TREATMENT HETEROGENEITY AND POTENTIAL OUTCOMES IN LINEAR MIXED EFFECTS MODELS

by

TROY E. RICHARDSON

B.S., Wheaton College, 2000
M.S., Kansas State University, 2001,
M.P.H., University of Kansas Medical Center, 2004

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2013

Abstract

Studies commonly focus on estimating a mean treatment effect in a population. However, in some applications the variability of treatment effects across individual units may help to characterize the overall effect of a treatment across the population. Consider a set of treatments, {T,C}, where T denotes some treatment that might be applied to an experimental unit and C denotes a control. For each of N experimental units, the duplet $\{r_{Ti}, r_{Ci}\}, i = 1,2, \dots, N$, represents the potential response of the i^{th} experimental unit if treatment were applied and the response of the experimental unit if control were applied, respectively. The causal effect of T compared to C is the difference between the two potential responses, $r_{Ti} - r_{Ci}$. Much work has been done to elucidate the statistical properties of a causal effect, given a set of particular assumptions. Gadbury and others have reported on this for some simple designs and primarily focused on finite population randomization based inference. When designs become more complicated, the randomization based approach becomes increasingly difficult.

Since linear mixed effects models are particularly useful for modeling data from complex designs, their role in modeling treatment heterogeneity is investigated. It is shown that an individual treatment effect can be conceptualized as a linear combination of fixed treatment effects and random effects. The random effects are assumed to have variance components specified in a mixed effects "potential outcomes" model when both potential outcomes, r_T, r_C, are variables in the model. The variance of the individual causal effect is used to quantify treatment heterogeneity. Post treatment assignment, however, only one of the two potential outcomes is observable for a unit. It is then shown that the variance component for treatment heterogeneity becomes non-estimable in an analysis of observed data. Furthermore, estimable variance components in the observed data model are demonstrated to arise from linear combinations of the non-estimable variance components in the potential outcomes model. Mixed effects models are considered in context of a particular design in an effort to illuminate the loss of information incurred when moving from a potential outcomes framework to an observed data analysis.

TREATMENT HETEROGENEITY AND POTENTIAL OUTCOMES IN LINEAR MIXED EFFECTS MODELS

by

TROY E. RICHARDSON

B.S., Wheaton College, 2000
M.S., Kansas State University, 2001,
M.P.H., University of Kansas Medical Center, 2004

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2013

Approved by:

Major Professor
Gary L. Gadbury, Ph.D.

Copyright

Abstract

Studies commonly focus on estimating a mean treatment effect in a population. However, in some applications the variability of treatment effects across individual units may help to characterize the overall effect of a treatment across the population. Consider a set of treatments, {T,C}, where T denotes some treatment that might be applied to an experimental unit and C denotes a control. For each of N experimental units, the duplet $\{r_{T_i}, r_{C_i}\}, i = 1,2, \dots, N$, represents the potential response of the i^{th} experimental unit if treatment were applied and the response of the experimental unit if control were applied, respectively. The causal effect of T compared to C is the difference between the two potential responses, $r_{T_i} - r_{C_i}$. Much work has been done to elucidate the statistical properties of a causal effect, given a set of particular assumptions. Gadbury and others have reported on this for some simple designs and primarily focused on finite population randomization based inference. When designs become more complicated, the randomization based approach becomes increasingly difficult.

Since linear mixed effects models are particularly useful for modeling data from complex designs, their role in modeling treatment heterogeneity is investigated. It is shown that an individual treatment effect can be conceptualized as a linear combination of fixed treatment effects and random effects. The random effects are assumed to have variance components specified in a mixed effects "potential outcomes" model when both potential outcomes, r_T, r_C, are variables in the model. The variance of the individual causal effect is used to quantify treatment heterogeneity. Post treatment assignment, however, only one of the two potential outcomes is observable for a unit. It is then shown that the variance component for treatment heterogeneity becomes non-estimable in an analysis of observed data. Furthermore, estimable variance components in the observed data model are demonstrated to arise from linear combinations of the non-estimable variance components in the potential outcomes model. Mixed effects models are considered in context of a particular design in an effort to illuminate the loss of information incurred when moving from a potential outcomes framework to an observed data analysis.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I am indebted to faculty and staff in the Department of Statistics at Kansas State University for their direction and encouragement over the course of the past seven years. In particular, I would like to thank Nora Bello, John Boyer, Suzanne Dubnicka, Jim Higgins, and Leigh Murray for their advice, support, and instruction. I would also like to express my appreciation for Walt Stroup in the Department of Statistics at the University of Nebraska. His time on sabbatical at Kansas State and his approach to statistical modeling have profoundly affected my understanding of statistics.

This work would not have been possible without the help and guidance of my committee, Gary Gadbury, Abigail Jager, Jim Neill, and David Renter. At some point in my graduate career, each of them, perhaps unknowingly, has shared a timely word of encouragement, a piece of much-needed advice, or an example worth imitating. I cannot thank them enough for their work on my behalf. I would like to offer a very special thanks to Gary Gadbury. His knowledge, experience and creativity in statistics are surpassed only by his kind, gracious, and steadfast demeanor. His comments and suggestions may have shaped the direction of this research, but his mentorship, friendship, and professional example have shaped the direction of this statistician.

Finally, I would like thank my wife, Christy. Her sacrifice and support made this work possible; her friendship and love made it fulfilling.

Chapter 1 - Introduction

Treatment heterogeneity refers to the variability of a treatment effect across individuals in a population. The term *treatment effect* implies a comparison of one level of treatment against another. To state that a treatment effect varies across individuals implies that this comparison of treatment levels is made *within an individual*. Although such variability has often been acknowledged as an important consideration in the application of experimental findings to prospective individual experimental units (EU), many experimental design settings preclude the comparison of treatment within an individual EU. Consequently, a measure of treatment heterogeneity is not directly estimable. Therefore, in order to facilitate some sort of decision about the use of treatment in individual EU's, general statistical information is gathered about the average or mean effect and then that same information is applied to the individual (cf. Marshall, 1997). It should be noted, however, that the mean effect may be misleading when the effect of a treatment varies widely across individuals. If individual treatment variation is large with respect to the mean, then the mean treatment effect may appear to be favorable for one treatment over another while the other treatment may be more favorable for a non-negligible proportion of the EU's in the population.

Crossover designs are one type of experimental design that allows observation of an "individual treatment effect" because an individual crosses over from one treatment to another after a washout period, thereby providing observable responses to each of the two treatments. Therefore, they have been recommended as a design that provides more capability of evaluating treatment heterogeneity in a study (cf., Senn, 2001). However, estimating treatment heterogeneity, even in crossover designs, can involve assumptions that are not always explicitly stated or apparent in random effects models. More about these assumptions will be discussed in the next chapter.

Another approach to assessing treatment heterogeneity is the use of a potential outcomes framework. *Potential outcomes* (Rubin, 1974) can help elucidate the role of treatment heterogeneity in a statistical analysis. In this framework, an unobservable, individual treatment effect is defined. It is the variance of this individual difference that is of primary interest.

This research explores issues that arise when estimating a variance of individual treatment effects. This variance serves to quantify the degree of treatment heterogeneity in a

population. Concepts presented here should be useful for applications where estimating this variance, in addition to estimating a mean effect, may be of interest.

1.1 Potential Outcomes

1.1.1 Causal Effect and the Fundamental Problem of the Causal Inference

Consider a set of treatments, $\{T, C\}$ say, where T denotes some treatment that might be applied to an EU and C denotes a control that also might be applied to an EU. For each EU, consider the duplet $\{r_T, r_C\}$, which represents the potential response of the experimental unit if treatment were applied and the response of the experimental unit if control were applied, respectively. The *true causal effect* of T compared to C, denoted d, is the difference between the two potential responses. That is,

$$d = r_T - r_C \tag{1.1}$$

Notice that it is important to use terminology such as "imagine", "consider", or "conceptualize" when discussing potential outcomes as it is impossible to simultaneously observe all potential outcomes for a given experimental unit at a particular time. Only one of the potential responses is actually observable. This constraint of a potential outcomes framework has been called the fundamental problem of causal inference (Holland, 1986).

Although direct observation of the true causal effect is unachievable, the potential outcomes framework is still a very viable pedagogical tool for conceptualizing varying responses to the application of different treatments. As discussed in the next chapter, much work has been done to elucidate the statistical properties of a causal effect, given a particular set of assumptions.

1.1.2 The Randomization Mechanism and Naïve Difference

As noted above, only one potential response may be observed for a given EU at a given time. The question then becomes which of the potential outcomes should be selected for the observable outcome and how should that choice be made. From a statistical perspective, the inherent answer is to permit random chance to select the observable responses from the potential responses.

Define a random indicator variable, W_j , such that

$$W_j = \begin{cases} 1, if\ j^{th}\ experimental\ unit\ receives\ T \\ 0, if\ j^{th}\ experimental\ unit\ receives\ C \end{cases}$$

Define the observable outcome of the j^{th} experimental unit, R_j, as follows:

$$R_j = r_{T_j} \cdot W_j + r_{C_j} \cdot (1 - W_j)$$

where r_{T_j} and r_{C_j} are the potential responses of the j^{th} experimental unit. In potential outcomes literature, the probability distribution of W_j is referred to as the *randomization mechanism*.

Once the samples have been selected, define the usual mean difference using the observable outcomes

$$\bar{D} = \bar{R}_{T\cdot} - \bar{R}_{C\cdot} = \frac{1}{n_T}\sum_{j=1}^{N} r_{T_j} \cdot W_j - \frac{1}{n_C}\sum_{j=1}^{N} r_{C_j}(1 - W_j) \qquad (1.2)$$

where $\bar{R}_{T\cdot}$ is the arithmetic average of the n_T responses for those units whose potential response under T was selected to be observed and $\bar{R}_{C\cdot}$ is the arithmetic average of the n_C responses of those units whose potential response under C was selected to be observed. We distinguish $\bar{D}$ from the true individual causal effect given in (1.1) by referring to $\bar{D}$ as the *naïve difference* or the *naïve effect*. In a usual two-sample completely randomized design (CRD), the mean in (1.2) would be an estimator for a population mean. Here, however, it is interpreted more generally as a naïve effect in a CRD because it is the only effect that could be attributed to individuals and would be a naïve version of the true quantity in (1.1). In later designs, such as the matched pairs design, the naïve effect would be a paired difference and would serve as a naïve version for the true effect for two individuals in a pair. More distinction between "naïve" individual effects versus their true counterparts will be discussed in later sections.

Employing the randomization mechanism effectively removes one-half of the potential outcomes to yield the observable data. Removing one-half of the data alters the dataset in such a way that certain quantities become inestimable. Consequently, information about pertinent effects is lost. A reasonable question to ask is "What information is no longer available after implementing the randomization mechanism?" Answering this question is a key component to relating a potential outcomes model to an observable data model. Throughout this paper, it is assumed that estimable effects in a potential outcomes model that are no longer estimable after implementing the randomization mechanism are not removed from the model but, are confounded together to produce the "residual" term in the corresponding observable data model. Thus an observable data model produced from a potential data model contains the estimable effects in the potential outcomes model that remain estimable after implementing the randomization mechanism and the "residual" term consisting of the effects from the potential data model that are confounded.

1.1.3 Inference Space and Statistical Properties of Potential Outcomes

In the potential outcomes framework, we conceptualize the experimental process as the selection of a finite set of N duplets (F) from an infinite population of duplets (Ω). Each duplet contains the set of potential responses for an EU. A randomization mechanism is then employed to the duplets in F to select which EU's have their potential response under treatment selected as the observable response and which EU's have their potential response under control selected as the observable response. As in the "usual" experimental setting, the end result is a collection of n_T EU's receiving T and n_C EU's receiving C, where $N = n_T + n_C$. From a broad inference space perspective, the duplets are independent of one another, and the potential responses within a duplet follow the joint distribution:

$$\begin{pmatrix} r_{T_j} \\ r_{C_j} \end{pmatrix} \sim \left\{ \begin{pmatrix} \mu_T \\ \mu_C \end{pmatrix}, \begin{bmatrix} \sigma_T^2 & \rho_{TC} \cdot \sigma_T \sigma_C \\ \rho_{TC} \cdot \sigma_T \sigma_C & \sigma_C^2 \end{bmatrix} \right\} \tag{1.3}$$

where ρ_{TC} is the correlation of the potential outcomes. It should be expected that the two potential responses are correlated as they are potential responses of the same individual under different treatment conditions. The correlation, however, is non-estimable due to the fundamental problem of causal inference.

Much work has been done to elucidate the statistical properties of d, defined in (1.1), under certain sets of assumptions. With $\bar{D}$ given in (1.2), it can be shown that

$$E_{\Omega}[\bar{D}] = E_{\Omega}[E_W(\bar{D}|F)] = E_{\Omega}[\bar{d}] = \mu_d \tag{1.4}$$

where $\bar{d} = \frac{1}{N}\sum_{j=1}^{N} d_j$, where $\mu_d = \mu_T - \mu_C$, and where the unconditional expectation is with respect to the distribution in (1.3) from which the finite set *F* is selected.

Similarly, based on the properties of conditional variance and assuming uniform randomization,

$$var_{\Omega}[\bar{D}] = var_{\Omega}[E_W(\bar{D}|F)] + E_{\Omega}[var_W(\bar{D}|F)] = var_{\Omega}[\bar{d}] + E_{\Omega}[var_W(\bar{D}|F)]. \tag{1.5}$$

Notice that

$$var_{\Omega}[\bar{D}] \geq var_{\Omega}[\bar{d}] \tag{1.6}$$

with equality if and only if $var_W(\bar{D}|F) = 0$. In other words, the equality holds if all of the variability in the estimator $\bar{D}$ for μ_d is in the selection of the finite set *F* from the broader population. The inequality incorporates random variability resulting from the treatment assignment mechanism

1.2 Overview of Research

Identifying the presence of treatment heterogeneity is the first step in determining whether one treatment compared to another is uniformly more efficacious for all EU's within a given population or whether the efficacy of one treatment compared with another depends on the EU under consideration. If treatment heterogeneity exists, then it would be reasonable to try and identify the most effective treatment for a particular EU, based on the individual characteristics of that EU.

Treatment heterogeneity has been clearly defined in terms of the variance of a true causal effect by Gadbury et. al (2001), among others, using a potential outcomes framework. The statistical properties of this variance compared with the variance of a naïve effect have also been

considered from a finite population perspective, where the naïve effect depends on the design. More details on this and other pertinent results from published literature will be presented in Chapter 2. From an infinite population perspective, Senn (2001) discusses an estimable subject-by-treatment variance in a repeated measures crossover model. Based on results presented in Chapter 3, relating this subject-by-treatment variance to the variance of the true causal effect requires additional assumptions.

To my knowledge, no one has tied the quantities defined in a potential outcomes framework that describe treatment heterogeneity to the components of an infinite-population linear model. Linear models and, in particular, linear mixed models are quite flexible for modeling data from complex experimental designs. Investigating treatment heterogeneity in data from complex designs using a randomization based approach on a finite population becomes nearly intractable for complex designs (cf. Ndum et al., 2012). In particular, some designs analyzed by linear mixed models produce an estimate of a subject-treatment interaction variance component, but it is not clear how this component relates to the variance of true effects and/or what assumptions are required to equate the two. It is the goal of this research to, first relate potential outcomes to a linear model in a two-sample completely randomized design (CRD) and detail the loss of information that occurs when moving from a potential outcomes framework to an observable model setting. In addition, I will describe new information gained about treatment heterogeneity by considering increasingly complex experimental designs. In particular, I will show that, while the variance of the true causal effect remains inestimable, it can be bounded above, and in some designs, bounded above and below, by linear combinations of estimable variance components associated with random effects from the observable linear model. The purpose of this research is to clearly delineate the assumptions necessary to equate treatment heterogeneity in a potential outcomes framework to estimable components of an observable data model.

Chapter 3 presents the results of this process carried out under the assumptions of independent random effects and Gaussian data. In Chapter 4, issues raised in Chapter 3 concerning correlation and the relationship of model sums of squares to finite population variance estimates are resolved. Chapters 5 and 6 contain papers prepared for submission in peer-reviewed journals. Both chapters include extensions of the research in Chapter 3 to more complex treatment structures. Chapter 6 also includes a discussion of the extension of this work

to generalized linear mixed models. SAS codes used in Chapters 3 through 6 are standard SAS codes, and are available upon request. I conclude with a discussion of ideas for future work in Chapter 7.

Chapter 2 - Review of Literature

This chapter reviews the statistical literature on potential outcomes, treatment heterogeneity and linear models. This is not intended to be an exhaustive review of the pertinent literature on these topics, but it is intended to serve as a summary of the key contributions addressing the question at hand, namely how to model data when individual treatment heterogeneity is suspected.

2.1 Potential Outcomes

In 1990, Terrance P. Speed and Dorota M. Dabrowska edited and translated from Polish into English a 1923 publication by Jerzy Neyman in which he states,

"…let us consider a field divided into m equal plots and let $U_1, U_2, \cdots U_m$ *be the true yields of a particular variety on each of these plots…If we could repeat the measurement of the yield on the same fixed plot under the same conditions, we could use the above definition of the true yield. However, since we can only repeat the measurement of a particular observable yield, and this measurement can be made with high accuracy, we have to suppose that the observable yield is essentially equal to* U_i*…"*

Thus, we likely have one of the first references to what has come to be known as potential outcomes. In his discussion following the Dabrowska and Speed translation of Neyman's 1923 work, Rubin (1990, p.479), often himself credited with first formalizing the potential outcomes framework (1974), states, "Without a doubt, Neyman (1923) is an important, but previously unposted milestone, in statistics. …with respect to his definition of causal effects, although the underlying implicit definition was relatively common prior to 1923, Neyman certainly appears to be the first to formalize it."

Rubin (1974) utilized this potential outcomes framework to first formally define the causal effect of a treatment versus control as the difference in potential outcomes for a particular EU. Rubin highlights three important points related to a causal effect. First, a causal effect requires a comparison of two treatments. This point is reiterated by Holland (1986) in his

discussion of Rubin's Model for Causal Inference. Second, the causal effect cannot be measured since potential outcomes cannot be measured simultaneously. Holland referred to this property as the Fundamental Problem of Causal Inference. Rosenbaum and Rubin (1983) later wrote that this Fundamental Problem of Causal Inference can be construed as a missing data problem since either the potential outcome under treatment or the potential outcome under control is missing.

Finally, Rubin maintained that an assumption he termed stable unit treatment value assumption (SUTVA-Rubin, 1980, 1986) must hold in order for a question to be well formulated enough to have causal answers. This was a generalization of ideas described by Cox (1958). SUTVA is the a priori assumption that the value of the response for a particular EU exposed to a particular level of treatment will be the same regardless of how the assignment of treatment to the EU is made, and regardless of what levels of treatment are assigned to other EU's under consideration. This assumption should hold for all EU's under consideration in a study. For the purposes of this research, it is assumed that SUTVA holds for all experimental designs under consideration.

Potential outcomes are contrasted to observable outcomes, which can be thought of as the realization of one of the potential outcomes via some selection process. As noted previously, the inherent selection process for choosing which of the potential outcomes is selected as the observable outcome is a random process. Rubin (1978; p.34) states that a treatment assignment should be made according to a defined randomization mechanism and "…not according to ad hoc decisions of the experimenters or the subjects of experiments." He proceeds to describe a process under which an experimenter could move from a conceptual collection of data to an observed dataset. The conceptual data set includes all covariates measured on all EU's and all possible values of variables affected by level of treatment assigned to EU under every possible level of treatment. The observable dataset contains only pieces of information found in the conceptual dataset. As part of this process, Rubin (1978) defines a random vector, which can take on one of $t + 1$ values $0,1,2 \ldots t$, where t is the number of treatment levels under consideration. The probability distribution of this random vector is referred to as the randomization mechanism. Furthermore, Rubin(1978, p.42-43) describes circumstances under which the randomization mechanism is *ignorable*.

Rosenbaum and Rubin (1983) refined Rubin's (1978) concept of ignorable treatment assignment when they defined a strongly ignorable treatment assignment. They argued that the

conditional independence, or lack thereof, of the potential responses and randomization mechanism given a vector of possible covariates that affect both treatment assignment and potential responses is a characteristic difference between randomized and non-randomized trials. If this conditional independence exists, then the treatment assignment mechanism is said to be strongly ignorable. A strongly ignorable treatment assignment mechanism is a hallmark of a properly designed, randomized experiment. Unless otherwise noted, a strongly ignorable treatment assignment is assumed for the purposes of this research.

2.2 Treatment Heterogeneity

In a 1997 feature article concerning the foundations of personalized medicine, Andrew Marshall (p. 954) wrote,

"…Medicine today is geared around taking statistical information about the general population and then applying it to the individual…"

If either unit homogeneity or a constant effect (Holland, 1986) are valid assumptions in the experimental process, then this method of prescribing a level of treatment for a particular EU is valid. Holland defined unit homogeneity as the assumption that the same level of treatment applied to distinct EU's yields an identical response for each EU. The definition of constant effect permits distinct EU's receiving the same level of treatment to exhibit varying responses; however, from a potential outcomes framework perspective, it is assumed that the difference in potential outcomes *within* an individual EU is constant *across* EU's in a population.

The decision of selecting a particular level of treatment for an individual EU becomes increasingly complex if the true, causal effect of treatment compared with control varies across units of a population. While valid estimates of the mean response are still obtainable, the utility of applying average results to individual EU's is called into question. Hwang et. al (1978) discuss a phenomenon they observed in bioequivalence studies which they termed subject-by-formulation interaction. They pointed out that two treatments that appear similar on the average could perform very differently in individual subjects. Others have investigated the same phenomenon, although they may have used different terminology. Cox (1992) used the term treatment-by-patient interaction and Gadbury et al. (2001) defined what they termed subject-

treatment (S-T) interaction. All of these ideas attempt to capture the idea that heterogeneity of treatment effects exist at the individual level.

A consequence of this heterogeneity is that different individuals or groups of individuals may respond to treatment in opposite directions, with treatment T having higher efficacy for some and treatment C having higher efficacy for others. At times, this form of treatment heterogeneity may be accounted for by group or subset identification. The term qualitative interaction (QI) has been used to describe this condition at the subset level (Peto, 1982). Gail and Simon (1985) developed a test to detect a QI, and when such tests are significant, optimal treatments may differ across subsets (Byar and Corle, 1977).

Currently, the study of subset interaction alone may be too restrictive in light of existing research objectives in areas such as personalized nutrition, health care, and behavioral therapy (Lewis and Burton-Freeman, 2010; Marshall, 1997). For example, Kent and Hayward (2007, p. 1209) report, "There remain important differences between individuals in each treatment group that can dramatically affect the likelihood of benefiting from or being harmed by a therapy." The possibility of quantifying individual treatment heterogeneity brings the hope of identifying patients who may respond more favorably to one treatment over another based on personal attributes of the patient. However, there are those who view evaluating treatment heterogeneity from an individual perspective as a formidable challenge. For example, Senn (2001, p.1479) stated that personalized care "…May be rather more difficult to realize than has been supposed…"

Many methods that estimate a variance associated with treatment heterogeneity are actually evaluating observable consequences of treatment heterogeneity (e.g., variability across subsets of a population) rather than assessing treatment heterogeneity at the individual level. Hence, there is the necessity for a framework that can accommodate a single, individual EU. The potential outcomes framework is one such a framework. Other approaches may make assumptions that are not verifiable in observed data. For example, crossover designs have been utilized to try and quantify individual treatment heterogeneity. In such a case, one assumption would be that an observed individual treatment effect in a crossover design is equal to the true individual effect of treatment. The issues involved with making this type of assumption were recently discussed in Poulson et al. (2012). Senn (2001) notes that a subject-by-treatment effect is estimable in an observable, repeated-measures crossover design in which EU's are measured

more than once on each treatment. Even so, in order to completely characterize the variability of response as either between-EU variability or subject-by-treatment variability, one must assume no variability in EU effect over time and no variability in subject-by-treatment effect over time.

Using potential outcomes and adapting their notation to match that defined in Chapter 1, Gadbury et al. (2001) used the definition of a 'true' individual effect from (1.1) to delineate assumptions about $var(d)$. They show, given that (r_T, r_C) originate from an infinite bivariate normal distribution defined in (1.3), then

$$\sigma_d^2 = var(d) = \sigma_T^2 + \sigma_C^2 - 2\sigma_T\sigma_C\rho_{TC}. \qquad (2.1)$$

Notice that (2.1) can be bounded by taking $\rho_{TC} = \pm 1$, and estimating all other parameters in (1.3) from the observed data. Furthermore, they show that the proportion of the population receiving a harmful effect, or a negative effect, from T, is given by $P(d < 0) = \Phi\left(\frac{-\mu_d}{\sigma_d}\right)$, and may also be bounded. These bounds are given by

$$\Phi\left(\frac{-\mu_d}{\sqrt{\sigma_T^2 + \sigma_C^2 - 2\sigma_T\sigma_C}}\right) \leq P(d < 0) \leq \Phi\left(\frac{-\mu_d}{\sqrt{\sigma_T^2 + \sigma_C^2 + 2\sigma_T\sigma_C}}\right). \qquad (2.2)$$

Note that, as in (2.1), the upper bound is achieved when $\rho_{TC} = -1$ and the lower bound is achieved when $\rho_{TC} = 1$. Without loss of generality, assume $\mu_d > 0$. Then, when $\rho_{TC} = 1$ and $\sigma_T = \sigma_C$, a condition which indicates a constant individual effect (Holland, 1986), then $P(d < 0) = 0$. Gadbury and Iyer (2000) provide maximum likelihood estimates for the parameters $\mu_d, \sigma_T,$ and σ_C so that large sample confidence intervals can be placed on lower and upper bounds for $P(d < 0)$ using estimates from the observed data. They also consider the role of a covariate in tightening the bounds.

For certain designs, treatment heterogeneity has been accommodated in a general linear model (LM) or a linear mixed model (LMM) by including a subject-by-treatment effect. Wilk and Kempthorne (1955) modeled a subject-by-treatment effect as a fixed effect. First, they assumed a value of zero for the fixed subject-by-treatment effect in all subjects and all treatment combinations. Subsequent analyses assumed that the sum of fixed subject-by-treatment effects

over all units in a population receiving a particular treatment combination was zero. Ghosh and Crosby (2005) utilized clustering techniques in a crossover design to generate subgroups which they then considered replicates of one "subject" in order to estimate differences in subject-by-treatment effects. Kramer et al. (2011) presented a method in which they subtracted the estimated fixed effects from the observations in a crossover design and applied principle component analysis to residuals in order to isolate a subject-by-treatment effect.

2.2.1 Statistical Properties of d: Broad vs. Narrow Scope of Inference

McLean et al. (1991) define two possible scopes of inference: "the narrow inference space" and "the broad inference space." The narrow inference space presumes that once a finite set of EU's is selected from an infinite set, inference is specific to the finite set. A broad scope of inference extends inference to the population from which the finite set is selected. Extending the narrow scope of inference to the broad scope of inference is valid only if the finite set is representative of the broader population.

Historically, statistical inference on parameters in a potential outcomes framework has often been carried out under the assumption of a finite population from which a sample was taken. Neyman (1923), Rubin (1974) and Gadbury (2001) showed that the expectation with respect to the randomization distribution of the naïve effect is the causal effect in a two-sample CRD. That is,

$$E_W\left(\bar{D}|F\right) = \bar{d}.$$

Based on the properties of conditional expectation, it is rather straight-forward to see that both the naïve effect and the true causal effect are unbiased estimators of the true super-population difference, μ_d, as shown in equation (1.4). When considering the variance of the naïve effect with respect to the randomization distribution in a two-sample CRD, Neyman (1935) observed and Gadbury (2001) showed that the "natural" estimate of the finite population variance of the naïve effect taken with respect to the randomization distribution and computed from observable data is biased. That is, if we denote the "natural" estimate of $var_W\left(\bar{D}|F\right)$ based on observable data as $\widehat{var}_W\left(\bar{D}|F\right)$, which Gadbury (2001) considered to be the common pooled estimator of $var_W\left(\bar{D}|F\right)$, and take its expectation, then

$$E\{\widehat{var}_W(\bar{D}|F)\} \geq var_W(\bar{D}|F).$$

Furthermore, Gadbury(2001) showed that the bias was a function of the finite population variance of the causal effect taken with respect to the randomization distribution, $var_W(d|F)$.

The description of statistical properties of the true causal effect has also been extended to include more complex experimental designs than simply the two-sample CRD. In Gadbury et al. (2004), a matched-pairs design was considered where outcomes were binary and in Albert et al. (2005) a blocked design was considered with, again, binary outcomes. The latter paper produced nonparametric estimates in a randomization based framework. For continuous outcomes, results for estimating individual treatment heterogeneity in designs beyond a two-sample CRD were derived in the context of finite population, randomization-based inference. This was done for a matched-pairs design and a balanced two-period-two treatment crossover design (Gadbury 2001; see Gadbury, 2010, for a summary of some results). It should be noted that randomization techniques for deriving estimators of an S-T variance become increasingly intractable as designs become more complex.

Dawid (2000) elegantly considered the potential outcome framework from a broad scope of inference perspective. He clearly defined the joint distributional assumptions commonly imposed on the bivariate potential outcomes, and delineated the Fundamental Problem of Causal Inference as a problem of identifying the correlation between potential outcomes within an experimental unit. Furthermore, he also discussed the assumption of unit-treatment additivity and how the failure of this assumption to hold leads to a non-uniform causal treatment effect across EU's. He even noted the relationship of the variance of the naïve effect and the variance of the true effect given in equation (1.6) from a broad scope of inference perspective. Unfortunately, it seems that the ambiguity produced by the Fundamental Problem of Causal Inference soured Dawid on the potential outcomes approach as a pedagogical tool to investigate the nature of causation. He favored a decision-analytic approach in which he used the identifiable marginal distributions of responses under both treatment and control in addition to a specified loss function to predict the response of a future EU. It should be noted that Dawid (2000) considered only the two-sample CRD and did not explore the potential outcomes model in more complex experimental designs.

2.2.2 Treatment Heterogeneity and Unit-Treatment Additivity

In 1947, Cochran described the consequences of carrying out the usual analysis of variance (ANOVA) when basic assumptions were not satisfied. Four basic assumptions were considered, the first being that treatment effects and environmental effects, like block effects in a randomized block design or row and column effects in a Latin square design, should be additive, not multiplicative. Cox (1958, pp. 14-17) extended this idea of additivity from treatment effects and environmental effects to treatment effects and subject effects. Cox wrote that many fundamental experiments assume that the observation obtained when applying a particular treatment to a particular unit is assumed to be an additive relationship of a quantity depending only on the particular unit and a quantity depending on the treatment assigned. He noted that, assuming fixed treatment effects, one consequence to this additive assumption of units and treatments was that the true, causal effect was constant across subjects. Later, Cox (1992) termed this assumption of additivity between unit and level of treatment *unit-treatment additivity*. The statistical model based on this assumption is frequently referred to as the additive treatment model. Adapting Cox's (1992; p.295) notation to fit the notation presented in Chapter 1, this additive treatment model can be written

$$r_{T_j} = r_{C_j} + d \tag{2.3}$$

where d is assumed to be constant and $j = 1,2, \dots, N$.

Due to the Fundamental Problem of Causal Inference, the assumption of unit-treatment additivity cannot be directly checked. While no specific measures exist to show that unit-treatment additivity holds, there are several indicators that unit-treatment additivity fails to hold. One such indicator is considered below.

One of the fundamental consequences of the unit-treatment additivity assumption holding is that the dispersion of potential responses around some measure of center is the same for the potential responses under treatment as the potential responses under control. Thus, if the variance of the responses under T and the variance of the responses under C are vastly dissimilar, then this may be an indication that unit-treatment additivity does not hold. Cox (1992) recommends a non-linear transformation of the responses in order for unit-treatment additivity to be achieved. One example of such a transformation is the natural-logarithm transformation. Consider the case where

$$r_{T_j} = \delta \cdot r_{C_j},$$

for some $\delta > 1$. This is not an additive model as defined above. If this assumed model is true, it is very easily shown that

$$var\left(r_{T_j}\right) = \delta^2 \cdot var\left(r_{C_j}\right) > var\left(r_{C_j}\right).$$

But, by applying the natural logarithm transformation, it is possible to achieve an additive treatment model on the log-scale as follows:

$$\ln\left(r_{T_j}\right) = \ln(r_{C_j}) + \ln(\delta).$$

The above scenario is just one possible way in which dissimilar variances between outcomes receiving treatment and outcomes receiving control indicate a failure of the additive treatment model assumption to be satisfied. Consider a second situation which amounts to a variation on the additive treatment model given in (2.3) in which d is permitted to vary according to the experimental unit, rather than remaining constant across all experimental units. In essence, each EU is permitted its own causal effect. Again utilizing notation defined in Chapter 1 with Cox's (1992) notation, this model might be written as

$$r_{T_j} = r_{C_j} + d_j, j = 1,2, \dots, N. \tag{2.4}$$

A model of this form may arise as a result of interaction between level of treatment and a unit's covariate. These are the circumstances under which Gadbury et. al (2001) defined S-T interaction.

From a finite population perspective, where d_j is considered a fixed quantity, one difficulty in working with a model like that in (2.4) is that the number of parameters under consideration can quickly escalate. In a situation where a typical null hypothesis might be of the form, $H_0: d_1 = d_2 = \cdots = d_N$, the results of the test or estimates of a set of confidence intervals may be incomprehensible. Typically, methods are sought that reduce the dimension of the vector of parameters under consideration. One such approach is to take an infinite population perspective and consider the $d_j's$ as a random sample from some distribution such that

$$d_j \sim (\mu_d, {\sigma_d}^2)$$

thus reducing the complexity of the problem by considering only two parameters: $\mu_d, {\sigma_d}^2$. It should be straightforward to see that if model in equation (2.4) is true, then

$$var\left(r_{T_j}\right) = var\left(r_{C_j}\right) + {\sigma_d}^2 + 2 \cdot cov\left(r_{C_j}, d_j\right).$$

The variances of the potential outcomes are equal if and only if

$$\begin{aligned} {\sigma_d}^2 + 2 \cdot cov\left(r_{C_j}, d_j\right) = \ & (\sigma_T^2 + \sigma_C^2 - 2\sigma_T\sigma_C\rho_{TC}) + (2\sigma_T\sigma_C\rho_{TC} - 2\sigma_C^2) = \\ & \sigma_T^2 - \sigma_C^2 = 0. \end{aligned} \tag{2.5}$$

Otherwise, heteroscedasticity of variances exists.

To clearly understand the relationship between treatment heterogeneity and heteroscedasticity of variances, consider equation (2.1) as a function of ρ_{TC}, the correlation between potential outcomes given in (1.3). Notice that (2.1) achieves a maximum when $\rho_{TC} = -1$ and a minimum when $\rho_{TC} = 1$. Also note that when $\rho_{TC} = 1$, ${\sigma_d}^2 = (\sigma_T - \sigma_C)^2$ and when $\rho_{TC} = -1$, ${\sigma_d}^2 = (\sigma_T + \sigma_C)^2$. So even though ${\sigma_d}^2$ is not identifiable in an observable model setting, it can be bounded as follows:

$$0 \leq (\sigma_T - \sigma_C)^2 \leq {\sigma_d}^2 \leq (\sigma_T + \sigma_C)^2 \tag{2.6}$$

It should be clear from (2.6) that ${\sigma_\mathrm{d}}^2 = 0$ when $\sigma_T^2 = \sigma_C^2$ *and* $\rho_{TC} = 1$. Thus ${\sigma_\mathrm{d}}^2 = 0$ implies homoscedasticity of variances and (2.5) holds. It should also be clear from (2.6) that if heteroscedasticity of variance exists, then ${\sigma_d}^2 > 0$.

It should be noted that it is possible for $\sigma_T^2 = \sigma_C^2$ and yet the unit-treatment additivity assumption to still be violated. Note that (2.5) implies homoscedasticity of variances regardless of the value of ρ_{TC}. Thus if $\sigma_T^2 = \sigma_C^2$ but $\rho_{TC} \neq 1$, then (2.5) still holds even though ${\sigma_\mathrm{d}}^2 \neq 0$, indicating the presence of treatment heterogeneity. If it were possible to estimate some quantity that indicated the existence of treatment heterogeneity, then this estimate might provide evidence that the unit-treatment additivity assumption is violated, even when the variances of the potential outcomes are equal.

2.2.3 The Role of Covariates in Identifying Treatment Heterogeneity

As it has already been noted, the nature of treatment heterogeneity and its impact on choice of treatment for an individual EU has interested researchers from a variety of fields for decades. In particular there is a wide assortment of subset treatment heterogeneity literature in clinical trials research. Subset treatment heterogeneity differs from individual treatment heterogeneity in that subset interaction (SI) occurs when the effects of T and/or C change based on the subset identifiable by an observable covariate (Milliken and Johnson, 1984, p. 113). As Poulson et al. (2012) point out, individual treatment heterogeneity can be construed as subset treatment heterogeneity with the size of the subset equal to 1 EU. Therefore, individual treatment heterogeneity might be considered one form of subset treatment heterogeneity and it would seem beneficial to consider methods developed to identify and interpret subset interaction based on observable covariates in an attempt to elucidate the nature of individual treatment heterogeneity.

Byar and Corle (1977) began to develop the use of multivariate regression methods to define subsets for which T or C may be superior; however, they cautioned that "The proof of any conclusions tentatively drawn must depend on future experiments designed specifically to test the results suggested by the analysis" (Byar and Corle , 1977; p. 458). Later, Peto (1982) distinguished between quantitative subset interaction and qualitative subset interaction, the former meaning a change in magnitude of effect only across subsets, and the latter taken to mean a change in magnitude and direction of effects across subsets. Gail and Simon (1985), Silvapulle (2001), and Li and Chan (2006) all developed formal tests for qualitative interaction based on subsets formed using values of observable covariates.

While no such formal test for the existence of individual treatment heterogeneity has been developed, covariate information has been used to gain additional information about model parameters that would indicate the presence of individual treatment heterogeneity. Gadbury et al. (2001) showed that using a continuous covariate, say Z, that is not affected by the treatment and that augments the potential outcomes, the overall variability of individual effects can be reduced. The results shown here have been adapted to accommodate the notation presented in Chapter 1. Assume that the distribution of d given $Z = z_0$ is normal with conditional mean

$$\mu_{d|Z=z_0} = \mu_T - \mu_C + (\beta_{TZ} - \beta_{CZ})(z_0 - \mu_Z) \tag{2.7}$$

and conditional variance

$$\sigma_{d|Z}^2 = \sigma_{T|Z}^2 + \sigma_{C|Z}^2 - 2\sigma_{T|Z}\sigma_{C|Z}\rho_{TC|Z}. \quad (2.8)$$

β_{TZ} and β_{CZ} in (2.7) are the slope coefficients between Z and r_T and Z and r_C, respectively, and $\rho_{TC|Z}$ in (2.8) is the partial correlation of r_T and r_C given Z. The conditional variances, $\sigma_{T|Z}^2$ and $\sigma_{C|Z}^2$, are allowed to be different across the two treatment groups but are assumed not to depend on the value of Z. Coupled with (2.7) and (2.8), Gadbury et al. (2001) showed that

$$\sigma_d^2 = \left(\sigma_{T|Z} - \sigma_{C|Z}\right)^2 + 2\sigma_{T|Z}\sigma_{C|Z}\left(1 - \rho_{TC|Z}\right) + (\beta_{TZ} - \beta_{CZ})^2\sigma_Z^2.$$

Therefore, if evidence showed that $\beta_{TZ} \neq \beta_{CZ}$, $\sigma_{d|Z}^2$ may be less than σ_d^2 making it possible to reduce the bounds on $P(d < 0)$ over $Z = z_0$. Thus similar to (2.2) the proportion of the population receiving a harmful effect under T for a particular value of $z = z_0$ may be bounded by

$$\Phi\left(\frac{-\mu_{d|Z=z_0}}{\sqrt{\sigma_{T|Z}^2+\sigma_{C|Z}^2-2\sigma_{T|Z}\sigma_{C|Z}}}\right) \leq P(d < 0)_{Z=z_0} \leq \Phi\left(\frac{-\mu_{d|Z=z_0}}{\sqrt{\sigma_{T|Z}^2+\sigma_{C|Z}^2+2\sigma_{T|Z}\sigma_{D|Z}}}\right), \quad (2.9)$$

by letting the partial correlation $\rho_{TC|Z}$ be 1 and -1, respectively. Confidence intervals for the bounds on $P(d < 0)_{Z=z_0}$ given in (2.9) can be derived using bootstrap samples from the observed data or using asymptotic properties of maximum likelihood estimators (cf. Gadbury et al., 2001).

Zhang et al. (2013) used covariate information to tighten the bounds given in Gadbury et al. (2004) for the proportion of a population experiencing a detrimental treatment effect when potential responses were binary instead of continuous (i.e.- a potential response under C indicating success and a potential response under T indicating failure). Methods were presented under three sets of assumptions pertaining to the conditional independence of potential responses given a set of covariates.

2.3 Linear Models

As stated in section 1.2 of Chapter 1, one of the purposes of this research is to tie the potential outcomes framework to a linear model. The following section briefly reviews the pertinent literature pertaining to the development of statistical linear models.

Statistical models are concerned with relating the observations from a set of data to a set of components that is believed to give rise to the dataset. Based on statistical models, an attempt is made to make inference about these components. In earliest forms, a statistical model required three parts: the observation, the deterministic component, and the random components. Deterministic components (also referred to as systematic components) are considered to be *determined* by the level of treatment assigned to a particular EU. These deterministic components are assumed to be fixed constants. The random components describe how each individual response varies about the systematic component. As Stroup (2013) notes, the random component is a characterization of the uniqueness of the individual EU. By carefully stating relevant assumptions, the most common form of a statistical model takes the following generic form (Gbur et. al,2012):

$$observation = deterministic\ component + random\ component$$

While, technically, statistical models are approximations and it is unlikely that data are generated according to such a pedestrian process, the development of more complex approximations based upon this simple linear relationship has provided meaningful methods (i.e. logistic regression, Generalized Linear Mixed Models, etc.) of analyzing data that are vastly different than those data typically presented in an introductory statistical setting.

2.3.1 General Linear Models (LM's) and Linear Mixed Models (LMM's)

A complete history of the origins of the statistical linear model is well beyond the scope of this dissertation. Even if it were to be attempted, it would be imprudent to think that this author would be able to offer much in the way of additional information to what has already been summarized by those who are far more qualified to give attention to the subject. The interested reader is referred to the following three works in particular for a rather detailed history of the

general linear model: Eisenhart (1947), Scheffe (1956), and Searle et. al (1992, chapter 2). The following section is simply a brief overview of what is contained therein.

It is interesting to note that statistical modeling seems to have originated in the field of astronomy. In particular, Scheffe (1956, p.255) notes "…Very explicit use of a variance-components model for the one-way layout is made by Airy (1861, Part IV), with all the subscript notation necessary for clarity…Airy assumes the following structure for the j^{th} observation on the i^{th} night:

$$y_{ij} = \mu + c_i + e_{ij}$$

where μ is the general mean or 'true' value and the $\{c_i\}$ and $\{e_{ij}\}$ are random effects…" Searle et al. (1992) detail additional contributions to statistical modeling and variance component estimation throughout the latter part of the 19th century and the early part of the 20th century including the likes of Tippett (1931), Fisher (1918, 1925; although he did not explicitly apply linear modeling) and Neyman (1935).

Eisenhart (1947) distinguished between two types of linear statistical models, which he termed Model I and Model II. The former has come to be known as the fixed-effect or general linear model (LM) while the latter has come to be known as the random-effects model. Under the assumptions of the LM, responses are independently distributed, Gaussian random variables with a common variance and a mean that is taken to be fixed constant. Means of the responses may possibly differ depending on which level or combination of treatment factors are applied to the EU, however any difference between two means of interest is also taken to be a fixed quantity. Under the assumptions of the random-effects model, all treatment factors that are thought to affect the value of a response are considered random variables with a common mean of zero, but possibly different variances for each factor. Thus all observations, regardless of level or combination of treatment factor applied to the EU, are thought to vary around one common mean. Statistical models containing both fixed and random effects have been termed linear mixed models (LMM).

Over the past 40-50 years, statistical modeling has become a foundation in most introductions to statistical analysis. As such, there is a vast body of literature detailing methods for estimation of mean treatment effects, variance component estimation, inference procedures,

and confidence interval estimation in LM' s random effect models and LMM's. Among the most notable are Searle (1971), Rao (1973), Graybill (1976), and Hocking (1985).

2.3.2 Generalized Linear Models(GLM's)

In the previous section, it was noted that the LM and the LMM had the following two defining characteristics: first, the random components of the model were assumed to follow a normal distribution; second, the *responses* were modeled as a linear combination of fixed and random effects. The natural sequela of such investigations is to consider a scenario in which the responses do not follow a normal distribution. The following sections summarize the pertinent literature pertaining to such an investigation.

Although analyses pertaining to certain instances of non-normal data existed dating back to the mid 1930's, usually incorporating some transformation of the data, it was Nelder and Wedderburn (1972) that clearly described a theory for modeling non-normally distributed data which they termed *Generalized* Linear Models (GLM), so as not to be confused with the general linear model (LM) of the previous section. They described a method in which they used iterative weighted linear regression to arrive at maximum likelihood estimates of distribution parameters for distributions that were members of the exponential family. Furthermore, they modeled the mean of the responses as a monotonic transformation of a linear model. It should be noted that the linear models in this context contained fixed effects only.

Wedderburn (1974) extended these results so that in order to obtain parameter estimates, one need not know the actual distribution of the data, but must specify a quasi-likelihood function which is a function the defines the relationship between the mean and variance of the distribution. Wedderburn (1974) showed that a quasi-likelihood function possessed properties similar to properties of log-likelihood functions and thus maximum quasi-likelihood estimates of the distribution parameters could be obtained using iterative estimation procedures. Finally, he demonstrated that estimates obtained using maximum likelihood estimation as in Nelder and Wedderburn (1972) were a particular case of the quasi-likelihood approach.

2.3.3 Generalized Linear Mixed Models (GLMM)

After Nelder and Wedderburn (1972) published their results on GLM's, the next logical progression was to try and extend the GLM to include both fixed effects and random effects in the monotonically transformed linear model. Models that included both fixed and random

effects for non-normal data have become known as Generalized Linear Mixed Models (GLMM). Many researchers sought to do this from the mid 1980's to the mid 1990's. Two papers in particular are noted here. Breslow and Clayton (1993) and Wolfinger and O'Connell (1993) both demonstrated that estimates for fixed effects and random effects could be found by solving what have been termed the general mixed model equations (cf. Littell et. al, 2006; Ch. 14) which are a type of extension of mixed model equations to a non-normal setup. In both papers, iterated techniques were used to arrive at solutions rather than more cumbersome numerical methods that had been used previously to estimate effects in GLMM's. The difference between Breslow and Clayton (1993) and Wolfinger and O'Connell (1993), as the latter pointed out, was the assumptions about the values that certain model parameters could take. By constraining the dispersion or scale parameter defined in Wolfinger and O'Connell to equal 1, they demonstrated an equivalent analysis to that produced by Breslow and Clayton (1993). Thus Wolfinger and O'Connell's (1993) method may be thought of as a generalization of Breslow and Clayton's (1993) method. It should also be noted that Wolfinger and O'Connell's work (1993) forms the basis of the theory underlying PROC GLIMMIX in SAS.

Chapter 3 - Completed Research: Gaussian Data

3.1 Model Assumptions and Simulation Methods

This chapter presents results for potential outcome models and observable data models in each of five common experimental designs, assuming Gaussian random effects. The five experimental designs included the Two-Sample CRD, the Matched Pair Design (RCBD); the Generalized Complete Block Design (GRCBD) containing two observations per level of treatment, T and C within each block; the Two-Period-Two-Treatment Crossover Design; and the Repeated Measures Crossover Design with Two Treatments where each level of treatment is randomly assigned to two of four total time periods for each EU. Some of the material presented in this chapter on the CRD and RCBD designs has been reviewed and published in Richardson and Gadbury (2012).

Stroup (2013) developed a method termed What Would Fisher Do (WWFD) to correctly identify the components of the LMM. This method was based on the contribution Fisher made to a discussion paper authored by Yates (1935). We adapted this method and applied it to the potential outcomes framework to identify the potential LMM's for the experimental designs presented in the subsequent sections. As in Wilk and Kempthorne (1955), we assume no technical error.

For each of the five experimental designs, models were considered for each of two variance/covariance structures. The first structure assumes that all random effects are mutually independent of one another and that each random effect has its own variance component that is common to both levels of treatment, T and C. This variance structure will be referred to as the *common variance* structure. The second structure still assumes mutual independence of random effects, however outcomes under treatment are permitted a distinct variance component from outcomes under control. This variance structure will be referred to as the *distinct variance* structure. Only pertinent results for the distinct variance structure will be given in this chapter. See Appendix A for a set of complete results, including the common variance structure. Under both sets of assumptions for all experimental designs, the expectation of all random effects is assumed to be zero. With this structure, the treatment heterogeneity variance for the particular design is derived using potential outcomes and is shown to be linear combinations of variance components. Then the model is defined in terms of observable data and, where appropriate, the

variance of a naïve version of a treatment heterogeneity is derived. The connections between the naïve version and the true variance of individual effects are then established. The assumptions required to equate the two, or to bound the latter by estimable quantities are stated.

Derived results are illustrated using simulated data. Using SAS statistical software, potential outcomes data were simulated for each experimental design, under relevant assumptions. A total of $S = 100$ simulations were performed. Within each simulation, data were simulated for three distinct sample sizes. Unless otherwise specified, it is assumed that there are N total EU's in an *observable* experiment. Consequently, there are tN responses in a potential outcomes framework, one response for each of t levels of treatment imagined to have been simultaneously applied to each of N EU's. For all experimental designs in the following sections, $t = 2$. The resulting number of responses in each potential outcome framework will be highlighted for each experimental design in the results sections below. Where applicable, n_T and n_C refer to the number of subjects per treatment level, T and C, respectively. For the purposes of these simulations, we assumed designs were balanced. That is, we assumed $n_T = n_C = n$.

PROC GLIMMIX was then utilized on the simulated data to obtain REML estimates of: (1) the difference in fixed treatment effects between the two potential outcomes, (2) the variances of the random effects included in the potential model, and (3) the variance of the difference in the two potential outcomes, denoted $var(d)$.

Next, one-half of the data were removed to simulate observed data under uniformly random treatment assignment. Of the observations that were removed, one-half were treatment potential responses, and one-half were control potential responses. PROC GLIMMIX was again utilized on the observed data to obtain REML estimates of: (1) the difference in fixed treatment effects between the two treatment groups, (2) the variances of identifiable random effects in the observable model, (3) the variance of the linear combination of non-identifiable random effects that constitute the residual term or error variance in the observable data model, and (4) the variance of the naive difference in observable data, denoted $var(D)$.

Boxplots of estimates resulting from the $S = 100$ simulations were plotted for each of the three sample sizes to examine the shape and spread of the distribution of parameter estimates. The mean, median, minimum, maximum and standard deviation of the 100 parameter estimates

were computed. Then the mean of the $S = 100$ simulations was compared to the simulated value for each of the respective estimates.

3.2 Two-Sample CRD

Table 3.1 gives the effects and assumptions for both potential and observable models. A direct relationship between the two models is established by defining

$$e_{ij} = s_j + s\tau_{ij} \tag{3.1}$$

since multiple observations per subject are "lost" when the randomization mechanism is invoked. Thus the residual term in the observable two-sample CRD consists of the confounded subject and subject-by-treatment effects from the potential model. If such confounding occurs, then

$$\sigma_{ei}^2 = \sigma_s^2 + \sigma_{si}^2;\ i = T, C. \tag{3.2}$$

by the independence assumptions given in Table 3.1. Under the assumption of unit-treatment additivity, $s\tau_{ij} = 0$ for all i and j and

$$e_{ij} = s_j.$$

Thus

$$\sigma_{eT}^2 = \sigma_{eC}^2 = \sigma_s^2$$

irrespective of the level of treatment assigned to the j^{th} EU.

Model	Model Parameters	Assumptions
Potential Model	$r_{ij} = \mu + \tau_i + s_j + s\tau_{ij}$, $i = T, C$; $j = 1,2, \dots, N\ EU's$	$s_j \sim iid\ N(0, \sigma_S^2)$ $\begin{bmatrix} s\tau_{Tj} \\ s\tau_{Cj} \end{bmatrix} \sim MVN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{sT}^2 & 0 \\ 0 & \sigma_{sC}^2 \end{pmatrix}\right]$ s_j and $s\tau_{ij}$ are independent.
Observable Model	$R_{ij} = \mu + \tau_i + e_{ij}$, $i = T, C$; $j = 1,2, \dots, n_i$ $EU's\ per\ level\ of\ trt$	$\begin{bmatrix} e_{Tj} \\ e_{Cj} \end{bmatrix} \sim MVN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{eT}^2 & 0 \\ 0 & \sigma_{eC}^2 \end{pmatrix}\right]$

Table 3.1 *Model effects and assumptions in a 2-sample CRD.*

Define the true causal effect to be the difference in potential outcomes for the j^{th} EU. That is

$$d_j = r_{Tj} - r_{Cj} \ . \tag{3.3}$$

Given the potential model and assumptions in Table 3.1, the variance of the true causal effect is readily seen to be

$$var(d_j) = var(\mu + \tau_T + s_j + s\tau_{Tj} - \mu - \tau_C - s_j - s\tau_{Cj}) =$$
$$var(\tau_T - \tau_C + s\tau_{Tj} - s\tau_{Cj}) = var(s\tau_{Tj} - s\tau_{Cj}) = \quad \sigma_{sT}^2 + \sigma_{sC}^2 \tag{3.4}$$

Since only one observation per EU is recorded, an individual naïve effect is undefined in the 2-sample CRD. However, it is possible to compute the variance of an average naïve effect, $\bar{D}$, defined in (1.2). Under the model assumptions given above, the variance of the average naïve effect is given by

$$var(\bar{D}.) = var\left(\frac{1}{n_T}\sum_{j=1}^{n_T} R_{T_j} - \frac{1}{n_C}\sum_{j=1}^{n_C} R_{C_j}\right) =$$
$$\frac{1}{{n_T}^2} var\left(\sum_{j=1}^{n_T} s_j + s\tau_{Tj}\right) + \frac{1}{{n_C}^2} var\left(\sum_{j=1}^{n_C} s_j + s\tau_{Cj}\right) =$$
$$\frac{\sigma_s^2 + \sigma_{sT}^2}{n_T} + \frac{\sigma_s^2 + \sigma_{sC}^2}{n_C} = \frac{\sigma_{eT}^2}{n_T} + \frac{\sigma_{eC}^2}{n_C} = \frac{\sigma_{eT}^2 + \sigma_{eC}^2}{n}, when\ n_T = n_C = n. \tag{3.5}$$

Writing the result in (3.5) in terms of quantities from the potential model in Table 3.1 gives,

$$var(\bar{D}.) = \frac{(2\sigma_s^2 + \sigma_{sC}^2 + \sigma_{sT}^2)}{n}$$
$$\Rightarrow$$
$$n \cdot var(\bar{D}.) = (2\sigma_s^2 + \sigma_{sC}^2 + \sigma_{sT}^2) = \sigma_{eT}^2 + \sigma_{eC}^2 = 2\sigma_s^2 + var(d_j) \tag{3.6}$$

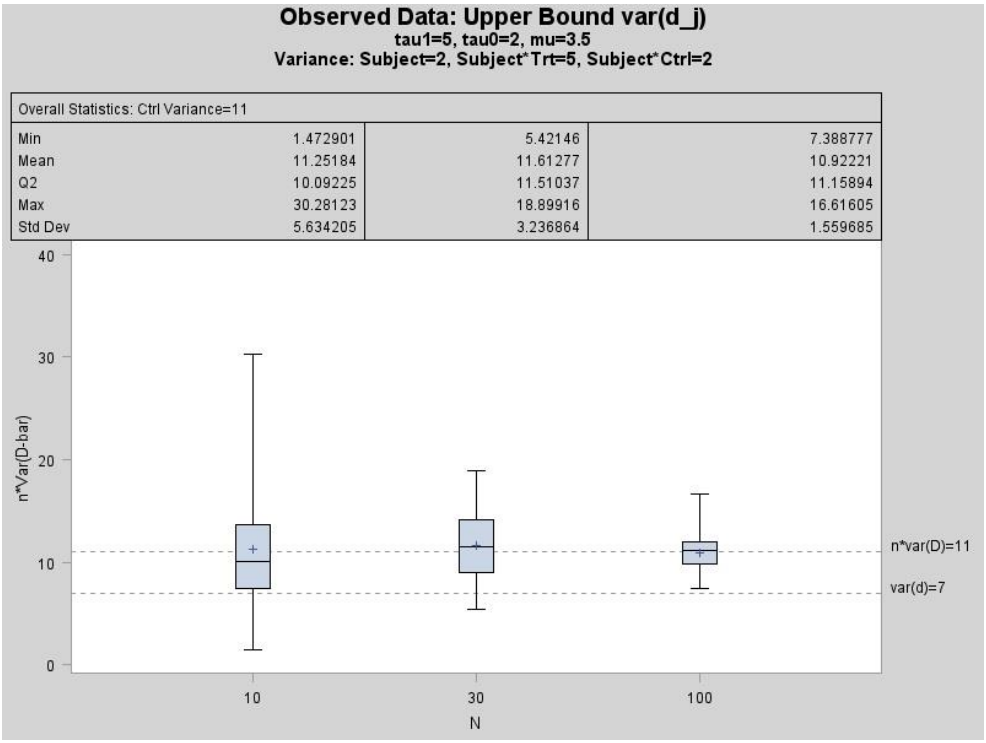

Figure 3.1 *Bounding the Individual Causal Effect: 2-Sample CRD.* $n \cdot var(\bar{D}.) \geq var(d_{ij})$. *Box plots of the* $S = 100$ *estimates of* $n \cdot var(\bar{D}.)$ *at N=10, 30, and 100 Dotted lines represent values used in the simulation design.*

when the design is balanced. $var(\bar{D}.)$ is estimable in observable data but the individual components are not. As demonstrated in Figure 3.1, one can see that $n \cdot var(\bar{D}.)$ is an estimable upper bound for $var(d_j)$, the variance of individual effects. Equation (3.6) also demonstrates that equality of $var(d_j)$ and $var(\bar{D}.)$ is achieved when $\sigma_S^2 = 0$. Recall that σ_S^2 is the variance attributed to EU's, so equality of $var(d_j)$ and $var(\bar{D}.)$ would require that all $j = 1,2, \ldots N$ EU's in the experiment be identical to one another in every respect except which level of treatment they were assigned to receive.

A comparison of $var(d_j)$ with $var(\bar{D}.)$ might seem a bit unusual since $var(\bar{D}.)$ is computed based on aggregate information from a sample and $var(d_j)$ is computed based on information available from a single EU. Therefore it is possible to define

$$\bar{d}. = \frac{1}{N}\sum_{j=1}^{N}\left(r_{T_j} - r_{C_j}\right) \tag{3.7}$$

and note

$$var(\bar{d}_\cdot) = \frac{1}{N^2} var\left[\sum_{j=1}^{N}\left(r_{T_j} - r_{C_j}\right)\right] = \frac{1}{N^2} var\left[\sum_{j=1}^{N}\left(s\tau_{Tj} - s\tau_{Cj}\right)\right] =$$

$$\frac{\sigma_{sT}^2 + \sigma_{sC}^2}{N} = \frac{var(d_j)}{N} \tag{3.8}$$

Combining the results of (3.6) and (3.8), note that

$$var(\bar{d}_\cdot) = \frac{var(d_j)}{N} \leq \frac{n \cdot var(\bar{D}_\cdot)}{N} = \frac{var(\bar{D}_\cdot)}{2}$$

$$\Rightarrow$$

$$var(\bar{d}_\cdot) \leq \frac{var(\bar{D}_\cdot)}{2} \tag{3.9}$$

when a two-sample CRD is balanced.

Figure 3.2 illustrates the results of (3.9). For each sample size, boxplots of the $S = 100$ values for $\frac{var(\bar{D}_\cdot)}{2}$ and $var(\bar{d}_\cdot)$ are shown. Estimates of $\frac{var(\bar{D}_\cdot)}{2}$ are shown in blue and estimates for $var(\bar{d}_\cdot)$ in red. For each sample size, the mean value of the 100 estimates of $\frac{var(\bar{D}_\cdot)}{2}$ is greater than the mean value of the 100 estimates of $var(\bar{d}_\cdot)$.

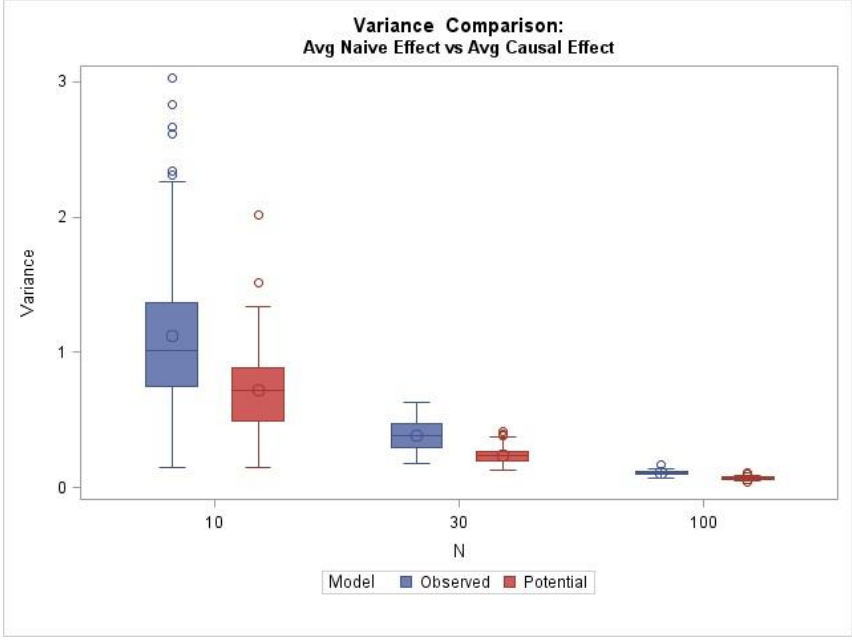

Figure 3.2 *Bounding the Average Causal Effect: 2-Sample CRD.*

Tables 3.2 (i), 3.2 (ii), and 3.2 (iii) give more specific results of all effects of interest based on $S = 100$ simulated data sets. Values represent the mean and standard error of estimates across the $S = 100$ data sets. Table 3.2 (i) gives results for the fixed treatment effect for the model fit to both potential and observable data, Table 3.2 (ii) shows the results for the random effects in the potential model and Table 3.2 (iii) the results for the random effects in the observable model. In all cases, as the sample size increased from 10 to 30 to 100, the variability of the effect estimates around the true simulated value decreased, and in all cases, the estimated value of the simulation parameter based on the $S = 100$ simulations is within 2 standard errors of the true value. Comparing the standard errors of the estimates between potential data and observable data in Tables 3.2 (ii) and 3.2 (iii) reveals a larger standard error for the observable estimates. This is to be expected as the observable estimates are computed from half the data, compared with the potential data.

Of particular note is that the estimates of $var(d_j)$ given in Table 3.2 (ii) seem to be reasonable estimates of the theoretical value derived in (3.4). In these simulations, $\sigma_{sT}^2 = 5$ and $\sigma_{sC}^2 = 2$. Thus by (3.4), $var(d_j) = \sigma_{sT}^2 + \sigma_{sC}^2 = 5 + 2 = 7$. Indeed, Table 3.2 (ii) demonstrates that the potential model estimates of $var(d_j)$ were within two standard errors of 7 for each of the three distinct sample sizes. Furthermore, notice that the estimates for σ_{eT}^2 and σ_{eC}^2 given in Table 3.2 (iii) also seem to be reasonable estimates of the theoretical value derived in (3.2), where it was assumed that the subject and subject-by-treatment effects from the potential model were confounded to form the residual term in the observable model. Assuming $\sigma_s^2 = 2, \sigma_{sT}^2 = 5$ and $\sigma_{sC}^2 = 2$, then $\sigma_{eT}^2 = 7$ and $\sigma_{eC}^2 = 4$ based on (3.2). The results in Table 3.2 (iii) demonstrate that the estimates of σ_{eT}^2 and σ_{eC}^2 are within two standard errors of 7 and 4, respectively, for each of the three sample sizes considered.

For the two-sample CRD, a comparison was made of two methods for computing estimates of both $var(d_j)$ and ρ_{TC}, the correlation seen in the distribution specified in equation (1.3). Recall that neither quantity is estimable in an observable model. As such, this comparison was made in the potential model only. Estimates of $var(d_j)$ were computed using one of two methods. The first method, termed *Model* $var(d_j)$ and denoted $\widehat{var}(d_j)$, was computed by summing the variance component estimates obtained from the PROC GLIMMIX procedure.

Fixed Effect (Potential)	Simulated Value	2N	Average ($S = 100$)	Std. Error ($S = 100$)
$\tau_T - \tau_C$	3	20	3.11	0.10
		60	3.04	0.05
		200	3.03	0.03

Fixed Effect (Obs.)	Simulated Value	N	Average ($S = 100$)	Std. Error ($S = 100$)
$\tau_T - \tau_C$	3	10	3.08	0.14
		30	3.08	0.09
		100	3.07	0.05

(i)

Potential Variance	Simulated Value	2N	Average ($S = 100$)	Std. Error ($S = 100$)
Subject	2	20	1.85	0.18
		60	2.15	0.11
		200	2.03	0.06
*Subject*Trt*	5	20	5.06	0.27
		60	5.25	0.17
		200	4.96	0.10
*Subject*Ctrl*	2	20	2.11	0.16
		60	1.98	0.10
		200	1.99	0.06
$var(d_j)$	7	20	7.18	0.31
		60	7.22	0.18
		200	6.95	0.11

(ii)

Observable Variance	Simulated Value	N	Average ($S = 100$)	Std. Error ($S = 100$)
Trt Residual	7	10	7.25	0.50
		30	7.33	0.27
		100	6.82	0.15
Ctrl Residual	4	10	4.00	0.27
		30	4.29	0.16
		100	4.10	0.08

(iii)

Table 3.2 *2-Sample CRD Simulation Results.*
Values represent the average and standard error of treatment effect estimates across $S = 100$ simulations in both the potential and observable data models for N=10, 30, and 100 for (i) Fixed Effects. (ii) Potential Random Effects. (iii) Observable Random Effects.

The second method entailed computing the difference in potential responses for each subject and then estimating the variance of these differences using PROC UNIVARIATE in SAS. This method is termed *Estimated* $var(d_j)$ and denoted $\widehat{var}(d_j)$. Table 3.3 gives the results of this comparison for one of the $S = 100$ simulations only. Results of the comparison in the 2-sample CRD demonstrate that the two methods of estimation yield identical estimates for all 3 sample sizes.

Estimation of ρ_{TC} was also carried out using one of two methods. The first method, denoted *Model Correlation*, estimated the intra-class correlation coefficient from the variance component estimates from PROC GLIMMIX. That is

$$\widetilde{\rho_{TC}} = \frac{\hat{\sigma}_S^2}{\sqrt{\hat{\sigma}_S^2 + \hat{\sigma}_{ST}^2} \cdot \sqrt{\hat{\sigma}_S^2 + \hat{\sigma}_{SC}^2}}.$$

Model correlation estimates are required to be non-negative by the assumptions given in Table 3.1. The default procedure of PROC GLIMMIX for handling negative variance component estimates is to replace the negative estimate equal to zero. Thus, anytime PROC GLIMMIX encountered a negative estimate of σ_S^2, the estimate of Model Correlation was also zero. The second estimate of ρ_{TC}, denoted $\widehat{\rho_{TC}}$, was computed by passing the simulated potential outcomes to PROC CORR in SAS where the Pearson correlation coefficient was computed. That is,

$$\widehat{\rho_{TC}} = \frac{\sum_{j=1}^{N}(r_{Tj} - \bar{r}_{T\cdot})(r_{Cj} - \bar{r}_{C\cdot})}{\sqrt{\sum_{j=1}^{N}(r_{Tj} - \bar{r}_{T\cdot}) \cdot \sum_{j=1}^{N}(r_{Cj} - \bar{r}_{C\cdot})}}.$$

This method is termed *Estimated Correlation* and permitted negative correlation coefficient estimates. Results given in Table 3.3 indicate that the two methods yielded identical estimates of correlation. This provides reassurance that the linear mixed effects model is providing estimates of the correlation in potential outcomes data that yields the same value as Pearson's computed correlation on the set of N bivariate potential outcomes. Recall that ρ_{TC} is the only quantity given in equation (2.1) that is nonestimable from observable data. Therefore since ρ_{TC} is nonestimable in an observable model, $var(d_j)$ is nonestimable in an observable model. As

2N	Model $var(d_j)$	Estimated $var(d_j)$	Model Correlation	Estimated Correlation
20	3.87	3.87	0.20	0.20
60	11.90	11.90	0.16	0.16
200	6.98	6.98	0.36	0.36

Table 3.3 *Different Methods of Estimation: 2-Sample CRD. Comparison for var(dj) and* ρ_{TC} .

such, any attempt to describe the loss of information incurred by moving from a potential model to an observable model ought to appropriately estimate ρ_{TC} in the potential data setting. The only times these two estimates of ρ_{TC} differed were when PROC GLIMMIX encountered a negative estimate of σ_S^2 and replaced the estimate with zero. The corresponding Estimated Correlation estimate was always negative in such situations. Specifying the potential LMM in such a way as to accommodate a negative correlation between potential outcomes under treatment and potential outcomes under control is discussed further in Chapter 4.

The connection between models for potential versus observable outcomes when evaluating individual treatment heterogeneity lacks some intuition in the CRD because there is not an actual naïve individual effect that can be defined, other than the sample mean difference. Other designs provide more intuition by having a naïve effect that makes more sense when attributing it to the individual.

3.3 RCBD

Table 3.4 gives the effects and model assumptions for the matched-pairs analysis. These results are easily extended to a conventional randomized complete block design, but for the purposes of these simulations, only the matched-pairs design is considered here. A direct relationship between the observable model and the potential model may be established by defining

$$e_{ijk} = s_{j(i)} + b\tau_{ik} + s\tau_{j(i)\,k} \tag{3.10}$$

since multiple observations per subject within a block and multiple observations under a specified treatment within a block are "lost" when the randomization mechanism is invoked. Thus the residual term in the observable matched-pairs design consists of the confounded subject-within-block, block-by-treatment and subject-within-block-by-treatment effects from the potential model. If such confounding occurs, then

$$\sigma_{ek}^2 = \sigma_s^2 + \sigma_{bt}^2 + \sigma_{sk}^2;\ k = T, C \tag{3.11}$$

Model	Model Parameters	Assumptions
Potential Model	$r_{ijk} = \mu + b_i + s_{j(i)} + \tau_k + b\tau_{ik} + s\tau_{j(i)k}$ $i = 1,2, \dots, B\ pairs;$ $j = 1,2\ EU's\ within\ a\ pair$ $k = T, C$	$b_i \sim iid\ N(0, \sigma_b^2)$ $s_{j(i)} \sim iid\ N(0, \sigma_s^2)$ $b\tau_{ik} \sim iid\ N(0, \sigma_{bt}^2)$ $\begin{bmatrix} s\tau_{j(i)T} \\ s\tau_{j(i)C} \end{bmatrix} \sim MVN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{sT}^2 & 0 \\ 0 & \sigma_{sC}^2 \end{pmatrix}\right]$ $b_i, s_{j(i)}, b\tau_{ik}$ and $s\tau_{j(i)k}$ are mutually independent.
Observable Model	$R_{ijk} = \mu + b_i + \tau_k + e_{ijk},$ $i = 1,2, \dots, B\ pairs;$ $j = 1\ EU\ within\ a\ pair\ receiving\ Trt\ k$ $k = T, C$	$b_i \sim iid\ N(0, \sigma_b^2)$ $\begin{bmatrix} e_{ijT} \\ e_{ijC} \end{bmatrix} \sim MVN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{eT}^2 & 0 \\ 0 & \sigma_{eC}^2 \end{pmatrix}\right]$ b_i, e_{iT}, and e_{iC} are mutually independent

Table 3.4 *Model effects and assumptions in a RCBD.*

under the assumptions given in Table 3.4. Furthermore, under the assumption of additivity, both unit-treatment additivity and block-treatment additivity, $b\tau_{ik} = s\tau_{j(i)\,k} = 0$ for all i and j and

$$e_{ijk} = s_{j(i)}.$$

Thus

$$\sigma_{eT}^2 = \sigma_{eC}^2 = \sigma_s^2$$

irrespective of the level of treatment assigned to the j^{th} EU.

Define the true causal effect to be the difference in potential outcomes for the j^{th} EU within the i^{th} pair. That is

$$d_{ij} = r_{ijT} - r_{ijC}\ . \tag{3.12}$$

Given the model assumptions in Table 3.4, the variance of the true effect is given by

$$var(d_{ij}) = var(\mu + \tau_T + b_i + s_{j(i)} + b\tau_{iT} + s\tau_{j(i)\,T} - \mu - \tau_C - b_i - s_{j(i)} - b\tau_{iC} - s\tau_{j(i)\,C})$$

$$= var\left(b\tau_{iT} - b\tau_{iC} + s\tau_{j(i)\,T} - s\tau_{j(i)\,C}\right) = 2\sigma_{bt}^2 + (\sigma_{sT}^2 + \sigma_{sC}^2) \quad (3.13)$$

The structure of the matched-pairs design lends itself to an intuitive definition of naïve effect. This is defined as the difference between the EU receiving treatment and the EU receiving control within the i^{th} pair and is given by

$$D_i = R_{iT} - R_{iC}. \quad (3.14)$$

D_i may be thought of as a naïve version of the true, individual causal effect for the two units in the i^{th} pair, which here would be given by $d_{i1}$and d_{i2}. Given the model assumptions in Table 3.4, the variance of the naïve effect is given by

$$\begin{aligned} var(D_i) &= var\left(\mu + \tau_T + b_i + s_{j(i)} + b\tau_{iT} + s\tau_{j(i)\,T} - \mu - \tau_C - b_i - s_{j'(i)} - b\tau_{iC} - s\tau_{j'(i)\,C}\right) \\ &= var\left(s_{j(i)} - s_{j'(i)} + b\tau_{iT} - b\tau_{iC} + s\tau_{j(i)\,T} - s\tau_{j'(i)\,C}\right) \\ &= (\sigma_{eT}^2 + \sigma_{eC}^2) \\ &= 2\sigma_S^2 + 2\sigma_{bt}^2 + (\sigma_{sT}^2 + \sigma_{sC}^2) \end{aligned} \quad (3.15)$$

where the final equality in (3.15) follows from equation (3.11).

Notice, D_i is the difference between the observable treatment value and the observable control value within the i^{th} block/pair. Denote different EU's within the same pair as j and j'. The difference in (3.14) is across EU's so the difference in random subject terms, $s_{j(i)} - s_{j'(i)}$, remains as a component of D_i . Contrast this to $var(d_{ij})$, where the subject effect is removed because the difference in potential outcomes is within the same EU. Also notice that based on (3.11), (3.13) and (3.15), $var(D_i)$ is an estimable upper bound for $var\left(d_{ij}\right)$ since

$$var(D_i) = 2\sigma_S^2 + var\left(d_{ij}\right)$$

$$\Rightarrow$$

$$var\left(d_{ij}\right) \leq var(D_i)$$

$$\Leftrightarrow$$

$$var\left(d_{ij}\right) \leq 2\sigma_S^2 + var\left(d_{ij}\right)$$

$$\Leftrightarrow$$

$$var(d_{ij}) \leq 2\sigma_s^2 + 2\sigma_{bt}^2 + (\sigma_{sT}^2 + \sigma_{sC}^2)$$

$$\Leftrightarrow$$

$$var(d_{ij}) \leq (\sigma_{eT}^2 + \sigma_{eC}^2). \tag{3.16}$$

The third line of equation (3.16) demonstrates that equality of $var(d_{ij})$ and $var(D_i)$ is achieved when $\sigma_S^2 = 0$. Recall that σ_S^2 is the variance attributed to EU's within a pair. It would be reasonable to expect that as the quality of matching improves, σ_S^2 decreases, and consequently $var(D_i)$ nears $var(d_{ij})$. If a perfect match of EU's within pair were achievable in an observable model setting so that $\sigma_S^2 = 0$, then the estimate of $var(D_i)$ from observed data could indeed be considered an estimate of $var(d_{ij})$. Otherwise, $var(D_i)$ serves as an estimable upper bound of $var(d_{ij})$.

Tables 3.5 (i), 3.5 (ii), and 3.5 (iii) give the results of all effects of interest based on $S = 100$ simulated data sets. Within each simulation, the following numbers of blocks of size $n = 2$ were considered: $B = 10, B = 30, and\ B = 100$. The resulting number of responses in the potential outcome framework is given by $2N = 2 \cdot Bn = 4B$ and the resulting number of EU's in the entire observable experiment was given by $N = Bn = 2B$. Values represent the mean and standard error of estimates across the $S = 100$ data sets. Table 3.5 (i) gives results for the fixed treatment effect for the model fit to both potential and observable data, Table 3.5 (ii) shows the results for the random effects in the potential model and Table 3.5 (iii) the results for the random effects in the observable model. In all cases, as the block size increased from 10 to 30 to 100, the variability of the effect estimates around the true simulated value decreased. For most effects under consideration with $B = 100$, the true simulated value is within one or two standard errors of the mean of the $S = 100$ estimates. All were within three standard errors of the mean across the $S = 100$ estimates at $B = 100$. This would indicate that as the block size increases, the REML estimates of these effects are reasonable estimates. Comparing the standard errors of the estimates between potential data and observable data in Tables 3.5 (ii) and 3.5 (iii) reveals a larger standard error for the observable estimates, as expected because they are computed from half the data versus the potential model.

Fixed Effect (Potential)	Simulated Value	2N	Average ($S = 100$)	Std. Error ($S = 100$)
$\tau_T - \tau_C$	7	40	6.97	0.12
		120	7.03	0.06
		400	7.06	0.03

Fixed Effect (Obs.)	Simulated Value	N	Average ($S = 100$)	Std. Error ($S = 100$)
$\tau_T - \tau_C$	7	20	6.90	0.16
		60	6.98	0.09
		200	7.06	0.05

(i)

Potential Variance	Simulated Value	2N	Average ($S = 100$)	Std. Error ($S = 100$)
Block	10	40	9.67	0.63
		120	9.68	0.42
		400	9.94	0.22
*Block*Trt*	3	40	3.06	0.25
		120	3.09	0.14
		400	3.06	0.07
Subject	4	40	3.94	0.23
		120	4.00	0.15
		400	3.83	0.08
*Subject*Trt*	6	40	5.57	0.30
		120	6.03	0.19
		400	6.00	0.09
*Subject*Ctrl*	2	40	2.03	0.18
		120	1.97	0.13
		400	2.07	0.07
$var(d_{ij})$	14	40	13.73	0.51
		120	14.19	0.26
		400	14.19	0.15

(ii)

Observable Variance	Simulated Value	N	Average ($S = 100$)	Std. Error ($S = 100$)
Block	10	20	9.32	0.68
		60	9.63	0.48
		200	9.93	0.24
Trt Residual	13	20	12.39	0.83
		60	13.34	0.51
		200	13.20	0.28
Ctrl Residual	9	20	9.88	0.73
		60	9.33	0.44
		200	9.00	0.20
$var(D_i)$	22	20	22.64	1.11
		60	22.68	0.63
		200	22.19	0.31

(iii)

Table 3.5 *Matched-Pairs/RCBD Simulation Results.*
Values represent the average and standard error of treatment effect estimates across $S = 100$ simulations in both the potential and observable data models for B=10, 30, and 100 of size 2 for (i) Fixed Effects. (ii) Potential Random Effects. (iii) Observable Random Effects.

Once again, it noteworthy that the estimates of $var(d_{ij})$ given in Table 3.5 (ii) correspond the theoretical value derived in (3.13). The relevant values used in these simulations were $\sigma_{bt}^2 = 3, \sigma_{sT}^2 = 6, \text{and } \sigma_{sC}^2 = 2$. By (3.13), $var(d_{ij}) = 2\sigma_{bt}^2 + (\sigma_{sT}^2 + \sigma_{sC}^2) = 2 \cdot 3 + (6 + 2) = 14$. Results in Table 3.5 (ii) demonstrate that the model estimates of $var(d_{ij})$ are reasonably close to 14. The estimates for σ_{eT}^2 and σ_{eC}^2 given in Table 3.5 (iii) also seem to be reasonable estimates of the theoretical value derived in (3.11), where it was assumed that subject-within-block, block-by-treatment and subject-within-block-by-treatment effects in the potential model are confounded to form the residual term in the observable model.

Figure 3.3 illustrates the result in (3.16). Dotted lines represent the true values used in the simulation. The upper line corresponds to the simulated value of $var(D_i)$ and the lower line corresponds to the value of $var(d_{ij})$. The difference between the upper and lower dotted line should be equal to $2\sigma_s^2$, as demonstrated above. Indeed, in these particular simulations, $\sigma_s^2 = 4$, thus the distance between the two dotted lines can be seen to be $2\sigma_s^2 = 2 \cdot 4 = 8$. Notice that as the block size increased from 10 to 30 to 100, the variability of the effect estimates around the true simulated value decreased. When $B = 100$, the true simulated value is within one standard error of the mean of the $S = 100$ estimates. This would indicate that as the block size increases,

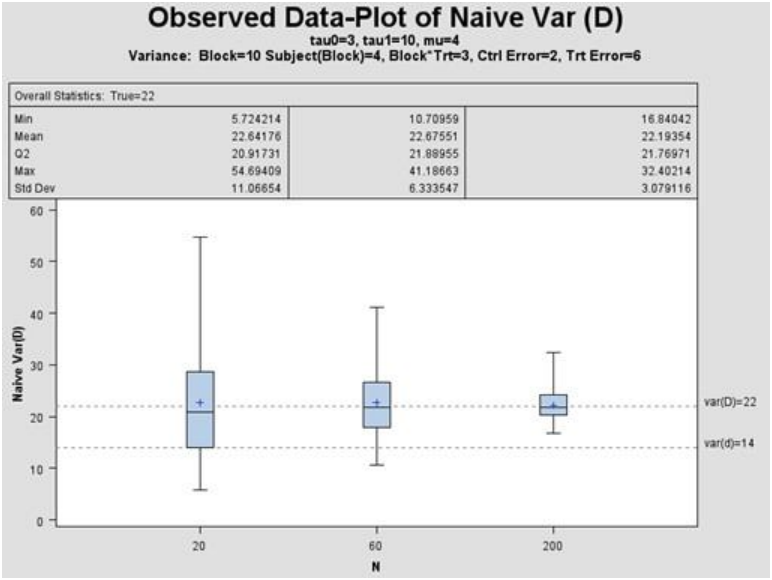

Figure 3.3 *Bounding the Individual Causal Effect: Matched-Pairs Design.* $var(D_i) \geq var(d_{ij})$. *Box plots of the* $S = 100$ *estimates of* $var(D_i)$ *at B=10, 30, and 100 blocks of size 2. Dotted lines represent values used in the simulation design.*

the REML estimates are reasonable estimates. In addition, notice the distributions of the effect estimates became more symmetric as the number of blocks increased.

Once again, two methods of estimating $var(d_{ij})$ were compared, the first method utilizing estimated model components to compute the estimate of $var(d_{ij})$, denoted by $\widehat{var}(d_{ij})$, and the second estimating $var(d_{ij})$ directly using the sample variance from the simulated data, $\widehat{var}(d_{ij})$. Table 3.6 gives the results of the comparison. Recall that in the two-sample CRD, these two methods of computation yielded identical results. However here, the two methods of computation yielded slightly differing values. To see why, consider the computation of $\widehat{var}(d_{ij})$ under the assumption that $\sigma^2_{sT} = \sigma^2_{sC} = \sigma^2_{st}$:

$$\widehat{var}(d_{ij}) = \frac{1}{2B-1}\sum_{i=1}^{B}\sum_{j=1}^{2}(d_{ij} - \bar{d}_{..})^2$$

$$\Leftrightarrow$$

$$(2B-1)\widehat{var}(d_{ij}) = \sum_{i=1}^{B}\sum_{j=1}^{2}(d_{ij} - \bar{d}_{..})^2 \qquad (3.17)$$

As shown in Appendix B.1, the sum of squares in (3.17) can be written as follows:

$$\sum_{i=1}^{B}\sum_{j=1}^{2}(d_{ij} - \bar{d}_{..})^2 = 2 \cdot (SS_{BT} + SS_{ST}) \qquad (3.18)$$

where SS_{BT} is the sum of squares of the block-by-treatment effect and SS_{ST} is the sum of squares due of subject-by-treatment effect. Both SS_{BT} and SS_{ST} are defined in Appendix B.1. Thus

$$\widehat{var}(d_{ij}) = \frac{1}{2B-1} \cdot 2 \cdot (SS_{BT} + SS_{ST}) = 2 \cdot \left(\frac{SS_{BT}}{2B-1} + \frac{SS_{ST}}{2B-1}\right). \qquad (3.19)$$

However, estimating $var(d_{ij})$ from (3.13) yields

$$\widehat{var}(d_{ij}) = 2 \cdot \hat{\sigma}^2_{bt} + (\hat{\sigma}^2_{st} + \hat{\sigma}^2_{sc}) = 2\left(\frac{MS_{BT} - MS_{ST}}{2}\right) + 2(MS_{ST}) =$$

2N	Model $var(d_{ij})$	Estimated $var(d_{ij})$
20	7.92	7.82
60	13.20	13.25
200	15.24	15.13

Table 3.6 *Different Methods of Estimation: Matched-Pairs Comparison of* $var(d_{ij})$.

$$MS_{BT} + MS_{ST} = \frac{SS_{BT}}{(B-1)} + \frac{SS_{ST}}{B} \tag{3.20}$$

where MS_{BT} is the mean square of the block-by-treatment effect and MS_{ST} is the mean square of the subject-by-treatment effect. Thus from (3.19) and (3.20), one can see that

$$2 \cdot \left(\frac{SS_{BT}}{2B-1} + \frac{SS_{ST}}{2B-1}\right) \neq \left(\frac{SS_{BT}}{(B-1)} + \frac{SS_{ST}}{B}\right) \Rightarrow \widehat{var}(d_{ij}) \neq \widetilde{var}(d_{ij}) \tag{3.21}$$

where the inequality is due to degrees of freedom associated with sums of squares terms in the linear model.

3.4 GRCBD

The potential model for the generalized complete block design in which each level of treatment is replicated more than once, is almost exactly the same model as the potential model in the matched pair analysis, with the caveat that our number of subjects within a block is now greater than 2. For the case considered here, blocks of size $n = 4$ are assumed. Consequently, everything that is estimable in the matched pair potential analysis is also estimable in the generalized complete block design setting. In addition, the variance of a random block-by-treatment effect becomes identifiable in the GRCBD since multiple observations per treatment are observable within the same block. Table 3.7 gives the effects and assumption for both the potential and observable models in the GCBD.

Because the random block-by-treatment effect becomes identifiable in a GRCBD, a direct relationship between the observable model and the potential model may be established by defining

$$e_{ijk} = s_{j(i)} + s\tau_{j(i)\,k} \tag{3.22}$$

Model	Model Parameters	Assumptions
Potential Model	$r_{ijk} = \mu + b_i + s_{j(i)} + \tau_k + b\tau_{ik} + s\tau_{j(i)\,k}$ $i = 1,2, \ldots, B$ *blocks*; $j = 1,2,..4$ *subjects within a block*; $k = T, C$	$b_i \sim iid\ N(0, \sigma_b^2)$ $s_{j(i)} \sim iid\ N(0, \sigma_s^2)$ $b\tau_{ik} \sim iid\ N(0, \sigma_{bt}^2)$ $\begin{bmatrix} s\tau_{j(i)\,T} \\ s\tau_{j(i)\,C} \end{bmatrix} \sim MVN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{sT}^2 & 0 \\ 0 & \sigma_{sC}^2 \end{pmatrix}\right]$ $b_i, s_{j(i)}$, $b\tau_{ik}$ and $s\tau_{j(i)\,k}$ are mutually independent.
Observable Model	$R_{ijk} = \mu + b_i + \tau_k + b\tau_{ik} + e_{ijk}$ $i = 1,2, \ldots, B$ *blocks*; $j = 1,2$ *subjects within a block on trt* k; $k = T, C$	$b_i \sim iid\ N(0, \sigma_b^2)$ $b\tau_{ik} \sim iid\ N(0, \sigma_{bt}^2)$ $\begin{bmatrix} e_{ijT} \\ e_{ijC} \end{bmatrix} \sim MVN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{eT}^2 & 0 \\ 0 & \sigma_{eC}^2 \end{pmatrix}\right]$ b_i, $b\tau_{ik}$, e_{ijT}, and e_{ijC} are mutually independent

Table 3.7 *Model effects and assumptions in a GRCBD.*

since multiple observations on subject within a block are "lost" when the randomization mechanism is invoked. Thus the residual term in the observable GRCBD consists of the confounded subject-within-block and subject-within-block-by-treatment effects from the potential model. If such confounding occurs, then

$$\sigma_{ek}^2 = \sigma_s^2 + \sigma_{sk}^2;\ k = T, C \tag{3.23}$$

under the assumptions given in Table 3.7. Furthermore, under the assumption of unit-treatment additivity, $s\tau_{j(i)\,k} = 0$ for all $i, j,$ and k and

$$e_{ijk} = s_{j(i)}.$$

Thus

$$\sigma_{eT}^2 = \sigma_{eC}^2 = \sigma_s^2$$

irrespective of the level of treatment assigned to the j^{th} EU.

The definitions of d_{ij} and $var(d_{ij})$ remain unchanged from those given in (3.12) and (3.13), respectively. What does change, however, is the definition of the naïve effect.

Recall that the structure of the matched-pairs design lent itself to an intuitive definition of naïve effect, D_i given in (3.14). However in the GRCBD with 4 EU's per block, there are 4 possible $D_i's$ that can be defined within a block given the random treatment assignment of EU's. Selecting which treatment observation and which control observation to use in the computation of D_i in order to accurately reflect the true value of D_i is not at all intuitive. It seems more reasonable to consider the average difference in outcomes for EU's assigned treatment and EU's assigned control. More formally, for the two units receiving treatment T, define

$$\bar{R}_{i\cdot T} = \frac{1}{2}\sum_{j=1}^{2} R_{ijT}$$

and, for the two receiving treatment C,

$$\bar{R}_{i\cdot C} = \frac{1}{2}\sum_{j=1}^{2} R_{ijC}$$

so that

$$\bar{D}_{i\cdot} = \bar{R}_{i\cdot T} - \bar{R}_{i\cdot C} \ . \tag{3.24}$$

Under the model assumptions given in Table 3.7

$$var(\bar{D}_{i\cdot}) = var\begin{pmatrix} \frac{1}{2}\sum_{\{j:k=T\}} (\mu + b_i + \tau_T + s_{j(i)} + b\tau_{iT} + s\tau_{j(i)\,T}) \\ -\frac{1}{2}\sum_{\{j:k=C\}} (\mu + b_i + \tau_C + s_{j'(i)} + b\tau_{iC} + s\tau_{j(i)\,C}) \end{pmatrix}$$

$$= var\begin{pmatrix} \mu + \tau_T + b_i + b\tau_{iT} + \frac{1}{2}\sum_{\{j:k=T\}} (s_{j(i)} + s\tau_{j(i)\,T}) \\ -\mu - \tau_C - b_i - b\tau_{iC} - \frac{1}{2}\sum_{\{j:k=C\}} (s_{j(i)} + s\tau_{j(i)\,C}) \end{pmatrix}$$

$$= var\left((b\tau_{iT} - b\tau_{iC}) + \frac{1}{2}\left[\sum_{\{j:k=T\}} \left(s_{j(i)} + s\tau_{j(i)\,T}\right) - \sum_{\{j:k=C\}} \left(s_{j(i)} + s\tau_{j(i)\,C}\right) \right] \right)$$

$$= 2\sigma_{bt}^2 + \frac{1}{2^2} var\left[\sum_{\{j:k=T\}} \left(s_{j(i)} + s\tau_{j(i)\,T}\right) \right] + \frac{1}{2^2} var\left[\sum_{\{j:k=C\}} \left(s_{j(i)} + s\tau_{j(i)\,C}\right) \right]$$

$$= 2\sigma_{bt}^2 + \frac{1}{4} \cdot 2(\sigma_s^2 + \sigma_{sT}^2) + \frac{1}{4} \cdot 2(\sigma_s^2 + \sigma_{sC}^2)$$

$$= 2\sigma_{bt}^2 + \frac{1}{2}(\sigma_s^2 + \sigma_{sT}^2) + \frac{1}{2}(\sigma_s^2 + \sigma_{sC}^2)$$

$$= 2\sigma_{bt}^2 + \sigma_s^2 + \left(\frac{\sigma_{sT}^2 + \sigma_{sC}^2}{2}\right)$$

$$= 2\sigma_{bt}^2 + \left(\frac{\sigma_{eT}^2 + \sigma_{eC}^2}{2}\right) \qquad (3.25)$$

where the final equality in (3.25) follows from (3.23). The variance in (3.25) is estimable, but of the individual components in the potential model given in Table 3.7, only σ_{bt}^2 is estimable. The linear combinations $(\sigma_s^2 + \sigma_{sT}^2)$ and $(\sigma_s^2 + \sigma_{sC}^2)$ are estimable, but the individual components are not. Multiplying both sides of the equality in (3.25) by four yields

$$4 \cdot var(\bar{D}_{i\cdot}) = 8\sigma_{bt}^2 + 4\sigma_s^2 + 2(\sigma_{sT}^2 + \sigma_{sC}^2)$$
$$= 6\sigma_{bt}^2 + (\sigma_{eT}^2 + \sigma_{eC}^2) + 2\sigma_{bt}^2 + 2\sigma_s^2 + (\sigma_{sT}^2 + \sigma_{sC}^2)$$
$$= 6\sigma_{bt}^2 + (\sigma_{eT}^2 + \sigma_{eC}^2) + 2\sigma_s^2 + var(d_{ij})$$
$$\Rightarrow$$
$$4 \cdot var(\bar{D}_{i\cdot}) - 6\sigma_{bt}^2 - (\sigma_{eT}^2 + \sigma_{eC}^2) = 2\sigma_{bt}^2 + (\sigma_{eT}^2 + \sigma_{eC}^2) = 2\sigma_s^2 + var(d_{ij}) \qquad (3.26)$$

and one can see that an estimable upper bound for $var(d_{ij})$ has been established, since σ_{bt}^2, σ_{eT}^2, and σ_{eC}^2 are all estimable in a GRCBD. Recall the definition of $var(d_{ij})$given in (3.13). Then based on the first line given in equation (3.26), equality of $var(d_{ij})$ and $2 \cdot var(\bar{D}_{i\cdot})$ is achieved when $\sigma_{bt}^2 = \sigma_s^2 = 0$. From Table 3.7, σ_{bt}^2 is the variance attributed to applying the k^{th} level of treatment to the i^{th} block and σ_s^2 is the variance attributed to EU's within a pair. Also recall that

$\sigma_{bt}^2 = 0$ is a consequence of the additivity of block and treatment effects. So if additivity of block and treatment effects holds, but unit-treatment additivity does not, *and* perfect matching within a block of all $n = 4$ EU's occurs, then twice the estimate of $var(\overline{D}_{i\cdot})$ from observed data could indeed be considered an estimate of $var(d_{ij})$. However if the assumption of additivty of block and treatment effects is valid, then $var(d_{ij})$ given in (3.13) reduces to $(\sigma_{sT}^2 + \sigma_{sC}^2)$, which is equivalent to $var(d_j)$, the variance of the causal effect defined for the two-sample CRD, given in (3.4)

Furthermore, from (3.13)

$$var(d_{ij}) = 2\sigma_{bt}^2 + (\sigma_{sT}^2 + \sigma_{sC}^2)$$
$$\Rightarrow$$
$$var(d_{ij}) - (\sigma_{sT}^2 + \sigma_{sC}^2) = 2\sigma_{bt}^2$$
$$\Rightarrow$$
$$2\sigma_{bt}^2 \leq var(d_{ij}) \tag{3.27}$$

and since σ_{bt}^2 is estimable in a GRCBD, an estimable lower bound has been established for $var(d_{ij})$. Combining the results of (3.26) and (3.27), one can see

$$2\sigma_{bt}^2 \leq var(d_{ij}) \leq 2 \cdot [var(\overline{D}_{i\cdot}) - \sigma_{bt}^2]$$
$$\Leftrightarrow$$
$$2\sigma_{bt}^2 \leq var(d_{ij}) \leq 2\sigma_s^2 + var(d_{ij})$$
$$\Leftrightarrow$$
$$2\sigma_{bt}^2 \leq var(d_{ij}) \leq 2\sigma_s^2 + 2\sigma_{bt}^2 + (\sigma_{sT}^2 + \sigma_{sC}^2)$$
$$\Leftrightarrow$$
$$2\sigma_{bt}^2 \leq var(d_{ij}) \leq 2\sigma_{bt}^2 + (\sigma_{eT}^2 + \sigma_{eC}^2) \tag{3.28}$$

In the matched-pairs analysis, the trivial lower bound of zero and a non-trivial estimable upper bound for $var(d_{ij})$ were demonstrated in (3.16). However, here in the GRCBD, both a non-trivial lower bound and upper bound have been established. The lower bound, $2\sigma_{bt}^2$, is non-trivial if the assumption of additivity of block and treatment effects fails to hold. It is important

to note that the upper bound in (3.16) and (3.28) are identical in terms of the potential model parameters given in Tables 3.4 and 3.7, respectively. This upper bound is $2\sigma_s^2 + 2\sigma_{bt}^2 + (\sigma_{sT}^2 + \sigma_{sC}^2)$. From an observable model perspective, the difference between the matched-pairs design and the GRCBD can be described by the respective difference in residual variances given in (3.11) and (3.23). According to (3.11), $2\sigma_s^2 + 2\sigma_{bt}^2 + (\sigma_{sT}^2 + \sigma_{sC}^2) = \sigma_{eT}^2 + \sigma_{eC}^2$ in the matched-pairs design, however, in the GRCBD, $2\sigma_s^2 + 2\sigma_{bt}^2 + (\sigma_{sT}^2 + \sigma_{sC}^2) = 2\sigma_{bt}^2 + (\sigma_{eT}^2 + \sigma_{eC}^2)$, according to (3.23).

Equations (3.25) and (3.26) can be extended to accommodate a balanced GRCBD with more than 4 EU's per block. The following equations give the general result for any balanced GRCBD with blocks of size n.

$$var(\overline{D}_{i\cdot}) = 2\sigma_{bt}^2 + \frac{4}{n}\sigma_s^2 + \frac{2}{n}(\sigma_{sT}^2+\sigma_{sC}^2) \tag{3.29}$$

and

$$\begin{gathered} n \cdot var(\overline{D}_{i\cdot}) = var\left(d_{ij}\right) + 2(n-1)\sigma_{bt}^2 + 4\sigma_s^2 + (\sigma_{sT}^2+\sigma_{sC}^2) \\ \Longrightarrow \\ n \cdot var(\overline{D}_{i\cdot}) - 2(n-1)\sigma_{bt}^2 - (\sigma_{eT}^2 + \sigma_{eC}^2) = var\left(d_{ij}\right) + 2\sigma_s^2 \\ = 2\sigma_{bt}^2 + (\sigma_{sT}^2 + \sigma_{sC}^2) + 2\sigma_s^2 \\ = 2\sigma_{bt}^2 + (\sigma_{eT}^2 + \sigma_{eC}^2) \end{gathered} \tag{3.30}$$

As in the 2-sample CRD, comparison of $var\left(d_{ij}\right)$ with $var(\overline{D}_{i\cdot})$ may not be intuitive since $var(\overline{D}_{i\cdot})$ is computed based on aggregate information from a sample and $var\left(d_{ij}\right)$ is computed based on information available from a single EU. Therefore define

$$\bar{d}_{i\cdot} = \frac{1}{4}\sum_{j=1}^{4} d_{ij} = \frac{1}{4}\sum_{j=1}^{4}(r_{ijT} - r_{ijC}) \tag{3.31}$$

to compare and contrast with $\overline{D}_{i\cdot}$. For $\bar{d}_{i\cdot}$ defined in (3.31),

$$var\left(\bar{d}_{i\cdot}\right) = var\left(\frac{1}{4}\sum_{j=1}^{4}(\mu + \tau_T + b_i + s_{j(i)} + b\tau_{iT} + s\tau_{j(i)\,T} - \mu - \tau_C - b_i - s_{j(i)} - b\tau_{iC} - s\tau_{j(i)\,C})\right) =$$

$$var\left(\frac{1}{4}\sum_{j=1}^{4} b\tau_{iT} - b\tau_{iC} + s\tau_{j(i)\,T} - s\tau_{j(i)\,C}\right) =$$

$$var\left((b\tau_{iT} - b\tau_{iC}) + \frac{1}{4}\sum_{j=1}^{4}\left(s\tau_{j(i)\,T} - s\tau_{j(i)\,C}\right)\right) =$$

$$2\sigma_{bt}^2 + \frac{1}{4^2}(4\sigma_{sT}^2 + 4\sigma_{sC}^2) =$$

$$2\sigma_{bt}^2 + \left(\frac{\sigma_{sT}^2 + \sigma_{sC}^2}{4}\right) \tag{3.32}$$

By the same reasoning used in using (3.27), it is easily seen that $2\sigma_{bt}^2$ is also an estimable lower bound for $var(\bar{d}_{i\cdot})$.

Comparing and contrasting $var(\bar{d}_{i\cdot})$ to $var(\bar{D}_{i\cdot})$, notice that $var(\bar{D}_{i\cdot})$ can be written

$$var(\bar{D}_{i\cdot}) = 2\sigma_{bt}^2 + \sigma_s^2 + \left(\frac{\sigma_{sT}^2+\sigma_{sC}^2}{2}\right)$$

$$= 2\sigma_{bt}^2 + \frac{4\sigma_s^2}{4} + 2\cdot\left(\frac{\sigma_{sT}^2+\sigma_{sC}^2}{4}\right)$$

$$= 2\sigma_{bt}^2 + \frac{2\sigma_s^2}{4} + \left(\frac{\sigma_{sT}^2+\sigma_{sC}^2}{4}\right) + \left(\frac{\sigma_{eT}^2 + \sigma_{eC}^2}{4}\right)$$

$$= var(\bar{d}_{i\cdot}) + \frac{2\sigma_s^2}{4} + \left(\frac{\sigma_{eT}^2 + \sigma_{eC}^2}{4}\right)$$

$$= 2\sigma_{bt}^2 + 2\cdot\left(\frac{\sigma_{eT}^2 + \sigma_{eC}^2}{4}\right) \tag{3.33}$$

where the final two equalities in (3.33) follows from (3.23) and (3.32). Writing $var(\bar{D}_{i\cdot})$ in this form, notice that

$$2\sigma_{bt}^2 \le var(\bar{d}_{i\cdot}) \le var(\bar{D}_{i\cdot}) - \left(\frac{\sigma_{eT}^2 + \sigma_{eC}^2}{4}\right)$$

$$\Leftrightarrow$$

$$2\sigma_{bt}^2 \le var(\bar{d}_{i\cdot}) \le 2\sigma_{bt}^2 + \left(\frac{\sigma_{eT}^2 + \sigma_{eC}^2}{4}\right) \tag{3.34}$$

thus, estimable upper and lower bounds of $var(\bar{d}_{i\cdot})$ have been established.

Equations (3.32), and (3.33) can also be extended to accommodate a GRCBD with more than 4 EU's per block. The following equations give the general result for any balanced GRCBD with blocks of size n.

$$var(\bar{d}_{i\cdot}) = 2\sigma_{bt}^2 + \left(\frac{\sigma_{sT}^2 + \sigma_{sC}^2}{n}\right) \tag{3.35}$$

and

$$var(\bar{D}_{i\cdot}) = var(\bar{d}_{i\cdot}) + \frac{4}{n}\sigma_s^2 + \left(\frac{\sigma_{sT}^2+\sigma_{sC}^2}{n}\right)$$
$$\Leftrightarrow$$
$$var(\bar{D}_{i\cdot}) = 2\sigma_{bt}^2 + 2\cdot\left(\frac{\sigma_{eT}^2 + \sigma_{eC}^2}{n}\right) \tag{3.36}$$

It would be reasonable to consider the behavior of $var(\bar{d}_{i\cdot})$ as n increases. From (3.35), notice

$$\lim_{n\to\infty} var(\bar{d}_{i\cdot}) = \lim_{n\to\infty}\left[2\sigma_{bt}^2 + \left(\frac{\sigma_{sT}^2 + \sigma_{sC}^2}{n}\right)\right] = 2\sigma_{bt}^2 \tag{3.37}$$

which shows that the variance of a block average converges to the variance component associated with a block-treatment random effect.

Notice that the results given in (3.28) and (3.34) are not the same result. The result in (3.28) is a statement with respect to individual treatment heterogeneity. The result from (3.34) is a statement about the average casual effect within a block. As such, there is no comparable result to (3.37) for $var(d_{ij})$. The variance of the true, individual causal effect given in (3.28) is a fixed population parameter, thus extending the matched-pairs design to a balanced GRCBD with blocks of size n only permits an estimable lower bound. Extending the design does not change either the value of $var(d_{ij})$ or the estimable upper bound.

As with the RCBD, $B = 10, B = 30, and\ B = 100$ blocks of size four were considered. Thus the resulting number of responses in the potential outcome framework is given by $2N = 2\cdot$

$Bn = 8B$ and the resulting number of EU's in the entire observable experiment was given by $N = Bn = 4B$.

Tables 3.8 (i), 3.8 (ii), and 3.8 (iii) give the results interest based on $S = 100$ simulated data sets. Values represent the mean and standard error of estimates across the $S = 100$ data sets. Table 3.8 (i) gives results for the fixed treatment effect for the model fit to both potential and observable data, Table 3.8 (ii) shows the results for the random effects in the potential model and Table 3.8 (iii) the results for the random effects in the observable model. As in the RCBD, as the block size increased from 10 to 30 to 100, the variability of the effect estimates around the true simulated value decreased. For most effects under consideration with $B = 100$, the true simulated value is within one or two standard errors of the mean of the $S = 100$ estimates. All were within three standard errors of the mean across the $S = 100$ estimates at $B = 100$. This would indicate that as the block size increases, the REML estimates of these effects are reasonable estimates. Comparing the standard errors of the estimates between potential data and observable data in Tables 3.8 (ii) and 3.8 (iii), notice that the standard errors for the observable estimates are larger. This is to be expected since these estimates are computed with one-half of the data available for the potential model estimates.

As in the two-sample CRD and matched-pairs design, the estimates of $var(d_{ij})$ given in Table 3.8 (ii) correspond to the theoretical value derived in (3.13). Furthermore, the estimates of $var(\bar{d}_{i\cdot})$ also correspond to the theoretical values derived in (3.32). The relevant values used in simulation to establish (3.13) and (3.32) were $\sigma_{bt}^2 = 3, \sigma_{sT}^2 = 6, \text{and } \sigma_{sC}^2 = 2$. Once again, the estimates for σ_{eT}^2 and σ_{eC}^2 given in Table 3.8 (iii) also seem to be reasonable estimates of the theoretical value derived in (3.23), where it was assumed that subject-within-block and subject-within-block-by-treatment effects in the potential model are confounded to form the residual term in the observable model. Relevant simulation values demonstrating the result of (3.23) are $\sigma_S^2 = 4, \sigma_{sT}^2 = 6, \text{and } \sigma_{sC}^2 = 2$.

Figure 3.4 demonstrates the results of (3.28). Dotted lines represent the true values used in simulation. The upper line corresponds to the simulated value of $var(d_{ij}) + 2\sigma_s^2$, the middle line corresponds to the value of $var(d_{ij})$, and the lower line represents the lower bound of $var(d_{ij})$, $2\sigma_{bt}^2$. The difference between the upper and middle dotted line should be equal to $2\sigma_s^2$, as demonstrated in (3.28). In these particular simulations, $\sigma_s^2 = 4$, thus the distance

Fixed Effect (Potential)	Simulated Value	2N	Average ($S = 100$)	Std. Error ($S = 100$)
$\tau_T - \tau_C$	7	80	6.98	0.08
		240	6.97	0.05
		800	6.99	0.03

Fixed Effect (Obs.)	Simulated Value	N	Average ($S = 100$)	Std. Error ($S = 100$)
$\tau_T - \tau_C$	7	40	6.91	0.12
		120	6.92	0.06
		400	6.98	0.04

(i)

Potential Variance	Simulated Value	2N	Average ($S = 100$)	Std. Error ($S = 100$)
Block	10	80	9.77	0.58
		240	9.95	0.34
		800	9.96	0.19
*Block*Trt*	3	80	3.37	0.20
		240	2.90	0.10
		800	2.95	0.06
Subject	4	80	4.07	0.15
		240	3.99	0.09
		800	4.00	0.05
*Subject*Trt*	6	80	6.01	0.19
		240	5.94	0.10
		800	5.86	0.06
*Subject*Ctrl*	2	80	2.11	0.14
		240	2.02	0.07
		800	2.07	0.04
$var(d_{ij})$	14	80	14.87	0.46
		240	13.76	0.23
		800	13.81	0.13
$var(\bar{d}_{i\cdot})$	8	80	8.78	0.40
		240	7.80	0.21
		800	7.87	0.12

(ii)

Observable Variance	Simulated Value	N	Average ($S = 100$)	Std. Error ($S = 100$)
Block	10	40	9.88	0.63
		120	10.01	0.38
		400	9.95	0.22
*Block*Trt*	3	40	3.46	0.31
		120	2.82	0.18
		400	2.91	0.11
Trt Residual	10	40	10.43	0.47
		120	9.92	0.24
		400	9.73	0.15
Ctrl Residual	6	40	5.96	0.35
		120	6.04	0.17
		400	6.21	0.08
$var(\bar{D}_{i\cdot})$	14	40	14.78	0.58
		120	13.58	0.35
		400	13.81	0.21
Upper Bound $var(d_{ij})$	22	40	23.31	0.57
		120	21.61	0.40
		400	21.75	0.24
Upper Bound $var(\bar{d}_{i\cdot})$	10	40	11.01	0.56
		120	9.63	0.34
		400	9.80	0.21
Lower Bound	6	40	6.92	0.61
		120	5.64	0.35
		400	5.81	0.22

(iii)

Table 3.8 *GRCBD Simulation Results.*
Values represent the average and standard error of treatment effect estimates across $S = 100$ simulations in both the potential and observable data models for B=10, 30, and 100 of size 4 for (i) Fixed Effects. (ii) Potential Random Effects. (iii) Observable Random Effects.

between the upper two dotted lines is $2\sigma_s^2 = 2 \cdot 4 = 8$. Indeed, note from Figure 3.4 that the same distance is seen to be $22 - 14 = 8$. Also notice that Figure 3.4 is nearly identical to Figure 3.3 from the matched-pairs design, except that Figure 3.4 now shows the estimable lower bound of (d_{ij}). The upper bounds of $var(d_{ij})$ in both Figure 3.3 and 3.4 occur at the same value, 22.

Figure 3.5 illustrates the result in (3.34). Dotted lines represent the true values used in simulation. The upper line corresponds to the simulated value of $var(\bar{D}_{i\cdot})$, the middle line

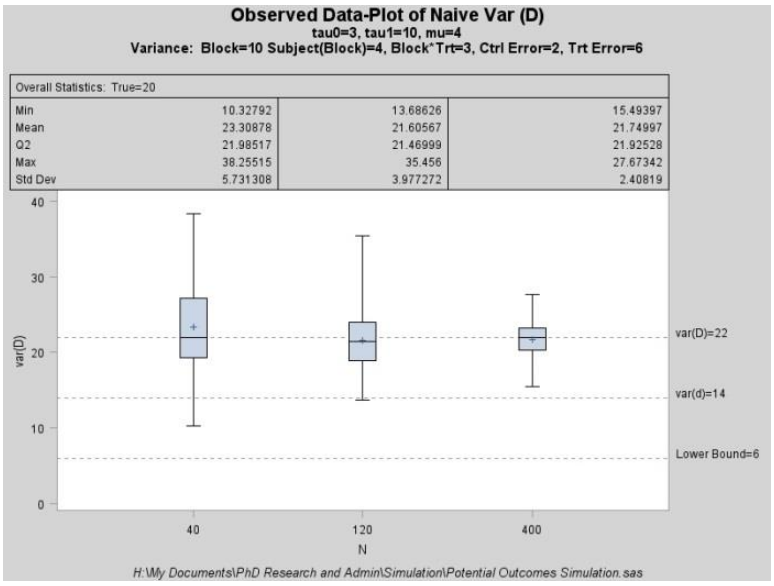

Figure 3.4 *Bounding the Individual Causal Effect: GRCBD.*
$2\sigma_{bt}^2 + (\sigma_{eT}^2 + \sigma_{eC}^2) \geq var(d_{ij}) \geq 2\sigma_{bt}^2$. *Box plots of the* $S = 100$ *estimates of* $2\sigma_{bt}^2 + (\sigma_{eT}^2 + \sigma_{eC}^2)$ *at B=10, 30, and 100 blocks of size 4. Dotted lines represent values used in the simulation design.*

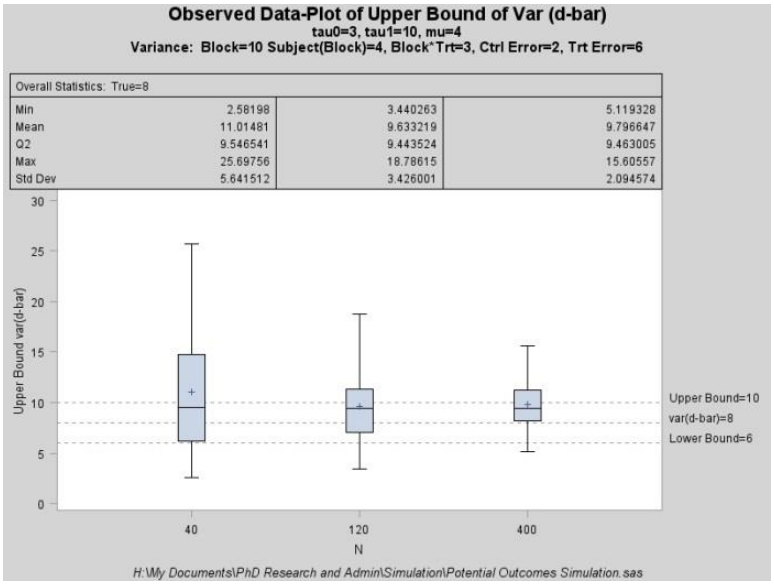

Figure 3.5 *Bounding the Average Causal Effect: GRCBD.*
$2\sigma_{bt}^2 + \frac{(\sigma_{eT}^2+\sigma_{eC}^2)}{4} \geq var(\bar{d}_{i\cdot}) \geq 2\sigma_{bt}^2$. *Boxplots of the* $S = 100$ *estimates of* $2\sigma_{bt}^2 + \frac{(\sigma_{eT}^2+\sigma_{eC}^2)}{4}$ *at B=10, 30, and 100 blocks of size 4. Dotted lines represent values used in the simulation design.*

corresponds to the value of $var(\bar{d}_{i\cdot})$, and the lower line represents the lower bound of $var(\bar{d}_{i\cdot})$, $2 \cdot \sigma_{bt}^2$. The difference between the upper and middle dotted line should be equal to $\sigma_s^2 + \left(\frac{\sigma_{sT}^2+\sigma_{sC}^2}{4}\right)$, as demonstrated in (3.33). In these particular simulations, $\sigma_s^2 = 4$, $\sigma_{sT}^2 = 6$, and $\sigma_{sC}^2 = 2$ thus the distance between the upper two dotted lines is $\sigma_s^2 + \left(\frac{\sigma_{sT}^2+\sigma_{sC}^2}{4}\right) = 4 + 2 = 6$. Indeed note from Figure 3.5 that the same distance is seen to be $12 - 6 = 6$. In both Figure 3.4 and Figure 3.5, the variability of the effect estimates around the true simulated value decreased as the block size increased from 10 to 30 to 100. When $B = 100$, the true simulated value is within one standard error of the mean of the $S = 100$ estimates. This would indicate that as the block size increases, the REML estimates are reasonable estimates. In addition, notice the distributions of the effect estimates became more symmetric as the number of blocks increased.

Table 3.9 gives the results of the comparison of $\widehat{var}(d_{ij})$ and $\widetilde{var}(d_{ij})$. As in matched-pairs designs, the two estimates do not coincide. To see why, consider the relationship between the estimates given in (3.21). Here we alter (3.21) slightly to reflect changes in degrees of freedom that occur due to the fact that there are now 4 EU's per block instead of 2 as in the matched-pairs design. Even with this slight alteration, the inequality in (3.21) still holds. That is,

$$2 \cdot \left(\frac{SS_{BT}}{4B-1} + \frac{SS_{ST}}{4B-1}\right) \neq \left(\frac{SS_{BT}}{(B-1)} + \frac{SS_{ST}}{3B}\right) \Rightarrow \widehat{var}(d_{ij}) \neq \widetilde{var}(d_{ij}). \qquad (3.38)$$

2N	Model $var(d_{ij})$	Estimated $var(d_{ij})$
80	12.65	12.60
240	12.30	12.15
800	11.66	11.57

Table 3.9 *Different Methods of Estimation: GRCBD. Comparison of var(d_{ij}).*

3.5 Two-Period-Two-Treatment Crossover Design

In an observable two-period-two-treatment crossover design, EU's are randomly assigned to one of two groups in which one group receives treatment at time period 1 followed by control at time period 2. The other group receives control at time period 1 followed by treatment at time period 2. Regardless of which sequence an EU receives, two responses are measured for each EU, one response under treatment and one response under control. Random assignment to the different sequences prevents the confounding of period effect and treatment effect. The model for the two-period-two-treatment crossover design can be thought of as an extension of the matched-pairs design in which the EU from the crossover design now takes on the role of the block in the matched-pairs design and the period from the crossover design takes on the role of the EU in the matched-pairs design. One significant disparity between the two designs is that periods and EU's are crossed in a crossover design (i.e. a response is measured in every EU at every period) while EU's are nested within blocks in a RCBD. Table 3.10 gives the effects and assumption for both the potential and observable models in the two-period-two-treatment crossover design, assuming no carry-over effect.

A direct relationship between the observable model and the potential model may be established by defining

Model	Model Parameters	Assumptions
Potential Model	$r_{ijk} = \mu + s_i + \pi_j + s\pi_{ij} + \tau_k + s\tau_{ik} + \pi\tau_{jk} + s\pi\tau_{ijk}$ $i = 1,2 \ldots N\ EU's;$ $j = 1,2\ periods;\quad k = T, C$	$s_i \sim iid\ N(0, \sigma_s^2)$ $s\pi_{ij} \sim iid\ N(0, \sigma_{sp}^2)$ $\begin{bmatrix} s\tau_{iT} \\ s\tau_{iC} \end{bmatrix} \sim MVN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{sT}^2 & 0 \\ 0 & \sigma_{sC}^2 \end{pmatrix}\right]$ $s\pi\tau_{ijk} \sim iid\ N(0, \sigma_{spt}^2)$ $s_i, s\pi_{ij}, s\tau_{ik}$ and $s\pi\tau_{ijk}$ are mutually independent.
Observable Model	$R_{ijk} = \mu + s_i + \pi_j + \tau_k + \pi\tau_{jk} + e_{ijk}$ $i = 1,2 \ldots N\ EU's;$ $j = 1, 2\ periods$ $k = T, C$	$s_i \sim iid\ N(0, \sigma_s^2)$ $\begin{bmatrix} e_{ijT} \\ e_{ijC} \end{bmatrix} \sim MVN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{eT}^2 & 0 \\ 0 & \sigma_{eC}^2 \end{pmatrix}\right]$ s_i, e_{ijT}, and e_{ijC} are mutually independent

Table 3.10 *Model effects and assumptions in a Two-Period-Two-Treatment Crossover.*

$$e_{ijk} = s\pi_{ij} + s\tau_{ik} + s\pi\tau_{ijk} \tag{3.39}$$

since multiple observations per subject-period combination and multiple observations of subject-treatment combinations are "lost" by invoking the randomization mechanism. Consequently, it is reasonable to conclude that the subject-by-period, the subject-by-treatment, and the subject-by-period-by-treatment effect from the potential model are confounded together in order to form the residual term in the observable model. If such confounding occurs, then

$$\sigma_{ek}^2 = \sigma_{sp}^2 + \sigma_{sk}^2 + \sigma_{spt}^2; k = T, C \tag{3.40}$$

under the assumptions given in Table 3.10. Furthermore, under the assumption of unit-treatment additivity, $s\pi_{ij} = s\tau_{ik} = s\pi\tau_{ijk} = 0$ for all i, j and k and

$$e_{ijk} = 0$$

irrespective of the level of treatment assigned to the i^{th} EU at the j^{th} period. This implies that the only random variability in responses in a two-period-two-treatment crossover is due to the random variability of EU's. The above assumption of unit-treatment additivity assumes additivity at each time period so that the true individual causal effect is constant across both time periods.

In every experimental design considered to this point, the observable data model generated by confounding effects from the potential model has agreed with some form of a "standard" model for that particular design. Considering the two-period-two-treatment crossover design, the observable model here may not be readily recognizable. A common, standard crossover model assuming no carry-over effects might look something like

$$R_{ijl} = \mu + \pi_j + \tau_k + \gamma_l + \tilde{e}_{ijl} \tag{3.41}$$

where R_{ijl} is the response of the i^{th} EU, $i = 1,2,..,N$, at the j^{th} time period, $j = 1,2$, on the k^{th} level of treatment, $k = T, C$ receiving the l^{th} treatment sequence, $l = 1,2$. Without loss of

generality, define γ_1 as the sequence effect resulting from a $\{T, C\}$ sequence and γ_2 as the sequence effect of resulting from a $\{C, T\}$ sequence. The indices k and l are related since the indices of treatment sequence, $l = 1,2$, arise from combinations of the indices of treatment, $k = T, C$, and time period, $j = 1,2$. Therefore the effect of $\gamma_1$may be thought of as the fixed effect of receiving treatment $k = T$ at time period $j = 1$ followed by treatment $k = C$ at time period $j = 2$. The effect of $\gamma_2$may be thought of as the fixed effect of receiving treatment $k = C$ at period $j = 1$ followed by treatment $k = T$ at period $j = 2$. Conversely, if the sequence and the time period are known, then the level of treatment applied at that time period is known. All other effects are previously defined in Table 3.10.

The observable data models given in Table 3.10 and (3.41) differ in the following respects: there is no sequence effect in the model given in Table 3.10 and there is no random subject effect in the model given in (3.41). The following is a brief explanation of the discrepancies. First, define $\tilde{e}_{ijl}$ in (3.41) as

$$\tilde{e}_{ijl} = s_i + e_{ijk}$$

where e_{ijk} is given in Table 3.10. It has been noted above that under the assumption of unit-treatment additivity,

$$e_{ijk} = 0.$$

Therefore,

$$\tilde{e}_{ijl} = s_i$$

and

$$var(R_{ijl}) = \sigma_s^2$$

if unit-treatment additivity holds.

Second, from a potential outcomes point of view, treatment sequence is an artifact of the implementation of the random treatment assignment mechanism producing a certain level of treatment at a particular time period. Assuming uniform randomization, one could argue, as we do here, that there is no reason to expect a significant effect due to groups of EU's other than the

fact that different treatment sequences were applied to the groups. Furthermore, notice that the indices of the period-by-treatment effect inform us of which treatment was applied at which time period. Therefore, a significant difference between the two groups to which treatment sequences were applied should be attributed to period-by-treatment effects instead of group effects. Without loss of generality, if we define γ_1 as the sequence effect resulting from a $\{T, C\}$ sequence and γ_2 as the sequence effect of resulting from a $\{C, T\}$ sequence, the observable data models given Table 3.10 and (3.41) are equivalent under the following assumptions:

$$\begin{aligned} &(i)\ \pi\tau_{1T} = \pi\tau_{2C} = \gamma_1 \\ &(ii)\ \pi\tau_{1C} = \pi\tau_{2T} = \gamma_2 \\ &(iii)\ \tilde{e}_{ijl} = s_i + e_{ijk} \end{aligned} \tag{3.42}$$

Therefore, the model in (3.41) may be thought of as a specific case of the model in Table 3.10

Using the potential and observable model in Table 3.10, the existence of a difference in sequence effects is something that can be tested, even if the assumptions in (3.42) do not hold. Consider the following null hypothesis of no mean sequence effect under our particular model:

$$H_0: \mu_{1T} - \mu_{2C} = \mu_{2T} - \mu_{1C} \tag{3.43}$$

In other words, as long as the effect of treatment is defined as the difference in observations under treatment and observations under control (cf., equations (1.1),(1.2), (3.9)), the null hypothesis given in (3.43) assumes that the effect of a $\{T, C\}$ sequence is the same as a $\{C, T\}$ sequence. By substituting the fixed effects from our model in Table 3.10, we can re-write the null hypothesis

$$H_0:\ \mu + \pi_1 + \tau_T + \pi\tau_{1T} - \mu - \pi_2 - \tau_C - \pi\tau_{2C} = \mu + \pi_2 + \tau_T + \pi\tau_{2T} - \mu - \pi_1 - \tau_C - \pi\tau_{1C}$$

$$\Rightarrow$$

$$H_0: (\pi_1 - \pi_2) + (\tau_T - \tau_C) + (\pi\tau_{1T} - \pi\tau_{2C}) = (\pi_2 - \pi_1) + (\tau_T - \tau_C) + (\pi\tau_{2T} - \pi\tau_{1C})$$

Re-arranging we write

$$H_0: 2 \cdot (\pi_1 - \pi_2) + (\pi\tau_{1T} - \pi\tau_{2T} + \pi\tau_{1C} - \pi\tau_{2C}) = 0$$

$$\Rightarrow$$

$$H_0\colon (\pi_1 - \pi_2) + \frac{1}{2}(\pi\tau_{1T} + \pi\tau_{1C}) - \frac{1}{2}(\pi\tau_{2T} + \pi\tau_{2C}) = 0$$

$$\Rightarrow$$

$$H_0\colon (\pi_1 - \pi_2) + (\overline{\pi\tau}_{1\cdot} - \overline{\pi\tau}_{2\cdot}) = 0$$

It is important to note that all of the parameters of interest in testing for a difference in mean sequence effects are estimable in both the observable and potential data models. For the purposes of the following simulations, values of $\pi_1, \pi_2, \pi\tau_{1T}, \pi\tau_{1C}, \pi\tau_{2C},$ and $\pi\tau_{2T}$ were chosen so that $(\pi_1 - \pi_2) + (\overline{\pi\tau}_{1\cdot} - \overline{\pi\tau}_{2\cdot}) = 0$. Consequently, no mean sequence effect was present.

Define the true causal effect to be the difference in potential outcomes for the i^{th} EU in the j^{th} time period. That is

$$d_{ij} = r_{ijT} - r_{ijC} \ . \tag{3.44}$$

Given the model assumptions in Table 3.10, the variance of the true effect is given by

$$\begin{aligned} var(d_{ij}) &= var\begin{pmatrix} \mu + s_i + \pi_j + s\pi_{ij} + \tau_T + s\tau_{iT} + \pi\tau_{jT} + s\pi\tau_{ijT} \\ -\mu - s_i - \pi_j - s\pi_{ij} - \tau_C - s\tau_{iC} - \pi\tau_{jC} - s\pi\tau_{ijC} \end{pmatrix} \\ &= var(\tau_T - \tau_C + s\tau_{iT} - s\tau_{iC} + \pi\tau_{jT} - \pi\tau_{jC} + s\pi\tau_{ijT} - s\pi\tau_{ijC}) \\ &= (\sigma_{sT}^2 + \sigma_{sC}^2) + 2\sigma_{spt}^2 \end{aligned} \tag{3.45}$$

As was the case with the RCBD, the structure of the two-period-two-treatment crossover design lends itself to an intuitive definition of naïve effect. This is defined as the difference between the response under treatment and the response under control for the i^{th} EU, irrespective of which time period treatment and control were applied. Thus

$$D_i = R_{ijT} - R_{ij'C}. \tag{3.46}$$

D_i may be thought of as a naïve version of the true individual causal effect for the two time periods in the i^{th} EU, which here would be given by $d_{i1}$and d_{i2}. Given the model assumptions in Table 3.10, the variance of the naïve effect is given by

$$\begin{aligned} var(D_i) &= var\left(\mu + s_i + \pi_j + \tau_T + \pi\tau_{jT} + e_{ijT} - \mu - s_i - \pi_{j'} - \tau_C - \pi\tau_{j'C} - e_{ij'C}\right) \\ &= var\left(\pi_j - \pi_{j'} + \tau_T - \tau_C + \pi\tau_{jT} - \pi\tau_{j'C} + e_{ijT} - e_{ij'C}\right) \\ &= 2 \cdot \left(\sigma_{sp}^2 + \sigma_{spt}^2\right) + (\sigma_{sT}^2 + \sigma_{sC}^2) = \sigma_{eT}^2 + \sigma_{eC}^2 \end{aligned} \tag{3.47}$$

where the final equality in (3.47) follows from (3.40). Since D_i is the difference between the observable treatment value and the observable control value within the i^{th} EU, this difference is across time periods within the same EU so the difference in the random subject-by-period effects, $s\pi_{ij} - s\pi_{ij'}$, remains as a component of D_i . Contrast this to d_{ij}, which is the difference between potential outcomes *within the* j^{th} *period* for the i^{th} EU. Since the potential outcomes are defined within the same period and the same subject, the subject-by-period effect is removed. Note that the variance in (3.47) is estimable, but none of the individual components from the potential model given in Table 3.10 are estimable. However, using equations (3.40), (3.45), and (3.47), $var(D_i)$ can be written

$$\begin{gathered} var(D_i) = 2 \cdot \sigma_{sp}^2 + var\left(d_{ij}\right) \\ \Rightarrow \\ var\left(d_{ij}\right) \leq var(D_i) \\ \Leftrightarrow \\ var\left(d_{ij}\right) \leq 2\sigma_{sp}^2 + var\left(d_{ij}\right) \\ \Leftrightarrow \\ var\left(d_{ij}\right) \leq 2\sigma_{sp}^2 + 2\sigma_{spt}^2 + (\sigma_{sT}^2 + \sigma_{sC}^2) \\ \Leftrightarrow \\ var\left(d_{ij}\right) \leq (\sigma_{eT}^2 + \sigma_{eC}^2) \end{gathered} \tag{3.48}$$

and an estimable upper bound of for $var\left(d_{ij}\right)$ has been established.

The third line of equation (3.48) demonstrates that equality of $var(d_{ij})$ and $var(D_i)$ is achieved when $\sigma_{sp}^2 = 0$. Recall that σ_{sp}^2 is the variance in a response attributed to that response being measured at the j^{th} time period in the i^{th} EU, regardless of which level of treatment was applied at that time period. If the effect of the j^{th} time period is π_j with probability 1 for $= 1,2 \ldots N$ EU's , then the estimate of $var(D_i)$ from observed data could indeed be considered an estimate of $var(d_{ij})$.

Tables 3.11 (i), 3.11 (ii), and 3.11 (iii) give the results of all effects of interest based on $S = 100$ simulated data sets. Within each simulation, $N = 10, N = 30,$ and $N = 100$ EU's were considered. In the potential model, a potential response is considered for each EU *at each time period*, thus the resulting number of responses in the potential outcome framework is given by $2 \cdot \mathrm{P} \cdot N = 2 \cdot 2N = 4N$, where P is the number of periods under consideration. For this particular design, $\mathrm{P} = 2$. The resulting number of responses in the entire observable experiment was given by $\mathrm{P} \cdot N = 2N$. Values represent the mean and standard error of estimates across the $S = 100$ data sets. Table 3.11 (i) gives results for the fixed treatment effect for the model fit to both potential and observable data, Table 3.11 (ii) shows the results for the random effects in the potential model and Table 3.11 (iii) the results for the random effects in the observable model. For most effects under consideration, the true simulated value is within one or two standard errors of the mean of the $S = 100$ estimates. All were within three standard errors of the mean across the $S = 100$ estimates at $N = 100$. This would indicate that as the number of EU's increases, the REML estimates of these effects are reasonable estimates. Comparing the standard errors of the estimates between potential data and observable data in Tables 3.11 (ii) and 3.11 (iii) reveals a larger standard error for the observable estimates, as expected because they are computed from half the data versus the potential model.

As has been consistent in all other designs considered, the estimates of $var(d_{ij})$ given in Table 3.11 (ii) correspond the theoretical value derived in (3.45). Relevant simulation values demonstrating (3.45) are $\sigma_{sT}^2 = 7, \sigma_{sC}^2 = 2,$ and $\sigma_{spt}^2 = 2$. The estimates for σ_{eT}^2 and σ_{eC}^2 given in Table 3.11 (iii) also seem to be reasonable estimates of the theoretical value derived in (3.40), where it was assumed that subject-by-period, the subject-by-treatment, and the subject-by-period-by-treatment effects in the potential model are confounded to form the residual term in

Fixed Effect (Potential)	Simulated Value	4N	Average ($S = 100$)	Std. Error ($S = 100$)
$\tau_T - \tau_C$	7	40	6.87	0.12
		120	6.96	0.06
		400	6.98	0.04

Fixed Effect (Obs.)	Simulated Value	2N	Average ($S = 100$)	Std. Error ($S = 100$)
$\tau_T - \tau_C$	7	20	6.85	0.18
		60	6.95	0.08
		200	6.98	0.05

(i)

Potential Variance	Simulated Value	4N	Average ($S = 100$)	Std. Error ($S = 100$)
Subject	10	40	9.12	0.65
		120	9.38	0.40
		400	9.81	0.23
*Subject*Period*	3	40	3.22	0.22
		120	3.20	0.10
		400	2.96	0.06
*Subject*Trt*	7	40	5.92	0.40
		120	6.77	0.27
		400	7.10	0.15
*Subject*Ctrl*	2	40	3.71	0.42
		120	2.70	0.25
		400	2.09	0.13
*Subject*Period*Trt*	2	40	1.93	0.08
		120	2.00	0.05
		400	2.03	0.03
$var(d_{ij})$	13	40	13.49	0.55
		120	13.47	0.32
		400	13.25	0.13

(ii)

Observable Variance	Simulated Value	2N	Average ($S = 100$)	Std. Error ($S = 100$)
Subject	10	20	9.54	0.71
		60	9.59	0.44
		200	9.81	0.25
Trt Residual	12	20	11.37	0.83
		60	11.83	0.46
		200	12.15	0.25
Ctrl Residual	7	20	8.70	0.76
		60	7.73	0.38
		200	7.15	0.20
$var(D_i)$	19	20	20.08	1.14
		60	19.55	0.53
		200	19.30	0.26

(iii)

Table 3.11 *Two-Period-Two-Treatment Crossover Simulation Results.*
Values represent the average and standard error of effect estimates across $S = 100$ simulations in both the potential and observable data models for N=10, 30, and 100 for (i) Fixed Treatment Effects. (ii) Potential Random Effects. (iii) Observable Random Effects.

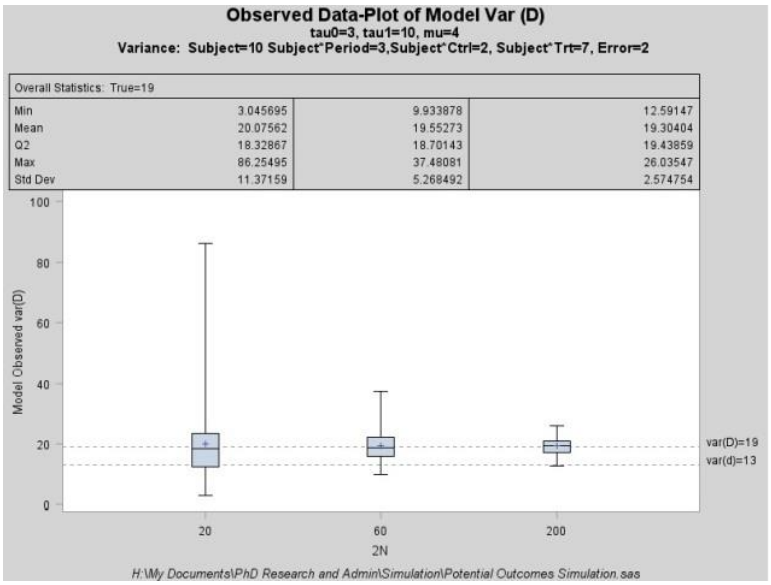

Figure 3.6 *Bounding the Individual Causal Effect: Two-Period-Two-Treatment Crossover.* $var(D_i) \geq var(d_{ij})$. *Boxplots of the* $S = 100$ *estimates of* $var(D_i)$ *at N=10, 30, and 100 EU's. Dotted lines represent values used in the simulation design.*

the observable model. Relevant simulation values for the result in (3.40) are $\sigma^2_{sp} = 3, \sigma^2_{sT} = 7, \sigma^2_{sT} = 2,$ and $\sigma^2_{spt} = 2.$

Figure 3.6 illustrates the result in (3.48). Dotted lines represent the true values used in simulation. The upper line corresponds to the simulated value of $var(D_i)$ and the lower line corresponds to the value of $var(d_{ij})$. The difference between the upper and lower dotted line should be equal to $2 \cdot \sigma^2_{sp}$, as demonstrated in (3.48). In these particular simulations, $\sigma^2_{sp} = 3$, thus the distance between the upper two dotted lines is $2 \cdot \sigma^2_{sp} = 2 \cdot 3 = 6$. Indeed note from Figure 3.6 that the same distance is seen to be $19 - 13 = 6$. As the number of EU's increased from 10 to 30 to 100, the variability of the effect estimates around the true simulated value decreased. When $N = 100$, the true simulated value is within one standard error of the mean of the $S = 100$ estimates. This would indicate that as the number of EU's increases, the REML estimates are reasonable estimates. In addition, notice the distributions of the effect estimates became more symmetric as the number of blocks increased.

Table 3.12 gives the results of the comparison of $\widehat{var}(d_{ij})$ and $\widetilde{var}(d_{ij})$. Recall that $\widetilde{var}(d_j)$, was computed using the variance component estimates obtained from the PROC GLIMMIX procedure and $\widehat{var}(d_j)$ represents the estimate of $var(d_j)$ based on the computed

4N	Model $var(d_{ij})$	Estimated $var(d_{ij})$
40	5.44	6.88
120	8.48	10.35
400	13.51	14.76

Table 3.12 *Different Methods of Estimation: Two-Period-Two-Treatment Crossover. Comparison of* $var(d_{ij})$.

difference in potential responses for each subject across both periods. As in matched-pairs designs and the GRCBD, the two estimates do not coincide. To this point, no result analogous to that shown in Appendix B.1 has been derived for the two-period-two-treatment crossover design. An extension of the result given in Appendix B.1 to a two-period-two-treatment crossover is discussed in Chapter 4.

3.6 Repeated Measures Two-Treatment Crossover Design

As was the case with the matched-pairs design compared with the GRBCD, the potential model for the repeated measures two-treatment crossover design is nearly identical to the potential model in the two-period-two-treatment crossover design, with the caveat that the number of periods crossed with each subject totals four or more periods. Each EU, therefore, receives each treatment at least twice. Consequently, everything that is estimable in the potential two-period-two-treatment crossover model is also estimable in the potential repeated measures two-treatment crossover setting. In the observable model, there are now multiple observations on each treatment for each EU that are observable, so the variance of a random subject-by-treatment effect is estimable.

Table 3.13 gives the effects and assumption for both the potential and observable models in the repeated measures two-treatment crossover.

A direct relationship between the observable model and the potential model may be established by defining

$$e_{ijk} = s\pi_{ij} + s\pi\tau_{ijk} \tag{3.49}$$

since multiple observations per subject-period combination are "lost" by invoking the randomization mechanism. Consequently, it is reasonable to conclude that the subject-by-period, and the subject-by-period-by-treatment effect from the potential model are confounded together in order to form the residual term in the observable model. If such confounding occurs, then

Model	Model Parameters	Assumptions
Potential Model	$r_{ijk} = \mu + s_i + \pi_j + s\pi_{ij} + \tau_k + s\tau_{ik} + \pi\tau_{jk} + s\pi\tau_{ijk}$ $i = 1,2 \ldots N\ EU's;$ $j = 1,4\ periods;\quad k = T, C$	$s_i \sim iid\ N(0, \sigma_s^2)$ $s\pi_{ij} \sim iid\ N(0, \sigma_{sp}^2)$ $\begin{bmatrix} s\tau_{iT} \\ s\tau_{iC} \end{bmatrix} \sim MVN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{sT}^2 & 0 \\ 0 & \sigma_{sC}^2 \end{pmatrix}\right]$ $s\pi\tau_{ijk} \sim iid\ N(0, \sigma_{spt}^2)$ $s_i, s\pi_{ij}, s\tau_{ik}$ and $s\pi\tau_{ijk}$ are mutually independent.
Observable Model	$R_{ijk} = \mu + s_i + \pi_j + \tau_k + \pi\tau_{jk} + + s\tau_{ik} + e_{ijk}$ $i = 1,2 \ldots N\ EU's;$ $j = 1,4\ periods$ $k = T, C$	$s_i \sim iid\ N(0, \sigma_s^2)$ $\begin{bmatrix} s\tau_{iT} \\ s\tau_{iC} \end{bmatrix} \sim MVN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{sT}^2 & 0 \\ 0 & \sigma_{sC}^2 \end{pmatrix}\right]$ $e_{ijk} \sim\ iid\ N(0, \sigma_e^2)$ s_i, $s\tau_{iT}, s\tau_{iC}$, and e_{ijk} are mutually independent

Table 3.13 *Model effects and assumptions in a Repeated Measures Two-Treatment Crossover.*

$$\sigma_e^2 = \sigma_{sp}^2 + \sigma_{spt}^2 \tag{3.50}$$

under the assumptions given in Table 3.13. Furthermore, under the assumption of unit-treatment additivity, $s\pi_{ij} = s\tau_{ik} = s\pi\tau_{ijk} = 0$ for all i, j and k and

$$e_{ijk} = 0$$

irrespective of the level of treatment assigned to the i^{th} EU at the j^{th} period. This implies that the random variability in responses in a repeated measures two-treatment crossover is due to the random variability of EU's and the random variability of EU's receiving a particular level of treatment.

The definition of d_{ij} and thus, the resulting $var(d_{ij})$ remain unchanged from that given in (3.44) and (3.45), respectively. However, a new definition of the naïve effect from that given in the two-period-two-treatment crossover design is required.

Recall that the structure of the two-period-two-treatment crossover design lent itself to an intuitive definition of naïve effect, D_i given in (3.46). However in the repeated measures two-treatment crossover design with each EU exhibiting a response at each of 4 different time periods, there are 4 possible $D_i's$ that can be defined for a given subject, depending on the random treatment assignment of treatment to periods. Selecting which treatment period and which control period to use in the computation of D_i in order to accurately reflect the true value of D_i is not intuitive. It seems more reasonable to consider the average difference in outcomes for periods assigned treatment and periods assigned control. More formally, for the two periods receiving treatment T, define

$$\bar{R}_{i\cdot T} = \frac{1}{2}\sum_{j=1}^{2} R_{ijT}$$

and, for the two periods receiving treatment C

$$\bar{R}_{i\cdot C} = \frac{1}{2}\sum_{j=1}^{2} R_{ijC}$$

so that

$$\bar{D}_{i\cdot} = \bar{R}_{i\cdot T} - \bar{R}_{i\cdot C} \ . \tag{3.51}$$

Given the model assumptions in Table 3.13,

$$var(\bar{D}_{i\cdot}) = var\left(\begin{array}{c} \frac{1}{2}\sum_{\{j:k=T\}}\left(\mu + s_i + \pi_j + s\pi_{ij} + \tau_T + s\tau_{iT} + \pi\tau_{j\,T} + s\pi\tau_{ijT}\right) \\ -\left[\frac{1}{2}\sum_{\{j':k=C\}}\left(\mu + s_i + \pi_{j'} + s\pi_{ij'} + \tau_C + s\tau_{iC} + \pi\tau_{j'C} + s\pi\tau_{ij'C}\right)\right] \end{array}\right)$$

$$= var\left(\begin{array}{c} \mu + s_i + \tau_T + s\tau_{iT} + \frac{1}{2}\sum_{\{j:k=T\}}\left(\pi_j + s\pi_{ij} + \pi\tau_{j\,T} + s\pi\tau_{ijT}\right) \\ -\left[\mu + s_i + \tau_C + s\tau_{iC} + \frac{1}{2}\sum_{\{j':k=C\}}\left(\pi_{j'} + s\pi_{ij'} + \pi\tau_{j'C} + s\pi\tau_{ij'C}\right)\right] \end{array}\right)$$

$$= var\left((s\tau_{iT} - s\tau_{iC}) + \frac{1}{2}\left[\sum_{\{j:k=T\}} (\pi_j + s\pi_{ij} + \pi\tau_{j\,T} + s\pi\tau_{ijT}) - \sum_{\{j':k=C\}} (\pi_{j'} + s\pi_{ij'} + \pi\tau_{j'C} + s\pi\tau_{ij'C}) \right] \right)$$

$$= (\sigma_{st}^2 + \sigma_{sc}^2) + \frac{1}{2^2} var\left[\sum_{\{j:k=T\}} (s\pi_{ij} + s\pi\tau_{ijT}) \right] + \frac{1}{2^2} var\left[\sum_{\{j':k=C\}} (s\pi_{ij'} + s\pi\tau_{ij'C}) \right]$$

$$= (\sigma_{sT}^2 + \sigma_{sC}^2) + \frac{1}{2^2} \cdot 2(\sigma_{sp}^2 + \sigma_{spt}^2) + \frac{1}{2^2} \cdot 2(\sigma_{sp}^2 + \sigma_{spt}^2)$$

$$= (\sigma_{sT}^2 + \sigma_{sC}^2) + \frac{1}{2}(\sigma_{sp}^2 + \sigma_{spt}^2) + \frac{1}{2}(\sigma_{sp}^2 + \sigma_{spt}^2)$$

$$= (\sigma_{sT}^2 + \sigma_{sC}^2) + 2 \cdot \left(\frac{\sigma_{sp}^2 + \sigma_{spt}^2}{2} \right)$$

$$= (\sigma_{sT}^2 + \sigma_{sC}^2) + \left(\sigma_{sp}^2 + \sigma_{spt}^2 \right)$$

$$= (\sigma_{sT}^2 + \sigma_{sC}^2) + \sigma_e^2 \qquad (3.52)$$

where the final equality in (3.52) follows from (3.50). The variance in (3.52) is estimable, but of the individual components from the potential model given in Table 3.13, only σ_{sT}^2 and σ_{sC}^2 are estimable. Multiplying both sides of the equality in (3.52) by four yields

$$4 \cdot var(\bar{D}_{i\cdot}) = 4(\sigma_{sT}^2 + \sigma_{sC}^2) + 4\left(\sigma_{sp}^2 + \sigma_{spt}^2 \right)$$

$$= 3(\sigma_{sT}^2 + \sigma_{sC}^2) + 2\sigma_e^2 + (\sigma_{sT}^2 + \sigma_{sC}^2) + 2\sigma_{sp}^2 + 2\sigma_{spt}^2$$

$$= 3(\sigma_{sT}^2 + \sigma_{sC}^2) + 2\sigma_e^2 + 2\sigma_{sp}^2 + var\left(d_{ij} \right)$$

$$\Rightarrow$$

$$4 \cdot var(\bar{D}_{i\cdot}) - 3(\sigma_{sT}^2 + \sigma_{sC}^2) - 2\sigma_e^2 = (\sigma_{sT}^2 + \sigma_{sC}^2) + 2\sigma_e^2 = var\left(d_{ij} \right) + 2\sigma_{sp}^2 \qquad (3.53)$$

and one can see that an estimable upper bound for $var\left(d_{ij} \right)$ has been established, since σ_{sT}^2, σ_{sC}^2, and σ_e^2 are all estimable in a repeated measures two-treatment crossover design. Recall the definition of $var\left(d_{ij} \right)$ from (3.45). From equation (3.53), notice that equality of $var\left(d_{ij} \right)$ and

the upper bound given in (3.53) is achieved when $\sigma_{sp}^2 = 0$. If $\sigma_{sp}^2 = 0$, then the estimate $(\sigma_{sT}^2 + \sigma_{sC}^2) + 2\sigma_e^2$ from observable data can indeed be considered an estimate of $var(d_{ij})$.

Furthermore, from (3.45)

$$var(d_{ij}) = (\sigma_{sT}^2 + \sigma_{sC}^2) + 2 \cdot \sigma_{spt}^2$$
$$\Rightarrow$$
$$var(d_{ij}) - 2 \cdot \sigma_{spt}^2 = (\sigma_{sT}^2 + \sigma_{sC}^2)$$
$$\Rightarrow$$
$$(\sigma_{sT}^2 + \sigma_{sC}^2) \leq var(d_{ij}) \tag{3.54}$$

and since σ_{sT}^2 and σ_{sC}^2 are estimable in a repeated measures two-treatment crossover design, an estimable lower bound has been established for $var(d_{ij})$. Combining the results of (3.53) and (3.54), one can see

$$(\sigma_{sT}^2 + \sigma_{sC}^2) \leq var(d_{ij}) \leq 4 \cdot var(\overline{D}_{i\cdot}) - 3(\sigma_{sT}^2 + \sigma_{sC}^2) - 2\sigma_e^2$$
$$\Leftrightarrow$$
$$(\sigma_{sT}^2 + \sigma_{sC}^2) \leq var(d_{ij}) \leq var(d_{ij}) + 2\sigma_{sp}^2$$
$$\Leftrightarrow$$
$$(\sigma_{sT}^2 + \sigma_{sC}^2) \leq var(d_{ij}) \leq (\sigma_{sT}^2 + \sigma_{sC}^2) + 2\sigma_{spt}^2 + 2\sigma_{sp}^2$$
$$\Leftrightarrow$$
$$(\sigma_{sT}^2 + \sigma_{sC}^2) \leq var(d_{ij}) \leq (\sigma_{sT}^2 + \sigma_{sC}^2) + 2\sigma_e^2 \tag{3.55}$$

In the two-period-two-treatment crossover analysis, the trivial lower bound of zero and a non-trivial estimable upper bound for $var(d_{ij})$ were demonstrated in (3.48). However, here in the repeated measures two-treatment crossover design, both a non-trivial lower bound and upper bound have been established. The lower bound, $(\sigma_{sT}^2 + \sigma_{sC}^2)$, is a partial description of treatment heterogeneity. More on this later. The upper bound in (3.48) and (3.55) are identical in terms of the potential model parameters given in Tables 3.10 and 3.13, respectively. This upper bound is $(\sigma_{sT}^2 + \sigma_{sC}^2) + 2\sigma_{spt}^2 + 2\sigma_{sp}^2$. From an observable model perspective, the difference between the

two-period-two-treatment crossover design and the repeated measures two-treatment crossover can be described by the respective difference in residual variances given in (3.40) and (3.50). According to (3.40), $(\sigma_{sT}^2 + \sigma_{sC}^2) + 2\sigma_{spt}^2 + 2\sigma_{sp}^2 = \sigma_{eT}^2 + \sigma_{eC}^2$ in the two-period-two-treatment crossover design, however, in the repeated measures two-treatment crossover design, $(\sigma_{sT}^2 + \sigma_{sC}^2) + 2\sigma_{spt}^2 + 2\sigma_{sp}^2 = (\sigma_{sT}^2 + \sigma_{sC}^2) + 2\sigma_e^2$, according to (3.50).

Equations (3.52) and (3.53) can be extended to accommodate a repeated measures two-treatment crossover design with more than 4 periods. The following equations give the general result for any balanced repeated measures two-treatment crossover design with P periods.

$$var(\bar{D}_{i\cdot}) = (\sigma_{sT}^2 + \sigma_{sC}^2) + \frac{4}{P} \cdot \left(\sigma_{sp}^2 + \sigma_{spt}^2\right) = (\sigma_{sT}^2 + \sigma_{sC}^2) + 4 \cdot \frac{\sigma_e^2}{P} \tag{3.56}$$

and

$$\begin{gathered} P \cdot var(\bar{D}_{i\cdot}) = (P-1)(\sigma_{sT}^2 + \sigma_{sC}^2) + 2\sigma_{sp}^2 + 2\sigma_e^2 + var\left(d_{ij}\right) \\ \Rightarrow \\ P \cdot var(\bar{D}_{i\cdot}) - (P-1)(\sigma_{sT}^2 + \sigma_{sC}^2) - 2\sigma_e^2 = var\left(d_{ij}\right) + 2\sigma_{sp}^2 \\ = (\sigma_{sT}^2 + \sigma_{sC}^2) + 2\sigma_{spt}^2 + 2\sigma_{sp}^2 \\ = (\sigma_{sT}^2 + \sigma_{sC}^2) + 2\sigma_e^2 \end{gathered} \tag{3.57}$$

As in the two-sample CRD and GRCBD, comparison of $var\left(d_{ij}\right)$ with $var(\bar{D}_{i\cdot})$ may not seem intuitive since $var(\bar{D}_{i\cdot})$ is computed based on aggregate information from a sample and $var\left(d_{ij}\right)$ is computed based on information available from a single EU. Therefore define

$$\bar{d}_{i\cdot} = \frac{1}{4}\sum_{j=1}^{4} d_{ij} = \frac{1}{4}\sum_{j=1}^{4}(r_{ijT} - r_{ijC}) \tag{3.58}$$

to compare and contrast with $\bar{D}_{i\cdot}$. Given the model assumptions in Table 3.13,

$$var\left(\bar{d}_{i\cdot}\right) = var\left(\begin{array}{l} \frac{1}{4}\sum_{j=1}^{4}(\mu + s_i + \pi_j + s\pi_{ij} + \tau_T + s\tau_{iT} + \pi\tau_{j\,T} + s\pi\tau_{ijT} \\ \qquad -\mu - s_i - \pi_j - s\pi_{ij} - \tau_C - s\tau_{iC} + \pi\tau_{jC} + s\pi\tau_{ijC}) \end{array} \right)$$

$$= var\left(\frac{1}{4}\sum_{j=1}^{4}\left(s\tau_{iT} - s\tau_{iC} + s\pi\tau_{ijT} - s\pi\tau_{ijC}\right)\right)$$

$$= var\left((s\tau_{iT} - s\tau_{iC}) + \frac{1}{4}\sum_{j=1}^{4}\left(s\pi\tau_{ijT} - s\pi\tau_{ijC}\right)\right)$$

$$= (\sigma_{sT}^2 + \sigma_{sC}^2) + \frac{1}{4^2}\left(4 \cdot 2 \cdot \sigma_{spt}^2\right)$$

$$= (\sigma_{sT}^2 + \sigma_{sC}^2) + 2 \cdot \frac{\sigma_{spt}^2}{4} = (\sigma_{sT}^2 + \sigma_{sC}^2) + \frac{\sigma_{spt}^2}{2} \tag{3.59}$$

Using a similar argument given (3.54), it is easily seen that $(\sigma_{sT}^2 + \sigma_{sC}^2)$ is also an estimable lower bound for $var(\bar{d}_{i\cdot})$.

Comparing and contrasting $var(\bar{d}_{i\cdot})$ to $var(\bar{D}_{i\cdot})$, notice that $var(\bar{D}_{i\cdot})$ can be written

$$var(\bar{D}_{i\cdot}) = (\sigma_{sT}^2 + \sigma_{sC}^2) + \sigma_e^2$$

$$= (\sigma_{sT}^2 + \sigma_{sC}^2) + \left(\sigma_{sp}^2 + \sigma_{spt}^2\right)$$

$$= (\sigma_{sT}^2 + \sigma_{sC}^2) + \frac{\sigma_{spt}^2}{2} + \frac{\sigma_{spt}^2}{2} + \sigma_{sp}^2$$

$$= var(\bar{d}_{i\cdot}) + \sigma_{sp}^2 + \frac{\sigma_{spt}^2}{2}$$

$$= var(\bar{d}_{i\cdot}) + \frac{\sigma_{spt}^2}{2} + \frac{\sigma_e^2}{2} \tag{3.60}$$

where the final equality in (3.60) follows from (3.50). Writing $var(\bar{D}_{i\cdot})$ in this form and noting the argument in (3.54) establishing an estimable lower bound for $var(\bar{d}_{i\cdot})$, one can see

$$(\sigma_{sT}^2 + \sigma_{sC}^2) \leq var(\bar{d}_{i\cdot}) \leq var(\bar{D}_{i\cdot}) - \frac{\sigma_e^2}{2}$$

$$\Leftrightarrow$$

$$(\sigma_{sT}^2 + \sigma_{sC}^2) \leq var(\bar{d}_{i\cdot}) \leq (\sigma_{sT}^2 + \sigma_{sC}^2) + \frac{\sigma_e^2}{2}. \tag{3.61}$$

By (3.61), estimable upper and lower bounds of $var(\bar{d}_{i\cdot})$ have been established.

Equations (3.59) and (3.60) can also be extended to accommodate a repeated measures two-treatment crossover design with more than 4 periods. The following equations give the general result for any balanced repeated measures two-treatment crossover design with *P* periods.

$$var(\bar{d}_{i\cdot}) = (\sigma_{sT}^2 + \sigma_{sC}^2) + 2 \cdot \frac{\sigma_{spt}^2}{P} \tag{3.62}$$

$$var(\bar{D}_{i\cdot}) = var(\bar{d}_{i\cdot}) + \frac{4}{P} \cdot \sigma_{sp}^2 + \frac{2}{P} \cdot \sigma_{spt}^2$$
$$\Leftrightarrow$$
$$var(\bar{D}_{i\cdot}) = var(\bar{d}_{i\cdot}) + \frac{2}{P} \cdot \sigma_{sp}^2 + \frac{2}{P} \cdot \sigma_{e}^2$$
$$\Leftrightarrow$$
$$var(\bar{D}_{i\cdot}) = (\sigma_{sT}^2 + \sigma_{sC}^2) + 4 \cdot \frac{\sigma_{e}^2}{P} \tag{3.63}$$

It would be reasonable to consider the behavior of $var(\bar{d}_{i\cdot})$ as *P* increases. From (3.62), notice

$$\lim_{P\to\infty} var(\bar{d}_{i\cdot}) = \lim_{P\to\infty} \left[(\sigma_{sT}^2 + \sigma_{sC}^2) + 2 \cdot \frac{\sigma_{spt}^2}{P} \right] = (\sigma_{sT}^2 + \sigma_{sC}^2) \tag{3.64}$$

which shows that the variance of an average effect for an EU converges to the sum of the variance component associated with subject-by-treatment random effects.

Notice that the results given in (3.55) and (3.61) are not the same result. The result in (3.55) is a statement with respect to individual treatment heterogeneity. The result from (3.61) is a statement about the average casual effect within a block. As such, there is no comparable result to (3.64) for $var(d_{ij})$. The variance of the true, individual causal effect given in (3.55) is a fixed population parameter, thus extending the two-period-two-treatment crossover design to a balanced repeated measures two-treatment crossover design with *P* periods only permits an estimable lower bound. Extending the design does not change either the value of $var(d_{ij})$ or the estimable upper bound.

Tables 3.14 (i), 3.14 (ii), and 3.14 (iii) give the results of all effects of interest based on $S = 100$ simulated data sets. In contrast to the other designs considered, the number of EU's was altered slightly to accommodate balance with respect to the six different possible treatment sequences. The number of sequences was a consequence of considering four time periods instead of two. So for this design only, $N = 12, N = 36, and\ N = 120$ EU's were considered in simulation. In the potential model, a potential response is considered for each EU *at each time period*, thus the resulting number of responses in the potential outcome framework is given by $2 \cdot \mathrm{P} \cdot N = 2 \cdot 4N = 8N$, where P is the number of periods under consideration. For this particular design, $\mathrm{P} = 4$. The resulting number of responses in the entire observable experiment was given by $\mathrm{P} \cdot N = 4N$. Values represent the mean and standard error of estimates across the $S = 100$ data sets. Table 3.14 (i) gives results for the fixed treatment effect for the model fit to both potential and observable data, Table 3.14 (ii) shows the results for the random effects in the potential model and Table 3.14 (iii) the results for the random effects in the observable model. For most effects under consideration, the true simulated value is within one or two standard errors of the mean of the $S = 100$ estimates. All were within three standard errors of the mean across the $S = 100$ estimates at $N = 100$. This would indicate that as the number of EU's increases, the REML estimates of these effects are reasonable estimates. Comparing the standard errors of the estimates between potential data and observable data in Tables 3.14 (ii) and 3.14 (iii) reveals a larger standard error for the observable estimates, as expected because they are computed from half the data versus the potential model.

As has been consistent in all other designs considered, the estimates of $var(d_{ij})$ given in Table 3.14 (ii) correspond the theoretical value derived in (3.45). Furthermore, the estimates of $var(\bar{d}_{i\cdot})$ given in Table 3.14 (ii) correspond the theoretical value derived in (3.59). Relevant simulation values demonstrating the results in (3.45) and (3.59) are $\sigma^2_{sT} = 7, \sigma^2_{sC} = 2, \text{and } \sigma^2_{spt} = 2$. The estimates for σ^2_{eT} and σ^2_{eC} given in Table 3.14 (iii) also seem to be reasonable estimates of the theoretical value derived in (3.50), where it was assumed that subject-by-period and the subject-by-period-by-treatment effects in the potential model are confounded to form the residual term in the observable model. Relevant simulation values for the result in (3.50) are $\sigma^2_{sp} = 3$ and $\sigma^2_{spt} = 2$.

Fixed Effect (Potential)	Simulated Value	8N	Average ($S = 100$)	Std. Error ($S = 100$)
$\tau_T - \tau_C$	7	96	6.91	0.10
		288	6.99	0.05
		960	7.02	0.03

Fixed Effect (Obs.)	Simulated Value	4N	Average ($S = 100$)	Std. Error ($S = 100$)
$\tau_T - \tau_C$	7	48	6.86	0.12
		144	6.99	0.06
		480	7.04	0.04

(i)

Potential Variance	Simulated Value	8N	Average ($S = 100$)	Std. Error ($S = 100$)
Subject	10	96	8.99	0.53
		288	9.65	0.32
		960	10.09	0.17
*Subject*Period*	3	96	3.05	0.11
		288	3.01	0.06
		960	3.04	0.03
*Subject*Trt*	7	96	6.45	0.42
		288	6.93	0.24
		960	6.83	0.13
*Subject*Ctrl*	2	96	2.69	0.31
		288	2.24	0.17
		960	2.12	0.10
*Subject*Period*Trt*	2	96	2.01	0.05
		288	2.02	0.03
		960	1.99	0.01
$var(d_{ij})$	13	96	13.17	0.45
		288	13.20	0.25
		960	12.93	0.12
$var(\bar{d}_{i\cdot})$	10	96	10.15	0.44
		288	10.17	0.24
		960	9.95	0.12

(ii)

Observable Variance	Simulated Value	4N	Average ($S = 100$)	Std. Error ($S = 100$)
Subject	10	48	8.98	0.60
		144	9.76	0.32
		480	10.06	0.17
*Subject*Trt*	7	48	6.75	0.53
		144	7.13	0.29
		480	6.92	0.18
*Subject*Ctrl*	2	48	3.18	0.42
		144	2.04	0.21
		480	2.08	0.14
Residual	5	48	5.06	0.17
		144	4.98	0.09
		480	5.00	0.05
$var(\bar{D}_{i\cdot})$	14	48	14.99	0.61
		144	14.15	0.31
		480	13.99	0.18
$Upper\ Bound\ var(d_{ij})$	19	48	20.05	0.64
		144	19.14	0.32
		480	18.99	0.19
$Upper\ Bound\ var(\bar{d}_{i\cdot})$	11.5	48	12.46	0.62
		144	11.66	0.31
		480	11.49	0.19
Lower Bound	9	48	9.93	0.63
		144	9.17	0.32
		480	9.00	0.19

(iii)

Table 3.14 *Repeated Measures Two-Treatment Crossover Simulation Results. Values represent the average and standard error of effect estimates across $S = 100$ simulations in both the potential and observable data models for N=12, 36, and 120 for (i) Fixed Treatment Effects. (ii) Potential Random Effects. (iii) Observable Random Effects.*

Figure 3.7 illustrates the result in (3.55). Dotted lines represent the true values used in simulation. The upper line corresponds to the simulated value of $(\sigma_{sT}^2 + \sigma_{sC}^2) + 2\sigma_e^2$, the middle line corresponds to the value of $var(d_{ij})$, and the lower line represents the lower bound of $var(d_{ij}), (\sigma_{sT}^2 + \sigma_{sC}^2)$. The difference between the upper and middle dotted line should be equal to $2\sigma_{sp}^2$, as demonstrated in (3.55). In these particular simulations, $\sigma_{sp}^2 = 3$ thus the anticipated distance between the upper two dotted lines is $2\sigma_{sp}^2 = 2 \cdot 3 = 6$. Indeed note from Figure 3.7 that the distance between the upper and middle dotted lines is seen to be $19 - 13 = 6$. Also

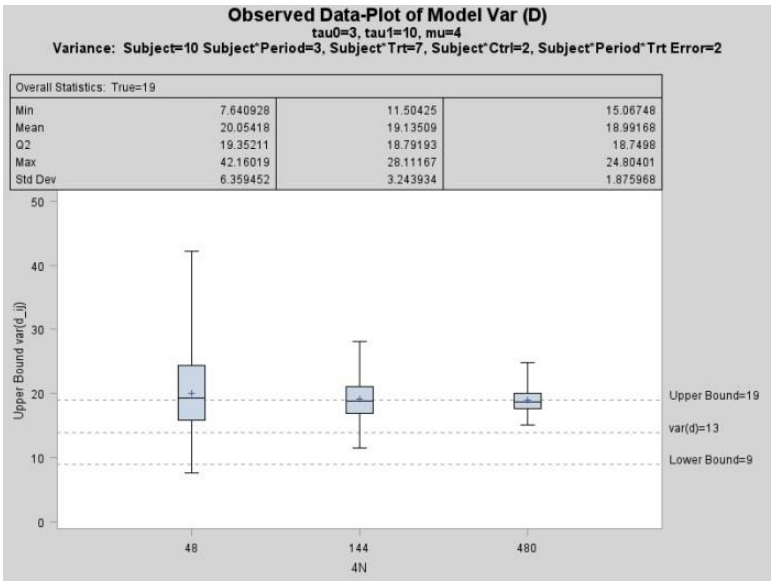

Figure 3.7 *Bounding the Individual Causal Effect: Repeated Measures Two-Treatment Crossover Design.* $(\sigma_{sT}^2 + \sigma_{sC}^2) + 2\sigma_e^2 \geq var(d_{ij}) \geq (\sigma_{sT}^2 + \sigma_{sC}^2)$. *Boxplots of the* $S = 100$ *estimates of* $(\sigma_{sT}^2 + \sigma_{sC}^2) + 2\sigma_e^2$ *at N=12, 36, and 120 EU's measured at 4 time periods. Dotted lines represent values used in the simulation design.*

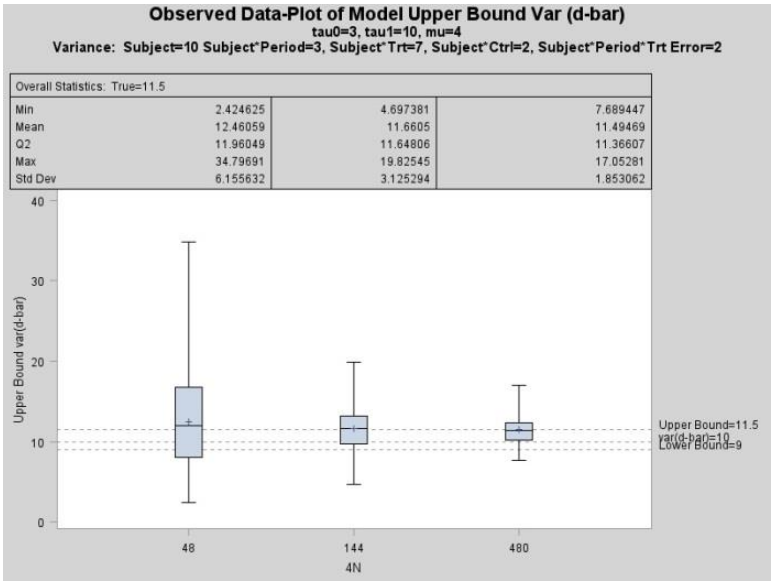

Figure 3.8 *Bounding the Average Causal Effect: Repeated Measures Two-Treatment Crossover Design.* $(\sigma_{sT}^2 + \sigma_{sC}^2) + \frac{\sigma_e^2}{2} \geq var(\bar{d}_{i\cdot}) \geq (\sigma_{sT}^2 + \sigma_{sC}^2)$. *Boxplots of the* $S = 100$ *estimates of* $(\sigma_{sT}^2 + \sigma_{sC}^2) + \frac{\sigma_e^2}{2}$ *at N=12, 36, and 120 EU's measured at 4 time periods. Dotted lines represent values used in the simulation design.*

notice that the upper bounds of $var(d_{ij})$ in both Figure 3.6 and 3.7 occur at the same value, 19. This confirms that the upper bound of $var(d_{ij})$is the same in both the two-period-two-treatment crossover design and the repeated measures two-treatment crossover design.

Figure 3.8 illustrates the result in (3.61). Dotted lines represent the true values used in simulation. The upper line corresponds to the simulated value of $(\sigma_{sT}^2 + \sigma_{sC}^2) + \frac{\sigma_{\tilde{e}}^2}{2}$, the middle line corresponds to the value of $var(\bar{d}_{i\cdot})$, and the lower line represents the lower bound of $var(\bar{d}_{i\cdot}), (\sigma_{sT}^2 + \sigma_{sC}^2)$. It can be shown that the difference between the upper and middle dotted line should be equal to $\frac{\sigma_{sp}^2}{2}$. In these particular simulations, $\sigma_{sp}^2 = 3$ thus the anticipated distance between the upper two dotted lines is $\frac{\sigma_{sp}^2}{2} = \frac{3}{2} = 1.5$. Indeed note from Figure 3.8 that the distance between the upper and middle dotted lines is seen to be $11.5 - 10 = 1.5$. In both Figure 3.7 and 3.8, as the number of EU's increased from 12 to 36 to 120, the variability of the effect estimates around the true simulated value decreased. When $N = 120$, the true simulated value is within one standard error of the mean of the $S = 100$ estimates. This would indicate that as the number of EU's increases, the REML estimates are reasonable estimates. In addition, notice the distributions of the effect estimates became more symmetric as the number of EU's increased.

It should not be overlooked that σ_{sT}^2 and σ_{sC}^2 were the inestimable quantities that identified the presence of treatment heterogeneity in the two-sample CRD when either σ_{sT}^2 or σ_{sC}^2 were non-zero. By considering a more complex experimental design, these previously inestimable quantities have become estimable, and treatment heterogeneity may be *partially* described from observable data in a repeated measures two-treatment crossover. However, the estimates of σ_{sT}^2 and σ_{sC}^2 still do not *completely* characterize treatment heterogeneity, at least not without additional assumptions. The assumptions required for treatment heterogeneity to be completely described and the consequences of those assumptions in the current experimental setting are now considered.

Recall that in (3.45), $var(d_{ij})$ was defined as the linear combination of $\sigma_{sT}^2, \sigma_{sC}^2$, and σ_{spt}^2. Estimable quantities from observable data include σ_{sT}^2 and σ_{sC}^2, but σ_{spt}^2 is not estimable in the observable data model, therefore $var(d_{ij})$ is not estimable. Under the assumption that $\sigma_{spt}^2 = 0$, however, $var(d_{ij})$ is completely characterized by σ_{sT}^2 and σ_{sC}^2 so that $var(d_{ij})$

becomes estimable in the observable data model. Coincidently, if $\sigma^2_{spt} = 0$, then $var(d_{ij})$ given in (3.45) and $var(d_j)$ given in (3.4) are equivalent, where $var(d_j)$ is the variance of the true causal effect for the j^{th} EU in a two-sample CRD.

But, practically speaking, what does it mean that $\sigma^2_{spt} = 0$? Recall, $s\pi\tau_{ijk}$ is the effect produced by applying the k^{th} level of treatment at the j^{th} period to the i^{th} EU. It may be helpful to contrast $s\pi\tau_{ijk}$ with the fixed effect $\pi\tau_{jk}$, which is the effect produced by applying the k^{th} level of treatment at the j^{th} period. Under the assumptions given in Table 3.13, the sum of these two effects yields the following random effect:

$$(\pi\tau_{jk} + s\pi\tau_{ijk}) \sim iid\ N(\pi\tau_{jk}, \sigma^2_{spt}); \quad i = 1,2..N, j = 1,2..4, k = T, C \tag{3.65}$$

So if no variability is produced by applying the k^{th} level of treatment at the j^{th} period across $i = 1,2..N$ EU's, then $\sigma^2_{spt} = 0$. In other words, applying the k^{th} level of treatment at the j^{th} period to the i^{th} EU yields the effect $\pi\tau_{jk}$ with probability 1 for $i = 1,2..N$ when $\sigma^2_{spt} = 0$. Since σ^2_{spt} is not estimable in an observable model, the validity of this assumption cannot be tested.

Table 3.15 gives the results of the comparison of $\widehat{var}(d_{ij})$ and $\widetilde{var}(d_{ij})$. As in matched-pairs designs, GRCBD, and the two-period-two-treatment crossover design, the two estimates do not coincide. To this point, no result analogous to that shown in Appendix B.1 has been derived for the Repeated measures two-treatment crossover design. These results will be discussed further in Chapter 4.

8N	Model $var(d_{ij})$	Estimated $var(d_{ij})$
96	9.40	12.29
288	16.43	18.88
960	15.53	17.36

Table 3.15 *Different Methods of Estimation: Repeated Measures Two-Treatment Crossover Design. Comparison of var(d_{ij}).*

3.7 Summary

In the preceding sections, models for potential outcomes were derived for each of five common experimental designs. All models assumed Gaussian responses. Pertinent model assumptions have been stated for each design. In the two-sample CRD, it was shown that for a non-negative correlation between potential outcomes under treatment and potential outcomes under control, estimates of this correlation using model components yielded identical results to those estimates obtained by computing Pearson's correlation on the set of N bivariate potential outcomes.

Using the potential models, a definition of treatment heterogeneity has been clearly defined in terms of potential model components. Simulations confirmed that using REML estimates of the potential model components to estimate treatment heterogeneity yielded reasonable results for all experimental designs.

Furthermore, "usual" observable models for each experimental design and the corresponding potential models were linked by defining the residual term in the observable model to be the sum of the confounded effects from the potential model. These potential model effects were confounded together by removing one-half of the data to mimic the implementation of a uniform randomization mechanism. Once again, simulations demonstrated that this relationship between observable and potential models was reasonable, as REML estimates of the observable residual were "close" to the sum of the confounded potential model effects used to produce the simulated potential data.

Naïve estimates of treatment heterogeneity were defined for each observable model and the variance of these naïve effects were given in terms of the variance of the appropriate potential model components. In all experimental designs, the variance of the naïve estimate of treatment heterogeneity served as an upper bound for the variance of the true, causal effect. In more complex designs (i.e.-GRCBD, and repeated measures two-treatment crossover), lower bounds for the variance of the true, causal effect were also established. Simulations confirmed both the existence and accuracy of these bounds. Furthermore, for each design, the assumptions required to equate the variance of the naïve effect and the variance of the true, causal effect were presented.

Finally, it was demonstrated that some inestimable quantities in relatively simple experimental designs become estimable by increasing the complexity of the design. In

particular, the variance associated with a subject-by-treatment effect becomes estimable by moving from a two-period-two-treatment crossover design to a repeated measures two-treatment crossover design. The fact that this variance is estimable in a repeated measures two-treatment crossover design has been noted previously Senn (2001); however, it was not clear how this component was related to the variance of true effects and/or what assumptions were required to equate the two. The results presented here clearly identify the relationship between the estimable variance of a subject-by-treatment effect and treatment heterogeneity and the appropriate assumptions required to equate the two have been described.

Chapter 4 - Proposed Research Completed

The following chapters represent the work that was proposed to complete this dissertation research. Topics in Chapter 4 clarify results presented in Chapter 3. Topics included in Chapters 5 and 6 serve as extensions of the research presented in Chapter 3. Further research ideas are also presented in Chapter 7.

4.1 Discrepancy of Model $var(d)$ and Estimated $var(d)$

For each of the five experimental designs presented in Chapter 3, two methods of computing the variance of the individual causal effect, $\widehat{var}(d)$ and $\widehat{var}(d)$, were compared. Both methods used estimates from the potential model only. $\widehat{var}(d)$ was termed *Model* $var(d)$ and was computed using the appropriate variance component estimates obtained from the PROC GLIMMIX procedure. $\widehat{var}(d)$ was termed *Estimated* $var(d)$ and was computed by estimating the variance of the difference in potential responses for each EU using PROC UNIVARIATE in SAS.

Of the five experimental designs, only the two-sample CRD yielded identical estimates between the two methods. Discrepancies observed in the matched-pairs design and GRCBD were shown in Appendix B.1 to be due to degrees of freedom associated with sums of squares terms in the linear model. However, the proof presented in Appendix B.1 assumed homoscedasticity of variances for potential outcomes. Furthermore, no comparable proof has yet been established for the crossover designs presented in Chapter 3.

The results given in Appendix B.1 can be extended to a two-period-two treatment crossover design by considering the computation of $\widehat{var}(d_{ij})$ under the homoscedastic assumption that $\sigma^2_{sT} = \sigma^2_{sC} = \sigma^2_{st}$:

$$\widehat{var}(d_{ij}) = \frac{1}{2N-1}\sum_{i=1}^{N}\sum_{j=1}^{2}(d_{ij} - \bar{d}_{..})^2$$

$$\Leftrightarrow$$

$$(2N-1)\widehat{var}(d_{ij}) = \sum_{i=1}^{B}\sum_{j=1}^{2}(d_{ij} - \bar{d}_{..})^2 \qquad (4.1)$$

As shown in Appendix B.2, the sum of squares in (4.1) can be written as follows:

$$\sum_{i=1}^{N}\sum_{j=1}^{2}\left(d_{ij}-\bar{d}_{..}\right)^2 = 2\cdot(SS_{ST}+SS_{PT}+SS_{SPT}) \tag{4.2}$$

where SS_{ST} is the sum of squares due to the subject-by-treatment effect, SS_{PT} is the sum of squares due to the period-by-treatment effect and SS_{SPT} is the sum of squares due to the subject-by-period-by-treatment effect. Theses sums of squares are defined in Appendix B.2. Thus

$$\widehat{var}\left(d_{ij}\right) = \frac{1}{2N-1}\cdot 2\cdot(SS_{ST}+SS_{PT}+SS_{SPT}) = 2\cdot\left(\frac{SS_{ST}}{2N-1}+\frac{SS_{PT}}{2N-1}+\frac{SS_{SPT}}{2N-1}\right). \tag{4.3}$$

However, estimating $var\left(d_{ij}\right)$ from (3.45) yields

$$\widetilde{var}\left(d_{ij}\right) = (\hat{\sigma}_{st}^2+\hat{\sigma}_{sc}^2)+2\cdot\hat{\sigma}_{spt}^2 = 2\left(\frac{MS_{ST}-MS_{SPT}}{2}\right)+2(MS_{SPT}) =$$

$$MS_{ST}+MS_{SPT} = \frac{SS_{ST}}{(N-1)}+\frac{SS_{SPT}}{(N-1)(P-1)(T-1)} \tag{4.4}$$

where MS_{ST} is the mean square of the subject-by-treatment effect and MS_{SPT} is the mean square of the subject-by-period-by-treatment effect. Thus from (4.3) and (4.4), one can see that

$$2\cdot\left(\frac{SS_{ST}}{2N-1}+\frac{SS_{PT}}{2N-1}+\frac{SS_{SPT}}{2N-1}\right) \neq \frac{SS_{ST}}{(N-1)}+\frac{SS_{SPT}}{(N-1)(P-1)(T-1)}$$

$$\Rightarrow\ \widehat{var}\left(d_{ij}\right) \neq \widetilde{var}\left(d_{ij}\right) \tag{4.5}$$

where the inequality is due to degrees of freedom associated with sums of squares terms in the linear model and the additional sums of squares due to a period-by-treatment effect in the computation of $\widehat{var}\left(d_{ij}\right)$.

4.2 Correlation

In section 3.2, two methods for computing the correlation between potential outcomes under treatment and potential outcomes under control in the two-sample CRD were also compared. One method used estimated variance components to compute an intra-class

coefficient. The second method used Pearson's computed correlation coefficient as an estimate of correlation. Recall that the estimates were identical so long as Pearson's computed correlation was non-negative. When Pearson's computed correlation returned a negative estimate, the corresponding intra-class correlation estimate was always zero. Based on the model assumptions given in Table 3.1, the intra-class correlation estimate is required to be non-negative since the covariance between the potential response under treatment and the potential response under control within the same EU is σ_S^2, the variance attributed to the EU regardless of the level of treatment applied. Also recall that equation (2.6) gave bounds for ${\sigma_d}^2$, the variance of the individual causal effect. The upper bound and lower bound were determined by assuming $\rho_{TC} = -1$ and $\rho_{TC} = 1$ respectively, where ρ_{TC} is the correlation between potential outcomes given in (1.3). However, if the correlation between potential outcomes is restricted to being non-negative, as is the case for intra-class correlation under the assumption of the model given in (3.1), then different bounds from those given in (2.6) would be achieved. That is,

$$0 \leq (\sigma_T - \sigma_C)^2 \leq {\sigma_d}^2 \leq ({\sigma_T}^2 + {\sigma_C}^2). \tag{4.6}$$

Further investigation of the discrepancy between (4.6) and (2.6) is warranted. Results are given in section 4.2.1.

A second issue that may be related to the nature of the correlation between potential responses relates to the assumption of unit-treatment additivity. Recall that if unit-treatment additivity holds, then the variance of the true causal effect, ${\sigma_d}^2$, is zero. Gadbury et. al (2001) demonstrated that based on the definition of ${\sigma_d}^2$ given in (2.1), ${\sigma_d}^2 = 0$ if and only if the following two conditions hold:

$$(i)\ \sigma_T^2 = \sigma_C^2$$
$$\text{and} \tag{4.7}$$
$$(ii)\ \rho_{TC} = 1$$

However, given the results of equation (3.4), ${\sigma_d}^2 = 0$ if and only if $\sigma_{sT}^2 = \sigma_{sC}^2 = 0$. Assuming the intra-class correlation definition

$$\rho_{TC} = \frac{\sigma_s^2}{\sqrt{\sigma_s^2 + \sigma_{sT}^2} \cdot \sqrt{\sigma_s^2 + \sigma_{sC}^2}} \tag{4.8}$$

$\sigma_{sT}^2 = \sigma_{sC}^2 = 0$ forces $\rho_{TC} = 1$. The results given here for ${\sigma_d}^2 = 0$ differ from those given in the literature. Resolving these differences in conditions under which ${\sigma_d}^2 = 0$ needs to be carefully considered. Part of resolving these differences will include a description of how the bounds for ${\sigma_d}^2$or conditions for estimability of ${\sigma_d}^2$ relate to correlation assumptions in the potential data model. This is discussed further in section 4.2.2.

4.2.1 Pearson Correlation vs. Intra-class Correlation: Determining Bounds

According the model and assumptions for the two-sample CRD given in Table 3.1, the joint distribution of the random effects in the potential LMM are

$$\begin{bmatrix} s_j \\ s\tau_{Tj} \\ s\tau_{Cj} \end{bmatrix} \sim MVN\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_s^2 & 0 & 0 \\ 0 & \sigma_{sT}^2 & 0 \\ 0 & 0 & \sigma_{sC}^2 \end{pmatrix}\right]$$

In order to resolve the discrepancy between (4.6) and (2.6), assume a more general multivariate normal distribution of the random effects in the potential LMM such as

$$\begin{bmatrix} s_j \\ s\tau_{Tj} \\ s\tau_{Cj} \end{bmatrix} \sim MVN\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_s^2 & \rho_{sT}\sigma_s\sigma_{sT} & \rho_{sC}\sigma_s\sigma_{sC} \\ \rho_{sT}\sigma_s\sigma_{sT} & \sigma_{sT}^2 & \rho_{sTC}\sigma_{sT}\sigma_{sC} \\ \rho_{sC}\sigma_s\sigma_{sC} & \rho_{sTC}\sigma_{sT}\sigma_{sC} & \sigma_{sC}^2 \end{pmatrix}\right]. \qquad (4.9)$$

According to the potential model given in Table 3.1

$$(i)\quad var(r_{ij}) = \sigma_s^2 + \sigma_{si}^2 + 2\rho_{si}\sigma_s\sigma_{si}; i = T, C$$

$$(ii)\ cov(r_{Tj}, r_{Cj}) = E({s_j}^2) + E(s_j s\tau_{Tj}) + E(s_j s\tau_{Cj}) + E(s\tau_{Tj} s\tau_{Cj})$$
$$= \sigma_s^2 + \rho_{sT}\sigma_s\sigma_{sT} + \rho_{sC}\sigma_s\sigma_{sC} + \rho_{sTC}\sigma_{sT}\sigma_{sC}.$$

so that

$$\begin{bmatrix} r_{Tj} \\ r_{Cj} \end{bmatrix} \sim MVN\left[\begin{pmatrix} \mu + \tau_T \\ \mu + \tau_C \end{pmatrix}, \begin{pmatrix} \sigma_s^2 + \sigma_{sT}^2 + 2\rho_{sT}\sigma_s\sigma_{sT} & \sigma_s^2 + \rho_{sT}\sigma_s\sigma_{sT} + \rho_{sC}\sigma_s\sigma_{sC} + \rho_{sTC}\sigma_{sT}\sigma_{sC} \\ \sigma_s^2 + \rho_{sT}\sigma_s\sigma_{sT} + \rho_{sC}\sigma_s\sigma_{sC} + \rho_{sTC}\sigma_{sT}\sigma_{sC} & \sigma_s^2 + \sigma_{sC}^2 + 2\rho_{sC}\sigma_s\sigma_{sC} \end{pmatrix}\right].$$

Now reconsider the intra-class correlation given in (4.8) under these revised assumptions about the random effects in the potential LMM

$$\rho_{TC} = \frac{\sigma_s^2 + \rho_{sT}\sigma_s\sigma_{sT} + \rho_{sC}\sigma_s\sigma_{sC} + \rho_{sTC}\sigma_{sT}\sigma_{sC}}{\sqrt{\sigma_s^2 + \sigma_{sT}^2 + 2\rho_{sT}\sigma_s\sigma_{sT}} \cdot \sqrt{\sigma_s^2 + \sigma_{sC}^2 + 2\rho_{sC}\sigma_s\sigma_{sC}}}. \quad (4.10)$$

Notice that the intra-class correlation is no longer restricted to being non-negative. That is, $-1 \leq \rho_{TC} \leq 1$ depending on the values of σ_s^2, ρ_{sT}, ρ_{sC}, and ρ_{sTC}. The intra-class correlation in (4.8) can be derived from (4.10) if the assumption $\rho_{sT} = \rho_{sC} = \rho_{sTC} = 0$ holds. Under this assumption, the value of ρ_{TC} must be non-negative.

By permitting a more general multivariate distribution on the random effects as in (4.9), a result synonymous with the result given in (2.6) may be obtained. Reconsider the $var(d_j)$, now under the more general assumptions given in (4.9):

$$var(d_j) = var(r_{Tj} - r_{Cj}) = \sigma_{sT}^2 + \sigma_{sC}^2 - 2\rho_{sTC}\sigma_{sT}\sigma_{sC}.$$

Upper and lower bounds are achieved by assuming $\rho_{sTC} = -1$ and $\rho_{sTC} = 1$, respectively, so that

$$0 \leq (\sigma_{sT} - \sigma_{sC})^2 \leq {\sigma_d}^2 \leq (\sigma_{sT} + \sigma_{sC})^2 \quad (4.11)$$

It should be noted that these bounds are not estimable in a two-sample CRD.

4.2.2 Pearson Correlation vs. Intra-class Correlation: Conditions for Zero Variance

From (4.11), ${\sigma_d}^2 = 0$ if and only if $\rho_{sTC} = 1$ and $\sigma_{sT}^2 = \sigma_{sC}^2$. Denote this common variance as σ_{st}^2. If we impose these conditions on the definition of ρ_{TC} given in (4.10), then

$$\rho_{TC} = \frac{\sigma_s^2 + (\rho_{sT} + \rho_{sC})\sigma_s\sigma_{st} + \sigma_{st}^2}{\sqrt{\sigma_s^2 + \sigma_{st}^2 + 2\rho_{sT}\sigma_s\sigma_{st}} \cdot \sqrt{\sigma_s^2 + \sigma_{st}^2 + 2\rho_{sC}\sigma_s\sigma_{st}}}.$$

Notice that even though $\rho_{sTC} = 1$ and $\sigma_{sT}^2 = \sigma_{sC}^2 = \sigma_{st}^2$, ρ_{TC} is not necessarily 1 and σ_T^2 is not necessarily equal to σ_C^2. In order to accomplish the necessary requirements for ${\sigma_d}^2 = 0$ from (2.6), the additional assumption that $\rho_{sT} = \rho_{sC}$ is required. Denote this common correlation as ρ_{st}. If we assume (i) $\rho_{sTC} = 1$, (ii) $\sigma_{sT}^2 = \sigma_{sC}^2 = \sigma_{st}^2$ and (iii) $\rho_{sT} = \rho_{sC} = \rho_{st}$ then

$$\rho_{TC} = \frac{\sigma_s^2 + 2\rho_{st}\sigma_s\sigma_{st} + \sigma_{st}^2}{\sqrt{\sigma_s^2 + \sigma_{st}^2 + 2\rho_{st}\sigma_s\sigma_{st}} \cdot \sqrt{\sigma_s^2 + \sigma_{st}^2 + 2\rho_{st}\sigma_s\sigma_{st}}} = 1$$

and

$$\sigma_T^2 = \sigma_C^2 = \sigma_s^2 + \sigma_{st}^2 + 2\rho_{st}\sigma_s\sigma_{st}$$

and the necessary assumptions required for ${\sigma_d}^2 = 0$ from (2.6) are met. It is worth noting that if $\rho_{sTC} = 1$ and $\sigma_{sT}^2 = \sigma_{sC}^2 = \sigma_{st}^2$, then ${\sigma_d}^2 = 0$ even if $\rho_{sT} \neq \rho_{sC}$ so that $\sigma_T^2 \neq \sigma_C^2$.

4.2.3 Summary

Using linear mixed models to delineate the assumptions necessary to equate treatment heterogeneity in a potential outcomes framework to estimable components of an observable data model yielded some surprising results, compared with those results published from a finite-population perspective. For complex designs in particular, the estimates of ${\sigma_d}^2$ using linear mixed model components did not always match the finite-population estimates of ${\sigma_d}^2$. Furthermore, intra-class correlation estimates based on LMM variance components matched Pearson correlation estimates for non-negative values only. By carefully considering the model assumptions used in linear mixed models and relating model variance component estimates to the finite-population estimate of ${\sigma_d}^2$ through the use of sums of squares, these discrepancies have been resolved.

Chapter 5 - Identifying Treatment Heterogeneity in Complex Designs: A Linear Mixed Effects Models Approach

The following chapter is a paper submitted to a peer-reviewed journal. The concepts discussed in this dissertation are further developed to accommodate a more complex treatment structure and then applied to an illustrative data example.

5.1.Abstract

A treatment's efficacy or safety is often assessed by a study of the mean effect of a treatment with respect to some reference treatment. If a treatment effect is highly variable across units in a population, then applying information about the mean effect to each individual unit cannot be recommended since there may exist a non-negligible portion of the population that experiences an individual effect in the opposite direction of the mean effect. This variability of a treatment effect is referred to as treatment heterogeneity.

Using a potential outcomes framework, treatment heterogeneity for several simple designs has been investigated using a randomization based approach. However, as experimental designs become more complicated, a randomization-based approach becomes increasingly intractable. We present an approach to derive a "potential outcomes" linear mixed effects model. From this model, treatment heterogeneity is conceptualized as a linear combination of potential model variance components. These variance components are non-estimable in observable data, but estimable bounds exist that depend on the experimental design and they arise from linear combinations of the non-estimable potential model variance components. A specific application of these results to a 2x2 factorial treatment structure in a 4-period cross-over experimental design is presented. Assumptions required for equating naïve estimates from observable data to those that could be obtained from potential outcomes data are discussed.

5.2. Introduction

Treatment heterogeneity refers to the variability of a treatment effect across individuals in a population. Although such variability has sometimes been acknowledged as an important consideration in experimental studies, decisions about the use of treatment generally make use of statistical information gathered about the mean effect and then apply that same information to the individual (cf. Marshall, 1997). When there is a high degree of treatment heterogeneity in a population, there may be a non-negligible proportion of the population responding differently to a treatment, and possibly in the opposite direction, from the average subject.

Quantifying the degree of treatment heterogeneity is facilitated by potential outcomes (Rubin, 1974). Consider a set of treatments, $\{T, C\}$, where T denotes some test treatment and C denotes a reference or perhaps a control treatment. For each subject there is a duplet, $\{r_T, r_C\}$, which represents the potential outcome to the test treatment and to the control treatment, respectively. At any particular time point, either r_T or r_C is observable for an individual so that the individual causal effect, $d = r_T - r_C$, cannot be observed – what Holland (1986) referred to as the fundamental problem of causal inference. As in Gadbury (2010) or in Poulson et al. (2012), treatment heterogeneity is quantified by $\sigma_d^2 = Var(d)$, a nonestimable quantity since there is no information in observable data on the correlation between r_T and r_C. If we suppose, as in Gadbury and Iyer (2000) or Poulson et al., (2012) that the duplets arise from an infinite population model given by

$$\begin{pmatrix} r_T \\ r_C \end{pmatrix} \sim \left\{ \begin{pmatrix} \mu_T \\ \mu_C \end{pmatrix}, \begin{bmatrix} \sigma_T^2 & \rho_{TC}\sigma_T\sigma_C \\ \rho_{TC}\sigma_T\sigma_C & \sigma_C^2 \end{bmatrix} \right\}, \tag{5.1}$$

then it is easy to see that $(\sigma_T - \sigma_C)^2 \leq \sigma_d^2 \leq (\sigma_T + \sigma_C)^2$. Thus, non-estimable treatment heterogeneity can be bounded by estimable quantities, resulting from setting the non-estimable correlation, ρ_{TC}, equal to 1 and -1. Bounds can be tightened using covariate information (Gadbury et al., 2001; Poulson et al., 2012), and estimates of treatment heterogeneity can be obtained using assumed conditional independence between potential outcomes given covariates (Zhang et al. 2013).

As experimental designs become more sophisticated, more information about treatment heterogeneity may become available. If a blocking or subsetting variable is available, then there

are methods that can detect the presence of treatment heterogeneity and potential qualitative interactions, the latter meaning that the direction of a treatment's effect differs across subsets (e.g., Byar and Corle 1977; Simon 1982; Gail and Simon 1985; Silvapulle 2001; Li and Chan 2006). In repeated measures designs or cross-over designs, there is a true individual treatment effect at each time period, and some have demonstrated the use of mixed-effects models fit to data from cross-over designs that estimated a subject-treatment (S-T) interaction variance (e.g., Hauck et al. 2000; Endrenyi and Tothfalusi 1999). However, the estimated variance computed from observed data may not equal a variance of true individual effects without certain assumptions and/or depending upon how one defines an individual effect in multiple period designs. In more complex designs, it is not always clear what these assumptions are and whether or not they are reasonable for the application. The relationship between an estimable S-T variance component and the true variance of an individual effect defined in a potential outcomes framework remains unclear.

In this paper a data example from a 2x2 treatment structure applied to a 4-period cross-over design is analyzed. These data were collected to investigate the effect of diet and plant sterols on blood low-density lipoprotein cholesterol levels. Dietary or nutritional recommendations for health that are reported in the literature and media can be a source of considerable confusion to the public. Discussions relating to this, though with different perspectives, can be found in a popular book by Campbell and Campbell (2005) and at The Weston A. Price Foundation (http://www.westonaprice.org/). Thus, it seems pertinent to consider an application area where an investigation into treatment heterogeneity may yield additional insights regarding a treatment's behavior on a population beyond what is told by a study of mean effects.

The data considered here resulted from a double-blind, randomized cross-over design, and were reported in Chen et al, 2009 and Kramer et al, 2011. The purpose of Chen et al, (2009) was to determine if the main effects of two levels of diet and two levels of supplemented plant sterols on low-density lipoprotein cholesterol (LDLC) blood concentrations were additive. In a subsequent publication (Kramer et al, 2011), these data were used as an illustrative example while investigating the use of multiplicative decomposition techniques to estimate a subject-by-diet interaction effect since experience suggested the LDLC responses to diet tend to be subject-specific.

The method proposed here first conceptualizes the potential outcomes in a design and quantifies treatment heterogeneity as a linear combination of variance components in a linear mixed effects model (LMM). Then the randomization mechanism is invoked to produce observable data and variance components that are no longer estimable in observable data, at least not without assumptions. A key step in comparing the potential and observable LMM's is the appropriate identification of the potential LMM. This is accomplished using a technique proposed by Fisher in a discussion of Yates' paper on complex experiments (1935) where Fisher demonstrated that the choice of an experimental design is the choice of how a topographical layout of the experiment is related to the treatment structure of an experiment. Stroup (2013) adapted Fisher's approach as a means of correctly identifying the appropriate components of an observable LMM, and termed the approach "What Would Fisher Do" (WWFD). Using Stroup's WWFD method, we further adapt Fisher's approach to accommodate a potential outcomes framework, and then consider what information is "lost" when the randomization mechanism is invoked, that is, we use the potential LMM as a template to arrive at the observable LMM. This process is an important step in the appropriate estimation of effects in the observable model as misspecification of the model in PROC GLIMMIX has been demonstrated to alter both model effect estimation and inference (Boykin et al., 2010).

For ease of illustration, the WWFD idea is first presented in the context of a straightforward two-sample completely randomized design (CRD). We then use the technique on the diet and plant sterol data example from a 2x2 treatment structure applied to a 4-period cross-over design, previously described. Considering the potential LMM clarifies the assumptions necessary to equate estimable variances to the variances of the individual effects. Furthermore, additional information regarding treatment heterogeneity that is not estimable in a traditional 2x2 factorial design at one time period becomes available due to the cross-over nature of this design. This additional information hints at what might be surmised about treatment heterogeneity with added time periods, if practical, and what assumptions would be required to directly estimate a treatment heterogeneity variance. We conclude with a discussion.

5.3. WWFD in a Two-Sample CRD

A simple two-sample CRD is used to illustrate basic principles that may be extended to accommodate more complex experimental designs. Consider a two-sample CRD in which a

random effect arising from the application of the j^{th} level of treatment to the i^{th} experimental unit (EU) is permitted. Potential outcomes consist of two sets of N responses, where each of the N EU's simultaneously contribute one response to each of the two sets so that EU's are crossed with sets. The structure for the potential outcomes framework and corresponding degrees of freedom given in Table 5.1 (i) are a completely topographical analysis in that the total degrees of freedom for the experiment are accounted for, independent of the treatment structure. The treatment structure and its corresponding degrees of freedom are given in Table 5.1 (ii). "Parallels" in Table 5.1 (ii) was a term used by Fisher to represent the number of times a level of treatment must be prepared to accommodate a given sample size. In this case, there are two levels of treatment and each level of treatment must be prepared N times, once for each EU; therefore, the degrees of freedom associated with Parallels is 2*(N-1). Both the Topographical and Treatment aspects completely account for the total degrees of freedom in the experiment. Per Fisher's instruction that the choice of an experimental design is the choice of which components from the topographical and treatment aspects are permitted to correspond, we combine these two aspects in Table 5.1 (i) and (ii) by choosing the degrees of freedom associated with "Trt" in Table 5.1 (ii) to correspond to the degrees of freedom associated with "Set" in Table 5.1 (i). That is, assume that any difference between sets is attributed to the level of treatment applied to that set and not to characteristics inherent to the set. Accordingly, we choose the degrees of freedom associated with "Parallels" in the Table 5.1 (ii) to be partitioned into the degrees of freedom associated with "EU" and "Set*EU" in the Table 5.1 (i). That is, we assume that differences in responses within a set are due to either inherent characteristics of the EU or the application of a level of treatment to a particular EU rather than differences in the preparation of a particular level of treatment. The resulting combined ANOVA table is given in Table 5.1 (iii) by replacing "Set" with "Trt" everywhere "Set" appears in Table 5.1 (i):

Topographical		Trt		Combined	
Source	d.f.	Source	d.f.	Source	d.f.
Set	2-1	Trt	2-1	Trt	2-1
EU	N-1	"parallels"	2(N-1)	EU	N-1
Set*EU	(2-1)*(N-1)			Trt*EU	(2-1)*(N-1)
Total	2N-1	Total	2N-1	Total	2N-1
(i)		(ii)		(iii)	

Table 5.1 *Potential WWFD ANOVA Structure: Two-Sample CRD*
(i) Topographical, (ii) Treatment, and (iii) Combined ANOVA structures for a Potential two-sample CRD.

Using the components of the combined ANOVA table above as a guide, the resulting potential LMM is

$$r_{ij} = \mu + s_i + \tau_j + s\tau_{ij} \qquad (5.2)$$
$$i = 1,2,\dots,N\ subjects;\ j = T, C$$

where s_i represents a random effect of the i^{th} EU, τ_j represents a fixed effect of the j^{th} level of treatment, and $s\tau_{ij}$ represents the random effect of the j^{th} level of treatment applied to the i^{th} EU. In a model assuming no technical error, $s\tau_{ij}$ would be considered the experimental error.

Under the "usual" set of experimental circumstances for random effects models, the following distributional properties of s_i and $s\tau_{ij}$ are assumed:

$$s_j \sim iid\ N(0, \sigma_s^2)$$
$$s\tau_{ij} \sim iid\ N(0, \sigma_{st}^2) \qquad (5.3)$$
$$s_j\ and\ s\tau_{ij}\ are\ mutually\ independent.$$

One can allow for different variance components for $s\tau_{ij}$for $j = T, C$, but this is unnecessary for illustrating the ideas here. Invoking the randomization mechanism effectively removes one-half of the data so that each EU is now represented only once within a set instead of being represented in both sets. This results in two distinct sets of responses with $n = \frac{N}{2}$ EU's in each set, assuming a balanced design. This effectively removes the "Set*EU" term from Table 5.1 (i) and replaces it with an "EU(set)" term. Also notice that the degrees of freedom associated with "Parallels" in Table 5.1 (ii) is reduced since each level of treatment need be prepared only n times instead of N. Table 5.2 (i) and (ii) demonstrate how the Topographical and Treatment structures are altered after the randomization mechanism is invoked.

Based on this new Combined ANOVA table given in Table 5.2 (iii), the observable LMM can be written

$$R_{ij} = \mu + \tau_j + \varepsilon_{ij}$$
$$i = 1,2,\dots,n_j, \qquad j = T, C$$

Topographical		Trt		Combined	
Source	d.f.	Source	d.f.	Source	d.f.
Set	2-1	Trt	2-1	Trt	2-1
EU(Set)	~~N-1~~ 2(n-1)	"parallels"	~~2(N-1)~~ 2(n-1)	EU(Trt)	~~N-1~~ 2(n-1)
Set*EU	(2-1)*(N-1)			Trt*EU	(2-1)*(N-1)
Total	~~2N-1~~ 2n-1	Total	~~2N-1~~ 2n-1	Total	~~2N-1~~ 2n-1
(i)		(ii)		(iii)	

Table 5.2 *Observable WWFD ANOVA Structure: Two-Sample CRD*
(i) Topographical, (ii) Treatment, and (iii) Combined ANOVA structures for an Observable two-sample CRD.

where n_j is the number of EU's per level of treatment, such that $N = n_T + n_C = 2n$ in a balanced two-sample CRD (i.e., $n_T = n_C = n$), and ε_{ij} is the usual error term in a two-sample CRD.

A direct relationship between the potential and observable models can be established by defining

$$\varepsilon_{ij} = s_i + s\tau_{ij}$$

Based on the distributional assumptions in (5.3), the error variance in the observable model, denoted σ_e^2, is given by

$$\sigma_e^2 = \sigma_s^2 + \sigma_{st}^2$$

There is not enough experimental material in the observable model framework to estimate all effects of interest specified in the potential model. In the observable model, only the linear combination of the variance components of subject and subject-by-treatment effects can be estimated. If the potential framework were feasible, both the variance of the subject effect and the variance of the subject-by-treatment effect would be estimable.

With the potential LMM given in (5.2), d_i is given by,

$$d_i = (\tau_T - \tau_C) + (s\tau_{iT} - s\tau_{iC}).$$

Using the distribution of d_i given in (5.3) and the model distributional assumptions in (5.3) gives

$$d_i \sim iid\ N[(\tau_T - \tau_C), 2{\sigma_{st}}^2].$$

If ${\sigma_{st}}^2 > 0$, then ${\sigma_d}^2 = 2{\sigma_{st}}^2 > 0$, and treatment heterogeneity exists. Bounds for ${\sigma_d}^2$ using this structure are different from the trivial ones given following equation (5.1). These bounds depend on the non-estimable individual variance components, $\sigma_s^2, \sigma_{st}^2$, that are estimable as a linear combination. Still, an estimable upper bound is given by $2\sigma_e^2$ but the lower bound is zero. The non-estimable correlation in (5.1) is now the intra-class correlation, $\frac{\sigma_s^2}{\sigma_s^2+\sigma_{st}^2}$, and the lower bound for this quantity is zero rather than -1. Allowing for a negative correlation between potential outcome variables requires specifying a bivariate distribution of random effects where the random effects are not independent. We have not seen this done when applying linear mixed effects models to data arising from experimental designs.

For this simple design, relating the quantities in an observable model to those in the potential model takes some thought. But it highlights the information that gets lost as one moves from potential to observable data and, thus, what quantities in a model become inestimable. The relationship between the potential model and observable model is not as explicit in more complicated designs, but the WWFD technique can still be used to relate quantities in a potential LMM to those in an observable LMM for any particular experimental design.

5.4. 2x2 Treatment Structure in a Cross-over Design: A Data Example

5.4.1. Data Description

Each of 22 subjects (13 male, 9 female) was assigned to receive each of four treatment combinations of diet and plant sterols in random order for a period of 28 days (Chen et al., 2009). There were no washout periods between 28-day intervals. Two levels of treatment were considered for each treatment factor. The levels of diet were a typical American diet (TAD) versus a recommended cholesterol-lowering Step-1 diet (STP). The levels of plant sterol (PSE) were 0 g/day and 3.3 g/day incorporated into the diet. At each period, the study design resembled a 2 x 2 factorial treatment structure with two levels of each treatment factor assigned to each subject. At the end of the four periods, each subject had received all combinations of the two treatments.

A number of blood compounds were measured, however only LDLC (mmol/L) measurements are discussed here. LDLC responses represent the average LDLC values from two samples taken at day 22 and day 24 of each 28-day period. Baseline (pre-experiment) measurements were taken the week prior to the initiation of the first, randomly-assigned treatment combination. The outcome is a change from baseline with negative values meaning a decrease from baseline.

5.4.2 Applying WWFD to this Design

The previous discussion related to a two-sample CRD can be extended to accommodate a factorial treatment structure with two treatment factors, $\{\alpha, \beta\}$, each having two levels, $\{T, C\}$. The entire set of possible treatment combinations in this 2x2 factorial experiment is the set $\{TT, TC, CT, CC\}$, where the level of α is given first followed by the level of β. For the LDLC data, a treatment level T denotes the STP diet for α and the 3.3 g/day dose of PSE for β (i.e., the respective C treatment levels are TAD for α and 0 g/day of PSE for β). A design consisting of more than two levels of treatment per treatment factor could also be accommodated.

For each EU, potential outcomes are a 4-tuple $\{r_{iTT}, r_{iTC}, r_{iCT}, r_{iCC}\}$, which represents the potential response of the i^{th} EU under each of the four possible treatment combinations arising from the factorial treatment structure, with only one being observable at a particular time. The observed response of the i^{th} EU at a particular time is given by,

$$R_i = [r_{iTT}U_iW_i] + [r_{iTC}U_i(1 - W_i)] + [r_{iCT}(1 - U_i)W_i] + [r_{iCC}(1 - U_i)(1 - W_i)],$$

where

$$U_i = \begin{cases} 1, if\ \alpha = T\ for\ the\ i^{th}\ EU \\ 0, if\ \alpha = C\ for\ the\ i^{th}\ EU \end{cases}$$

and

$$W_i = \begin{cases} 1, if\ \beta = T\ for\ the\ i^{th}\ EU \\ 0, if\ \beta = C\ for\ the\ i^{th}\ EU. \end{cases}$$

We assume uniform randomization and independence of U and W.

To extend the 2x2 factorial potential outcomes framework to a 4-period cross-over design, we assume a unique set of four potential responses at every time period, one response per

treatment combination per period. There are a total of 16 potential responses per EU are permitted across the entire experiment. We suppose that the randomization mechanism randomly selects a sequence of responses across the four periods for every EU so that every EU receives every treatment combination once across the four periods.

Using Stroup's WWFD method, the following potential LMM for a 2x2 factorial treatment structure in a 4-Period cross-over design is obtained:

$$\begin{gathered} r_{ijkl} = \mu + s_i + \pi_j + s\pi_{ij} + \alpha_k + \beta_l + \alpha\beta_{kl} + s\alpha_{ik} + s\beta_{il} + s\alpha\beta_{ikl} + \pi\alpha_{jk} + \pi\beta_{jl} \\ +\pi\alpha\beta_{jkl} + s\pi\alpha_{ijk} + s\pi\beta_{ijl} + s\pi\alpha\beta_{ijkl} \\ i = 1,2,\dots,N\ EU's;\ j = 1,2,3,4\ periods; k = T,C\ levels\ of\ diet; l = T,C\ levels\ of\ PSE \end{gathered} \tag{5.4}$$

where s_i represents a random effect of the i^{th} EU; π_j represents the fixed effect of the j^{th} period; $s\pi_{ij}$ represents a random interaction effect of the i^{th} EU measured at the j^{th} period; α_k represents a fixed effect of the k^{th} level of diet; β_l represents a fixed effect of the l^{th} level of PSE; $\alpha\beta_{kl}$ represents a fixed interaction effect of the k^{th} level of diet combined with the l^{th} level of PSE; $s\alpha_{ik}$ represents a random interaction effect of the k^{th} level of diet applied to the i^{th} EU; $s\beta_{il}$ represents a random interaction effect of the l^{th} level of PSE applied to the i^{th} EU; $s\alpha\beta_{ikl}$ represents a random interaction effect of the k^{th} level of diet combined with the l^{th} level of PSE applied to the i^{th} EU; $\pi\alpha_{jk}$ represents a fixed interaction effect of the k^{th} level of diet applied at the j^{th} period; $\pi\beta_{jl}$ represents a fixed interaction effect of the l^{th} level of PSE applied at the j^{th} period; $\pi\alpha\beta_{jkl}$ represents a fixed interaction effect of the k^{th} level of diet combined with the l^{th} level of PSE applied at the j^{th} period; $s\pi\alpha_{ijk}$ represents a random interaction effect of the k^{th} level of diet applied to the i^{th} EU at the j^{th} period; $s\pi\beta_{ijl}$ represents a random interaction effect of the l^{th} level of PSE applied to the i^{th} EU at the j^{th} period; and $s\pi\alpha\beta_{ijkl}$ represents a random interaction effect of the k^{th} level of diet combined with the l^{th} level of PSE applied to the i^{th} EU at the j^{th} period, and should be considered experimental error.

The distributional assumptions of the random effects are as follows:

$$\begin{gathered} s_i \sim iid\ N(0, \sigma_s^2) \\ s\pi_{ij} \sim iid\ N(0, \sigma_{s\pi}^2) \\ s\alpha_{ik} \sim iid\ N(0, \sigma_{s\alpha}^2) \\ s\beta_{il} \sim iid\ N\left(0, \sigma_{s\beta}^2\right) \end{gathered} \tag{5.5}$$

$$s\alpha\beta_{ikl} \sim iid\ N(0, \sigma^2_{s\alpha\beta})$$
$$s\pi\alpha_{ikl} \sim iid\ N(0, \sigma^2_{s\pi\alpha})$$
$$s\pi\beta_{ikl} \sim iid\ N(0, \sigma^2_{s\pi\beta})$$
$$s\pi\alpha\beta_{ikl} \sim iid\ N(0, \sigma^2_{s\pi\alpha\beta})$$

$s_i, s\pi_{ij}, s\alpha_{ik}, s\beta_{il},\ s\alpha\beta_{ikl}, s\pi\alpha_{ikl}, s\pi\beta_{ikl}\ and\ s\pi\alpha\beta_{ikl}$ are mutually independent.

The resulting observable LMM for this design is

$$R_{ijkl} = \mu + s_i + \pi_j + \alpha_k + \beta_l + \alpha\beta_{kl} + s\alpha_{ik} + s\beta_{il} + \pi\alpha_{jk} + \pi\beta_{jl} + \pi\alpha\beta_{jkl} + \varepsilon_{ijkl} \qquad (5.6)$$
$$i = 1,2,\dots,N\ EU's;\ j = 1,2,3,4\ periods; k = T, C\ levels\ of\ diet; l = T, C\ levels\ of\ PSE$$

where ε_{ijk} is comprised of the confounded potential model effects for which there is not enough experimental material in the observable model framework to estimate. In the observable model, ε_{ijk} is considered experimental error. All other effects maintain the same definition as in the potential LMM.

A direct relationship between the observable model and the potential model is established by defining

$$\varepsilon_{ijkl} = s\pi_{ij} + s\alpha\beta_{ikl} + s\pi\alpha_{ijk} + s\pi\beta_{ijl} + s\pi\alpha\beta_{ijkl}.$$

Given the distributional assumptions specified in (5.5),

$$\sigma^2_e = \sigma^2_{s\pi} + \sigma^2_{s\alpha\beta} + \sigma^2_{s\pi\alpha} + \sigma^2_{s\pi\beta} + \sigma^2_{s\pi\alpha\beta}.$$

According to the model given in (5.4), a true causal effect at each of the four periods can be defined as

$$d_{ij|k=T} = r_{ijTT} - r_{ijTC},$$
$$d_{ij|k=C} = r_{ijCT} - r_{ijCC},$$
$$d_{ij|l=T} = r_{ijTT} - r_{ijCT},$$
$$d_{ij|l=C} = r_{ijTC} - r_{ijCC}$$

for the i^{th} EU at the j^{th} period. For each EU there are 16 causal effects across the four periods.

Based on the model assumptions given in (5.5) , $var(d_{ij|k=T}) = var(d_{ij|k=C})$ and $var(d_{ij|l=T}) = var(d_{ij|l=C})$. Denote these two variances $var(d_{ij|Diet})$ and $var(d_{ij|PSE})$, respectively. Writing these variances in terms of the potential LMM variance components yields

$$\begin{aligned} var(d_{ij|Diet}) &= 2(\sigma^2_{s\beta} + \sigma^2_{s\alpha\beta} + \sigma^2_{s\pi\beta} + \sigma^2_{s\pi\alpha\beta}) \\ \text{and} \\ var(d_{ij|PSE}) &= 2(\sigma^2_{s\alpha} + \sigma^2_{s\alpha\beta} + \sigma^2_{s\pi\alpha} + \sigma^2_{s\pi\alpha\beta}). \end{aligned} \tag{5.7}$$

$\sigma^2_{s\alpha}$ and $\sigma^2_{s\beta}$ are estimable in observable data because there are multiple observations per EU on a particular level of diet (but differing levels of PSE) and multiple observations per EU on a particular level of PSE (but differing levels of diet). This permits an estimable lower bound since

$$\begin{aligned} 2\sigma^2_{s\beta} &\leq var(d_{ij|Diet}) \\ \text{and} \\ 2\sigma^2_{s\alpha} &\leq var(d_{ij|PSE}) \end{aligned} \tag{5.8}$$

The cross-over nature of this design permits the definition of an observable, naïve version of individual effects. Four naïve differences are

$$\begin{aligned} D_{i|k=T} &= R_{ijTT} - R_{ij'TC}, \\ D_{i|k=C} &= R_{ij''CT} - R_{ij'''CC}, \\ D_{i|l=T} &= R_{ijTT} - R_{ij''CT}, \\ D_{i|l=C} &= R_{ij'TC} - R_{ij'''CC}. \end{aligned} \tag{5.9}$$

Two distinct variances for the naïve individual effects defined in (5.9) emerge from this design. They are, $var(D_{i|k=T}) = var(D_{i|k=C})$ and $var(D_{i|l=T}) = var(D_{i|l=C})$ based on the assumptions in (5.5). Denote these variances as $var(D_{i|Diet})$ and $var(D_{i|PSE})$, respectively, then,

$$\begin{aligned} var(D_{i|Diet}) &= var(R_{ijkT} - R_{ij'kC}) = 2(\sigma^2_{s\beta} + \sigma^2_e) \\ \text{and} \\ var(D_{i|PSE}) &= var(R_{ijTl} - R_{ij'Cl}) = 2(\sigma^2_{s\alpha} + \sigma^2_e) \end{aligned} \tag{5.10}$$

where *j* and *j'* indicate two different periods. Estimable upper bounds of the true variances of the individual effects can be established since

$$\begin{aligned} 2\left(\sigma_{s\beta}^2 + \sigma_e^2\right) &= 2\left(\sigma_{s\beta}^2 + \sigma_{s\pi}^2 + \sigma_{s\alpha\beta}^2 + \sigma_{s\pi\alpha}^2 + \sigma_{s\pi\beta}^2 + \sigma_{s\pi\alpha\beta}^2\right) \\ &= 2\sigma_{s\pi}^2 + 2\sigma_{s\pi\alpha}^2 + var\left(d_{ij|Diet}\right) \end{aligned}$$

and (5.11)

$$\begin{aligned} 2(\sigma_{s\alpha}^2 + \sigma_e^2) &= 2\left(\sigma_{s\alpha}^2 + \sigma_{s\pi}^2 + \sigma_{s\alpha\beta}^2 + \sigma_{s\pi\alpha}^2 + \sigma_{s\pi\beta}^2 + \sigma_{s\pi\alpha\beta}^2\right) \\ &= 2\sigma_{s\pi}^2 + 2\sigma_{s\pi\beta}^2 + var\left(d_{ij|PSE}\right). \end{aligned}$$

Combining the results in (5.8) and (5.11), we have

$$2\sigma_{s\beta}^2 \le var\left(d_{ij|Diet}\right) \le var\left(D_{i|Diet}\right)$$

and (5.12)

$$2\sigma_{s\alpha}^2 \le var\left(d_{ij|PSE}\right) \le var\left(D_{i|PSE}\right)$$

where the upper bounds are given in (5.10). The difference between the upper and lower bounds is equal to $2\sigma_e^2$.

Comparing the lower bounds established in (5.12) with the results of a traditional 2x2 factorial design carried out at a single time period yields an important distinction. In a standard 2x2 factorial design, a single observable response is permitted for each EU at a single time period under only one level of diet combined with only one level of PSE. By construction of the design, then, none of the variance components given in (5.7) are individually estimable from observable data in this design. Consequently, the most that can be stated about the lower bound of the variance of an individual effect is that it is non-negative. Thus, the extension of the 2x2 factorial to a cross-over design yields additional information regarding treatment heterogeneity and provides an estimable lower bound.

If it were possible and practical to extend the design to permit eight time periods instead of four, and each of the four diet-by-PSE combinations were randomly assigned to two of the eight time periods, then the subject-by-diet-by-PSE variance component would become estimable. So by extending the design to a repeated measures cross-over design, previously non-estimable components of the variances given in (5.7) become estimable. Additional discussion

regarding repeated measures cross-over designs for evaluating treatment heterogeneity can be found in Senn (2001).

5.4.3 Results of Analysis

Using PROC GLIMMIX, we analyzed the LDLC data according to the model given in (5.6) where α_k represents a fixed diet effect and β_l represents a fixed PSE effect. Table 5.3(i) gives the results for tests of fixed effects in the model. Table 5.3 (ii) presents the estimates of interest and standard errors for both fixed and random effects. A negative value represents a reduction in LDLC levels.

Our results demonstrate that the STP diet significantly lowers mean LDLC compared with the TAD diet (P = 0.012), and the introduction of 3.3 g/day of PSE significantly reduces mean LDLC compared with 0 g/day of PSE (P < 0.0001). The interaction between diet and PSE is not significant. Period-by-treatment interactions, Period-by-Diet, Period-by-PSE, and Period-by-Diet-by-PSE, are also not significant. These results are consistent with those published by Chen et al, although estimates and P-values are slightly different. Chen et al. accounted for individual differences by including a base-line LDLC measurement in the model and a random subject effect. The remaining residuals were fit with a one parameter autoregressive correlation structure.

Further analyses not shown here demonstrated that adding a base-line LDLC measurement to the model affected the estimate of the EU variance component but not the estimates of the EU-by-diet, the EU-by-PSE, or the residual variances. The estimates of the variance components in Table 5.3 (ii) give rise to estimable bounds of the variance of individual effects established in (5.12). Gadbury and Iyer (2000) describe a process by which the proportion of EU's in a population experiencing an unfavorable response can be estimated, assuming a normal distribution for individual effects. In this case, an unfavorable response would be considered an elevation in blood LDLC levels even though, on average, a reduction in LDLC levels was observed. Without loss of generality, assume $\mu_d < 0$. Then, the proportion of EU's experiencing an unfavorable response in this particular experimental setting is given by

$$P_{-} = P(d > 0) = 1 - \Phi\left(\frac{-\mu_d}{\sqrt{var(d)}}\right),$$

where μ_d is the mean effect of one level of treatment compared to the other. Table 5.4 gives the estimated upper and lower bounds of the variance of the individual causal effects. In addition to the estimable bounds we used $B = 50$ bootstrap (Efron and Tibshirani, 1994) samples to compute the bootstrap standard error of both the upper and lower bounds. These bootstrap standard errors, as well as the corresponding estimates of P_-, are given in Table 5.4. The difference between the estimable upper and lower bounds for both variances of interest is 0.0223. This is twice the estimate of σ_e^2 given in Table 5.3 (ii). Also notice that for both variances of interest, the estimate of the lower bound is more than two bootstrap standard errors above zero. Thus, the data suggest that treatment heterogeneity exists for both Diet and PSE effects.

Type III Tests of Fixed Effects		
Fixed Effect	**F-Value**	**P-value**
Period	2.05	0.1601
Diet	7.52	0.0122
PSE	70.44	<.0001
Diet ∗ PSE	0.21	0.6543
Period ∗ Diet	1.02	0.4181
Period ∗ PSE	1.50	0.2642
Period ∗ Diet ∗ PSE	1.36	0.3009

(i)

Estimates: LDLC (mmol/L)		
Difference	**Estimate**	**Std. Error**
Diet: STP-TAD	-0.1637	0.0597
PSE: 3.3 g/day – 0 g/day	-0.4491	0.0535

Variance Component	**Estimate**	**Std. Error**
σ_s^2	0.2095	0.0754
$\sigma_{s\alpha}^2$	0.0332	0.0125
$\sigma_{s\beta}^2$	0.0256	0.0104
σ_e^2	0.0112	0.0041

(ii)

Table 5.3 *SAS PROC GLIMMIX Results.*
(i)Type III Tests for Fixed Effects. (ii) Estimates of the difference in levels of Diet and PSE with standard errors and estimates of the variance of random effects with estimated standard errors.

Individual Effect	Estimate	Bootstrap SE	Estimate of $P_{\_}$	Bootstrap SE $P_{\_}$
$var(d_{ij\|Diet})$				
Lower Bound	0.0511	0.0220	0.0235	0.0288
Upper Bound	0.0734	0.0206	0.0487	0.0339
$var(d_{ij\|PSE})$				
Lower Bound	0.0665	0.0288	0.2628	0.0809
Upper Bound	0.0888	0.0271	0.2914	0.0647

Table 5.4 *Estimable Bounds for the Variance of Individual Effects.*
Estimates of the upper and lower bounds given in equation (12) with bootstrap standard errors and estimates of P_ with bootstrap standard errors.

5.5. Discussion and Conclusion

In cases where treatment heterogeneity is suspected, it would be prudent design experiments in such a way as to investigate the presence of treatment heterogeneity in addition to estimating a mean effect before a claim of the superiority of one treatment over another is established (Longford, 1999). The variance of an individual effect is the parameter of interest when assessing treatment heterogeneity, with a non-zero value indicating the presence of treatment heterogeneity. If the estimate of the lower bound is substantially greater than zero, one might conclude that treatment heterogeneity is present. Likewise, if an estimable upper bound is very close to zero then one might conclude that the treatment is having a similar effect on individuals across a population. Experimental designs in which an estimable lower and/or upper bound can be established permit the investigation of treatment heterogeneity essentially "without cost" in the sense that no new data are needed to confirm the presence of treatment heterogeneity. Furthermore, a comparison of the observable LMM and potential LMM for a given experimental design delineates the information about causal effects that is lost in moving from potential to observable data, and what assumptions about non-estimable quantities (or design modifications) are needed to evaluate treatment heterogeneity in observable data.

We demonstrated that the extension of a traditional 2x2 factorial treatment structure to a four-period cross-over design permits the estimation of both an upper and lower bound of the variance of an individual effect, defined in a potential outcomes framework. Given the estimated bounds of the individual effects and the bootstrap standard errors, it was reasonable to conclude that treatment heterogeneity exists when considering the effect of diet (TAD vs. STP) and PSE (0 mg/day vs.3.3 mg/day). Furthermore, we estimated the proportion of EU's experiencing an

unfavorable effect. The point estimates of P_- indicate that there could exist certain EU's for which the TAD is a more favorable diet than the STP and certain EU's for which 0 g/day of PSE could be more favorable than 3.3 g/day of PSE even though, on average, the STP diet and the 3.3 g/day of PSE appeared to be more favorable for lowering LDLC levels. Poulson et al. (2012) called this an individual qualitative interaction (IQI). However, after considering standard errors of estimated bounds for P_-, statistically it appears that only diet may have an IQI.

The difference between the estimable upper and lower bounds of the variance of individual effects is comparatively small. This is because the variability explained by the random residuals is less than the variability explained by the other random effects, based on the estimates given in Table 5.3 (ii) ($\hat{\sigma}_e{}^2 = 0.0112$). The majority of variability in responses is accounted for by the variability due to the random EU effect ($\hat{\sigma}_s{}^2 = 0.2095$). In other words, while treatment heterogeneity likely exists, the amount of total variability in responses explained by the variability of individual effect is small compared to the variability inherent to EU's selected from a given population.

Consideration of (5.7), (5.8), (5.10), and (5.11) reveals the required assumptions to equate the variance of individual effects with the corresponding naïve estimates available from observable data. From (5.7) and (5.8), $2\sigma^2_{s\beta} = var\left(d_{ij|Diet}\right)$ and $2\sigma^2_{s\alpha} = var\left(d_{ij|PSE}\right)$ if we are willing to assume that $\sigma^2_{s\alpha\beta} = \sigma^2_{s\pi\alpha} = \sigma^2_{s\pi\beta} = \sigma^2_{s\pi\alpha\beta} = 0$; that is , if we are willing to assume that the diet-by-sterol effect, the period-by-diet effect, the period-by-sterol effect, and the period-by-diet-by-sterol effect are all constant across EU's in a population. If we are willing to make these assumptions, then the estimable lower bounds of $var\left(d_{ij|Diet}\right)$ and $var\left(d_{ij|PSE}\right)$ become estimates of the variance of the respective individual effects. If we assume $\sigma^2_{s\pi} = \sigma^2_{s\pi\alpha} = \sigma^2_{s\pi\beta} = 0$, that is, if the period effect, the period-by-diet effect, and the period-by-sterol effect are all constant across EU's in a population, then the estimable upper bounds become estimates of $var\left(d_{ij|Diet}\right)$ and $var\left(d_{ij|PSE}\right)$. Though the potential LMM helps to clarify what assumptions are needed to equate estimated bounds with estimated treatment heterogeneity, these assumptions cannot be directly tested using observable data from this design.

Senn (2001) noted that studies are rarely designed to separate information on an individual effect from other sources of variability. In the two-sample CRD and traditional 2x2 factorial designs, only estimable upper bounds can be established. While knowledge of this

upper bound informs the "worst-case" scenario regarding the estimate of P_- (e.g.- a larger variance yields a larger P_-), it is the estimable lower bound that informs the presence of treatment heterogeneity. If treatment heterogeneity is suspected and a design permitting an estimable lower bound of the variance of an individual effect is possible, then estimating the degree of treatment heterogeneity in addition to a mean treatment effect should be of value when characterizing a treatment effect across an entire population.

While the statistical methods presented here may be used to quantify the degree of treatment heterogeneity in these data, they cannot explain the source of the treatment heterogeneity. Further research is required to investigate the possible causes of treatment heterogeneity in LDLC response to different diets and amounts of PSE. The data example was used for illustration and not to confirm a superiority of one treatment over another. The sample size was small and other issues such as treatment compliance were not considered.

5.6 Supplementary Material

The following supplementary material is included for the benefit of the reader, and describes how the potential LMM was generated. This material was omitted from the body of the paper due to space concerns.

Topographical		Trt		Combined	
Source	d.f.	Source	d.f.	Source	d.f.
Replicate	4-1=3	Period	4-1=3	Period	4-1=3
Row(Rep)	4(2-1)=4	Diet	2-1=1	Diet	2-1=1
		Period*Diet	3x1=3	Period*Diet	3x1=3
Col(Rep)	4(2-1)=4	PSE	2-1=1	PSE	2-1=1
		Period*PSE	3x1=3	Period*PSE	3x1=3
Row(Rep)*Col(Rep)	4(1x1)=4	Diet*PSE	1x1=1	Diet*PSE	1x1=1
		Period*Diet*PSE	3x1=3	Period*Diet*PSE	3x1=3
Subject	N-1	Parallels	16(N-1)	Subject	N-1
Subject*Rep	3(N-1)			Subject*Period	3(N-1)
Subject* Row(Rep)	4(N-1)			Subject* Diet	(N-1)
				Subject*Period*Diet	3(N-1)
Subject* Col(Rep)	4(N-1)			Subject* PSE	(N-1)
				Subject*Period*PSE	3(N-1)
Subject* Row(Rep)*Col(Rep)	4(N-1)			Subject* Diet*PSE	(N-1)
				Subject*Period*Diet*PSE	3(N-1)
Total	16N-1	Total	16N-1	Total	16N-1
(i)		(ii)		(iii)	

Potential Model WWFD:2x2 Factorial in a Repeated Measures Cross-over Design.

	Part of 4-Tuple Receiving:			
EU	**TT**	**TC**	**CT**	**CC**
1	TT	TC	CT	CC
2	TT	TC	CT	CC
...	...	...	...	...
N-1	TT	TC	CT	CC
N	TT	TC	CT	CC

Period 1

...

	Part of 4-Tuple Receiving:			
EU	**TT**	**TC**	**CT**	**CC**
1	TT	TC	CT	CC
2	TT	TC	CT	CC
...	...	...	...	...
N-1	TT	TC	CT	CC
N	TT	TC	CT	CC

Period 4

2x2 Treatment Structure in a 4-Period Cross-over Plot Plans. An abbreviated representation of the plot plan for the potential outcomes framework.

Chapter 6 - Identifying Treatment Heterogeneity in GLMM's

6.1 GLMM: Logistic Regression

For all five experimental designs in Chapter 3, a Gaussian distribution of responses was assumed. Thus, the results given in Chapter 3 are confined to the LMM setting. The obvious question remains whether or not the ideas presented in Chapter 3 can be extended to a non-Gaussian distribution. The $Binomial(n, \pi)$ distribution will be considered for the purposes of this research, where n represents the number of independent Bernoulli trials and π is the probability of success in a binomial process. I will pursue modeling potential outcomes using logistic regression, although many of the ideas presented here for the binomial process should be extendable to any of the distributions in the exponential family for which GLMM theory holds.

The first step in extending the results from Chapter 3 to a logistic regression setting is to clearly define what is meant by treatment heterogeneity. While considering binary outcomes in a matched-pairs design, Gadbury et. al (2004) used the same definition of treatment heterogeneity as has already been presented in Chapter 3. That is, they defined the causal effect as the difference in the potential outcome under treatment and the potential outcome under control. Adapting their notation to fit that given in Chapter 1,

$$d = r_T - r_C, \tag{6.1}$$

they showed that in the binary data setting, the causal effect may take on one of three possible values, $-1, 0,$ or 1. Gadbury et. al (2004) defined probabilities for each of these three possible outcomes, noting that $P(d = -1)$ represented the probability of an individual experiencing a detrimental treatment effect. They further defined the average causal effect,

$$E(d) = E(r_T) - E(r_C) = P(r_T = 1) - P(r_C = 1)$$

and demonstrated that $E(d)$ could be estimated from observable data.

In a binomial setting, the definition of d given in (6.1) could take on one of $2n + 1$ possible discrete values. Depending on the number of Bernoulli trials, assigning probabilities to

each of these possible values may quickly become cumbersome. I propose defining the individual causal effect in a slightly different manner. Instead of defining a causal effect as a difference as in (6.1), define it in terms of an odds ratio (or)

$$or = \frac{\pi_T / (1-\pi_T)}{\pi_C / (1-\pi_C)} \Rightarrow logit(or) = logit(\pi_T) - logit(\pi_C) \tag{6.2}$$

where π_T is the potential probability of success for an individual EU receiving treatment and π_C is the potential probability of success if the same EU had received control.

Treatment heterogeneity, then, permits each EU its own probability of success under treatment and its own probability of success under control. Consider a set of $j = 1,2, \ldots N$ EU's, each exhibiting a set of potential responses, $\{r_{Tj}, r_{Cj}\}$, such that

$$r_{ij} \sim Binomial(n, \pi_{ij});\ i = T, C. \tag{6.3}$$

Extending our definition of causal effect in (6.2), we have

$$or_j = \frac{\pi_{Tj} / (1-\pi_{Tj})}{\pi_{Cj} / (1-\pi_{Cj})} \Rightarrow ln(or_j) = logit(\pi_{Tj}) - logit(\pi_{Cj}) \tag{6.4}$$

so that each EU is permitted its own individual causal effect.

Once treatment heterogeneity and an individual causal effect have been clearly defined, it should be relatively straight-forward to see that logistic regression is an intuitive approach to

modeling the $\pi_{ij}'s$ referenced in equation (6.3). Preliminary results for the 2-Sample CRD are presented here. For convenience, results are presented under the assumption of homoscedasticity of variances.

Table 6.1 gives the logistic regression model assumptions for both the potential and observable data models. A direct relationship between the two models is established by defining

$$\tilde{s}_{ij} = s_j + s\tau_{ij} \tag{6.5}$$

Logistic Regression Model	Model Parameters	Assumptions
Potential Model	$r_{ij}\|s_j, s\tau_{ij} \sim Binomial(n, \pi_i)$ $logit(\pi_{ij}) = \mu + \tau_i + s_j + s\tau_{ij},$ $i = T, C;$ $j = 1,2, \dots, N\ EU's$	$s_j \sim iid\ N(0, \sigma_s^2)$ $s\tau_{ij} \sim iid\ N(0, \sigma_{st}^2)$ s_j and $s\tau_{ij}$ are independent.
Observable Model	$r_{ij}\|\tilde{s}_{ij} \sim Binomial(n, \pi_i)$ $logit(\pi_{ij}) = \mu + \tau_i + \tilde{s}_{ij},$ $i = T, C;$ $j = 1,2, \dots, n_i$ $EU's\ per\ level\ of\ trt$	$\tilde{s}_{ij} \sim iid\ N(0, \sigma_e^2)$

Table 6.1 *Logistic Regression Model effects and assumptions in a 2-sample CRD.*

since multiple observations per subject are "lost" when the randomization mechanism is invoked. Thus the residual term, $\tilde{s}_{ij}$, in the observable 2-sample CRD consists of the confounded subject and subject-by-treatment effects from the potential model. If such confounding occurs, then

$$\sigma_e^2 = \sigma_s^2 + \sigma_{st}^2.$$

by the independence assumptions given in Table 6.1. As in Chapter 3, the assumption of unit-treatment additivity in combination with those specified in Table 6.1 mean
$s\tau_{ij} = 0$ for all i and j and

$$\tilde{s}_{ij} = s_j.$$

Thus

$$\sigma_e^2 = \sigma_s^2$$

irrespective of the level of treatment assigned to the j^{th} EU. This means that if unit-treatment additivity holds in a 2-sample CRD, then the only variability in $logit(\pi_{ij})$ is due to characteristics inherent to the EU's so that $ln(or_j) = logit(\pi_{Tj}) - logit(\pi_{Cj})$ is constant for all $j = 1,2, \ldots N$ EU's.

Defining causal effect as in (6.4) and under the potential model assumptions given in Table 6.1,

$$ln(or_j) = (\tau_T - \tau_C) + (s\tau_{Tj} - s\tau_{Cj}) \sim iid\ N(\mu_d, 2 \cdot \sigma_{st}^2) \tag{6.6}$$

where μ_d is defined as in (1.4). Exponentiating (6.6) gives

$$or_j = exp[(\tau_T - \tau_C) + (s\tau_{Tj} - s\tau_{Cj})] \sim lognormal(\mu_d, 2 \cdot \sigma_{st}^2). \tag{6.7}$$

This implies that in the logistic regression setting, treatment heterogeneity can be quantified in terms of the scale parameter associated with or_j instead of $var(d_j)$, as was done in the two-sample CRD in section 3.2.

As with Gaussian responses presented in section 3.2, an average naïve effect must be used to estimate the individual causal effect given in (6.4), as an individual naïve effect is undefined in the two-sample CRD. Define the average naïve effect in the logistic regression setting as

$$ln(\overline{OR}.) = (\tau_T - \tau_C) + \left(\frac{1}{n_T}\sum_{j=1}^{n_T} s_j + s\tau_{Tj}\right) - \left(\frac{1}{n_C}\sum_{j=1}^{n_C} s_j + s\tau_{Cj}\right). \tag{6.8}$$

According to the assumptions given in Table 6.1,

$$\overline{OR}. \sim lognormal\left(\mu_d, 2 \cdot \frac{(\sigma_S^2 + \sigma_{st}^2)}{n}\right) \tag{6.9}$$

if (6.8) is exponentiated, and the 2-sample CRD is balanced.

Define ${\sigma_{or}}^2$ as the scale parameter given in (6.7) and ${\sigma_{\overline{OR}.}}^2$ as the scale parameter given in (6.9) and notice that

$$ {\sigma_{or}}^2 \leq n \cdot {\sigma_{\overline{OR}.}}^2 $$

so $n \cdot {\sigma_{\overline{OR}.}}^2$ is an estimable upper bound for ${\sigma_{or}}^2$.

One difficulty in obtaining estimates of σ_S^2 and σ_{st}^2 in logistic regression is that variance components in PROC GLIMMIX are not estimated using REML techniques. By default, PROC GLIMMIX utilizes pseudo-likelihood (PL) methods to estimate model parameters. However, PL methods can produce estimates that are biased (Pinheiro and Chao, 2006). Initial results in the logistic regression setting verified the presence of bias in model parameter estimates.

Integral approximation techniques exist in PROC GLIMMIX that serve as alternative methods to PL-estimation. LaPlace approximation and adaptive Gauss-Hermite quadrature are both still capable of producing biased results, but the bias is typically smaller using these estimation techniques compared with PL-estimation (Pinheiro and Chao, 2006). Adaptive Gauss-Hermite quadrature was utilized in producing the preliminary results that follow. Although a relationship between the lognormal distribution and the logistic regression model given in Table 6.1 likely exists based on (6.7) and (6.9), it is unclear how to properly relate the adaptive Gauss-Hermite quadrature estimates of σ_S^2 and σ_{st}^2 to the estimated scale parameter of a lognormal distribution, given a set of observable binomial data. Describing this relationship remains a topic of further investigation. It is encouraging, however, that reasonable estimates of model variance components can be obtained for both the potential and observable data models.

Tables 6.2 (i), 6.2 (ii), and 6.2 (iii) give more specific results of some of the effects of interest based on $S = 100$ **s**imulated data sets. Values represent the mean and standard error of estimates across the $S = 100$ data sets. Table 6.2 (i) gives results for the fixed treatment effect for the model fit to both potential and observable data, Table 6.2 (ii) shows the results for some of the random effects in the potential model and Table 6.2 (iii) some of the results for the random effects in the observable model. In all cases, as the sample size increased from 10 to 30 to 100, the variability of the effect estimates around the true simulated value decreased, and in most cases, the estimated value of the simulation parameter based on the $S = 100$ simulations is within 3 standard errors of the true value.

Of particular note, the estimates for σ_e^2 given in Table 6.2 (iii) estimate the theoretical value derived in (6.5), where it was assumed that the subject and subject-by-treatment effects from the potential model were confounded to form the residual term in the observable model. Letting $\sigma_s^2 = 2$ and $\sigma_{st}^2 = 1$, then $\sigma_e^2 = 3$ based on (6.5). The results in Table 6.2 (iii) demonstrate that the estimates of σ_e^2 are reasonably close to 3, for $N = 100$.

Fixed Effect (Potential)	Simulated Value	2N	Average ($S = 100$)	Std. Error ($S = 100$)
$\tau_T - \tau_C$	3	20	3.03	0.06
		60	3.00	0.03
		200	2.99	0.01

Fixed Effect (Obs.)	Simulated Value	N	Average ($S = 100$)	Std. Error ($S = 100$)
$\tau_T - \tau_C$	3	10	3.13	0.12
		30	3.03	0.07
		100	3.04	0.03

(i)

Potential Random Effect	Simulated Value	2N	Average ($S = 100$)	Std. Error ($S = 100$)
σ_s^2	2	20	1.82	0.12
		60	1.91	0.07
		200	1.98	0.04
σ_{st}^2	1	20	0.94	0.06
		60	0.99	0.03
		200	1.01	0.02

(ii)

Observable Random Effect	Simulated Value	N	Average ($S = 100$)	Std. Error ($S = 100$)
σ_e^2	3	10	2.33	0.13
		30	2.82	0.08
		100	2.97	0.05

(iii)

Table 6.2 *2-Sample CRD Logistic Regression Simulation Results.*
Values represent the average and standard error of treatment effect estimates across $S = 100$ simulations in both the potential and observable data models for N=10, 30, and 100 for (i) Fixed Effects. (ii) Potential Random Effects. (iii) Observable Random Effects.

6.2 Treatment Heterogeneity in Generalized Linear Mixed Models

The following section is a paper being prepared for submission to a peer-reviewed journal. The basis of this paper is an extension of the concepts previously discussed in Chapter 6, using a real data example for illustration. For completeness, the paper is presented in tact so the reader may note that some material setting the framework for the problem is repeated here.

6.2.1 Abstract

For continuous data, quantifying treatment heterogeneity is facilitated through potential outcomes by considering the variance of an individual effect, defined as the difference in potential outcomes. As the complexity of an experimental design increases, using the same definition of individual effect for discrete data becomes increasingly intractable. In this paper, the definition of individual effect is altered slightly to accommodate a potential outcomes analysis for a generalized linear mixed model (GLMM). Treatment heterogeneity is conceptualized as a linear combination of potential model variance components, modeled on the link scale. These variance components are non-estimable in observable data, but estimable bounds that arise from linear combinations of the non-estimable potential model variance components exist and depend on the experimental design.

These methods are presented in the context of a 2x2 treatment structure applied to a randomized complete block design with repeated measures where responses are assumed to follow a binomial distribution. Only data from a single period are considered for analysis. The data were collected as part of investigation of the effect of vaccine (VAC) administration and direct-fed microbial (DFM) on the fecal shedding of *E. coli* O157:H7 in a commercial setting.

6.2.2. Introduction

Treatment heterogeneity refers to the variability of a treatment effect across individuals in a population. Studies often focus on estimation of a mean treatment effect (cf. Marshall, 1997), but when there is a high degree of treatment heterogeneity in a population, there may be a non-negligible proportion of the population responding differently to a treatment, and possibly in the opposite direction, from the average subject.

Quantifying the degree of treatment heterogeneity is facilitated by potential outcomes (Rubin, 1974). Consider two treatments, $\{T, C\}$, where T denotes some test treatment and C denotes a reference or perhaps a control treatment. For each subject, imagine a duplet of

responses, $\{r_T, r_C\}$, which represents the potential outcome to the test treatment and to the control treatment, respectively. The individual causal effect can be defined as $d = r_T - r_C$, which cannot be observed since either r_T or r_C , but not both, may be observed at any particular time point. When responses are continuous, treatment heterogeneity has been quantified by $\sigma_d^2 = Var(d)$, a nonestimable quantity since there is no information in observable data on the correlation between r_T and r_C. However, bounds for σ_d^2 can be defined that are estimable in observed data (cf. Gadbury and Iyer 2000, Poulson et al., 2012). Kaiser and Gadbury (2013) recently made use of this result in evaluating the presence of treatment heterogeneity in published weight loss clinical trials. Using a technique called What Would Fisher Do (WWFD, Stroup 2013) applied to a potential outcomes framework, Richardson and Gadbury (2012, 2013-*in progress*) used a linear mixed model (LMM) approach to evaluate treatment heterogeneity in complex designs. They were able to elucidate the necessary assumptions required to equate the variance of naïve estimates of treatment heterogeneity from observable data in complex designs with the variance of the true individual effects.

In this paper, a data example from a 2 x 2 factorial treatment structure applied to a randomized complete block design (RCBD) with four experimental units (EU's) per block is analyzed. These data were collected to investigate the effect of vaccine (VAC) and direct-fed microbial (DFM) on the fecal shedding of *E. coli* O157:H7 in a commercial setting (Cull et al, 2012). The actual data were collected from a RCBD with repeated measures where each of four treatment combinations of VAC and DFM were applied to one of four pens blocked by allocation date since seasonal effects associated with degree of fecal shedding (i.e.—higher shedding in summer) have been well documented. For purposes of simplicity, we consider data from a single period only, however, these methods may be extended to accommodate repeated measures across four periods. Pre-harvest interventions that reduce fecal shedding of *E. coli* O157:H7 have important food safety and commercial economic implications. Blanket administration of treatment based only on average effects when there may exist a non-negligible portion of a population that experiences an unfavorable individual effect is not a trivial matter. Thus, quantifying the degree of treatment heterogeneity associated with these treatments beyond an average affect seems appropriate.

Thirty fresh fecal samples were collected from pens each week over a period of four consecutive weeks. Fecal samples were assessed for the presence (positive) or absence

(negative) of the *E. coli* O157:H7 bacteria. Pen-level proportions were fit using a generalized linear mixed model (GLMM) assuming a binomial distribution on thirty independent trials with a logit link function. Outcomes were the proportion of positive samples from each collection.

When the potential responses are not continuous, a different approach to treatment heterogeneity may be required. Gadbury et al (2004) considered binary response and defined d as a multinomial response taking on one of three distinct values, 0, 1, or -1. They established bounds for the probability of an EU experiencing an unfavorable effect of the test treatment compared with the reference in a matched pairs design. Albert et al (2005) extended those results to a blocked design with binary outcomes. Zhang et al (2013) further extended these results from Gadbury et al (2004) to incorporate information on treatment heterogeneity from known covariates and repeated measures.

The method proposed here compares a GLMM derived under a potential outcomes framework to the usual observable GLMM. A comparison of the potential and observable GLMM's reveals components associated with treatment heterogeneity that are estimable in the potential GLMM but not in the observable GLMM, at least not without non-trivial assumptions. A key step in comparing the potential and observable GLMM's is the appropriate identification of the potential GLMM. This is accomplished by adapting Stroup's WWFD (2013) technique to accommodate a potential outcomes framework.

In the subsequent sections, we i) use Stroup's WWFD technique to develop a potential GLMM linear predictor and corresponding observable GLMM linear predictor; ii) re-define treatment heterogeneity in terms of GLMM components; iii) establish estimable bounds for model parameters quantifying treatment heterogeneity; and iv) apply these results to the *E.coli* data, first at one collection period and then across the four collection periods.

6.2.3. Potential and Observable GLMM Models

In order to accommodate a GLMM analysis, the traditional potential outcomes framework is slightly altered. Imagine a collection of non-continuous potential responses as in (1), except for each potential response, there exists an underlying parameter (or set of parameters) giving rise to a non-continuous potential response. These underlying parameters may be EU-specific and may possibly differ depending on the level of treatment the EU receives.

For the *E. coli* data in particular, imagine that each pen in each block is afforded a 4-tuple of potential binomial responses at each collection period. These potential responses are based on a 4-tuple of underlying binomial probabilities, one for each VAC-DFM combination, on thirty independent Bernoulli trials. When the randomization mechanism is invoked in a potential GLMM, one potential response is selected as the observable response. This is tantamount to selecting one of the underlying potential parameters as the observable parameter under which the observable response is generated. We use the potential GLMM as a template to arrive at the observable GLMM. This process is an important step in the appropriate estimation of effects in the observable model as misspecification of the model in PROC GLIMMIX has been demonstrated to alter both model effect estimation and inference (Boykin et al., 2010).

The WWFD method is based on a discussion by Fisher of Yates' paper on complex experiments (1935) where Fisher demonstrates that the choice of an experimental design is the choice of how a topographical layout of the experiment is related to the treatment structure of an experiment. The potential responses are given for the j^{th} pen in the i^{th} block receiving the k^{th} level of VAC combined with the l^{th} level of DFM; $i = 1,2, \ldots ,10$ allocation dates; $j = 1,2,3,4$ pens; $k = T, C$ levels of VAC; and $l = T, C$ levels of DFM.

The potential outcomes framework results in four simultaneous replicate sets of 40 responses, one replicate set receiving each of the four VAC-DFM combinations. Figure 6.1 (i) gives a plot plan for the potential outcomes layout of this experiment. Notice that every pen and every block is represented in every replicate set. By virtue of the factorial treatment structure, every block and every pen receive every level of VAC, every level of DFM and every level of VAC-DFM treatment combination in the potential outcomes structure. Thus, from a potential outcomes perspective, block and pen are crossed with each main effect and treatment combination.

A topographical layout of the experiment at a single time period is given in Table 6.3 (i). Table 6.3 (ii) gives the factorial treatment layout of the experiment, accounting for the total degrees of freedom in the experiment. "Parallels" was a term used by Fisher and may be thought of the number of times a particular VAC-DFM combination needs to be replicated in order to carry out the entire experiment. In this case, there are 4 VAC-DFM combinations and each combination must be replicated 40 times to accommodate the potential outcomes framework.

Using the combined ANOVA table in Table 6.3 (iii) as a guide, the linear predictor for the potential outcomes experiment is given by

$$logit(\pi_{ijkl}) = \mu + b_i + p_{j(i)} + \alpha_k + \tau_l + b\alpha_{ik} + b\tau_{il} + \alpha\tau_{kl} + b\alpha\tau_{ikl} + p\alpha_{j(i)k} + p\tau_{j(i)l} + p\alpha\tau_{j(i)kl} \quad (6.10)$$

$i = 1,2, \ldots, 10$ allocation dates; $j = 1,2,3,4$ pens; $k = T, C$ levels of VAC; and $l = T, C$ levels of DFM

where π_{ijkl} represents the binomial probability of the j^{th} pen in the i^{th} block receiving the combination of the k^{th} level of VAC and the l^{th} level of DFM; b_i is the random effect of the i^{th} block (i.e.—allocation date); $p_{j(i)}$ is the random effect of the j^{th} pen in the i^{th} block; α_k represents the fixed effect of the k^{th} level of VAC; τ_l represents the fixed effect of the l^{th} level of DFM; $b\alpha_{ik}$ represents the random effect arising from the application of k^{th} level of VAC to the i^{th} block; $b\tau_{il}$ represents the random effect arising from the application of l^{th} level of DFM to the i^{th} block; $\alpha\tau_{kl}$ represents the fixed interaction effect of the k^{th} level of VAC combined with the l^{th} level of DFM**;** $b\alpha\tau_{ikl}$ represents the random interaction effect arising from the application of the k^{th} level of VAC combined with the l^{th} level of DFM to the i^{th} block; $p\alpha_{j(i)k}$ represents the random interaction of the k^{th} level of VAC applied to the j^{th} pen in the i^{th} block**;** $p\tau_{j(i)l}$ represents the random interaction of the l^{th} level of DFM applied to the j^{th} pen in the i^{th} block**;** and $p\alpha\tau_{j(i)kl}$ represents the random interaction effect arising from the application of the k^{th} level of VAC combined with the l^{th} level of DFM to the j^{th} pen in the i^{th} block.

For a distribution in which the estimation of a scale parameter is of interest, the final term in the model, $p\alpha\tau_{j(i)kl}$, would be considered the residual or error term, and would be utilized in the estimation of the error variance. However, for a distribution belonging to the one-parameter exponential family, like the binomial distribution, there is no scale parameter to estimate. Consequently, this final source of variability must play either a different role than that of the Gaussian residual term, or no role at all (Stroup, 2013). It is common practice to assume no variability can be attributed to the final term and remove it from the model. For now, it will be left in the model in order for the potential model to account for all degrees of freedom in the experiment. Further discussion for the interested reader can be found in Stroup (2013, pp. 112-114). The distributional assumptions in probability distribution form (Stroup, 2013) are as follows:

Data:

$$r_{ijkl}|\boldsymbol{q} \sim Binomial\left(30, \pi_{ijkl}\right)$$

where $\boldsymbol{q}$ is a vector of the following random effects:

$$\begin{gathered} b_i \sim iid\ N(0, \sigma_b^2) \\ p_{j(i)} \sim iid\ N\left(0, \sigma_p^2\right) \\ b\alpha_{ik} \sim iid\ N(0, \sigma_{b\alpha}^2) \\ b\tau_{il} \sim iid\ N(0, \sigma_{b\tau}^2) \\ b\alpha\tau_{ikl} \sim iid\ N(0, \sigma_{b\alpha\tau}^2) \\ p\alpha_{j(i)k} \sim iid\ N(0, \sigma_{p\alpha}^2) \\ p\tau_{j(i)l} \sim iid\ N(0, \sigma_{p\tau}^2) \\ p\alpha\tau_{j(i)kl} \sim iid\ N(0, \sigma_{p\alpha\tau}^2) \end{gathered} \tag{6.11}$$

$b_i, p_{j(i)}, b\alpha_{ik}, b\tau_{il},\ b\alpha\tau_{ikl}, p\alpha_{j(i)k}, p\tau_{j(i)l}\ and\ p\alpha\tau_{j(i)kl}$ are mutually independent.

Previously published results (Richardson and Gadbury, 2012; Richardson and Gadbury 2013-*in progress*) have shown that an observable model can be derived from a potential model by considering the information lost after invoking the randomization mechanism resulting in the removal of a portion of potential data. Figure 6.1 (ii) represents a plot plan after the randomization mechanism has been invoked and three-fourths of the potential data have been removed. By removing three-fourths of the data, the following information is lost:

(i) Multiple observations per block on the same DFM-VAC combination

(ii) Multiple observations per pen within a block

(iii) Multiple observations per pen within a block on the same level of VAC

(iv) Multiple observations per pen within a block on the same level of DFM

The resulting observable linear predictor is given by:

$$logit(\pi_{ikl}) = \mu + b_i + \alpha_k + \tau_l + b\alpha_{ik} + b\tau_{il} + \alpha\tau_{kl} + \widetilde{b\alpha\tau}_{ikl} \tag{6.12}$$

$i = 1,2, \ldots, 10$ allocation dates; $k = T, C$ levels of VAC; and $l = T, C$ levels of DFM

where the pen within block can be identified by the VAC-DFM combination if the randomization scheme is known. All other effects are defined as in the potential outcomes framework.

A direct relationship between the observable model and the potential model is established by defining

$$\widetilde{b\alpha\tau}_{ikl} = p_{j(i)} + b\alpha\tau_{ikl} + p\alpha_{j(i)k} + p\tau_{j(i)l} + p\alpha\tau_{j(i)kl}.$$

Given the distributional assumptions specified in (6.11),

$$\tilde{\sigma}^2_{bat} = \sigma^2_p + \sigma^2_{b\alpha\tau} + \sigma^2_{p\alpha} + \sigma^2_{p\tau} + \sigma^2_{p\alpha\tau}. \quad (6.13)$$

Table 6.4 gives the WWFD result for this experiment design. Since there is no scale parameter to estimate, a usual observable GLMM approach attributes any remaining variability in the linear predictor after fitting the VAC and DFM main effects, VAC-by-DFM interaction, and block-by-VAC and block-by-DFM interactions to block-by-VAC-by-DFM interaction. By first considering the potential GLMM linear predictor, the assumptions required to substantiate this assertion become clear. In particular, by considering the variance components in (6.13), $\tilde{\sigma}^2_{bat} = \sigma^2_{b\alpha\tau}$ only when $\sigma^2_p = \sigma^2_{p\alpha} = \sigma^2_{p\tau} = \sigma^2_{p\alpha\tau} = 0$. So, if one is willing to assume no variability due to pen, no variability in VAC effect across pens, no variability in DFM effect across pens, and no variability in VAC-DFM interaction across pens, then the block-by-treatment interaction effect completely explains any remaining variability after the main effects (fixed), the interaction effect (fixed) and the random block and block-by-main effect interactions have been included in the model.

6.2.4. GLMM Individual Effects

Previous work with binary potential outcomes (Gadbury et al., 2004; Zhang et al. 2013) has utilized the traditional definition of an individual effect. Extending the traditional definition of individual effect to the binomial distribution results in $(2n + 1)$ possible values of $d = r_T - r_C$, where n is the number of Bernoulli trials (i.e. $n = 30$ for the *E. coli* dataset). For large values of n, using the approach described by Gadbury et al. (2004) may be rather cumbersome, and an alternative definition of individual effect may facilitate a more intuitive investigation of treatment heterogeneity.

Rather than defining an individual effect on the data scale (i.e.-the difference between two potential responses belonging to the same EU), define an individual effect on the model or link scale. For a binomial response assuming a logistic regression GLMM model, the resulting individual effect is an individual log-odds ratio of T compared with C. In the *E. coli* data

example, the 2x2 factorial treatment structure facilitates the following two individual effects for the j^{th} pen in the i^{th} block:

$$ln(or_{ij|l}) = logit(\pi_{ijTl}) - logit(\pi_{ijCl})$$

and (6.14)

$$ln(or_{ij|k}) = logit(\pi_{ijkT}) - logit(\pi_{ijkC})$$

where $ln(or_{ij|l})$ represents the individual effect of VAC conditioned on a given level of DFM and $ln(or_{ij|k})$ represents the individual effect of DFM conditioned on a given level of VAC. Using this modified definition of individual effect, the variance of the individual log-odds ratio, ${\sigma_{or}}^2$, quantifies the degree of treatment heterogeneity in an experiment in that a positive value of ${\sigma_{or}}^2$ indicates the presence of treatment heterogeneity. If no treatment heterogeneity exists, then the variability of the individual log-odds is zero.

Based on the model assumptions given in (6.11), ${\sigma^2}_{or|l=T} = {\sigma^2}_{or|l=C}$ and ${\sigma^2}_{or|k=T} = {\sigma^2}_{or|k=C}$. Denote these variances ${\sigma^2}_{or|DFM}$ and ${\sigma^2}_{or|VAC}$, respectively. Writing the individual effects given in (6.14) in terms of the potential GLMM linear predictor components and considering the variance of the individual log-odds ration based on the model assumptions given in (6.11) yields

$${\sigma^2}_{or|DFM} = 2(\sigma_{b\alpha}^2 + \sigma_{b\alpha\tau}^2 + \sigma_{p\alpha}^2 + \sigma_{p\alpha\tau}^2)$$

and (6.15)

$${\sigma^2}_{or|VAC} = 2(\sigma_{b\tau}^2 + \sigma_{b\alpha\tau}^2 + \sigma_{p\tau}^2 + \sigma_{p\alpha\tau}^2)$$

where $\sigma_{b\alpha}^2$ and $\sigma_{b\tau}^2$ are estimable from observable data since an observable data set contains multiple observations per EU on a particular level of VAC (but differing levels of DFM) and multiple observations per EU on a particular level of DFM (but differing levels of VAC). Thus estimable lower bounds can be established by noting

$$2\sigma_{b\alpha}^2 \leq {\sigma^2}_{or|DFM}$$

and (6.16)

$$2\sigma_{b\tau}^2 \leq {\sigma^2}_{or|VAC}$$

The factorial nature of this design permits two observable naïve estimates of the individual effects given in (6.14). The two naïve estimates are, again, defined on the model scale:

$$ln\left(OR_{i|l}\right) = logit(\pi_{iTl}) - logit(\pi_{iCl})$$
and
$$ln\left(OR_{i|k}\right) = logit(\pi_{ikT}) - logit(\pi_{ikC}) \tag{6.17}$$

where the differences in log-odds are across pens within the same block.

To compare the variances of the naïve effects in in (6.17) to the variances of the individual effects defined in (6.14), notice that based on the distributional assumptions given in (6.11) and the relationship between the potential linear predictor and the observable linear predictor in (6.13), $\sigma^2{}_{OR|l=T} = \sigma^2{}_{OR|l=C}$ and $\sigma^2{}_{OR|k=T} = \sigma^2{}_{OR|k=C}$, where $\sigma^2{}_{OR}$ represents the variance of a naïve effect. Denote these variances $\sigma^2{}_{OR|DFM}$ and $\sigma^2{}_{OR|VAC}$, respectively. The variances of the naïve effects written in terms of model variance components are:

$$\begin{aligned}\sigma^2{}_{OR|DFM} &= 2(\sigma^2_{b\alpha} + \tilde{\sigma}^2_{b\alpha\tau}) = 2\left(\sigma^2_{b\alpha} + \sigma^2_{p} + \sigma^2_{b\alpha\tau} + \sigma^2_{p\alpha} + \sigma^2_{p\tau} + \sigma^2_{p\alpha\tau}\right) \\ &= \sigma^2{}_{or|DFM} + 2\left(\sigma^2_{p} + \sigma^2_{p\tau}\right)\end{aligned}$$
and
$$\begin{aligned}\sigma^2{}_{OR|VAC} &= 2(\sigma^2_{b\tau} + \tilde{\sigma}^2_{b\alpha\tau}) = 2\left(\sigma^2_{b\tau} + \sigma^2_{p} + \sigma^2_{b\alpha\tau} + \sigma^2_{p\alpha} + \sigma^2_{p\tau} + \sigma^2_{p\alpha\tau}\right) \\ &= \sigma^2{}_{or|VAC} + 2\left(\sigma^2_{p} + \sigma^2_{p\alpha}\right)\end{aligned} \tag{6.18}$$

Combining the results of (6.16) and (6.18), notice

$$2\sigma^2_{b\alpha} \le \sigma^2{}_{or|DFM} \le \sigma^2{}_{OR|DFM}$$
and
$$2\sigma^2_{b\tau} \le \sigma^2{}_{or|VAC} \le \sigma^2{}_{OR|VAC}. \tag{6.19}$$

where $\sigma^2{}_{OR|DFM} = 2(\sigma^2_{b\alpha} + \tilde{\sigma}^2_{b\alpha\tau})$ and $\sigma^2{}_{OR|VAC} = 2(\sigma^2_{b\tau} + \tilde{\sigma}^2_{b\alpha\tau})$. Equation (6.19) demonstrates that non-trivial, estimable upper and lower bounds for the variances of an individual log-odds ratio can be established for this experimental design. The difference between the upper and lower bounds is $2\tilde{\sigma}^2_{b\alpha\tau}$.

6.2.5. Results of Data Analysis

Using PROC GLIMMIX, we analyzed the *E.coli* data from the first of four periods, according to the model given in (6.12) where α_k represents a fixed VAC effect and τ_l represents a fixed DFM effect. Table 6.5(i) gives the results for tests of fixed effects in the model. Table 6.5 (ii) presents the odd-ratio estimates of interest for fixed effects with standard errors and estimates of the random effects with standard errors.

Results from the analysis demonstrate that the probability of detecting a positive fecal sample in pens that were vaccinated were significantly lower (P=0.0038) than pens that were not vaccinated. There was no significant effect on the odds-ratio for the effect of DFM, neither was there a significant interaction effect. These results are consistent with those published by Cull et al (2012) even though we are only considering one period instead of four in this analysis. As such, estimates and P-values given here will differ from those reported by Cull et al (2012).

The estimates of the variance components in Table 6.5 (ii) give rise to estimable bounds of the variance of individual effects established in (6.19). Table 6.6 gives the estimated upper and lower bounds of the variance of the individual causal effects. In addition to the estimable bounds we used $B = 50$ bootstrap (Efron and Tibshirani, 1994) samples to compute the bootstrap standard error of both the upper and lower bounds. These bootstrap standard errors are also given in Table 6.6. The difference between the estimable upper and lower bounds for both variances of interest is 1.6036. This is twice the estimate of $\tilde{\sigma}^2_{b\alpha\tau}$ given in Table 6.5 (ii). For both variances of interest, the estimates of the lower bounds are within one bootstrap standard error of zero. Additionally, the estimate of the upper bound for the individual effect of VAC given DFM is within two bootstrap errors of zero. These estimates, together with a non-significant VAC-by-DFM interaction, suggest that it would be reasonable to conclude no treatment heterogeneity for VAC. For DFM at a given level of VAC, a conclusion of treatment heterogeneity is possible since the estimate of the upper bound is more than two bootstrap standard errors above zero, however, based on equation (6.18), one must be willing to assume $\sigma^2_p = \sigma^2_{p\alpha} = 0$ in order for the variability of the individual log-odds to equal the variability of the observed log-odds. In other words, if one is willing to assume no variability in individual log-odds due to the pens a block and no variability in individual log-odds due to different pens receiving the same level of VAC, then one could reasonably treatment heterogeneity of DFM. Even though the potential GLMM helps clarify what assumptions are needed to equate estimated bounds with estimated treatment

heterogeneity, these assumptions cannot be directly tested using observable data from this design.

6.2.6. Discussion and Conclusion

In cases where treatment heterogeneity is suspected, quantifying the degree of treatment heterogeneity in addition to estimating a mean effect should be undertaken before a claim of the superiority of one treatment over another is established (Longford, 1999). Treatment heterogeneity has frequently been assessed using finite population, randomization-based approaches. These techniques have been utilized for both continuous (Gadbury et al, 2001, Poulson et al, 2012) and non-continuous (Gadbury et al, 2004; Albert et al, 2004; Zhang et al, 2013) responses. However, as the complexity of an experimental design increases, assessing treatment heterogeneity becomes increasingly intractable (Ndum, 2012).

Since linear mixed models (LMM's) and GLMM's are particularly useful for modeling data from complex designs, their role in modeling treatment heterogeneity is investigated. In order to accommodate a potential outcomes analysis for a GLMM setting, we slightly altered the definition of an individual effect so that the individual effect is defined on the link or model scale. Once this has been done, the variance of an individual effect is the parameter of interest when assessing treatment heterogeneity, with a non-zero value indicating the presence of treatment heterogeneity. If the estimate of the lower bound is substantially greater than zero, one might conclude that treatment heterogeneity is present. Likewise, if an estimable upper bound is very close to zero then one might conclude that the treatment is having a similar effect on individuals across a population.

We demonstrated that both an upper and lower bound of the variance of an individual effect can be achieved for 2x2 factorial treatment structure applied to a RCBD. Given the estimated bounds of the individual effects and the bootstrap standard errors, there is not enough evidence from the current data to conclude treatment heterogeneity in the effect of VAC on fecal shedding. It should also be stated that this is not the same as concluding treatment homogeneity. But given that the main effect of VAC was significant (P=0.0038), it seems reasonable to conclude that the effect of VAC is favorable and that the effect does not vary significantly across units in a population.

The conclusion regarding the heterogeneity of a treatment effect for DFM is not as clear. While the estimated lower bound is reasonably close to zero, one can argue that treatment heterogeneity could exist since the estimable upper bound is more than two boot-strapped standard errors above zero. In this case, it seems prudent to consider what assumptions are required in order to equate ${\sigma^2}_{or|VAC}$ with its estimable upper bound. Based on the relationship of ${\sigma^2}_{or|VAC}$ and ${\sigma^2}_{OR|VAC}$ given in (6.18), ${\sigma^2}_{or|VAC} = {\sigma^2}_{OR|VAC}$ when $\sigma_p^2 = \sigma_{p\alpha}^2 = 0$. This means that in order for ${\sigma^2}_{or|VAC}$ to achieve its estimable upper bound, we need to be willing to assume that there is no variability due to pen-within-block and no variability due to the application of VAC to a particular pen. It should be noted that there is no way to test the validity of the assumption that $\sigma_p^2 = \sigma_{p\alpha}^2 = 0$ from the current data.

Given that $n = 30$ in this experiment and with so many possible values of the usual computation of $d = r_T - r_C$ (i.e-61 possible values), a normal approximation seems like a reasonable approach. In other words, one might consider the following distribution on the potential responses:

$$r_{ijkl}|\boldsymbol{q} \sim N\left(\mu_{ijkl}, \sigma^2\right)$$

where $\boldsymbol{q}$ is a vector of random effects, the estimate of μ_{ijkl} would typically serve as the estimate of $n\pi_{ijkl}$ and the distributional assumptions of the random effects remain unchanged from those given in (6.11). However, if estimates of μ_{ijkl} can be interpreted as the corresponding estimate of $n\pi_{ijkl}$, then the variance of these estimates should also be related the estimates of π_{ijkl}. Using the normal approximation, the variance of the estimates of μ_{ijkl} would be related to σ^2, the meaning of which is ambiguous. Furthermore, using the normal approximation convolutes the interpretation of treatment heterogeneity. Recall, $\sigma_d^2 = Var(d)$ quantifies the degree of treatment heterogeneity using the usual definition of individual effect. Using a normal approximation introduces $2\sigma^2$ into the computation of σ_d^2 and its upper and lower bounds, since d is defined as the difference between to potential responses for the same EU. Permitting $2\sigma^2$ into the computation of σ_d^2 introduces an ambiguous source of variability that is related neither to the variability of the true conditional distribution of the potential responses nor the random effects specified in the linear predictor. This is not trivial, especially if the marginal distribution

of the data is not approximately normal. Let $f(\boldsymbol{q})$ denote the joint distribution of random effects in the linear predictor. The marginal distribution of the data can be determined as follows:

$$f(\boldsymbol{r}) = \int \cdots \int f(\boldsymbol{r}|\boldsymbol{q}) f(\boldsymbol{q}) d\boldsymbol{q}$$

where $\boldsymbol{r}$ is a vector containing potential responses. When $f(r|\boldsymbol{q})$ is a binomial distribution, the integral of the resulting joint distribution, $f(\boldsymbol{r}|\boldsymbol{q})f(\boldsymbol{q})$**,** cannot be directly evaluated to obtain a marginal distribution. Simulation studies have shown this marginal distribution can be heavily skewed either to the right depending on the value of the binomial probability and the amount of variability introduced into the process by the random effects specified in $\boldsymbol{q}$. (Stroup, 2013) Imposing a normal distribution on the conditional distribution of the data given the random effects, and including the resulting "approximate" variance in the computation of σ_d^2 may lead to misleading conclusions about the existence of treatment heterogeneity.

As in the case of the heterogeneity of the DFM effect, a comparison of the observable GLMM and potential GLMM for a given experimental design delineates the information about causal effects that is lost in moving from potential to observable data, and what assumptions about non-estimable quantities (or design modifications) are needed to evaluate treatment heterogeneity in observable data. Furthermore, for experimental designs in which an estimable lower and/or upper bound can be established, the investigation of treatment heterogeneity is essentially "without cost" in the sense that no new data are needed to confirm the presence of treatment heterogeneity.

Studies are rarely designed to separate information on an individual effect from other sources of variability (Senn, 2001). For many simple designs, only estimable upper bounds of the variance of an individual effect can be established. If treatment heterogeneity is suspected, careful thought and planning should be undertaken to design an experiment in such a way that an estimable lower bound can be established since an estimable lower bound significantly greater than zero confirms the presence of treatment heterogeneity.

While the statistical methods presented here may be used to quantify the degree of treatment heterogeneity in these data, they cannot explain the source of the treatment heterogeneity. If we concluded that that treatment heterogeneity existed, further research would

be required to investigate the possible causes of treatment heterogeneity. The data example was used for illustration and not to confirm a superiority of one treatment over another. The sample size was small and other issues such as treatment compliance were not considered.

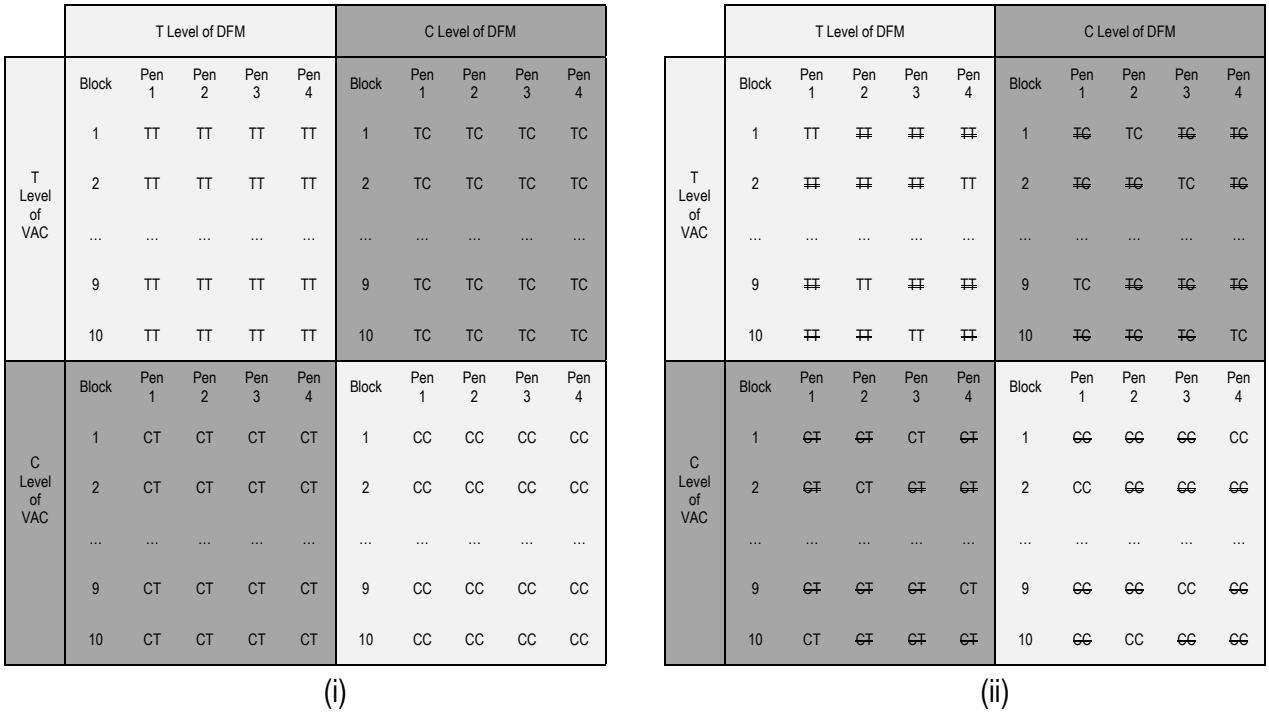

Figure 6.1 *Plot Plans: 2x2 factorial treatment structure in a RCBD.*
Plot plans for (i) the potential outcomes framework and (ii) the observable model framework.

Topographical		Trt		Combined	
Source	d.f.	Source	d.f.	Source	d.f.
Rows of Replicate Sets	2-1=1	VAC	2-1=1	VAC	2-1=1
Columns of Replicate Sets	2-1=1	DFM	2-1=1	DFM	2-1=1
Row*Column	1x1=1	VAC*DFM	1x1=1	VAC*DFM	1x1=1
Block	10-1=9	Parallels	4(40-1)=156	Block	10-1=9
Block*Row	9x1=9			Block*VAC	9x1=9
Block*Column	9x1=9			Block*DFM	9x1=9
Block*Row*Column	9x1x1=9			Block*VAC*DFM	9x1x1=9
Pen(Block)	10(4-1)=30			Pen(Block)	10(4-1)=30
Row*Pen(Block)	1x30=30			VAC* Pen(Block)	1x30=30
Column*Pen(Block)	1x30=60			DFM* Pen(Block)	1x30=30
Row*Column*Pen(Block)	1x1x30=30			VAC*DFM* Pen(Block)	1x1x30=30
Total	160-1=159	Total	160-1=159	Total	160-1=159
(i)		(ii)		(iii)	

Table 6.3 *Potential WWFD ANOVA Structure: 2x2 Factorial in RCBD*
(i) Topographical, (ii) Treatment, and (iii) Combined ANOVA structures for a Potential 2x2 factorial treatment structure in a RCBD.

Topographical		Trt		Combined	
Source	d.f.	Source	d.f.	Source	d.f.
Row	2-1=1	VAC	2-1=1	VAC	2-1=1
Column	2-1=1	DFM	2-1=1	DFM	2-1=1
Row*Column	1x1=1	VAC*DFM	1x1=1	VAC*DFM	1x1=1
Block	10-1=9	Parallels	4(40-1)=156 4(10-1)=36	Block	10-1=9
Block*Row	9x1=9			Block*VAC	9x1=9
Block*Column	9x1=9			Block*DFM	9x1=9
Block*Row*Column++	9x1x1=9			Block*VAC*DFM++	9x1x1=9
Pen(Block)	10(4-1)=30			Pen(Block)	10(4-1)=30
Row*Pen(Block)	1x30=30			VAC* Pen(Block)	1x30=30
Column*Pen(Block)	1x30=60			DFM* Pen(Block)	1x30=30
Row*Column*Pen(Block)	1x1x30=30			VAC*DFM* Pen(Block)	1x1x30=30
Total	160-1=159 40-1=39	Total	160-1=159 40-1=39	Total	160-1=159 40-1=39

(i) (ii) (iii)

Table 6.4 *Observable WWFD ANOVA Structure: 2x2 Factorial in RCBD*
(i) Topographical, (ii) Treatment, and (iii) Combined ANOVA structures for an Observable 2x2 factorial treatment structure in a RCBD.
++Assumes no pen-within-block variability and a uniform treatment effect of VAC, DFM and VAC-DFM combination on every pen within a block.

Type III Tests of Fixed Effects		
Fixed Effect	**F-Value**	**P-value**
VAC	14.94	0.0038
DFM	0.04	0.8385
VAC ∗ DFM	3.02	0.1163

(i)

Estimates:			
Fixed Effect	**Odds Ratio**	**Lower**	**Upper**
VAC (T vs. C)	0.26	0.12	0.57
DFM (T vs. C)	0.92	0.38	2.22

Variance Component	**Estimate**	**Std. Error**
σ_b^2	0.6942	0.5842
$\sigma_{b\alpha}^2$	0.0468	0.3886
$\sigma_{b\tau}^2$	0.1803	0.4552
$\tilde{\sigma}_{b\alpha\tau}^2$	0.8018	0.6068

(ii)

Table 6.5 *SAS PROC GLIMMIX Results*
(i)Type III Tests for Fixed Effects. (ii) Estimates of the odds ratios of observing a sample positive for E. Coli with upper and lower confidence limits and estimates of the variance of random effects with estimated standard errors.

Individual Effect	Estimate	Bootstrap SE
$\sigma^2_{or\|DFM}$		
Lower Bound	0.0936	0.4511
Upper Bound	1.6972	0.8523
$\sigma^2_{or\|VAC}$		
Lower Bound	0.3606	0.5791
Upper Bound	1.9642	0.8602

Table 6.6 *Estimable Bounds for the Variance of Individual Effects.*
Estimates of the upper and lower bounds given in equation (6.19) with bootstrap standard errors.

Chapter 7 - Future Work

The following sections present ideas for future work based on the research presented in Chapters 1-6.

7.1 The Role of the Randomization Mechanism

To this point, all results have been predicated on a uniform randomization mechanism. That is, assuming a balanced CRD experiment comparing two treatments, the marginal probability of assignment is

$$P(W_j = 1) = P(W_j = 0) = \frac{1}{2} \tag{7.1}$$

for any of the $j = 1,2,..N$ EU's, where W_j is the indicator variable defined in Chapter 1 to represent the random assignment of EU's to level of treatment. In a randomized experiment, (7.1) holds regardless of the values of the EU's potential outcomes and regardless of the values of either observed or unobserved covariates. Furthermore, in a randomized experiment, the treatment and control groups are usually comparable in every respect except for the level of treatment applied to the group. The reason for this is that the law of large numbers ensures that for a randomized experiment that is "large enough", values of both observed and unobserved covariates tend toward the mean value of the population from which the treatment groups were drawn.

It has been well established (Fisher, 1935; Rosenbaum and Ruben, 1983; Rosenbaum, 2010) that studies in which uniform randomization is either impractical or infeasible do not possess these same characteristics that tend to balance the treatment group and control group in randomized experiments. It is very common among studies in which randomization is not uniform to find significant differences between the treatment group compared with the control group in attributes that affect the outcome of the study. Therefore, there is no reason to suspect that the probability of being assigned to either treatment or control is independent of covariate values, or even of potential outcomes. In other words, certain values of a covariate may make an

EU more likely to be assigned to either treatment or control. If groups receiving treatment and control differ in ways besides the level of treatment assigned and these differences matter for the outcomes of the study, then the study is said to be biased. When all sources of bias in a study are overt, (i.e.- the pertinent covariates have been collected and recorded), then the bias can be controlled by making adjustments such as matching or stratification, under the assumption of a strongly ignorable treatment assignment. If the bias is hidden, however, then no adjustment can be made. A sensitivity analysis which seeks to describe the magnitude of the hidden bias that must be present in the study in order to explain any associations seen in that study should be included in the results of any study for which randomization is not uniform. There is a wide body of literature that discusses matching techniques and the intricacies of sensitivity analyses in studies containing bias. Unless these topics become a part of the current research, that literature will not be considered at this point. The interested reader is referred to Rosenbaum's (2010) text on designing observational studies for a noteworthy summary of the topics discussed here.

Up to this point, the discussion regarding randomization and bias has still been predicated on the assumption of an additive treatment model defined in equation (2.3). Few forays have been attempted that consider a model that contains both treatment heterogeneity and non-uniform randomization. One such attempt, however, was provided by Rosenbaum (1999) in which a dilated treatment effect model was defined and a sensitivity analysis was performed under the assumptions of this dilated treatment effect model. A dilated treatment effect model is a model that permits a type of treatment heterogeneity in which it is assumed that the potential responses under treatment, r_T, are systematically larger and more dispersed than potential responses under control, r_C. The difference between r_T and r_C is assumed to be a non-negative, non-decreasing function of r_C. This assumption has serious implications on the correlation between r_T and r_C, namely that the correlation is non-negative. It seems reasonable that the current research could be extended to investigate the nature of treatment heterogeneity in studies for which randomization is not uniform without the imposed structure of a dilated treatment effect.

7.2 Estimating Treatment Heterogeneity in Observable Data

Consider, again, the potential model for a 2-sample CRD given in Table 3.1. For simplicity purposes, consider a common variance on $s\tau_{ij}{}'s$ so that $\sigma^2_{sT} = \sigma^2_{sC} = \sigma^2_{st}$. Define

$$t_{ij} = \tau_i + s\tau_{ij};$$

$$i = T, C; j = 1,2 \dots N \tag{7.2}$$

so that the potential model in Table 3.1 becomes

$$\begin{gathered} r_{ij} = \mu + t_{ij} + s_j; \\ i = T, C; j = 1,2 \dots N \\ t_{ij} \sim iid\ N(\tau_i, \sigma_{st}^2)\ ;\ s_j \sim iid\ N(0, \sigma_s^2);\ t_{ij}\ and\ s_j\ are\ independent \end{gathered} \tag{7.3}$$

This is recognizable as a random model containing two random effects, where the expectation of t_{ij} is possibly non-zero. Assumptions about σ_{st}^2 and σ_s^2 required to equate the variance of the individual causal effect and the naïve effect have already been discussed. However, it seems reasonable given the model in (7.3) that there may be other constraints placed on the model that might permit both σ_s^2 and σ_{st}^2 to become estimable from observable data. For example, if the constraint under H_0: $\tau_T = \tau_C$, were to be imposed, could σ_s^2 and σ_{st}^2 then be estimated? Or what if τ_i and σ_{st}^2 were considered hyper-parameters from some specified prior distribution on t_{ij}? What kind of estimate of σ_{st}^2 would the variance of the posterior distribution then be if a Bayesian approach were adapted? Answers to questions like these seem tangible now that a potential data model has been defined and its relationship to the "usual" observable data model has been clearly established.

7.3 The Role of a Covariate

Gadbury and Iyer (2000) demonstrated the use of a single covariate obtained on a population of units in bounding measures of treatment heterogeneity in a two-sample CRD with maximum likelihood estimates (MLE's) obtained from observable data. They further discussed assumptions of the conditional model required for a lack of treatment heterogeneity to exist. Gadbury et. al (2001) performed sensitivity analyses over the range of possible values of conditional and unconditional correlation. Denoting the single covariate Z, they considered the population of potential responses to be drawn from the following trivariate Gaussian population

$$\begin{pmatrix} r_T \\ r_C \\ Z \end{pmatrix} \sim MVN\left\{ \begin{pmatrix} \mu_T \\ \mu_C \\ \mu_Z \end{pmatrix}, \begin{bmatrix} \sigma_T^2 & \rho_{TC}\sigma_T\sigma_C & \rho_{TZ}\sigma_T\sigma_Z \\ \rho_{TC}\sigma_T\sigma_C & \sigma_C^2 & \rho_{CZ}\sigma_C\sigma_Z \\ \rho_{TZ}\sigma_T\sigma_Z & \rho_{CZ}\sigma_C\sigma_Z & \sigma_Z^2 \end{bmatrix} \right\}. \quad (7.4)$$

By extending the models in Chapter 3 to develop analysis of covariance (ANCOVA) models and using the results of the proposed research in section 4.2 on correlation, I would like to reframe the work of Gadbury and Iyer (2000) and Gadbury et. al (2001) in light of the potential outcomes linear mixed models developed in Chapter 3. More specifically, I would like to consider how information from a single covariate might alter the estimable bounds of $\sigma_d{}^2$defined in terms of model variance components, if at all. Furthermore, I would like to investigate the assumptions in a potential ANCOVA model that are required in order for a lack of treatment heterogeneity to exist.

References

Airy, G.B. *On the Algebraical and Numerical Theory of Errors of Observations and the Combinations of Observations*. MacMillan: London, 1861.

Albert, J.M., Gadbury, G.L., and Mascha, E.J. (2005). Assessing Treatment Effect Heterogeneity in Clinical Trials with Blocked Binary Outcomes. *Biometrical Journal.* **47(5)**: 662-673.

Boykin, D., Camp, M.J., Johnson, L, Kramer, M., Meek, D., Palmquist, D., Vinyard, B., and West, M. (2010). Generalized Linear Mixed Model Estimation Using PROC GLIMMIX: Results from Simulations when the Data and Model Match, and when the Model is Misspecified. *In* Proceedings of the 22nd Annual Conference on Applied Statistics in Agriculture (*ed*. Weixing Song), Kansas State University, April 2010: 137-170.

Breslow, N.E. and Clayton, D.G. (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*. **88**: 9-25.

Byar, D. P. and Corle, D. K. (1977). Selecting optimal treatment in clinical trials using covariate information. *Journal of Chronic Diseases*. **30(7)**: 445 – 459.

Campbell, T.C., and Campbell, T.M.(2005). *The China Study: Startling Implications for Diet, Weight Loss and Long-term Health*. BenBella Books.

Chen, S.C., Judd, J.T., Kramer, M., Meijer, G.W., Clevidence, B.A., and Baer, D.J. (2009). Phytosterol Intake and Dietary Fat Reduction are Independent and Additive in their Ability to Reduce Plasma LDL Cholesterol. *Lipids*, **44**: 273-281.

Cochran, W.G. (1947). Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics*. **3(1)**: 22-38.

Cox, D. R., *The Planning of Experiments*. Wiley: New York, 1958.

Cox, D. R., (1992). Causality: Some statistical aspects. *Journal of the Royal Statistical Society. Series A.* **155(2)**: 291 – 301.

Cull, C.A., Paddock, Z.D., Nagaraja, T.G., Bello, N.M., Babcock, A.H., and Renter, D.G. (2012). Efficacy of a vaccine and a direct-fed microbial against fecal shedding of *Escherichia Coli* 0157:H7 in a randomized pen-level field trial of commercial feedlot cattle. *Vaccine*, **30**: 6210-6215.

Dawid, A. P. (2000). Causal inference without counterfactuals. *Journal of the American Statistical Association*. **95**: 407 – 424.

Efron, B., and Tibshirani, R.L. (1994), *An Introduction to the Bootstrap*, Chapman and Hall.

Eisenhart, C. (1947). The assumptions underlying the analysis of variance. *Biometrics*. **3**: 1-21.

Endrenyi, L., and Tothfalusi, L. (1999). Subject-by-Formulation Interaction in Determinations of Individual Bioequivalence: Bias and Prevalence. *Pharmaceutical Research*, **16**: 186–190.

Fisher, R.A. (1918). The correlation between relatives on the supposition of Medelian inheritance. *Transactions of the Royal Society of Edinburgh*. **52**: 399-433.

Fisher, R.A. *Statistical Methods for Research Workers, 1st ed*. Oliver & Boyd: Edinburgh, 1925.

Fisher, R.A. *Design of Experiments*. Oliver & Boyd: Edinburgh, 1935.

Gadbury, G.L. (1998). Causal inference in randomized experiments and observational studies. PhD Thesis, Colorado State University, Fort Collins, CO.

Gadbury, G.L. and Iyer, H.K. (2000). Unit-Treatment Interaction and Its Practical Consequences. *Biometrics.* **56**: 882-885.

Gadbury G. (2001). Randomization Inference and bias of standard errors. *American Statistician*. **55**: 310-313.

Gadbury, G.L., Iyer, H.K., and Allison, D. (2001). Evaluating subject-treatment interaction when comparing two treatments. *Journal of Biopharmaceutical Statistics.* **11**: 313-333.

Gadbury, G.L., Iyer, H.K., Albert, J.M. (2004). Individual treatment effects in randomized trials with binary outcomes. *Journal of Statistical Planning and Inference*. **121**: 163-174.

Gadbury, G.L. (2010). Subject-treatment interaction. In *Encyclopedia of Biopharmaceutical Statistics, 3rd. Ed., Revised and Expanded.* Edited by Shein-Chung Chow. London: Informa Healthcare: 1316-1321.

Gail, M. and Simon, R. (1985). Testing for qualitative interactions between treatment effects and patient subsets. *Biometrics*. **41**: 361 – 372.

Gbur, E.E., Stroup, W.W., McCarter, K.S., Durham, S., Young, L.J., Christman, M., West, M., and Kramer, M. *Analysis of Generalized Linear Mixed Models in the Agricultural and Natural Resources Sciences*. American Society of Agronomy, Soil Science Society of America, and Crop Science Society of America: Madison, WI, 2012.

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

Graybill, F.A. *Theory and Application of the Linear Model.* Duxbury: North Scituate, MA, 1976.

Hauck, W. W., Hyslop, T., Mei-Ling, C., Patnaik, R., and Williams, R. L. (2000). Subject-by-Formulation Interaction in Bioequivalence: Conceptual and Statistical Issues. *Pharmaceutical Research*, **17**: 375–380.

Hinkelmann K. and Kempthorne O. *Design and Analysis of Experiments, Volume 1 : Introduction to Experimental Design*. Hoboken, NJ: John Wiley and Sons, Inc, 2008.

Hocking, R.R. *The Analysis of Linear Models*. Brooks/Cole: Monterey, CA, 1985.

Holland, P.W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*. **81**: 945-960

Hwang, S., Huber, P. B., Hesney, M., Kwan, K. C. (1978). Bioequivalence and Interchangeability. *Journal of Pharmaceutical Sciences*. **67**: iv.

Kaiser, K.A. and Gadbury, G.L. (2013). Estimating the Range of Obesity Treatment Response Variability in Humans: Methods and Illustrations. *Human Heredity*. In press.

Kent, D. M. and Hayward, R. A., (2007). Limitations of applying summary results of clinical trials to individual patients: The need for risk stratification. *Journal of the American Medical Association*. **298**: 1209 – 1212.

Kramer, M., Chen, S.C., Gebauer, S.K., Baer, D.J. (2011). Estimating the subject by treatment interaction in non-replicated crossover diet studies. *In* Proceedings of the 23rd Annual Conference on Applied Statistics in Agriculture (*ed*. Weixin Yao), Kansas State University, April 2011: 96-110.

Lewis, K. D. and Burton-Freeman, B. M., (2010). The role of innovation and technology in meeting individual nutritional needs. *The Journal of Nutrition*. **140**: 426S – 436S.

Li, J. and Chan, I. S. F. (2006). Detecting qualitative interactions in clinical trials: An extension of range test. *Journal of Biopharmaceutical Statistics*. **16**: 831 – 841.

Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., and Schabenberger, O. *SAS for Mixed Models, 2nd ed.* SAS Institute, Inc.: Cary, NC, 2006.

Longford, N. T. (1999). Selection bias and treatment heterogeneity in clinical trials. *Statistics in Medicine.* **18:** 1467 – 1474.

Marshall, A. (1997). Laying the foundations for personalized medicines. *Nature Biotechnology*. **15**: 954 – 957.

McLean, R.A., Sanders, W.L., and Stroup, W.W. (1991). A Unified Approach to Mixed Linear Models. *American Statistician*. **45(1)**: 54-64.

Milliken, G. A. and Johnson, D. E. (1984). *Analysis of Messy Data*. Lifetime Learning Publications. Belmont, CA.

Ndum, E.A., Albert, J.A., and Gadbury, G.L. (2012). Individual Treatment Heterogeneity in a Three Period Two Treatment Cross-over Design. *JP Journal of Biostatistics*. **8**: 1-35.

Nelder, J.A. and Wedderburn, R.W.M. (1972) Generalized Linear Models. *Journal of the Royal Statistical Society. Series A.* **135**: 370 – 384.

Neyman, J., Dabrowska, D. M., Speed, T. P. (1990). On the application of probability theory to agricultural experiments. Essay on Principles. Section 9. *Statistical Science*. **5(4)**: 465 – 472.

Neyman, J. (1935). Statistical Problems in Agricultural Experimentation (with discussion). *Supplement to the Jouarn1al of the Royal Statistical Society, Series B.* **2**: 107-180.

Peto, R. (1982). Statistical aspects of cancer trials. In *Treatment of Cancer*, K.E. Hanlan (ed.), Chapman and Hall: London, 867 – 871.

Pinheiro, J.C. and Chao, E.C. (2006). Efficient Laplacian and Adaptive Gaussian Quadrature Algorithms for Multilevel Generalized Linear Mixed Models. *Journal of Computational and Graphical Statistics*. **15**: 58-81.

Poulson, R. S., Gadbury, G.L., and Allison, D.B. (2012). Treatment Heterogeneity and Individual Crossover Interaction. *American Statistician*. **66(1)**: 16-24.

Rao, C.R. *Linear Statistical Inference and its Applications, 2nd ed*. John Wiley and Sons, Inc : New York, 1973.

Richardson, T.E. and Gadbury G.L. (2012). Treatment heterogeneity and potential outcomes in linear mixed effects models. *In* Proceedings of the 24th Annual Conference on Applied Statistics in Agriculture (*ed*. Weixing Song), Kansas State University, April 2012, 215-232.

Rosenbaum, P.R. and Rubin D.B. (1983). The central role of propensity score in observational studies for causal effects. *Biometrika*. **70(1)**: 41-55.

Rosenbaum, P.R. (1999). Reduced Sensitivity to Hidden Bias at Upper Quartiles in Observational Studies with Dilated Treatment Effects. *Biometrics*. **55**: 560-564.

Rosenbaum, P.R. *Design of Observational Studies*. Springer: New York, 2010.

Rubin, D.B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*. **66**:688-701.

Rubin, D. B. (1978). Bayesian inference for causal effects: the role of randomization. *The Annals of Statistics*. **6**: 34 – 58.

Rubin, D.B. (1980). Comment on Basu's 'Randomization analysis of experimental data' (1980). *Journal of the American Statistical Association*. **75**: 591-593.

Rubin, D.B. (1986). Which ifs have causal answers, Comment on Holland's 'Statistics and Causal Inference' (1986). *Journal of the American Statistical Association*. **81**: 961-962.

Rubin, D.B. (1990). Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies. *Statistical Science*. **5(4):** 472 – 480.

Scheffe, H. (1956). Alternative models for the analysis of variance. *The Annals of Mathematical Statistics*. **27**: 251-271.

Searle, S.R. *Linear Models*. John Wiley and Sons, Inc : New York, 1971.

Searle, S.R., Casella, G., and McCulloch, C.E. *Variance Components*. John Wiley and Sons, Inc : Hoboken, NJ, 1992.

Senn, S, (2001). Individual therapy: New dawn or false dawn? *Drug Information Journal*. **35**, 1479 – 1494.

Silvapulle, M. J. (2001). Tests for qualitative interaction: exact critical values and robust tests. *Biometrics*. **57**: 1157 – 1165.

Simon, R., (1982). Patient Subsets and Variation in Therapeutic Efficacy. *British Journal of Pharmacology*, **14**: 473–482.

Stroup, W.W. (2011). GLMM and "the Basics" – Paradigm Shift or Just My Imagination? *24th Annual Conference on Applied Statistics in Agriculture.* Manhattan, KS.

Stroup, W.W. *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. Chapman & Hall/CRC: Boca Raton, FL, 2013.

Tippett, L.H.C. *The Methods of Statistics, 1st ed*. Williams and Norgate: London. 1931.

Wedderburn, R.W.M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika.* **61**: 439-447.

Wilk, M.B. and Kempthorne, O. (1955). Fixed, Mixed, and Random Models. *JASA.* **50**:1144-1167.

Wolfinger, R. and O'Connell, M. (1993). Generalized Linear Mixed Models: A Pseudo-Likelihood Approach. *Journal of Statistical Computation and Simulation*. **48**: 233-243.

Yates, F. (1935). Complex Experiments. *Supplement to the Jouarn1al of the Royal Statistical Society.* **2(2)**: 181-247.

Zhang, Z., Wang, C., Nie, L., and Soon, G. (2013). Assessing the heterogeneity of treatment effects via potential outcomes of individual patients. *Journal of the Royal Statistical Society: Series C (Applied Statistics),* **62**: doi: 10.1111/rssc.12012.

Appendix A- Results for Gaussian Data: Common Variance

A.1 2-Sample CRD

Model	Model Parameters	Assumptions
Potential Model	$r_{ij} = \mu + \tau_i + s_j + s\tau_{ij},$ $i = T, C;$ $j = 1,2, \dots, N\ EU's$	$s_j \sim iid\ N(0, \sigma_s^2)$ $s\tau_{ij} \sim iid\ N(0, \sigma_{st}^2)$ s_j and $s\tau_{ij}$ are independent.
Observable Model	$R_{ij} = \mu + \tau_i + e_{ij},$ $i = T, C;$ $j = 1,2, \dots, n_i$ $EU's\ per\ level\ of\ trt$	$e_{ij} \sim iid\ N(0, \sigma_e^2)$

Table A.1.1 *Model effects and assumptions in a 2-sample CRD.*

Fixed Effect (Potential)	Simulated Value	2N	Average ($S = 100$)	Std. Error ($S = 100$)
$\tau_T - \tau_C$	3	20	3.03	0.06
		60	3.01	0.03
		200	3.00	0.02

Fixed Effect (Obs.)	Simulated Value	N	Average ($S = 100$)	Std. Error ($S = 100$)
$\tau_T - \tau_C$	3	10	3.17	0.16
		30	3.03	0.09
		100	3.02	0.06

(i)

Potential Variance	Simulated Value	2N	Average ($S = 100$)	Std. Error ($S = 100$)
Subject	4.71	20	4.76	0.25
		60	4.83	0.13
		200	4.75	0.08
*Subject*Trt*	1.57	20	1.52	0.07
		60	1.53	0.04
		200	1.56	0.02
$var(d_j)$	3.14	20	3.04	0.14
		60	3.06	0.07
		200	3.11	0.05

(ii)

Observable Variance	Simulated Value	N	Average ($S = 100$)	Std. Error ($S = 100$)
Residual	6.28	10	6.33	0.32
		30	6.31	0.15
		100	6.23	0.09

(iii)

Table A.1.2 *2-Sample CRD Simulation Results.*
Values represent the average and standard error of treatment effect estimates across $S = 100$ simulations in both the potential and observable data models for N=10, 30, and 100 for (i) Fixed Effects. (ii) Potential Random Effects. (iii) Observable Random Effects.

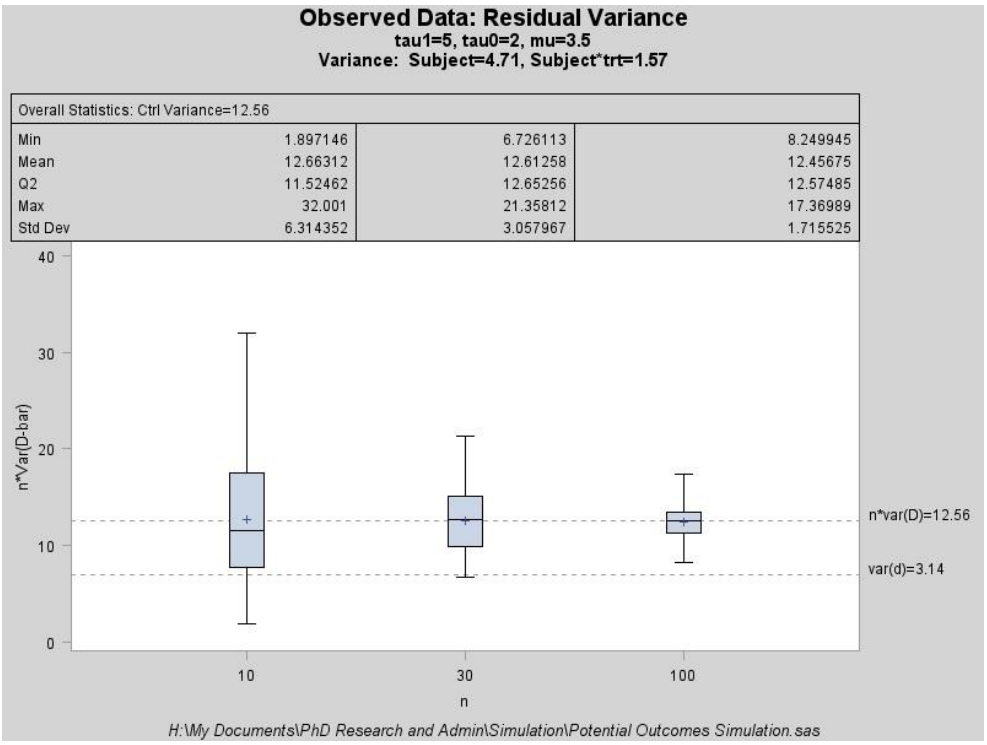

Figure A.1.1 $n \cdot var(\bar{D}.) \geq var(d_{ij})$. *Box plots of the* $S = 100$ *estimates of* $n \cdot var(\bar{D}.)$ *at N=10, 30, and 100 Dotted lines represent values used in the simulation design.*

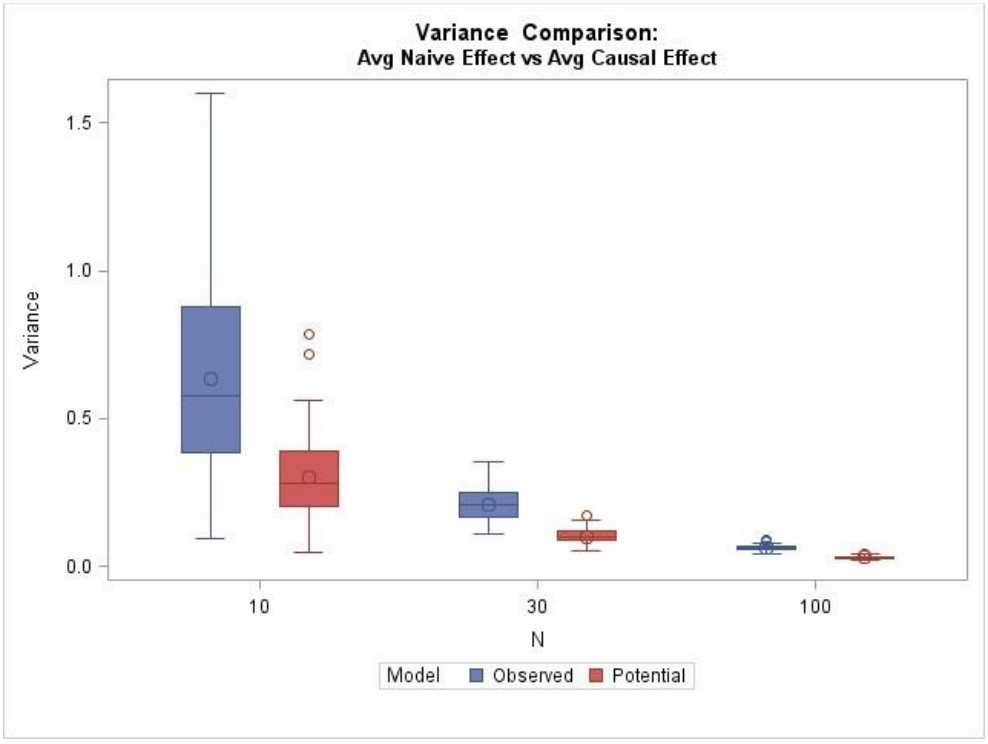

Figure A.1.2 $\frac{var(\bar{D}.)}{2}$ *vs.* $var(\bar{d}.)$. *One-half the variance of the average naïve effect is an upper bound for the variance of the average true causal effect.*

A.2 Matched-Pairs

Model	Model Parameters	Assumptions
Potential Model	$r_{ijk} = \mu + b_i + s_{j(i)} + \tau_k + b\tau_{ik} + s\tau_{j(i)k}$ $i = 1,2, \dots, B\ pairs;$ $j = 1,2\ EU's\ within\ a\ pair$ $k = T, C$	$b_i \sim iid\ N(0, \sigma_b^2)$ $s_{j(i)} \sim iid\ N(0, \sigma_s^2)$ $b\tau_{ik} \sim iid\ N(0, \sigma_{bt}^2)$ $s\tau_{j(i)k} \sim iid\ N(0, \sigma_{st}^2)$ $b_i, s_{j(i)}, b\tau_{ik}$ and $s\tau_{j(i)k}$ are mutually independent.
Observable Model	$R_{ijk} = \mu + b_i + \tau_k + e_{ijk},$ $i = 1,2, \dots, B\ pairs;$ $j = 1\ EU\ within\ a\ pair\ receiving\ Trt\ k$ $k = T, C$	$b_i \sim iid\ N(0, \sigma_b^2)$ $e_{ijk} \sim iid\ N(0, \sigma_e^2)$ $b_i$and e_{ijk} are independent

Table A.2.1 *Model effects and assumptions in a RCBD.*

Fixed Effect (Potential)	Simulated Value	2N	Average ($S = 100$)	Std. Error ($S = 100$)
$\tau_T - \tau_C$	7	40	6.90	0.10
		120	6.97	0.05
		400	6.98	0.03

Fixed Effect (Obs.)	Simulated Value	N	Average ($S = 100$)	Std. Error ($S = 100$)
$\tau_T - \tau_C$	7	20	6.78	0.14
		60	6.94	0.08
		200	6.99	0.05

(i)

Potential Variance	Simulated Value	2N	Average ($S = 100$)	Std. Error ($S = 100$)
Block	10	40	9.92	0.69
		120	9.62	0.40
		400	9.83	0.23
*Block*Trt*	3	40	3.19	0.22
		120	3.17	0.12
		400	3.06	0.05
Subject	4	40	4.25	0.26
		120	4.24	0.13
		400	3.95	0.07
*Subject*Trt*	2	40	1.99	0.08
		120	2.02	0.05
		400	2.03	0.03
$var(d_{ij})$	10	40	10.36	0.43
		120	10.37	0.25
		400	10.18	0.10

(ii)

Observable Variance	Simulated Value	N	Average ($S = 100$)	Std. Error ($S = 100$)
Block	10	20	9.89	0.69
		60	9.65	0.43
		200	9.81	0.24
Residual	9	20	12.39	0.83
		60	13.34	0.51
		200	13.20	0.28
$var(D_i)$	18	20	19.74	1.08
		60	18.79	0.50
		200	18.13	0.24

(iii)

Table A.2.2 *Matched-Pairs/RCBD Simulation Results. Values represent the average and standard error of treatment effect estimates across* $S = 100$ *simulations in both the potential and observable data models for B=10, 30, and 100 of size 2 for (i) Fixed Effects. (ii) Potential Random Effects. (iii) Observable Random Effects.*

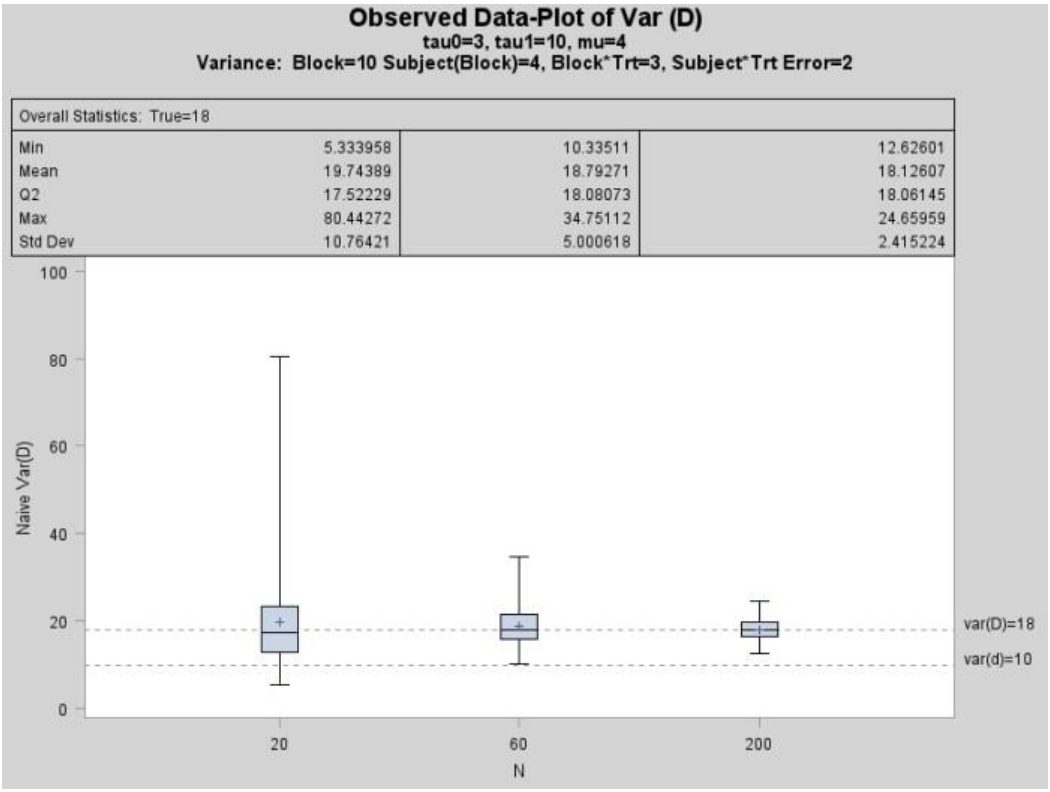

Figure A.2.1 $var(D_i) \geq var(d_{ij})$. *Box plots of the* $S = 100$ *estimates of* $var(\bar{D}_{i\cdot})$ *at B=10, 30, and 100 blocks of size 2. Dotted lines represent values used in the simulation design.*

A.3 GRCBD

Model	Model Parameters	Assumptions
Potential Model	$r_{ijk} = \mu + b_i + s_{j(i)} + \tau_k + b\tau_{ik} + s\tau_{j(i)\,k}$ $i = 1,2, \dots, B$ *blocks*; $j = 1,2,..4$ $EU's$ *within a block*; $k = T, C$	$b_i \sim iid\ N(0, \sigma_b^2)$ $s_{j(i)} \sim iid\ N(0, \sigma_s^2)$ $b\tau_{ik} \sim iid\ N(0, \sigma_{bt}^2)$ $s\tau_{j(i)\,k} \sim iid\ N(0, \sigma_{st}^2)$ $b_i, s_{j(i)}$, $b\tau_{ik}$ and $s\tau_{j(i)\,k}$ are mutually independent.
Observable Model	$R_{ijk} = \mu + b_i + \tau_k + b\tau_{ik} + e_{ijk}$ $i = 1,2, \dots, B$ *blocks*; $j = 1,2$ $EU's$ *within a block on trt* k; $k = T, C$	$b_i \sim iid\ N(0, \sigma_b^2)$ $b\tau_{ik} \sim iid\ N(0, \sigma_{bt}^2)$ $e_{ijk} \sim iid\ N(0, \sigma_e^2)$ b_i, $b\tau_{ik}$, and e_{ijk} are mutually independent

Table A.3.1 *Model effects and assumptions in a GRCBD.*

Fixed Effect (Potential)	Simulated Value	2N	Average ($S = 100$)	Std. Error ($S = 100$)
$\tau_T - \tau_C$	7	80	7.15	0.08
		240	7.06	0.05
		800	7.02	0.03

Fixed Effect (Obs.)	Simulated Value	N	Average ($S = 100$)	Std. Error ($S = 100$)
$\tau_T - \tau_C$	7	40	7.01	0.10
		120	6.98	0.06
		400	7.00	0.03

(i)

Potential Variance	Simulated Value	2N	Average ($S = 100$)	Std. Error ($S = 100$)
Block	10	80	10.08	0.63
		240	10.23	0.36
		800	10.05	0.19
*Block*Trt*	3	80	3.19	0.19
		240	2.98	0.09
		800	3.01	0.05
Subject	4	80	3.93	0.14
		240	3.90	0.08
		800	3.97	0.04
*Subject*Trt*	2	80	2.08	0.06
		240	2.03	0.03
		800	2.01	0.02
$var(d_{ij})$	10	80	10.55	0.38
		240	10.03	0.19
		800	10.03	0.10
$var(\bar{d}_{i\cdot})$	7	80	7.42	0.37
		240	6.98	0.18
		800	7.02	0.10

(ii)

Observable Variance	Simulated Value	N	Average ($S = 100$)	Std. Error ($S = 100$)
Block	10	40	10.04	0.65
		120	10.25	0.38
		400	10.06	0.18
*Block*Trt*	3	40	3.55	0.26
		120	3.09	0.16
		400	2.98	0.08
Residual	6	40	5.73	0.19
		120	5.80	0.12
		400	5.95	0.06
$var(\bar{D}_{i\cdot})$	12	40	12.93	0.57
		120	11.97	0.30
		400	11.91	0.15
$Upper\ Bound\ var(d_{ij})$	18	40	18.56	0.57
		120	17.78	0.30
		400	17.86	0.15
$Upper\ Bound\ var(\bar{d}_{i\cdot})$	9	40	9.97	0.57
		120	9.08	0.30
		400	8.93	0.15
Lower Bound	6	40	7.11	0.57
		120	6.17	0.30
		400	5.96	0.15

(iii)

Table A.3.2 *GRCBD Simulation Results. Values represent the average and standard error of treatment effect estimates across $S = 100$ simulations in both the potential and observable data models for B=10, 30, and 100 of size 4 for (i) Fixed Effects. (ii) Potential Random Effects. (iii) Observable Random Effects.*

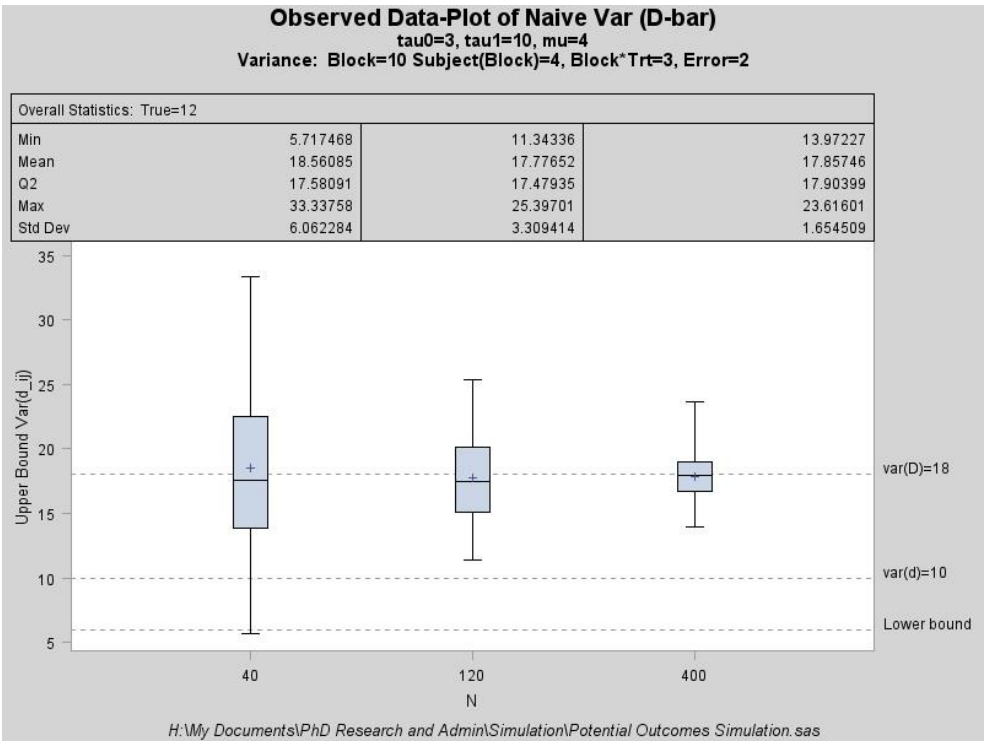

Figure A.3.1 $2\sigma_{bt}^2 + 2\sigma_e^2 \geq var(d_{ij}) \geq 2\sigma_{bt}^2$. *Boxplots of the* $S = 100$ *estimates of* $2\sigma_{bt}^2 + 2\sigma_e^2$ *at B=10, 30, and 100 blocks of size 4. Dotted lines represent values used in the simulation design.*

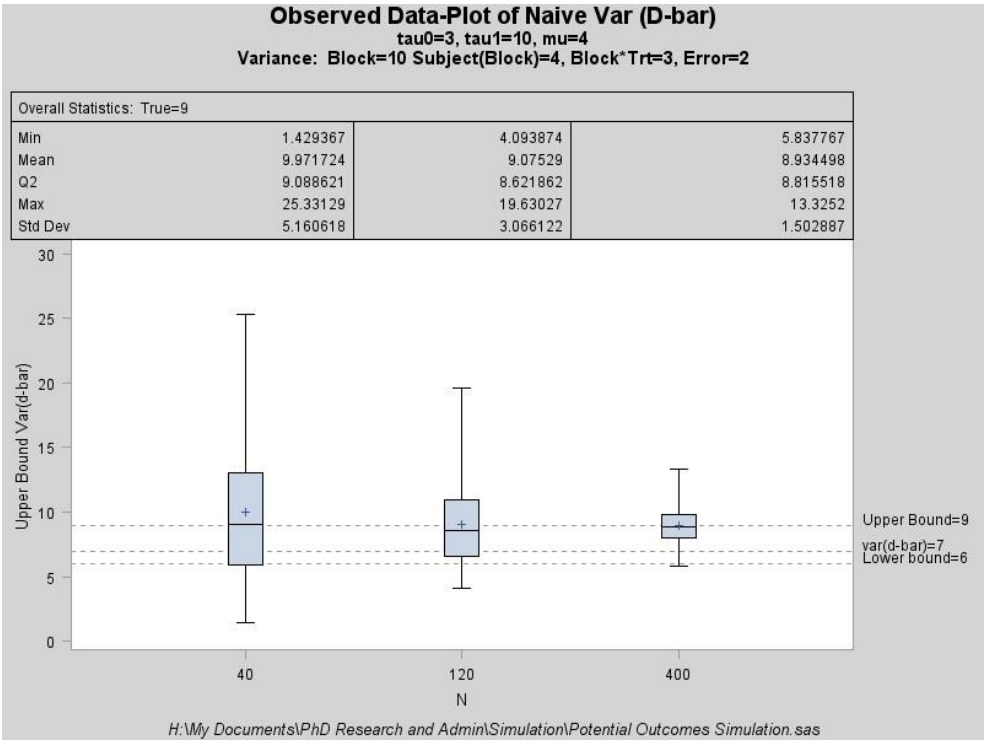

Figure A.3.2 $2\sigma_{bt}^2 + \frac{\sigma_e^2}{2} \geq var(\bar{d}_{i\cdot}) \geq 2\sigma_{bt}^2$. *Boxplots of the* $S = 100$ *estimates of* $2\sigma_{bt}^2 + \frac{\sigma_e^2}{2}$ *at B=10, 30, and 100 blocks of size 4. Dotted lines represent values used in the simulation design.*

A.4 Two-Period-Two-Treatment Crossover

Model	Model Parameters	Assumptions
Potential Model	$r_{ijk} = \mu + s_i + \pi_j + s\pi_{ij} + \tau_k + s\tau_{ik} + \pi\tau_{jk} + s\pi\tau_{ijk}$ $i = 1,2 \ldots N\ EU's;$ $j = 1,2\ periods; \quad k = T, C$	$s_i \sim iid\ N(0, \sigma_s^2)$ $s\pi_{ij} \sim iid\ N(0, \sigma_{sp}^2)$ $s\tau_{ik} \sim iid\ N(0, \sigma_{st}^2)$ $s\pi\tau_{ijk} \sim iid\ N(0, \sigma_{spt}^2)$ $s_i, s\pi_{ij}, s\tau_{ik}$ and $s\pi\tau_{ijk}$ are mutually independent.
Observable Model	$R_{ijk} = \mu + s_i + \pi_j + \tau_k + \pi\tau_{jk} + e_{ijk}$ $i = 1,2 \ldots N\ EU's;$ $j = 1, 2\ periods$ $k = T, C$	$s_i \sim iid\ N(0, \sigma_s^2)$ $e_{ijk} \sim iid\ N(0, \sigma_e^2)$ s_i and e_{ijk} are independent

Table A.4.1 *Model effects and assumptions in a Two-Period-Two-Treatment Crossover.*

Fixed Effect (Potential)	Simulated Value	4N	Average ($S = 100$)	Std. Error ($S = 100$)
$\tau_T - \tau_C$	7	40	6.84	0.14
		120	6.95	0.07
		400	6.98	0.05

Fixed Effect (Obs.)	Simulated Value	2N	Average ($S = 100$)	Std. Error ($S = 100$)
$\tau_T - \tau_C$	7	20	6.73	0.17
		60	6.93	0.09
		200	6.99	0.06

(i)

Potential Variance	Simulated Value	4N	Average ($S = 100$)	Std. Error ($S = 100$)
Subject	10	40	10.06	0.75
		120	9.53	0.44
		400	9.78	0.26
*Subject*Period*	3	40	3.19	0.22
		120	3.19	0.10
		400	2.96	0.06
*Subject*Trt*	7	40	7.20	0.42
		120	7.32	0.23
		400	7.17	0.10
*Subject*Period*Trt*	2	40	1.97	0.08
		120	2.01	0.05
		400	2.03	0.03
$var(d_{ij})$	18	40	18.34	0.83
		120	18.66	0.47
		400	18.40	0.19

(ii)

Observable Variance	Simulated Value	2N	Average ($S = 100$)	Std. Error ($S = 100$)
Subject	10	20	10.29	0.81
		60	9.53	0.47
		200	9.70	0.27
Residual	12	20	12.59	0.69
		60	12.52	0.34
		200	12.26	0.16
$var(D_i)$	24	20	25.18	1.39
		60	25.04	0.69
		200	24.51	0.32

(iii)

Table A.4.2 *Two-Period-Two-Treatment Crossover Simulation Results. Values represent the average and standard error of effect estimates across $S = 100$ simulations in both the potential and observable data models for N=10, 30, and 100 for (i) Fixed Treatment Effects. (ii) Potential Random Effects. (iii) Observable Random Effects.*

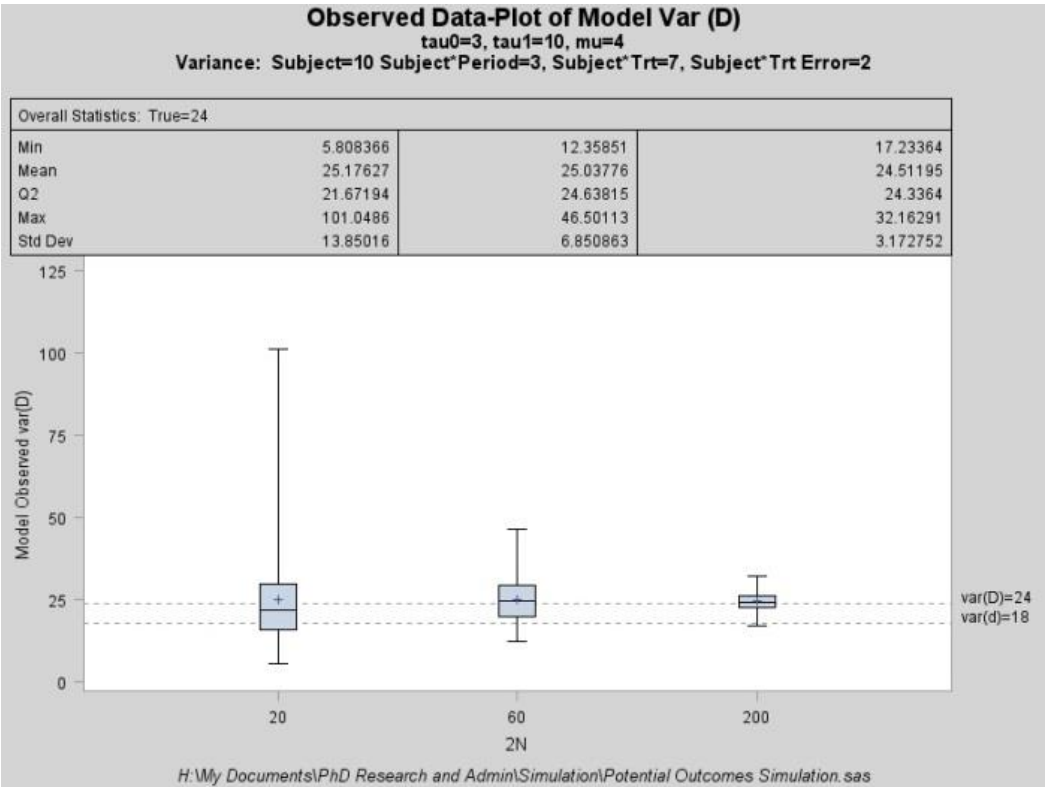

Figure A.4.1 $var(D_i) \geq var(d_{ij})$. *Boxplots of the* $S = 100$ *estimates of* $var(D_i)$ *at N=10, 30, and 100 EU's. Dotted lines represent values used in the simulation design.*

A.5 Repeated Measures Two-Treatment Crossover

Model	Model Parameters	Assumptions
Potential Model	$r_{ijk} = \mu + s_i + \pi_j + s\pi_{ij} + \tau_k + s\tau_{ik} + \pi\tau_{jk} + s\pi\tau_{ijk}$ $i = 1,2 \ldots N\ EU's;$ $j = 1,4\ periods;\quad k = T, C$	$s_i \sim iid\ N(0, \sigma_s^2)$ $s\pi_{ij} \sim iid\ N(0, \sigma_{sp}^2)$ $s\tau_{ik} \sim iid\ N(0, \sigma_{st}^2)$ $s\pi\tau_{ijk} \sim iid\ N(0, \sigma_{spt}^2)$ $s_i, s\pi_{ij}, s\tau_{ik}$ and $s\pi\tau_{ijk}$ are mutually independent.
Observable Model	$R_{ijk} = \mu + s_i + \pi_j + \tau_k + \pi\tau_{jk} + + s\tau_{ik} + e_{ijk}$ $i = 1,2 \ldots N\ EU's;$ $j = 1,4\ periods$ $k = T, C$	$s_i \sim iid\ N(0, \sigma_s^2)$ $s\tau_{ik} \sim iid\ N(0, \sigma_{st}^2)$ $e_{ijk} \sim iid\ N(0, \sigma_e^2)$ s_i, $s\tau_{ik}$, and e_{ijk} are mutually independent

Table A.5.1 *Model effects and assumptions in a Repeated Measures Two-Treatment Crossover.*

Fixed Effect (Potential)	Simulated Value	8N	Average ($S = 100$)	Std. Error ($S = 100$)
$\tau_T - \tau_C$	7	96	6.87	0.12
		288	7.00	0.06
		960	7.03	0.04

Fixed Effect (Obs.)	Simulated Value	4N	Average ($S = 100$)	Std. Error ($S = 100$)
$\tau_T - \tau_C$	7	40	6.82	0.14
		120	7.00	0.07
		400	7.05	0.04

(i)

Potential Variance	Simulated Value	8N	Average ($S = 100$)	Std. Error ($S = 100$)
Subject	10	96	9.28	0.64
		288	9.77	0.37
		960	10.17	0.19
*Subject*Period*	3	96	3.05	0.11
		288	3.01	0.06
		960	3.04	0.03
*Subject*Trt*	7	96	7.00	0.33
		288	7.21	0.18
		960	7.02	0.09
*Subject*Period*Trt*	2	96	2.02	0.05
		288	2.02	0.03
		960	1.99	0.01
$var(d_{ij})$	18	96	18.03	0.65
		288	18.45	0.37
		960	18.02	0.19
$var(\bar{d}_{i\cdot})$	15	96	15.00	0.65
		288	15.43	0.37
		960	15.04	0.19

(ii)

Observable Variance	Simulated Value	4N	Average ($S = 100$)	Std. Error ($S = 100$)
Subject	10	48	9.40	0.69
		144	9.94	0.36
		480	10.15	0.20
*Subject*Trt*	7	48	7.02	0.44
		144	7.13	0.22
		480	7.04	0.13
Residual	5	48	5.18	0.18
		144	5.01	0.09
		480	5.00	0.05
$var(\bar{D}_{i\cdot})$	19	48	19.22	0.85
		144	19.26	0.43
		480	19.08	0.25
$Upper\ Bound\ var(d_{ij})$	24	48	24.41	0.85
		144	24.26	0.44
		480	24.08	0.25
$Upper\ Bound\ var(\bar{d}_{i\cdot})$	16.5	48	16.63	0.86
		144	16.75	0.43
		480	16.58	0.25
$Lower\ Bound$	14	48	14.04	0.88
		144	14.25	0.44
		480	14.08	0.25

(iii)

Table A.5.2 Repeated Measures Two-Treatment Crossover Simulation Results.
Values represent the average and standard error of effect estimates across $S = 100$ simulations in both the potential and observable data models for N=10, 30, and 100 for (i) Fixed Treatment Effects. (ii) Potential Random Effects. (iii) Observable Random Effects.

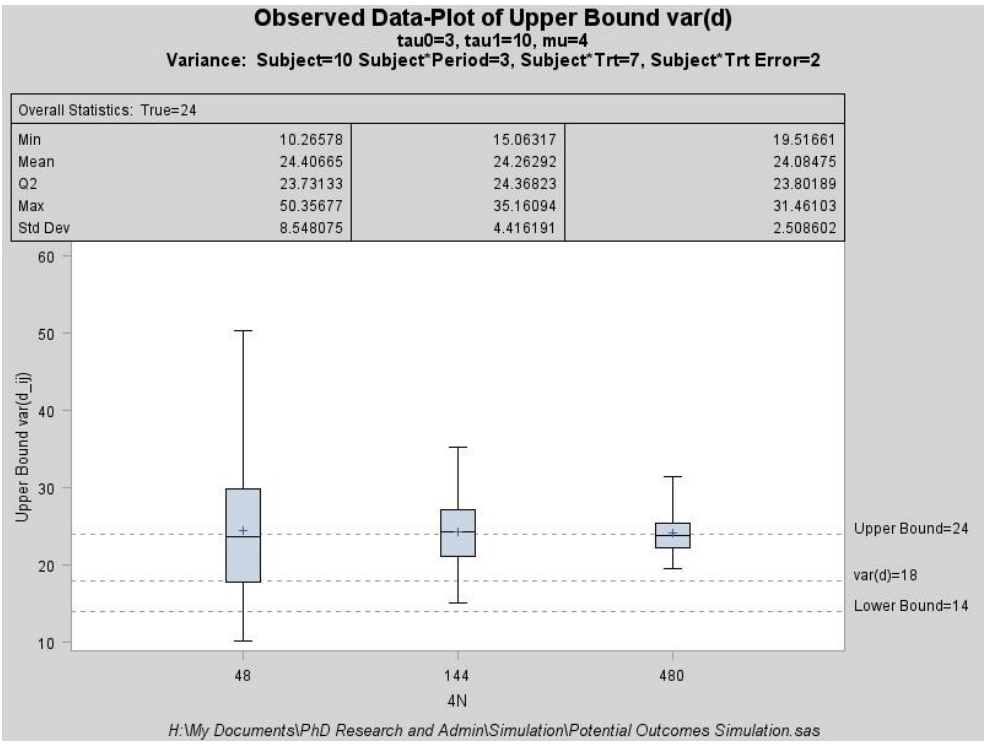

Figure A.5.1 $2\sigma_{st}^2 + 2\sigma_e^2 \geq var(d_{ij}) \geq 2\sigma_{st}^2$. *Boxplots of the* $S = 100$ *estimates of* $2\sigma_{st}^2 + 2\sigma_e^2$ *at N=10, 30, and 100 EU's measured at 4 time periods. Dotted lines represent values used in the simulation design.*

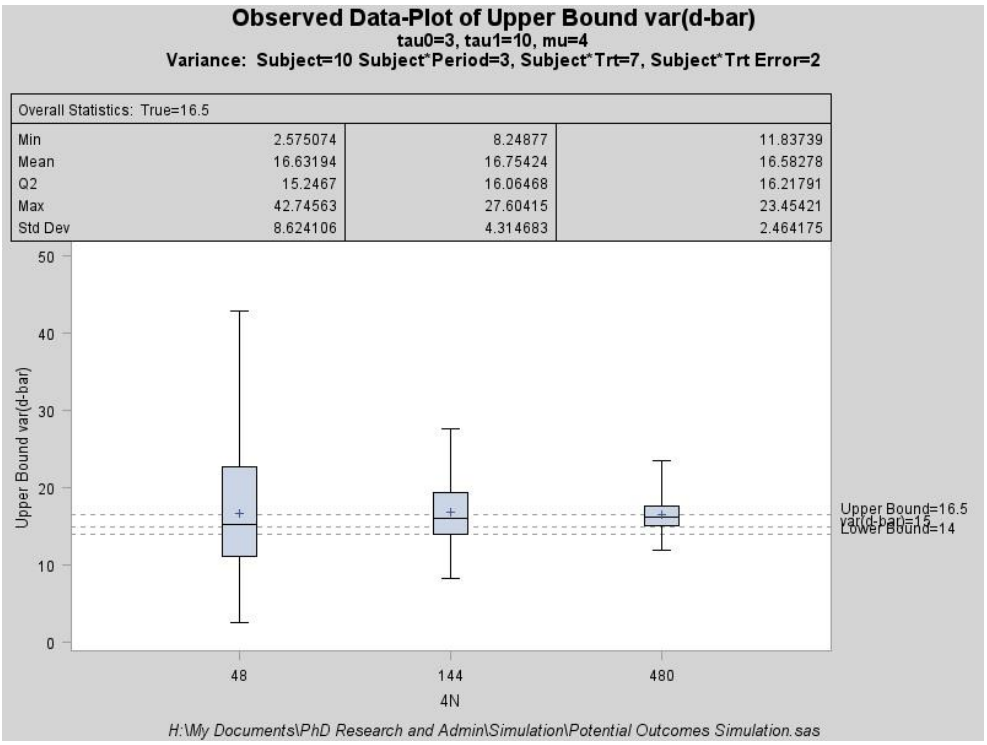

Figure A.5.2 $2\sigma_{st}^2 + \frac{\sigma_e^2}{2} \geq var(\bar{d}_{i.}) \geq 2\sigma_{st}^2$. *Boxplots of the* $S = 100$ *estimates of* $2\sigma_{st}^2 + \frac{\sigma_e^2}{2}$ *at N=10, 30, and 100 EU's measured at 4 time periods. Dotted lines represent values used in the simulation design.*

Appendix B

B.1: Proof of (3.18)

Consider the ANOVA table for the potential model given in Table 3.4. Without loss of generality, assume that $k = 1 \Leftrightarrow k = T$ and $k = 2 \Leftrightarrow k = C$.

Source	DF	Sum of Squares
Block	*(B-1)*	$SS_B = \sum_{i=1}^{B}\sum_{j=1}^{2}\sum_{k=1}^{2}(\bar{r}_{i\cdot\cdot} - \bar{r}_{\cdot\cdot\cdot})^2$
Subj(Block)	*B(2-1)=B*	$SS_{EU(B)} = \sum_{i=1}^{B}\sum_{j=1}^{2}\sum_{k=1}^{2}(\bar{r}_{ij\cdot} - \bar{r}_{i\cdot\cdot})^2$
Trt	*(2-1)=1*	$SS_{Trt} = \sum_{i=1}^{B}\sum_{j=1}^{2}\sum_{k=1}^{2}(\bar{r}_{\cdot\cdot k} - \bar{r}_{\cdot\cdot\cdot})^2$
*Blk*Trt*	*(B-1)*(2-1)=(B-1)*	$SS_{BT} = \sum_{i=1}^{B}\sum_{j=1}^{2}\sum_{k=1}^{2}(\bar{r}_{i\cdot k} - \bar{r}_{i\cdot\cdot} - \bar{r}_{\cdot\cdot k} + \bar{r}_{\cdot\cdot\cdot})^2$
*Subj(Blk)*Trt*	*B(2-1)* (2-1)=B*	$SS_{ST} = \sum_{i=1}^{B}\sum_{j=1}^{2}\sum_{k=1}^{2}(r_{ijk} - \bar{r}_{ij\cdot} - \bar{r}_{i\cdot k} + \bar{r}_{i\cdot\cdot})^2$
TOTAL	**4B-1**	$SS_{Total} = \sum_{i=1}^{B}\sum_{j=1}^{2}\sum_{k=1}^{2}(r_{ijk} - \bar{r}_{\cdot\cdot\cdot})^2$

(i) Claim: For fixed i^{th} pair and j^{th} EU within a pair,

$$\sum_{k=1}^{2}(r_{ijk} - \bar{r}_{ij\cdot} - \bar{r}_{i\cdot k} + \bar{r}_{i\cdot\cdot})^2 = \frac{1}{2}(d_{ij} - \bar{d}_{i\cdot})^2$$

(ii) Claim: For fixed i^{th} pair ,

$$\sum_{k=1}^{2}(\bar{r}_{i\cdot k} - \bar{r}_{i\cdot\cdot} - \bar{r}_{\cdot\cdot k} + \bar{r}_{\cdot\cdot\cdot})^2 = \frac{1}{2}(\bar{d}_{i\cdot} - \bar{d}_{\cdot\cdot})^2$$

(iii) Claim:

$$2(SS_{BT} + SS_{ST}) = \sum_{i=1}^{B}\sum_{j=1}^{2}(d_{ij} - \bar{d}_{\cdot\cdot})^2$$

Proof of (i):

$$\sum_{k=1}^{2}\left(r_{ijk}-\bar{r}_{ij\cdot}-\bar{r}_{i\cdot k}+\bar{r}_{i\cdot\cdot}\right)^2=\sum_{k=1}^{2}\left[\left(r_{ijk}-\bar{r}_{i\cdot k}\right)-\left(\bar{r}_{ij\cdot}-\bar{r}_{i\cdot\cdot}\right)\right]^2=$$

$$\left[\left(r_{ijT}-\bar{r}_{i\cdot T}\right)-\left(\bar{r}_{ij\cdot}-\bar{r}_{i\cdot\cdot}\right)\right]^2+\left[\left(r_{ijC}-\bar{r}_{i\cdot C}\right)-\left(\bar{r}_{ij\cdot}-\bar{r}_{i\cdot\cdot}\right)\right]^2=$$

$$\left\{\left(r_{ijT}-\bar{r}_{i\cdot T}\right)-\left[\frac{1}{2}\left(r_{ijT}+r_{ijC}\right)-\frac{1}{2}\left(\bar{r}_{i\cdot T}+\bar{r}_{i\cdot C}\right)\right]\right\}^2+\left\{\left(r_{ijC}-\bar{r}_{i\cdot C}\right)-\left[\frac{1}{2}\left(r_{ijT}+r_{ijC}\right)-\frac{1}{2}\left(\bar{r}_{i\cdot T}+\bar{r}_{i\cdot C}\right)\right]\right\}^2=$$

$$\left\{\left(r_{ijT}-\bar{r}_{i\cdot T}\right)-\left[\frac{1}{2}\left(r_{ijT}-\bar{r}_{i\cdot T}\right)+\frac{1}{2}\left(r_{ijC}-\bar{r}_{i\cdot C}\right)\right]\right\}^2+\left\{\left(r_{ijC}-\bar{r}_{i\cdot C}\right)-\left[\frac{1}{2}\left(r_{ijT}-\bar{r}_{i\cdot T}\right)+\frac{1}{2}\left(r_{ijC}-\bar{r}_{i\cdot C}\right)\right]\right\}^2=$$

$$\left\{\frac{1}{2}\left(r_{ijT}-\bar{r}_{i\cdot T}\right)-\frac{1}{2}\left(r_{ijC}-\bar{r}_{i\cdot C}\right)\right\}^2+\left\{\frac{1}{2}\left(r_{ijC}-\bar{r}_{i\cdot C}\right)-\frac{1}{2}\left(r_{ijT}-\bar{r}_{i\cdot T}\right)\right\}^2=$$

$$\left\{\frac{1}{2}\left[\left(r_{ijT}-\bar{r}_{i\cdot T}\right)-\left(r_{ijC}-\bar{r}_{i\cdot C}\right)\right]\right\}^2+\left\{-\frac{1}{2}\left[\left(r_{ijT}-\bar{r}_{i\cdot T}\right)-\left(r_{ijC}-\bar{r}_{i\cdot C}\right)\right]\right\}^2=$$

$$2\cdot\frac{1}{4}\left\{\left[\left(r_{ijT}-r_{ijC}\right)-\left(\bar{r}_{i\cdot T}-\bar{r}_{i\cdot C}\right)\right]^2\right\}=$$

$$\frac{1}{2}\left(d_{ij}-\bar{d}_{i\cdot}\right)^2 \quad \blacksquare$$

Proof of (ii):

$$\sum_{k=1}^{2}\left(\bar{r}_{i\cdot k}-\bar{r}_{i\cdot\cdot}-\bar{r}_{\cdot\cdot k}+\bar{r}_{\cdot\cdot\cdot}\right)^2=\sum_{k=1}^{2}\left[\left(\bar{r}_{i\cdot k}-\bar{r}_{\cdot\cdot k}\right)-\left(\bar{r}_{i\cdot\cdot}-\bar{r}_{\cdot\cdot\cdot}\right)\right]^2=$$

$$\left[\left(\bar{r}_{i\cdot T}-\bar{r}_{\cdot\cdot T}\right)-\left(\bar{r}_{i\cdot\cdot}-\bar{r}_{\cdot\cdot\cdot}\right)\right]^2+\left[\left(\bar{r}_{i\cdot C}-\bar{r}_{\cdot\cdot C}\right)-\left(\bar{r}_{i\cdot\cdot}-\bar{r}_{\cdot\cdot\cdot}\right)\right]^2=$$

$$\left\{\left(\bar{r}_{i\cdot T}-\bar{r}_{\cdot\cdot T}\right)-\left[\frac{1}{2}\left(\bar{r}_{i\cdot T}+\bar{r}_{i\cdot C}\right)-\frac{1}{2}\left(\bar{r}_{\cdot\cdot T}+\bar{r}_{\cdot\cdot C}\right)\right]\right\}^2+\left\{\left(\bar{r}_{i\cdot C}-\bar{r}_{\cdot\cdot C}\right)-\left[\frac{1}{2}\left(\bar{r}_{i\cdot T}+\bar{r}_{i\cdot C}\right)-\frac{1}{2}\left(\bar{r}_{\cdot\cdot T}+\bar{r}_{\cdot\cdot C}\right)\right]\right\}^2=$$

$$\left\{\left(\bar{r}_{i\cdot T}-\bar{r}_{\cdot\cdot T}\right)-\left[\frac{1}{2}\left(\bar{r}_{i\cdot T}-\bar{r}_{\cdot\cdot T}\right)+\frac{1}{2}\left(\bar{r}_{i\cdot C}-\bar{r}_{\cdot\cdot C}\right)\right]\right\}^2+\left\{\left(\bar{r}_{i\cdot C}-\bar{r}_{\cdot\cdot C}\right)-\left[\frac{1}{2}\left(\bar{r}_{i\cdot T}-\bar{r}_{\cdot\cdot T}\right)+\frac{1}{2}\left(\bar{r}_{i\cdot C}-\bar{r}_{\cdot\cdot C}\right)\right]\right\}^2=$$

$$\left\{\frac{1}{2}\left(\bar{r}_{i\cdot T}-\bar{r}_{\cdot\cdot T}\right)-\frac{1}{2}\left(\bar{r}_{i\cdot C}-\bar{r}_{\cdot\cdot C}\right)\right\}^2+\left\{\frac{1}{2}\left(\bar{r}_{i\cdot C}-\bar{r}_{\cdot\cdot C}\right)-\frac{1}{2}\left(\bar{r}_{i\cdot T}-\bar{r}_{\cdot\cdot T}\right)\right\}^2=$$

$$\left\{\frac{1}{2}\left[\left(\bar{r}_{i\cdot T}-\bar{r}_{\cdot\cdot T}\right)-\left(\bar{r}_{i\cdot C}-\bar{r}_{\cdot\cdot C}\right)\right]\right\}^2+\left\{-\frac{1}{2}\left[\left(\bar{r}_{i\cdot T}-\bar{r}_{\cdot\cdot T}\right)-\left(\bar{r}_{i\cdot C}-\bar{r}_{\cdot\cdot C}\right)\right]\right\}^2=$$

$$2\cdot\frac{1}{4}\left\{\left[\left(\bar{r}_{i\cdot T}-\bar{r}_{i\cdot C}\right)-\left(\bar{r}_{\cdot\cdot T}-\bar{r}_{\cdot\cdot C}\right)\right]^2\right\}=$$

$$\frac{1}{2}\left(\bar{d}_{i\cdot}-\bar{d}_{\cdot\cdot}\right)^2 \quad \blacksquare$$

Proof of (iii):

$$2(SS_{BT} + SS_{ST}) = 2 \cdot \left[\sum_{i=1}^{B} \sum_{j=1}^{2} \sum_{k=1}^{2} (\bar{r}_{i\cdot k} - \bar{r}_{i\cdot\cdot} - \bar{r}_{\cdot\cdot k} + \bar{r}_{\cdot\cdot\cdot})^2 + \sum_{i=1}^{B} \sum_{j=1}^{2} \sum_{k=1}^{2} \left(r_{ijk} - \bar{r}_{ij\cdot} - \bar{r}_{i\cdot k} + \bar{r}_{i\cdot\cdot} \right)^2 \right] =$$

$$2 \cdot \left[\sum_{i=1}^{B} \sum_{j=1}^{2} \frac{1}{2} \left(d_{ij} - \bar{d}_{i\cdot} \right)^2 + \sum_{i=1}^{B} \sum_{j=1}^{2} \frac{1}{2} \left(\bar{d}_{i\cdot} - \bar{d}_{\cdot\cdot} \right)^2 \right] by\ (i) and\ (ii) =$$

$$2 \cdot \left[\frac{1}{2} \sum_{i=1}^{B} \sum_{j=1}^{2} \left(d_{ij} - \bar{d}_{i\cdot} \right)^2 + \left(\bar{d}_{i\cdot} - \bar{d}_{\cdot\cdot} \right)^2 \right] =$$

$$\sum_{i=1}^{B} \sum_{j=1}^{2} \left[\left(d_{ij} - \bar{d}_{i\cdot} \right)^2 + \left(\bar{d}_{i\cdot} - \bar{d}_{\cdot\cdot} \right)^2 \right] + 2 \cdot 0 =$$

$$\sum_{i=1}^{B} \sum_{j=1}^{2} \left[\left(d_{ij} - \bar{d}_{i\cdot} \right)^2 + \left(\bar{d}_{i\cdot} - \bar{d}_{\cdot\cdot} \right)^2 \right] + 2 \cdot \sum_{i=1}^{B} \sum_{j=1}^{2} \left[\left(d_{ij} - \bar{d}_{i\cdot} \right) \left(\bar{d}_{i\cdot} - \bar{d}_{\cdot\cdot} \right) \right] =$$

$$\sum_{i=1}^{B} \sum_{j=1}^{2} \left[\left(d_{ij} - \bar{d}_{i\cdot} \right)^2 + 2 \cdot \left(d_{ij} - \bar{d}_{i\cdot} \right) \left(\bar{d}_{i\cdot} - \bar{d}_{\cdot\cdot} \right) + \left(\bar{d}_{i\cdot} - \bar{d}_{\cdot\cdot} \right)^2 \right] =$$

$$\sum_{i=1}^{B} \sum_{j=1}^{2} \left[\left(d_{ij} - \bar{d}_{i\cdot} \right) + \left(\bar{d}_{i\cdot} - \bar{d}_{\cdot\cdot} \right) \right]^2 =$$

$$\sum_{i=1}^{B} \sum_{j=1}^{2} \left(d_{ij} - \bar{d}_{\cdot\cdot} \right)^2 \quad \blacksquare$$

B.2: Proof of (4.2)

Consider the ANOVA table for the potential model given in Table 3.10. Without loss of generality, assume that $k = 1 \Leftrightarrow k = T$ and $k = 2 \Leftrightarrow k = C$.

Source	DF	Sum of Squares
Subject	*(N-1)*	$SS_S = \sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}(\bar{r}_{i\cdot\cdot} - \bar{r}_{\cdot\cdot\cdot})^2$
Period	*(2-1)=1*	$SS_P = \sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}(\bar{r}_{\cdot j\cdot} - \bar{r}_{\cdot\cdot\cdot})^2$
*Subject*Period*	*(N-1)(2-1)= (N-1)*	$SS_{SP} = \sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}(\bar{r}_{ij\cdot} - \bar{r}_{i\cdot\cdot} - \bar{r}_{\cdot j\cdot} + \bar{r}_{\cdot\cdot\cdot})^2$
Trt	*(2-1)=1*	$SS_{Trt} = \sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}(\bar{r}_{\cdot\cdot k} - \bar{r}_{\cdot\cdot\cdot})^2$
*Subj*Trt*	*(N-1)(2-1)= (N-1)*	$SS_{ST} = \sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}(\bar{r}_{i\cdot k} - \bar{r}_{i\cdot\cdot} - \bar{r}_{\cdot\cdot k} + \bar{r}_{\cdot\cdot\cdot})^2$
*Period*Trt*	*(2-1)(2-1)=1*	$SS_{PT} = \sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}(\bar{r}_{\cdot jk} - \bar{r}_{\cdot j\cdot} - \bar{r}_{\cdot\cdot k} + \bar{r}_{\cdot\cdot\cdot})^2$
*Subj*Period*Trt*	*(N-1) (2-1)(2-1)=(N-1)*	$SS_{SPT} = \sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}(\bar{r}_{ijk} - \bar{r}_{ij\cdot} - \bar{r}_{i\cdot k} - \bar{r}_{\cdot jk} + \bar{r}_{i\cdot\cdot} + \bar{r}_{\cdot j\cdot} + \bar{r}_{\cdot\cdot k} - \bar{r}_{\cdot\cdot\cdot})^2$
TOTAL	**4N-1**	$SS_{Total} = \sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}(r_{ijk} - \bar{r}_{\cdot\cdot\cdot})^2$

(i) Claim:

$$SS_{PT} = \sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}(\bar{r}_{\cdot jk} - \bar{r}_{\cdot j\cdot} - \bar{r}_{\cdot\cdot k} + \bar{r}_{\cdot\cdot\cdot})^2 = \sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\bar{r}_{\cdot jk}{}^2 - \sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\bar{r}_{\cdot j\cdot}{}^2 - \sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\bar{r}_{\cdot\cdot k}{}^2 + \sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\bar{r}_{\cdot\cdot\cdot}{}^2$$

(ii) Claim:

$$SS_{SPT} = \sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}(\bar{r}_{ijk} - \bar{r}_{ij\cdot} - \bar{r}_{i\cdot k} + \bar{r}_{i\cdot\cdot})^2 - SS_{PT}$$

(iii) Claim:

$$2 \cdot (SS_{ST} + SS_{PT} + SS_{SPT}) = \sum_{i=1}^{N}\sum_{j=1}^{2}(d_{ij} - \bar{d}_{\cdot\cdot})^2$$

Proof of (i):

$$SS_{PT} = \sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\left(\bar{r}_{\cdot jk} - \bar{r}_{\cdot j\cdot} - \bar{r}_{\cdot\cdot k} + \bar{r}_{\cdot\cdot\cdot}\right)^2 =$$

$$\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\left(\bar{r}_{\cdot jk} - \bar{r}_{\cdot j\cdot} - \bar{r}_{\cdot\cdot k} + \bar{r}_{\cdot\cdot\cdot}\right)\left(\bar{r}_{\cdot jk} - \bar{r}_{\cdot j\cdot} - \bar{r}_{\cdot\cdot k} + \bar{r}_{\cdot\cdot\cdot}\right) =$$

$$\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\left(\bar{r}_{\cdot jk}^{\,2} - 2\bar{r}_{\cdot jk}\bar{r}_{\cdot j\cdot} - 2\bar{r}_{\cdot jk}\bar{r}_{\cdot\cdot k} + 2\bar{r}_{\cdot jk}\bar{r}_{\cdot\cdot\cdot} + \bar{r}_{\cdot j\cdot}^{\,2} + 2\bar{r}_{\cdot j\cdot}\bar{r}_{\cdot\cdot k} - 2\bar{r}_{\cdot j\cdot}\bar{r}_{\cdot\cdot\cdot} + \bar{r}_{\cdot\cdot k}^{\,2} - 2\bar{r}_{\cdot\cdot k} + \bar{r}_{\cdot\cdot\cdot}^{\,2}\right) =$$

$$\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\bar{r}_{\cdot jk}^{\,2} - (2-1)\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\bar{r}_{\cdot j\cdot}^{\,2} - (2-1)\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\bar{r}_{\cdot\cdot k}^{\,2} + (2-2+1)\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\bar{r}_{\cdot\cdot\cdot}^{\,2} =$$

$$\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\bar{r}_{\cdot jk}^{\,2} - \sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\bar{r}_{\cdot j\cdot}^{\,2} - \sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\bar{r}_{\cdot\cdot k}^{\,2} + \sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\bar{r}_{\cdot\cdot\cdot}^{\,2} \quad \blacksquare$$

Proof of (ii):

$$SS_{SPT} = \sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\left(\bar{r}_{ijk} - \bar{r}_{ij\cdot} - \bar{r}_{i\cdot k} - \bar{r}_{\cdot jk} + \bar{r}_{i\cdot\cdot} + \bar{r}_{\cdot j\cdot} + \bar{r}_{\cdot\cdot k} - \bar{r}_{\cdot\cdot\cdot}\right)^2 =$$

$$\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\left[\left(\bar{r}_{ijk} - \bar{r}_{ij\cdot} - \bar{r}_{i\cdot k} + \bar{r}_{i\cdot\cdot}\right) - \left(\bar{r}_{\cdot jk} - \bar{r}_{\cdot j\cdot} - \bar{r}_{\cdot\cdot k} + \bar{r}_{\cdot\cdot\cdot}\right)\right]^2$$

=

$$\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\left(\bar{r}_{ijk} - \bar{r}_{ij\cdot} - \bar{r}_{i\cdot k} + \bar{r}_{i\cdot\cdot}\right)^2 - 2\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\left(\bar{r}_{ijk} - \bar{r}_{ij\cdot} - \bar{r}_{i\cdot k} + \bar{r}_{i\cdot\cdot}\right)\left(\bar{r}_{\cdot jk} - \bar{r}_{\cdot j\cdot} - \bar{r}_{\cdot\cdot k} + \bar{r}_{\cdot\cdot\cdot}\right) +$$

$$\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\left(\bar{r}_{\cdot jk} - \bar{r}_{\cdot j\cdot} - \bar{r}_{\cdot\cdot k} + \bar{r}_{\cdot\cdot\cdot}\right)^2$$

=

$$\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\left(\bar{r}_{ijk} - \bar{r}_{ij\cdot} - \bar{r}_{i\cdot k} + \bar{r}_{i\cdot\cdot}\right)^2$$

$$-2\left[\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\bar{r}_{\cdot jk}^{\,2} - (2-1)\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\bar{r}_{\cdot j\cdot}^{\,2} - (2-1)\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\bar{r}_{\cdot\cdot k}^{\,2} + (5-4)\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\bar{r}_{\cdot\cdot\cdot}^{\,2}\right] =$$

$$+\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\left(\bar{r}_{\cdot jk} - \bar{r}_{\cdot j\cdot} - \bar{r}_{\cdot\cdot k} + \bar{r}_{\cdot\cdot\cdot}\right)^2$$

$$=$$

$$\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\left(\bar{r}_{ijk}-\bar{r}_{ij\cdot}-\bar{r}_{i\cdot k}+\bar{r}_{i\cdot\cdot}\right)^2$$

$$-2\left[\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\bar{r}_{\cdot jk}{}^2-\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\bar{r}_{\cdot j\cdot}{}^2-\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\bar{r}_{\cdot\cdot k}{}^2+\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\bar{r}_{\cdot\cdot\cdot}{}^2\right]=$$

$$+\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\left(\bar{r}_{\cdot jk}-\bar{r}_{\cdot j\cdot}-\bar{r}_{\cdot\cdot k}+\bar{r}_{\cdot\cdot\cdot}\right)^2$$

$$=$$

$$\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\left(\bar{r}_{ijk}-\bar{r}_{ij\cdot}-\bar{r}_{i\cdot k}+\bar{r}_{i\cdot\cdot}\right)^2-(2-1)\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\left(\bar{r}_{\cdot jk}-\bar{r}_{\cdot j\cdot}-\bar{r}_{\cdot\cdot k}+\bar{r}_{\cdot\cdot\cdot}\right)^2\ by\ (i)\ of\ Appendix\ B.2=$$

$$\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\left(\bar{r}_{ijk}-\bar{r}_{ij\cdot}-\bar{r}_{i\cdot k}+\bar{r}_{i\cdot\cdot}\right)^2-\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\left(\bar{r}_{\cdot jk}-\bar{r}_{\cdot j\cdot}-\bar{r}_{\cdot\cdot k}+\bar{r}_{\cdot\cdot\cdot}\right)^2$$

$$=\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\left(\bar{r}_{ijk}-\bar{r}_{ij\cdot}-\bar{r}_{i\cdot k}+\bar{r}_{i\cdot\cdot}\right)^2-SS_{PT}$$ ■

Proof of (iii):

$$2\cdot(SS_{ST}+SS_{PT}+SS_{SPT})=$$

$$2\cdot\left[\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}(\bar{r}_{i\cdot k}-\bar{r}_{i\cdot\cdot}-\bar{r}_{\cdot\cdot k}+\bar{r}_{\cdot\cdot\cdot})^2+SS_{PT}+\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\left(\bar{r}_{ijk}-\bar{r}_{ij\cdot}-\bar{r}_{i\cdot k}+\bar{r}_{i\cdot\cdot}\right)^2-SS_{PT}\right]$$

$$by\ (ii)\ of\ Appendix\ B.2$$

$$=$$

$$2\cdot\left[\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}(\bar{r}_{i\cdot k}-\bar{r}_{i\cdot\cdot}-\bar{r}_{\cdot\cdot k}+\bar{r}_{\cdot\cdot\cdot})^2+\sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2}\left(\bar{r}_{ijk}-\bar{r}_{ij\cdot}-\bar{r}_{i\cdot k}+\bar{r}_{i\cdot\cdot}\right)^2\right]=$$

$$2\cdot\left[\sum_{i=1}^{N}\sum_{j=1}^{2}\frac{1}{2}\left(\bar{d}_{i\cdot}-\bar{d}_{\cdot\cdot}\right)^2+\sum_{i=1}^{N}\sum_{j=1}^{2}\frac{1}{2}\left(d_{ij}-\bar{d}_{i\cdot}\right)^2\right]\ by\ (i)and\ (ii)of\ Appendix\ B.1=$$

$$\sum_{i=1}^{N}\sum_{j=1}^{2}\left(d_{ij}-\bar{d}_{\cdot\cdot}\right)^2\ by\ (iii)of\ Appendix\ B.1$$ ■