$.

**Proof**: For design $ABC/BCA/CAB$, when random error terms follow an AR(1) process, Model 6.1 can be written in matrix notation as follows:

$$Y = \mu 1_{6nq} + X_1\pi + X_2D + X_3T + X_d\tau + Z_0s + Z_1\zeta + Z_2\xi + Z_3\omega + \epsilon, \qquad (6.2)$$

where

$$X_1 = 1_n \otimes I_3 \otimes 1_{2q}, \qquad X_2 = 1_n \otimes I_6 \otimes 1_q, \qquad X_3 = 1_n \otimes 1_6 \otimes I_q,$$

$$X_d = (X_{d11}^T, X_{d12}^T, \cdots X_{d3n_3}^T)^T \otimes 1_q \quad \text{with}$$

$$X_{d1k} = \begin{pmatrix} 0&1&0&0&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&0&0&1 \end{pmatrix}^T,\ X_{d2k} = \begin{pmatrix} 0&0&0&0&0&1 \\ 0&1&0&0&0&0 \\ 0&0&0&1&0&0 \end{pmatrix}^T \text{ and } X_{d3k} = \begin{pmatrix} 0&0&0&1&0&0 \\ 0&0&0&0&0&1 \\ 0&1&0&0&0&0 \end{pmatrix}^T,$$

$$Z_0 = I_n \otimes 1_{6q}, \quad Z_1 = I_n \otimes I_3 \otimes 1_{2q}, \quad Z_2 = I_n \otimes I_6 \otimes 1_q, \quad Z_3 = I_n \otimes 1_6 \otimes I_q$$

and $s$, $\zeta$, $\xi$ and $\omega$ are independently normally distributed with their elements being independently normally distributed with mean 0 and variances $\sigma_s^2$, $\sigma_{sp}^2$, $\sigma_{sd}^2$ and

$\sigma^2_{st}$, respectively. Furthermore, the variance of the random error terms $Var(\epsilon) = \sigma^2_\epsilon(I_n \otimes I_6 \otimes A)$ and the $q \times q$ matrix $A$ is given by $A = \begin{pmatrix} 1 & r & \cdots & r^{q-1} \\ r & 1 & \cdots & r^{q-2} \\ \vdots & \vdots & \ddots & \vdots \\ r^{q-1} & r^{q-2} & \cdots & 1 \end{pmatrix}$.

Let the proposed estimator of $\tau_A - \tau_B$ in the statement of Theorem 6.1 be $a^TY$ with $a^T = (a^T_{11}, \cdots, a^T_{1n_1}, a^T_{21}, \cdots, a^T_{2n_2}, a^T_{31}, \cdots, a^T_{3n_3})$, where $a^T_{jk} = (a^T_{01jk}, a^T_{11jk}, a^T_{02jk}, a^T_{12jk}, a^T_{03jk}, a^T_{13jk})$. Thus $a_{hijk}$ is the $q \times 1$ vector $(a_{hijk1}, \cdots, a_{hijkq})^T$ with $a_{hijkm} = \frac{1}{3(n_1n_2+n_1n_3+n_2n_3)} \lambda \frac{\psi_{ij}}{n_j} \phi_m$, where $\lambda = 1$ for $h = 1$ and $\lambda = -\rho$ for $h = 0$. It is easy to verify that $a^TY$ is a linear unbiased estimator for $\tau_A - \tau_B$. To prove that $a^TY$ is also the BLUE of $\tau_A - \tau_B$, by Theorem 5.1 we need to show that $Cov(a^TY, l^TY) = 0$ for all $l$ such that $E(l^TY) = 0$.

Let $l^T = (l^T_{11}, \cdots, l^T_{1n_1}, l^T_{21}, \cdots, l^T_{2n_2}, l^T_{31}, \cdots, l^T_{3n_3})$, where $l^T_{jk}$=$(l^T_{01jk}, l^T_{11jk}, l^T_{02jk}, l^T_{12jk}, l^T_{03jk}, l^T_{13jk})$ and $l_{hijk}$ is a $q \times 1$ vector. Based on Model 6.2, we have that $E(l^TY) = 0$ if and only if $l^TX_2 = 0^T$, $l^TX_3 = 0^T$ and $l^TX_d = 0^T$. Similar to the proof for Theorem 5.2, we have

$$\sum_{j=1}^{3} \sum_{k=1}^{n_j} (l^T_{01jk}1_q, l^T_{11jk}1_q, l^T_{02jk}1_q, l^T_{12jk}1_q, l^T_{03jk}1_q, l^T_{13jk}1_q) = 0^T, \tag{6.3}$$

$$\sum_{j=1}^{3} \sum_{k=1}^{n_j} \sum_{h=0}^{1} \sum_{i=1}^{3} l^T_{hijk} = 0^T,$$

$$\sum_{k=1}^{n_1} l^T_{111k}1_q + \sum_{k=1}^{n_2} l^T_{132k}1_q + \sum_{k=1}^{n_3} l^T_{123k}1_q = 0, \tag{6.4}$$

$$\sum_{k=1}^{n_1} l^T_{121k}1_q + \sum_{k=1}^{n_2} l^T_{112k}1_q + \sum_{k=1}^{n_3} l^T_{133k}1_q = 0, \tag{6.5}$$

$$\sum_{k=1}^{n_1} l^T_{131k}1_q + \sum_{k=1}^{n_2} l^T_{122k}1_q + \sum_{k=1}^{n_3} l^T_{113k}1_q) = 0.$$

Furthermore, $Cov(a^TY, l^TY) = \sum_{j=1}^{3} \sum_{k=1}^{n_j} \{\sigma^2_s(a^T_{jk}J_{6q}l_{jk}) + \sigma^2_{sp}[a^T_{jk}(I_3 \otimes J_{2q})l_{jk}] + \sigma^2_{sd}[a^T_{jk}(I_6 \otimes J_q)l_{jk}] + \sigma^2_{st}[a^T_{jk}(J_6 \otimes I_q)l_{jk}] + \sigma^2_\epsilon[a^T_{jk}(I_6 \otimes A)l_{jk}]\}$. Again, it is easy to obtain that $\sum_{j=1}^{3} \sum_{k=1}^{n_j} \sigma^2_s(a^T_{jk}J_{6q}l_{jk}) = 0$ and $\sum_{j=1}^{3} \sum_{k=1}^{n_j} \sigma^2_{st}[a^T_{jk}(J_6 \otimes I_q)l_{jk}] = 0$ because the coefficients of the proposed estimator add to zero for each subject.

Furthermore, after some algebraic manipulations as in the proof for Theorem 5.2 and use the results that $\rho = \frac{\sigma_{sp}^2}{\sigma_{sp}^2+\sigma_{sd}^2+\frac{1+r}{q-(q-2)r}\sigma_\epsilon^2}$, we have

$$
\begin{aligned}
& Cov(a^TY, l^TY) \\
= & \tfrac{1}{3(n_1n_2+n_1n_3+n_2n_3)}[(1-\rho)\sigma_{sp}^2 + \sigma_{sd}^2 + \tfrac{1+r}{q-(q-2)r}\sigma_\epsilon^2] \\
& \{\textstyle\sum_{k=1}^{n_1}[(n_2-n_1+n)l_{111k}^T1_q + (n_1-n_3-n)l_{121k}^T1_q + (n_3-n_2)l_{131k}^T1_q)] \\
& + \textstyle\sum_{k=1}^{n_2}[(n_2-n_1-n)l_{112k}^T1_q + (n_1-n_3)l_{122k}^T1_q + (n_3-n_2+n)l_{132k}^T1_q)] \\
& + \textstyle\sum_{k=1}^{n_3}[(n_2-n_1)l_{113k}^T1_q + (n_1-n_3+n)l_{123k}^T1_q + (n_3-n_2-n)l_{133k}^T1_q)]\} \\
= & \tfrac{1}{3(n_1n_2+n_1n_3+n_2n_3)}[(1-\rho)\sigma_{sp}^2 + \sigma_{sd}^2 + \tfrac{1+r}{q-(q-2)r}\sigma_\epsilon^2] \\
& \{\textstyle\sum_{j=1}^{3}\sum_{k=1}^{n_j}[(n_2-n_1)l_{11jk}^T1_q + (n_1-n_3)l_{12jk}^T1_q + (n_3-n_2)l_{13jk}^T1_q] \\
& + n(\textstyle\sum_{k=1}^{n_1} l_{111k}^T1_q + \sum_{k=1}^{n_2} l_{132k}^T1_q + \sum_{k=1}^{n_3} l_{123k}^T1_q) \\
& - n(\textstyle\sum_{k=1}^{n_1} l_{121k}^T1_q + \sum_{k=1}^{n_2} l_{112k}^T1_q + \sum_{k=1}^{n_3} l_{133k}^T1_q)\} \\
= & \, 0.
\end{aligned}
$$

The last equality holds because $\sum_{j=1}^{3}\sum_{k=1}^{n_j} l_{11jk}^T1_q = 0$, $\sum_{j=1}^{3}\sum_{k=1}^{n_j} l_{12jk}^T1_q = 0$ and $\sum_{j=1}^{3}\sum_{k=1}^{n_j} l_{13jk}^T1_q = 0$ from Equation 6.3, $\sum_{k=1}^{n_1} l_{111k}^T1_q + \sum_{k=1}^{n_2} l_{132k}^T1_q + \sum_{k=1}^{n_3} l_{123k}^T1_q = 0$ from Equation 6.4, and $\sum_{k=1}^{n_1} l_{121k}^T1_q + \sum_{k=1}^{n_2} l_{112k}^T1_q + \sum_{k=1}^{n_3} l_{133k}^T1_q = 0$ from Equation 6.5.

Therefore, we have shown that $Cov(a^TY, l^TY) = 0$ for all $l$ such that $E(l^TY) = 0$. Thus, $a^TY$ is the BLUE for $\tau_A - \tau_B$.

## 6.3 Four treatments in four periods

For uniform designs with four periods, four treatments and four sequences, we use the following design as an example (rows are sequences):

$$
\begin{array}{cccc}
A & B & C & D \\
B & C & D & A \\
C & D & A & B \\
D & A & B & C
\end{array}
.
$$

The model for retaining the baseline measurements as part of the response vector for this $4 \times 4$ design can be formulated as:

$$Y_{hijkm} = \mu + \pi_i + D_{hi} + T_m + h\tau_{t(i,j)} + s_{jk} + \zeta_{ijk} + \omega_{jkm} + \xi_{hijk} + \epsilon_{hijkm}, \qquad (6.6)$$
$$i = 1,2,3,4,\ j = 1,2,3,4,\ k = 1,2,\ldots,n_j,\ m = 1,2,\ldots,q,\ h = 0,1,$$

where $Y_{hijkm}$ corresponds to $X_{ijkm}$ when $h = 0$ and to $Y_{ijkm}$ when $h = 1$, respectively. All the random terms have the same distribution as in Model 5.2, and all the fixed effects have the same interpretation as in Model 5.2 too.

**Theorem 6.2** For Model 6.6, we assume that the random error terms follow an AR(1) process, variance components and autocorrelation coefficient $r$ are known and all the time points are equally spaced. We denote $n_1$, $n_2$, $n_3$ and $n_4$ as the numbers of subjects for sequences 1, 2, 3 and 4, respectively and $n_1 + n_2 + n_3 + n_4 = n$. Then, we can obtain the BLUE of $\tau_A - \tau_B$ for this design as $\sum_{m=1}^{q} \phi_m \sum_{i=1}^{4} \sum_{j=1}^{4} \varphi_j \psi_{ij} (\bar{Y}_{ij.m} - \rho \bar{X}_{ij.m})$, where

$$\varphi_j = \begin{cases} \frac{1}{4(2n_1n_2n_3+2n_1n_2n_4+2n_1n_3n_4+2n_2n_3n_4+n_2n_1^2+2n_3n_1^2+n_4n_1^2+2n_1n_3^2+n_2n_3^2+n_4n_3^2)} & \text{if } j = 1,\ 3 \\ \frac{1}{4(2n_1n_2n_3+2n_1n_2n_4+2n_1n_3n_4+2n_2n_3n_4+n_1n_2^2+n_3n_2^2+2n_4n_2^2+n_1n_4^2+2n_2n_4^2+n_3n_4^2)} & \text{if } j = 2,\ 4 \end{cases},$$

$\rho = \frac{\sigma_{sp}^2}{\sigma_{sp}^2+\sigma_{sd}^2+\frac{1+r}{q-(q-2)r}\sigma_\epsilon^2}$, $\phi_m = \begin{cases} \frac{1}{q-(q-2)r} & \text{if } m = 1,\ q \\ \frac{1-r}{q-(q-2)r} & \text{if } m = 2,\cdots,q-1 \end{cases}$ and $\psi_{ij}$ is shown in Table 6.2.

Table 6.2 Coefficients $\psi_{ij}$ of the BLUE of $\tau_A - \tau_B$ for design $ABCD/BCDA/CDAB/DABC$

| Sequence | Period | $\psi_{ij}$ |
|---|---|---|
| j=1 | i=1 | $n_1(3n_1n_2 + 4n_1n_3 + n_1n_4 + 3n_2n_3 + 2n_2n_4 + n_3n_4 + 2n_3^2)$ |
| | i=2 | $-n_1(n_1n_2 + 4n_1n_3 + 3n_1n_4 + n_2n_3 + 2n_2n_4 + 3n_3n_4 + 2n_3^2)$ |
| | i=3 | $-n_1(n_1n_2 - n_1n_4 + n_2n_3 - 2n_2n_4 - n_3n_4 + 2n_3^2)$ |
| | i=4 | $-n_1(n_1n_2 - n_1n_4 + n_2n_3 + 2n_2n_4 - n_3n_4 - 2n_3^2)$ |
| j=2 | i=1 | $-n_2(3n_1n_2 + 2n_1n_3 + 3n_1n_4 + n_2n_3 + 4n_2n_4 + n_3n_4 + 2n_4^2)$ |
| | i=2 | $n_2(n_1n_2 + 2n_1n_3 + n_1n_4 - n_2n_3 - n_3n_4 - 2n_4^2)$ |
| | i=3 | $n_2(n_1n_2 - 2n_1n_3 + n_1n_4 - n_2n_3 - n_3n_4 + 2n_4^2)$ |
| | i=4 | $n_2(n_1n_2 + 2n_1n_3 + n_1n_4 + 3n_2n_3 + 4n_2n_4 + 3n_3n_4 + 2n_4^2)$ |
| j=3 | i=1 | $-n_3(-n_1n_2 + n_1n_4 - n_2n_3 - 2n_2n_4 + n_3n_4 + 2n_1^2)$ |
| | i=2 | $n_3(n_1n_2 - n_1n_4 + n_2n_3 - 2n_2n_4 - n_3n_4 + 2n_1^2)$ |
| | i=3 | $n_3(n_1n_2 + 4n_1n_3 + 3n_1n_4 + n_2n_3 + 2n_2n_4 + 3n_3n_4 + 2n_1^2)$ |
| | i=4 | $-n_3(3n_1n_2 + 4n_1n_3 + n_1n_4 + 3n_2n_3 + 2n_2n_4 + n_3n_4 + 2n_1^2)$ |
| j=4 | i=1 | $-n_4(n_1n_2 + 2n_1n_3 + n_1n_4 - n_2n_3 - n_3n_4 - 2n_2^2)$ |
| | i=2 | $n_4(3n_1n_2 + 2n_1n_3 + 3n_1n_4 + n_2n_3 + 4n_2n_4 + n_3n_4 + 2n_2^2)$ |
| | i=3 | $-n_4(n_1n_2 + 2n_1n_3 + n_1n_4 + 3n_2n_3 + 4n_2n_4 + 3n_3n_4 + 2n_2^2)$ |
| | i=4 | $-n_4(n_1n_2 - 2n_1n_3 + n_1n_4 - n_2n_3 - n_3n_4 + 2n_2^2)$ |

Notice here, if $n_1 = n_2 = n_3 = n_4$, the BLUE of $\tau_A - \tau_B$ will reduce to $\frac{1}{4}\{\sum_{m=1}^{q} \phi_m[\bar{Y}_{11.m} - \bar{Y}_{21.m} - \bar{Y}_{12.m} + \bar{Y}_{42.m} + \bar{Y}_{33.m} - \bar{Y}_{43.m} + \bar{Y}_{24.m} - \bar{Y}_{34.m} - \rho(\bar{X}_{11.m} - \bar{X}_{21.m} - \bar{X}_{12.m} + \bar{X}_{42.m} + \bar{X}_{33.m} - \bar{X}_{43.m} + \bar{X}_{24.m} - \bar{X}_{34.m})]\}$.

**Proof**: For design $ABCD/BCDA/CDAB/DABC$, when random error terms follow an AR(1) process, Model 6.6 can be written in matrix notation as follows:

$$Y = \mu 1_{8nq} + X_1\pi + X_2D + X_3T + X_d\tau + Z_0s + Z_1\zeta + Z_2\xi + Z_3\omega + \epsilon, \qquad (6.7)$$

where

$$X_1 = 1_n \otimes I_4 \otimes 1_{2q}, \qquad X_2 = 1_n \otimes I_8 \otimes 1_q, \qquad X_3 = 1_n \otimes 1_8 \otimes I_q,$$

$$X_d = (X_{d11}^T, X_{d12}^T, \cdots X_{d4n_4}^T)^T \otimes 1_q \text{ with } X_{d1k}=\begin{pmatrix} 0&1&0&0&0&0&0&0\\ 0&0&0&1&0&0&0&0\\ 0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&1 \end{pmatrix}^T,$$

$$X_{d2k}=\begin{pmatrix} 0&0&0&0&0&0&0&1\\ 0&1&0&0&0&0&0&0\\ 0&0&0&1&0&0&0&0\\ 0&0&0&0&0&1&0&0 \end{pmatrix}^T, \; X_{d3k}=\begin{pmatrix} 0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&1\\ 0&1&0&0&0&0&0&0\\ 0&0&0&1&0&0&0&0 \end{pmatrix}^T, \text{ and } \; X_{d4k}=\begin{pmatrix} 0&0&0&1&0&0&0&0\\ 0&0&0&0&0&0&1&0\\ 0&0&0&0&0&0&0&1\\ 0&1&0&0&0&0&0&0 \end{pmatrix}^T,$$

$$Z_0 = I_n \otimes 1_{8q}, \quad Z_1 = I_n \otimes I_4 \otimes 1_{2q}, \quad Z_2 = I_n \otimes I_8 \otimes 1_q, \quad Z_3 = I_n \otimes 1_8 \otimes I_q$$

and $s$, $\zeta$, $\xi$ and $\omega$ are independently normally distributed with their elements being independently normally distributed with mean 0 and variance $\sigma_s^2$, $\sigma_{sp}^2$, $\sigma_{sd}^2$ and $\sigma_{st}^2$, respectively. Furthermore, the variance of the random error terms $Var(\epsilon) = \sigma_\epsilon^2(I_n \otimes I_8 \otimes A)$ and the $q \times q$ matrix $A$ is given by $A = \begin{pmatrix} 1 & r & \cdots & r^{q-1} \\ r & 1 & \cdots & r^{q-2} \\ \vdots & \vdots & \ddots & \vdots \\ r^{q-1} & r^{q-2} & \cdots & 1 \end{pmatrix}$.

Let the proposed estimator of $\tau_A - \tau_B$ in the statement of Theorem 6.2 be $a^TY$ with $a^T$=($a_{11}^T$, $\cdots$, $a_{1n_1}^T$, $a_{21}^T$, $\cdots$, $a_{2n_2}^T$, $a_{31}^T$, $\cdots$, $a_{3n_3}^T$, $a_{41}^T$, $\cdots$, $a_{4n_4}^T$), where $a_{jk}^T$ =($a_{01jk}^T$, $a_{11jk}^T$, $a_{02jk}^T$, $a_{12jk}^T$, $a_{03jk}^T$, $a_{13jk}^T$, $a_{04jk}^T$, $a_{14jk}^T$). Thus $a_{hijk}$ is the $q \times 1$ vector $(a_{hijk1}, \cdots, a_{hijkq})^T$ with $a_{hijkm} = \lambda\varphi_j \frac{\psi_{ij}}{n_j}\phi_m$, where $\lambda = 1$ for $h = 1$ and $\lambda = -\rho$ for $h = 0$. It is easy to verify that $a^TY$ is a linear unbiased estimator for $\tau_A - \tau_B$. To prove that $a^TY$ is also the BLUE of $\tau_A - \tau_B$, by Theorem 5.1 we need to show that $Cov(a^TY, l^TY) = 0$ for all $l$ such that $E(l^TY) = 0$.

Let $l^T = (l_{11}^T, \cdots, l_{1n_1}^T, l_{21}^T, \cdots, l_{2n_2}^T, l_{31}^T, \cdots, l_{3n_3}^T, l_{41}^T, \cdots, l_{4n_4}^T)$, where $l_{jk}^T$=($l_{01jk}^T$, $l_{11jk}^T$, $l_{02jk}^T$, $l_{12jk}^T$, $l_{03jk}^T$, $l_{13jk}^T$, $l_{04jk}^T$, $l_{14jk}^T$) and $l_{hijk}$ is a $q \times 1$ vector. Based on Model 6.7, we have that $E(l^TY) = 0$ if and only if $l^TX_2 = 0^T$, $l^TX_3 = 0^T$ and $l^TX_d = 0^T$. Similar to the proof for Theorem 6.1, we have

$$\sum_{j=1}^{4}\sum_{k=1}^{n_j}(l_{01jk}^T1_q, l_{11jk}^T1_q, l_{02jk}^T1_q, l_{12jk}^T1_q, l_{03jk}^T1_q, l_{13jk}^T1_q, l_{04jk}^T1_q, l_{14jk}^T1_q) = 0^T, \qquad (6.8)$$

$$\sum_{j=1}^{4}\sum_{k=1}^{n_j}\sum_{h=0}^{1}\sum_{i=1}^{4} l_{hijk}^T = 0^T,$$

$$\sum_{k=1}^{n_1} l_{111k}^T1_q + \sum_{k=1}^{n_2} l_{142k}^T1_q + \sum_{k=1}^{n_3} l_{133k}^T1_q + \sum_{k=1}^{n_4} l_{124k}^T1_q = 0, \qquad (6.9)$$

$$\sum_{k=1}^{n_1} l_{121k}^T 1_q + \sum_{k=1}^{n_2} l_{112k}^T 1_q + \sum_{k=1}^{n_3} l_{143k}^T 1_q + \sum_{k=1}^{n_4} l_{134k}^T 1_q = 0, \tag{6.10}$$

$$\sum_{k=1}^{n_1} l_{131k}^T 1_q + \sum_{k=1}^{n_2} l_{122k}^T 1_q + \sum_{k=1}^{n_3} l_{113k}^T 1_q + \sum_{k=1}^{n_4} l_{144k}^T 1_q = 0, \tag{6.11}$$

$$\sum_{k=1}^{n_1} l_{141k}^T 1_q + \sum_{k=1}^{n_2} l_{132k}^T 1_q + \sum_{k=1}^{n_3} l_{123k}^T 1_q + \sum_{k=1}^{n_4} l_{114k}^T 1_q = 0.$$

We also have

$$\begin{aligned}
&Cov(a^T Y, l^T Y) \\
&= \textstyle\sum_{j=1}^{4} \sum_{k=1}^{n_j} \{\sigma_s^2 (a_{jk}^T J_{8q} l_{jk}) + \sigma_{sp}^2 [a_{jk}^T (I_4 \otimes J_{2q}) l_{jk}] + \sigma_{sd}^2 [a_{jk}^T (I_8 \otimes J_q) l_{jk}] \\
&\quad + \sigma_{st}^2 [a_{jk}^T (J_8 \otimes I_q) l_{jk}] + \sigma_\epsilon^2 [a_{jk}^T (I_8 \otimes A) l_{jk}]\} \\
&= \textstyle\sum_{j=1}^{4} \sum_{k=1}^{n_j} \{\sigma_{sp}^2 [a_{jk}^T (I_4 \otimes J_{2q}) l_{jk}] + \sigma_{sd}^2 [a_{jk}^T (I_8 \otimes J_q) l_{jk}] + \sigma_\epsilon^2 [a_{jk}^T (I_8 \otimes A) l_{jk}]\}.
\end{aligned}$$

Similarly, after some algebraic manipulations, we have

$$\begin{aligned}
&Cov(a^T Y, l^T Y) \\
&= [(1-\rho)\sigma_{sp}^2 + \sigma_{sd}^2 + \tfrac{1+r}{q-(q-2)r}\sigma_\epsilon^2] \\
&\quad \{\textstyle\sum_{k=1}^{n_1} \varphi_1 [(3n_1n_2 + 4n_1n_3 + n_1n4 + 3n_2n_3 + 2n_2n_4 + n_3n_4 + 2n3^2) l_{111k}^T 1_q \\
&\qquad -(n_1n_2 + 4n_1n_3 + 3n_1n4 + n_2n_3 + 2n_2n_4 + 3n_3n_4 + 2n3^2) l_{121k}^T 1_q \\
&\qquad -(n_1n_2 - n_1n_4 + n_2n_3 - 2n_2n_4 - n_3n_4 + 2n_3^2) l_{131k}^T 1_q \\
&\qquad -(n_1n_2 - n_1n_4 + n_2n_3 + 2n_2n_4 - n_3n_4 - 2n_3^2) l_{141k}^T 1_q] \\
&\quad + \textstyle\sum_{k=1}^{n_2} \varphi_2 [-(3n_1n_2 + 2n_1n_3 + 3n_1n_4 + n_2n_3 + 4n_2n_4 + n_3n_4 + 2n_4^2) l_{112k}^T 1_q \\
&\qquad +(n_1n_2 + 2n_1n_3 + n_1n_4 - n_2n_3 - n_3n_4 - 2n_4^2) l_{122k}^T 1_q \\
&\qquad +(n_1n_2 - 2n_1n_3 + n_1n_4 - n_2n_3 - n_3n_4 + 2n_4^2) l_{132k}^T 1_q \\
&\qquad +(n_1n_2 + 2n_1n_3 + n_1n_4 + 3n_2n_3 + 4n_2n_4 + 3n_3n_4 + 2n_4^2)_{142k}^T 1_q] \\
&\quad + \textstyle\sum_{k=1}^{n_3} \varphi_3 [-(-n_1n_2 + n_1n_4 - n_2n_3 - 2n_2n_4 + n_3n_4 + 2n_1^2) l_{113k}^T 1_q \\
&\qquad +(n_1n_2 - n_1n_4 + n_2n_3 - 2n_2n_4 - n_3n_4 + 2n_1^2) l_{123k}^T 1_q \\
&\qquad +(n_1n_2 + 4n_1n_3 + 3n_1n_4 + n_2n_3 + 2n_2n_4 + 3n_3n_4 + 2n_1^2) l_{133k}^T 1_q \\
&\qquad -(3n_1n_2 + 4n_1n_3 + n_1n_4 + 3n_2n_3 + 2n_2n_4 + n_3n_4 + 2n_1^2) l_{143k}^T 1_q]\}
\end{aligned}$$

$$
\begin{aligned}
&+\textstyle\sum_{k=1}^{n_4}\varphi_4[-(n_1n_2+2n_1n_3+n_1n_4-n_2n_3-n_3n_4-2n_2^2)l_{114k}^T1_q \\
&\qquad +(3n_1n_2+2n_1n_3+3n_1n_4+n_2n_3+4n_2n_4+n_3n_4+2n_2^2)l_{124k}^T1_q \\
&\qquad -(n_1n_2+2n_1n_3+n_1n_4+3n_2n_3+4n_2n_4+3n_3n_4+2n_2^2)l_{134k}^T1_q \\
&\qquad -(n_1n_2-2n_1n_3+n_1n_4-n_2n_3-n_3n_4+2n_2^2)l_{144k}^T1_q]\} \\
&=[(1-\rho)\sigma_{sp}^2+\sigma_{sd}^2+\tfrac{1+r}{q-(q-2)r}\sigma_\epsilon^2] \\
&\quad\{-\varphi_4(n_1n_2-n_1n_4+n_2n_3+2n_2n_4-n_3n_4-2n_3^2)\textstyle\sum_{j=1}^4\sum_{k=1}^{n_j}l_{11jk}^T1_q \\
&\quad +\varphi_3(n_1n_2-2n_1n_3+n_1n_4-n_2n_3-n_3n_4+2n_4^2)\textstyle\sum_{j=1}^4\sum_{k=1}^{n_j}l_{12jk}^T1_q \\
&\quad +\varphi_2(n_1n_2-n_1n_4+n_2n_3-2n_2n_4-n_3n_4+2n_1^2)\textstyle\sum_{j=1}^4\sum_{k=1}^{n_j}l_{13jk}^T1_q \\
&\quad -\varphi_1(n_1n_2+2n_1n_3+n_1n_4-n_2n_3-n_3n_4-2n_2^2)\textstyle\sum_{j=1}^4\sum_{k=1}^{n_j}l_{14jk}^T1_q \\
&\quad +\alpha_1(\textstyle\sum_{k=1}^{n_1}l_{111k}^T1_q+\sum_{k=1}^{n_2}l_{142k}^T1_q+\sum_{k=1}^{n_3}l_{133k}^T1_q+\sum_{k=1}^{n_4}l_{124k}^T1_q) \\
&\quad -\alpha_2(\textstyle\sum_{k=1}^{n_1}l_{121k}^T1_q+\sum_{k=1}^{n_2}l_{112k}^T1_q+\sum_{k=1}^{n_3}l_{143k}^T1_q+\sum_{k=1}^{n_4}l_{134k}^T1_q) \\
&\quad -\alpha_3(\textstyle\sum_{k=1}^{n_1}l_{131k}^T1_q+\sum_{k=1}^{n_2}l_{122k}^T1_q+\sum_{k=1}^{n_3}l_{113k}^T1_q+\sum_{k=1}^{n_4}l_{144k}^T1_q)\},
\end{aligned}
$$

where $\alpha_1 = \frac{n}{4(n_1+n_3)(n_2+n_4)}$, $\alpha_2 = \frac{n}{2(n_1n_2+2n_1n_3+n_1n_4+n_2n_3+2n_2n_4+2n_3n_4)}$ and $\alpha_3 = \frac{n_2n_1^2-2n_3n_1^2+n_4n_1^2+n_1n_2^2+n_3n_2^2-2n_4n_2^2-2n_1n_3^2+n_2n_3^2+n_4n_3^2+n_1n_4^2-2n_2n_4^2+n_3n_4^2}{4(n_1+n_3)[n_1n_2^2+n_3n_2^2+2n_4n_2^2+n_1n_4^2+2n_2n_4^2+n_3n_4^2+2(n_1n_2n_3+n_1n_2n_4+n_1n_3n_4+n_2n_3n_4)]}$.

By applying Equations 6.8 and 6.9 - 6.11, if follows that the last expression equals zero. Therefore, we have shown that $Cov(a^TY, l^TY) = 0$ for all $l$ such that $E(l^TY) = 0$. Thus, $a^TY$ is the BLUE for $\tau_A - \tau_B$.

## 6.4 Discussion

In this chapter, we studied selected $3 \times 3$ and $4 \times 4$ uniform crossover designs. We obtained the BLUEs of a treatment contrast for the method of retaining baseline measurements as part of the response vector. Similar results can also be obtained for other methods discussed in Chapter 5. Even though we did not provide a comparison for the different methods for designs with more than two treatments in this chapter, we would still expect that retaining baseline measurements as part of the response vector will have the highest efficiency for most of the scenarios.

For the designs with three treatments and three periods, we could also use all six sequences in Section 6.2. However, the closed form for the BLUE of $\tau_A - \tau_B$ for that design is rather complicated unless that the numbers of the subjects in each sequence are the same. So, if the numbers of subjects for each sequence are different, simply using the average from the corresponding observations for the particular treatment contrast will lead to incorrect results. Of course, nowadays, we can use PROC MIXED in SAS to obtain the estimate for the treatment contrast of interest, and SAS will take care of unequal sequence replications.

Lack of balance resulting from unequal numbers of subjects in each sequence is, however, not uncommon in clinical trials. Even when the design calls for the same number of subjects for each sequence, we may wind up with unequal numbers due to dropout of subjects. This is especially true for designs with a larger number of periods. Thus, crossover designs with large number of periods should be avoided at the design stage.

## 6.5 References

1. Hedayat, A. S. and Yang, M. (2005) Optimal and efficient crossover designs for comparing test treatments with a control treatment. *The Annals of Statistics* 33, 915–943.

2. Koch, G. G., Amara, I. A., Brown, B. W., Colton, T. and Gillings, D.B (1989) A two-period crossover design for the comparison of two active treatments and a placebo. *Statistics in Medicine* 8, 487-504.

3. Pigeon, J. G. and Raghavarao, D. (1987) Crossover designs for comparison of treatments with a control. *Biometrika* 74, 321-328.

4. Yang, M and Hedayat, A. S. (2006) Efficient crossover designs for comparing test treatments with a control treatment. *Discrete Mathematics* 306, 3112-3124.

5. Yang, M. and Park, M. (2007) Efficient crossover designs for comparing test treatments with a control treatment when p=3. *Journal of Statistics and Planning Inference*. In press.

6. Yang, M. and Stufken, J. (2007) Optimal and efficient crossover designs for comparing test treatments to a control treatment under various models. In press.