INDIVIDUAL TREATMENT EFFECT HETEROGENEITY IN MULTIPLE TIME POINTS TRIALS

by

EDWIN ANDONG NDUM

B.S., UNIVERSITY OF BUEA, CAMEROON, 1998
M.S., UNIVERSITY OF NORTH DAKOTA, USA, 2005

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

2009

Abstract

In biomedical studies, the treatment main effect is often expressed in terms of an "average difference." A treatment that appears superior based on the average effect may not be superior for all subjects in a population if there is substantial "subject-treatment interaction." A parameter quantifying subject-treatment interaction is inestimable in two sample completely randomized designs. Crossover designs have been suggested as a way to estimate the variability in individual treatment effects since an "individual treatment effect" can be measured. However, variability in these observed individual effects may include variability due to the treatment plus inherent variability of a response over time. We use the "Neyman - Rubin Model of Causal Inference" (Neyman, 1923; Rubin, 1974) for analyses.

This dissertation consists of two parts: The quantitative and qualitative response analyses. The quantitative part focuses on disentangling the variability due to treatment effects from variability due to time effects using suitable crossover designs. Next, we propose a variable that defines the variance of a true individual treatment effect in a two crossover designs and show that they are not directly estimable but the mean effect is estimable. Furthermore, we show the variance of individual treatment effects is biased under both designs. The bias depends on time effects. Under certain design considerations, linear combinations of time effects can be estimated, making it possible to separate the variability due to time from that due to treatment.

The qualitative section involves a binary response and is centered on estimating the average treatment effect and bounding a probability of a negative effect, a parameter which relates to the individual treatment effect variability. Using a stated joint probability distribution

of potential outcomes, we express the probability of the observed outcomes under a two treatment, two periods crossover design. Maximum likelihood estimates of these probabilities are found using an iterative numerical method. From these, we propose bounds for an inestimable probability of negative effect. Tighter bounds are obtained with information from subjects that receive the same treatments over the two periods. Finally, we simulate an example of observed count data to illustrate estimation of the bounds.

INDIVIDUAL TREATMENT EFFECT HETEROGENEITY IN MULTIPLE TIME POINTS TRIALS

by

EDWIN ANDONG NDUM

B.S., UNIVERSITY OF BUEA, CAMEROON, 1998
M.S., UNIVERSITY OF NORTH DAKOTA, USA, 2005

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

2009

Approved by:

Major Professor
Dr. Gary Gadbury

Abstract

In biomedical studies, the treatment main effect is often expressed in terms of an "average difference." A treatment that appears superior based on the average effect may not be superior for all subjects in a population if there is substantial "subject-treatment interaction." A parameter quantifying subject-treatment interaction is inestimable in two sample completely randomized designs. Crossover designs have been suggested as a way to estimate the variability in individual treatment effects since an "individual treatment effect" can be measured. However, variability in these observed individual effects may include variability due to the treatment plus inherent variability of a response over time. We use the "Neyman - Rubin Model of Causal Inference" (Neyman, 1923; Rubin, 1974) for analyses.

This dissertation consists of two parts: The quantitative and qualitative response analyses. The quantitative part focuses on disentangling the variability due to treatment effects from variability due to time effects using suitable crossover designs. Next, we propose a variable that defines the variance of a true individual treatment effect in a two crossover designs and show that they are not directly estimable but the mean effect is estimable. Furthermore, we show the variance of individual treatment effects is biased under both designs. The bias depends on time effects. Under certain design considerations, linear combinations of time effects can be estimated, making it possible to separate the variability due to time from that due to treatment.

The qualitative section involves a binary response and is centered on estimating the average treatment effect and bounding a probability of a negative effect, a parameter which relates to the individual treatment effect variability. Using a stated joint probability distribution

of potential outcomes, we express the probability of the observed outcomes under a two treatment, two periods crossover design. Maximum likelihood estimates of these probabilities are found using an iterative numerical method. From these, we propose bounds for an inestimable probability of negative effect. Tighter bounds are obtained with information from subjects that receive the same treatments over the two periods. Finally, we simulate an example of observed count data to illustrate estimation of the bounds.

Table of Contents

List of Figures

List of Tables

Dedication

To *St. Padre Pio of Petrelcina*. My family and I have lived, and will continue living, under your guidance, support and constant supervision from Heaven. May you continue showing the way forward and lead us to our LORD and GOD, JESUS CHRIST.

To my dear parents: Mrs. Esther Neh Ndum, Mrs. Martina and Mr. Thaddeus Andong Ndum. May the seeds you sowed yield great fruits. This dissertation is a product of my Dad's philosophy: Education first. You all will live forever in our minds and we (the children) will make your names shine.

To my beloveth wife, Marceline Endah Ndum and baby, Janelle Neh Ndum. To you all, I owe more than words can capture. You were, and will for ever be, the roots on which the stem of my success hangs.

CHAPTER 1 - Introduction

1.0: An Overview

In clinical trials and other scientific studies comparing two or more treatments, the treatment effect is often expressed in terms of an "average" effect although the importance of variability of the effect has been recognized. A treatment that appears superior based on a general population average effect may not be superior for all subjects in a population. Less focus has been put on assessing the variability of the individual treatment effects or "subject-treatment" interaction (Gadbury, 2004) within the population. If substantial, this variance is worth considering in efficacy and safety measures. This dissertation focuses on estimating the individual treatment effect variability and the probability of a negative treatment effect for both the quantitative and qualitative responses using crossover designs. The "Rubin Model of Causal Inference" (Holland, 1986) which employs the "potential outcomes" framework is used.

1.1: The Potential Outcomes Framework

Briefly, let X_i and Y_i denote the response when unit (subject) i receives treatment T and control C (say) respectively. The bivariate pair (X_i, Y_i) are potential outcomes (Rubin, 2005, Neyman, 1923) for unit i. Only one of X_i or Y_i is observed for the i^{th} unit at a given time since we cannot expose a subject to both treatments at the same time. This is called the "fundamental problem of causal inference" (Holland, 1986). The unobservable outcome in the

pair (X_i, Y_i) is sometimes called counterfactual (Glymour, 1986). Note that this bivariate specification holds only when we are comparing two treatments. For a study comparing t treatments, the potential outcomes would be a vector containing t outcomes (rather than two) and only one of the t outcomes would be observable for a given subject at a particular time. The next section, expands on the Rubin model.

1.1.1: The Rubin Model for Causal Inference

Often called the *Neyman-Rubin Model of causal inference*, the framework originated with Neyman's (1923) model (in the context of completely randomized experiments) whereby each unit had two potential outcomes with only one of the two observable. Later Rubin (1974, 2005) and others developed the model into a general framework for causal inference in relation to behavioral science. Holland (1986) also wrote an influential paper using this model emphasizing the philosophical aspects of the framework. On the basis of the work done by Neyman and Rubin, the model is sometimes referred to as the "Neyman-Rubin Model" or sometimes "Neyman-Rubin-Holland Model" or simply the "Rubin Model." Suppose we are to compare these two treatments, the Rubin Model specifies that the true treatment effect for unit i is given as $D_i = X_i - Y_i$. This treatment effect applies to both quantitative and qualitative responses. This Rubin Model assumes the "stable unit treatment value assumption" (SUTVA, Rubin 1980, 1990). Essentially, SUTVA has two assumptions: (1) there is only one version of a specific treatment, either T or C, assigned to all subjects (for example, two or more manufacturers are assumed to produce the exact drug assigned to the subjects) and (2) there is no interference between subjects – that is, the value of each subject's potential outcome does not depend on the

treatment assigned to other subjects. When SUTVA is violated, "an experiment will not yield unbiased estimates of the causal effect of interest" (Sekhon, 2007, p.5). An added assumption in this dissertation is that, the potential outcomes is not affect by "how or whether we try to learn about it" Rubin (2005, p.323). In general, when the causal inference assumptions are defied, randomization of subjects to treatments and the subsequent analysis becomes very complicated. In this dissertation, we will be using the Rubin Model along with the SUTVA conditions. Furthermore, we assume there are carryover effect, no covariates and no missing values or if there are, then, the values are missing completely at random (MCAR, Little and Rubin 2002).

1.1.2: The Definition of Individual Effects, Treatment Effect Homogeneity/Heterogeneity and Subject-Treatment Interaction

Using the Rubin Model $D_i = X_i - Y_i$ for the i^{th} individual, *the individual treatment effect* may be defined as the difference in the response on an individual subject as a result of receiving treatment T versus C at a given time. This is unlike the average effect which is the mean response due to both treatments. Since a subject receives one treatment at a time, this individual treatment effect is not observable. When the focus is on an overall mean effect, the difference $D_i = X_i - Y_i$ is implicitly assumed constant for all individuals in the population when the mean effect is being tested using Fisher's Randomization Test (Fisher, 1935; Rubin, 1980). This assumption is what is referred to as treatment (effect) homogeneity (Longford, 1999). Kravitz et al. (2004, p.660) defined treatment heterogeneity as "…patient diversity in risk of disease, responsiveness to treatment, vulnerability to adverse effects, and utility for different outcomes." They further argue that individual treatment effect heterogeneity can lead to outcomes with a

mixture of "substantial benefit for some, little benefit for many and harm for a few" Kravitz et al. (2004, p.661). The variable treatment effect for each subject results in what is referred to as *subject-treatment interaction* (Marshall, 1997; Longford, 1999). Senn (2001, p. 1481) defines subject-treatment interaction as "the extent to which the difference between treatments differ from one patient to another" or equivalently, "the extent to which the difference between patients being given the same treatment depends on treatment given."

This dissertation consists of two parts: The quantitative and qualitative parts. The quantitative part of the dissertation seeks to estimate individual treatment effect variability and to separate such variability from variability due to time effects in multiple time point trials. The method of potential outcomes will be used to achieve this goal. Meanwhile, in the qualitative parts, focus will be placed on the average treatment effect and the "probability of negative effect" – a component which implicitly reflects individual treatment effect variability. If substantial, the individual effect variability or the probability of negative effect is worth considering in conclusions about effectiveness and safety of the treatment being analyzed.

1.2: Background

The effects of many treatments across individuals may vary widely. When such variation is present, there may be non-negligible proportion of a population that has an adverse effect of a treatment despite studies showing the effect of treatment to be beneficial, "on average." Complicating the detection of the individual effect variability is the fact that some response measures, such as blood pressure, vary widely at different time points. Determining whether a change in a response is due to the effect of a treatment or just due to natural variation of a

response over time can be challenging. A case in point is the controversy surrounding Dietary Approaches to Stop Hypertension (DASH), a salt and blood pressure or hypertension study (Obarzanek et al, 2003).

The DASH study is a widely published clinical trial that suggests that systolic blood pressure (SBP) could be reduced by eating diets rich in fruits, vegetables and with low-fat diary. Five institutions collaborated in the original study in which there were two treatments, a DASH diet and a control diet, each delivered at three levels of salt at 8g(high-H) a day or 140mmol/d, 6g a day or 104mmol/d (Medium-N: government's recommendation) and 4g(Low-L) a day or 62mmol/d. The response variable was the mean of 5 pairs of SBP measurements for each of the 188 participants taken over the final 9 days of each 30-day feeding period. Obarzanek et al (2003) concluded that most of the variability in SBP was caused by "other factors' than salt intake. They also pointed out that the variability depends on the group of individuals involved, suggesting a case for the introduction of covariates.

DASH study is one of many that may involve individual treatment effect heterogeneity or variability. Considering the controversies and limitations of the DASH-Sodium Trial, the knowledge and the ability to estimate variability in an individual treatment effect using the appropriate design is of critical importance. A treatment that appears superior based on the average effect may not be superior for all subjects in a population if there is substantial individual treatment effect variability expressed in terms of "subject-by-treatment" interaction. This interaction may consist of component factor-by-treatment interactions like "gene-by-treatment" interaction, "social status-by-treatment" interaction and so forth.

Cross-over designs have been suggested as a way to estimate the variability in individual treatment effects since some degree of a treatment effect's "separability" from effects of time can

be achieved. The DASH-Sodium results suggests that variability in observed individual effects may include variability due to the treatment plus inherent variability of a response over time and would require special types of cross designs to identify and estimate.

Disentangling variability due to treatment effects from variability due to time effects is challenging. Essentially, we will analyze the individual treatment effects as a random variable (rather than a constant effect) for a specified population of subjects and it suffices to look at the variance as well as mean effect parameters (Longford, 1999).

We present a method of potential outcomes analyses using various two treatment designs. For instance, we use the two treatments, three periods crossover design – a class of repeated crossover design or the "n-of-1" trial (Senn, 2001). This dissertation work builds on earlier work by Gadbury and others (2000, 2001, 2004). A parameter quantifying subject-treatment interaction is inestimable in two treatments, two period balanced crossover designs. The two treatments, three periods design used here extends the initial work on the two treatments, two periods design by Gadbury et al. (2004). The design used here permits certain inseparable effects or a combination of effects to be measured or estimated.

The first part of the research (Chapter 2 and 3) is based on quantitative treatment response variables. In the next Section 1.3, we present previous work involving the complete randomized design and the two treatments, two periods crossover designs. Section 1.4 introduces some population types to be used in the analyses presented in chapter 2 and 3. Chapter 2 deals with an extended two treatments, two periods design, a design whereby some subjects stay on the same treatments over the two periods. In chapter 3, we extend the analyses to a three period design for quantitative response and in chapter 4, we further the work done with qualitative

responses in Gadbury et al (2004), summarily presented in Section 1.3.3. Chapter 5 sums up the dissertation work and lays out some future challenges and research opportunities.

1.3: Previous Work on Subject-Treatment Interaction

Senn (2001, Vol. 35) outlined the various error terms and sources of variability (Table 1.1) that are identifiable with different types of designs (Table 1.2). With a 2 treatments, 2 period cross-over design, it is impossible to separate the variability due to patient-by-treatment interaction from the within-patient variation even in the absence of carryover effects but a repeated period crossover design will make it possible for such effects to be separable. In the table below, Senn (2001, Vol. 35) describes and lists the effects that are identifiable plus the various errors terms.

Table 1.1: *Sources of Variation in Clinical Trials*

Label	Source	Description
A	Between Treatments	The average difference between treatments over all randomizations (and hence over all patients). The 'true' mean difference between treatments
B	Between patients	The average difference between patients. (Averaged over both experimental and control treatments.)
C	Patient-by- treatment interaction	The extent to which the difference between treatments differ from one patient to another. (Equivalently, the extent to which the difference between patients being given the same treatment depends on treatment given.)
D	Within-patient error	The variability shown from treatment period to treatment period when the same patient is given the same Treatment

Source: Senn, S. (2001, p.1481). *"Individual Therapy: New Dawn or False Dawn?"* British Medical Journal (BMJ), Vol. 35

Table 1.2**:** *Indentifiability and Clinical Trials*

Type of Trial	Description	Identifiable Effects	Error Term
Parallel	Each patient receives one treatment	A	B + C+ D
Cross-over	Each patient receives each treatment in one period only	A and B	C + D
Repeated period cross-over (Sets of n-of-1 trials)	Each patient receives each treatment in at least two periods	A and B and C	D

Source: Stephen Senn (2001, p. 1481). *"Individual Therapy: New Dawn or False Dawn?"* British Medical Journal (BMJ), Vol. 35. Total Error E = A+B+C+D

In another paper, Senn (2001, Vol. 329) further echoes the ideas presented on Table 1.2 and recommends random effect models in the analysis of repeated periods cross-over design to identify individual effect variability represented by the subject-by-treatment interaction, though as will be shown, assumptions are still needed and these assumptions are not always obvious without the structure of potential outcomes. Thus, it appears worthy to consider a repeated period crossover design using potential outcomes. One particular example where a repeated period crossover design was used to study subject-by-treatment interaction is the double blind randomized comparison of paracetamol 1g b.i.d. (bis in die – twice a day) and diclofenac 50 mg b.i.d. osteoarthritis study reported by March et al.(1994), although their analysis did not used the random effect model nor were potential outcomes considered.

1.3.1: Two treatment completely randomized designs

Let X_i and Y_i denote quantitative or categorical outcomes when unit (subject) i receives treatment T and C respectively for $i = 1, 2, ..., N$. The set of *N* potential outcomes has the form

given below (left bracket), which after treatment assignment, produces observed outcomes of the form shown (right bracket), and where the "?" represents an unobservable potential outcome (Gadbury et al., 2004).

$$\begin{pmatrix} X_1 & Y_1 \\ \vdots & \vdots \\ X_N & Y_N \end{pmatrix} \xrightarrow{Treatment\ \ Assignment} \begin{pmatrix} X_1 & ? \\ ? & Y_2 \\ \vdots & \vdots \\ ? & Y_{N-1} \\ X_N & ? \end{pmatrix}$$

This two treatments randomized complete design assumes SUTVA. That is, subject's response to a particular treatment stays the same regardless of what treatment other subjects receive or whether there may be different types of treatments.

The individual treatment effect $D_i = X_i - Y_i$ cannot be observed because only one of the X_i or Y_i is observed for an individual at a particular time. So, some have proposed crossover design, whereby, the treatment effect for an individual can be observed. But the observed treatment effects also contain time effects. The next section explores this in a 2 treatments, 2 periods crossover design.

1.3.2: Initial Work on Two Period (TC CT) Cross-Over Designs for Quantitative Response

Gadbury (2001) developed some initial results for a two treatment balanced cross over design. Accordingly, consider two treatments labeled T and *C* in a 2 period design

		Period	
		1	2
Sequence	1	*T*	*C*
	2	*C*	*T*

Assume a finite population of $2n$ subjects used to define the potential outcomes. We assign n subjects to each sequence. Potential outcome framework for the $2n$ subjects is given by

Subject	*Time* 1		*Time* 2	
1	$X_1 - t_1$	$Y_1 - \tau_1$	$X_1 + t_1$	$Y_1 + \tau_1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$2n$	$X_{2n} - t_{2n}$	$Y_{2n} - \tau_{2n}$	$X_{2n} + t_{2n}$	$Y_{2n} + \tau_{2n}$

with potential outcomes $\left(X_i - t_1,\ Y_i - \tau_1\right)$ for period 1 and $\left(X_i + t_2,\ Y_i + \tau_2\right)$ for period 2.

Furthermore, define the "true" mean individual treatment effect for the i^{th} subject as the average of the two true treatment effects over the two time periods so that the time effects cancel, that is, the "true" finite population mean treatment effect is given as $\bar{D} = \bar{X} - \bar{Y}$, where $\bar{X} = (1/2n)\sum_{i=1}^{2n} X_i$ and $\bar{Y} = (1/2n)\sum_{i=1}^{2n} Y_i$.

The true finite population variance of the individual treatment effects, denoted S_D^2, is given as

$$S_D^2 = Var(X - Y) = Var(D) = \frac{1}{2n}\sum_{i=1}^{2n}(D_i - \bar{D})^2$$

The observed treatment effect for the i^{th} subject is

$$d_i = \left[\left(X_i - t_i\right) - \left(Y_i + \tau_i\right)\right]T_i + \left[\left(X_i + t_i\right) - \left(Y_i - \tau_i\right)\right]\left(1 - T_i\right)$$

where T_i represent the random assignment to sequence with $T_i = 1$ or 0 for assignment to T-C or to C-T, respectively.

It was shown that, the estimated "observed" mean treatment effect $\bar{d}$ is unbiased for $\bar{D}$ with respect to the randomization distribution for T_i. Gadbury (2001) defined a reasonable estimator of S_D^2 as S_d^2 where , $S_d^2 = \frac{1}{2n}\sum_{i=1}^{2n}\left(d_i - \bar{d}\right)^2$ and showed that,

$$E\left(S_d^2\right) = S_D^2 + \frac{2(n-1)}{2n-1}S_{t+\tau}^2 + \left(\bar{t} + \bar{\tau}\right)^2$$

where, $S_{t+\tau}^2$ is the finite population variance of the sum of time effect terms, $t+\tau$ and $\bar{t}$ and $\bar{\tau}$ are the finite population averages of t and τ. The bias term is given as

$$bias = \frac{2(n-1)}{2n-1}S_{t+\tau}^2 + \left(\bar{t} + \bar{\tau}\right)^2$$

This bias is always positive and will only be zero if $t+\tau = 0$, in which case, S_d^2 estimates S_D^2 exactly. In the next part, we present the previous work on qualitative (binary) response variable.

1.3.3: Initial Work on Two Period (TC – CT) Cross-Over Designs for Binary Response

Suppose that the outcome 1 denotes a "success" and 0, a "failure", the following table provides the assumed bivariate distribution of potential outcomes in an infinite population, as presented in Gadbury et al. (2004):

(x,y)	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
$P(X=x, Y=y)$	π_1	π_2	π_3	π_4

where $\sum_{i=1}^{4}\pi_i = 1$.

They noted that the individual treatment effect variable $D = X - Y$ is discrete with possible values 0, -1, and 1 with probabilities $\pi_1 + \pi_4$, π_2 and π_3 respectively. Subject-Treatment

interaction is present in the population unless one of these three probabilities, $\pi_1 + \pi_4$, π_2 and π_3, is equal to one. A detrimental or unfavorable effect means that $D = -1$. Thus, the proportion of the population experiencing a negative effect is π_2. They showed that, the mean treatment effect is given by $E(D) = E(X - Y) = \pi_3 - \pi_2$. The population parameters π_i, $(i = 1,2,3,4)$ are by themselves, nonestimable. In addition, the constructed bounds for the risk involved in administering the treatment *T* to the population were given as

$$\max(0,\ \pi_2 - \pi_3) \le \pi_2 \le \min(1 - (\pi_3 + \pi_4),\ \pi_2 + \pi_4).$$

Gadbury et al., (2004) also considered a matched-pairs design and showed that a design that includes some pairs receiving the same treatment can tighten the bounds for π_2, though the tightness depends on 'quality of matching' criteria that cannot directly be assessed from observable outcomes.

Remark 1.3.1

In the next chapter , we focus on the TC – CT – TT – CC design for quantitative response variable. Thus, in chapter 2, we let some subject stay on the same treatment (TT and CC) over the two periods. This is an extension of the TC – CT crossover design presented in Sections 1.3.2 above. These subjects will provide the additional information necessary to estimate the bias $bias = \frac{2(n-1)}{2n-1} S_{t+\tau}^2 + (\bar{t} + \bar{\tau})^2$ and subsequently, the true variance of the individual treatment effects, S_D^2, using a certain assumption. These estimations were not possible with the previous

TC – CT crossover design because $S^2_{t+\tau}$ (and hence $bias = \frac{2(n-1)}{2n-1} S^2_{t+\tau} + (\bar{t} + \bar{\tau})^2$) could not be estimable, without even more restrictive and perhaps implausible assumptions.

In both chapters 2 and 3 involving quantitative response variable, we will be working with three types of populations described in the next subsection. In chapter 4 that considers a binary response, the multinomial population model is used.

1.4: A List of Populations Used For the Quantitative Analysis

For a smooth understanding of the quantitative analyses, we list three types of populations used herein.

1). First, a finite population of potential outcomes from which we define a true individual effect, D, with finite population parameters $\bar{D}$ and S^2_D. We begin the analyses with this finite population of potential outcomes.

2). Second, a population distribution of time effects designated t and τ. The population will have parameters μ_t, μ_τ, σ^2_t and σ^2_τ. This population will be used when estimation of a bias term comes to focus.

3). Third, a "super – population" (Smith and Sugden, 1988) distribution for treatment effects. This population will be used to determine the distribution of D with population parameters μ_D and σ^2_D. The population of treatment effects will be used in the illustrative examples.

Again, note that these lists of populations will be useful with the quantitative response only. The qualitative analysis will make use of a binomial or multinomial population model.

CHAPTER 2 - A Two Period, Two Treatment Design for Quantitative Responses

2.1: A Two Period TC – CT – TT – CC Design for Quantitative Responses

Previously, Gadbury (2001) worked on the two period TC and CT design. In this section, we will extend this design to include TT and CC.

Accordingly, consider two treatments labeled T and *C* in the following 2 periods designs:

		Period 1	Period 2
Sequence	1	*T*	*C*
	2	*C*	*T*

Design 2.1.1: *Two sequence-two periods.*

		Period 1	Period 2
Sequence	1	*T*	*C*
	2	*C*	*T*
	3	*T*	*T*
	4	*C*	*C*

Design 2.1.2: *Four sequence-two periods.*

Using a slightly different estimator of the true individual treatment effect variability, Gadbury (2001) developed some initial results for Design 2.1.1(see section 1.3.2). However, it is impossible to estimate S_D^2 with Design 2.1.1. But, if we allow some subjects to stay on the same treatments as shown in Design 2.1.2 (sequences 3 and 4), estimation of S_D^2 is possible, with certain assumptions, through estimation of linear combinations of time effect parameters. Note that we cannot observe the treatment effects for subjects in sequences 3 and 4 of Design 2.1.2.

Those subjects will provide the time effect information necessary to estimate S_D^2 from Design 2.1.1.

Following the Gadbury (2001), assume we have finite population of $2n$ subjects from which we define our potential outcomes framework as shown below.

Subject	Period 1		Period 2	
1	$X_1 + t_{11}$	$Y_1 + \tau_{11}$	$X_1 + t_{12}$	$Y_1 + \tau_{12}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$2n$	$X_{2n} + t_{2n1}$	$Y_{2n} + \tau_{2n1}$	$X_{2n} + t_{2n2}$	$Y_{2n} + \tau_{2n2}$

with potential outcomes $\left(X_i + t_{i1},\ Y_i + \tau_{i1}\right)$ for period 1 and $\left(X_i + t_{i2},\ Y_i + \tau_{i2}\right)$ for period 2 $\left(i = 1, 2, ..., 2n\right)$. X_i and Y_i are the average responses to treatments T and C, respectively, over the two time periods for subject $i = 1, 2, ..., 2n$; t_{ij} (associated with treatment T) and τ_{ij} (associated with treatment C) are the time effect parameters for subject $i = 1, 2, ..., 2n$ in period $j = 1, 2$. We assume $t_{i1} + t_{i2} = 0$ and $\tau_{i1} + \tau_{i2} = 0$.

Remark 2.1

The symbols of the time parameters used here is a slight deviation from those in Gadbury (2001) where time parameters are simply denoted t_i and τ_i. The reason for specifying the time parameters as t_{i1}, t_{i2} and $\tau_{i1},\ \tau_{i2}$ is to synchronize the symbols with those of a more complex design (to be seen in chapter 3). Nonetheless, the results will not be affected by this change since it may be assumed that $t_i = t_{i1} = -t_{i2}$ and $\tau_i = \tau_{i1} = -\tau_{i2}$

Using Design 2.1.2, the observed outcome framework is given as

			Period 1	Period 2
Sequence	T	C	$X_i + t_{i1}$	$Y_i + \tau_{i2}$
	C	T	$Y_i + \tau_{i1}$	$X_i + t_{i2}$
	T	T	$X_i + t_{i1}$	$X_i + t_{i2}$
	C	C	$Y_i + \tau_{i1}$	$Y_i + \tau_{i2}$

Define the true individual treatment effect on the i^{th} subject as $D_i = X_i - Y_i + \sum_{j=1}^{2} t_{ij} - \sum_{j=1}^{2} \tau_{ij} = X_i - Y_i$. D_i is not observable for any i. The true finite population mean effect of treatment, $\bar{D}$ is given as $\bar{D} = \bar{X} - \bar{Y}$ where $\bar{X} = (1/2n)\sum_{i=1}^{2n} X_i$ and $\bar{Y} = (1/2n)\sum_{i=1}^{2n} Y_i$. Define the true finite population variance of individual treatment effects, denoted S_D^2, as

$$S_D^2 = \frac{1}{2n}\sum_{i=1}^{2n}(D_i - \bar{D})^2$$

Let γ_{i1} be an indicator variable which takes the value 1 when the i^{th} subject is in sequence 1, for $i = 1, 2, ..., 2n$. Observe that, γ_{i1} ~ Binomial$(1, 1/2)$. The observed treatment effect for the i^{th} subject under Design 2.1.2 is

$$d_i = \left[\left(X_i + t_{i1}\right) - \left(Y_i + \tau_{i2}\right)\right]\gamma_{i1} + \left[\left(X_i + t_{i2}\right) - \left(Y_i + \tau_{i1}\right)\right]\left(1 - \gamma_{i1}\right).$$

This simplifies to

$$d_i = \left(X_i - Y_i\right) + \left(t_{i2} - \tau_{i1}\right) + \left(t_{i1} - t_{i2} + \tau_{i1} - \tau_{i2}\right)\gamma_{i1}$$

Thus, $P(\gamma_{i1}=1)=1/2$. In addition, $E(\gamma_{i1})=1/2$ and $Var(\gamma_{i1})=1/4$.

If $i=i'$ then $E(\gamma_{i1}\gamma_{i'1})=E(\gamma_{i1}^2)=E(\gamma_{i1})=1/2$.

For $i\neq i'$, $E(\gamma_{i1}\gamma_{i'1})=P(\gamma_{i1}=1,\gamma_{i'1}=1)=\frac{1}{2}\left(\frac{n-1}{2n-1}\right)$ when subjects i and i' are in sequence 1 and $E(\gamma_{i1}(1-\gamma_{i'1}))=\frac{1}{2}\left(\frac{n}{2n-1}\right)$ when subjects i and i' are in sequence 1 and sequence 2, respectively.

Proposition 2.1

The observed mean treatment effect, $\bar{d}$, is an unbiased estimate of the true mean treatment effect $\bar{D}$. That is, $E(\bar{d})=\bar{D}$, where expectation is taken over all possible randomizations γ_{i1}.

Proof:

$$\begin{aligned}E_\gamma(\bar{d}) &= \frac{1}{2n}\sum_{i=1}^{n}\left((X_i-Y_i)+(t_{i2}-\tau_{i1})+(t_{i1}-t_{i2}+\tau_{i1}-\tau_{i2})\gamma_{i1}\right)\\ &= \frac{1}{2n}\sum_{i=1}^{2n}(X_i-Y_i)\\ &= \bar{X}-\bar{Y}\\ &= \bar{D}\end{aligned}$$ ■

Note that the proof was established with the fact that $E_\gamma(\gamma_{i1})=1/2$ and the assumptions that $\sum_{j=1}^{2} t_{ij}=0$ and $\sum_{j=1}^{2}\tau_{ij}=0$ for subject $i=1,2,...,2n$ and period $j=1,2$.

Define the observed individual treatment effect variability, denoted S_d^2, as

$$S_d^2 = \frac{1}{2n-1}\sum_{i=1}^{2n}\left(d_i - \bar{d}\right)^2 .$$

Proposition 2.2

For each subject, the observe treatment effect variability is not an unbiased estimator of the true individual treatment variability. That is,

$$E\left(\frac{2n-1}{2n}S_d^2\right) = S_D^2 + Bias$$

where $Bias = \frac{2n-1}{2n}\left(S_{t+\tau}^2 + \left(\overline{t+\tau}\right)^2\right)$, $S_{t+\tau}^2$ and $\overline{t+\tau} = \frac{1}{2n}\sum_{i=1}^{2n}\left(t_{i1}+\tau_{i1}\right)$ are the finite population variance and finite population mean of the sum of time effect terms $t+\tau$. Expectation is taken over all possible randomization of the $2n$ subject.

Proof: *See Appendix A*

Remark 2.2:

With Design 2.1.1, it is impossible to estimate S_D^2 due to the presence of $S_{t+\tau}^2$ and $\overline{t+\tau}$ in the bias formula. $S_{t+\tau}^2$ and $\overline{t+\tau}$ cannot be estimable because the combination of time effect parameters, $t+\tau$, cannot be observed for any individual. In order for S_D^2 to be estimated, consider a design where some subjects stay on the same treatments as given in Design 2.1.2. In the next section, Design 2.1.2 is used to estimate S_D^2 with a particular assumption.

Estimation of S_D^2

In Design 2.1.2, we let some subjects stay on same treatments as provided by sequences TT and CC. These subjects provide no information about the individual treatment effect and are used here to provide useful information about the combination of time effects $t+\tau$. This information will be used to estimate the finite population variance $S_{t+\tau}^2$ and the finite population mean $\overline{t+\tau}$ parameters. Sequences TC and CT considered together will give us same information as obtained above.

Assume a total of $2n$ subject where $i=1,2,...n$ subjects are assigned to each of sequences TT and CC. The observed outcome is

			Period 1	Period 2
Sequence	T	T	X_i+t_{i1}	X_i+t_{i2}
	C	C	$Y_i+\tau_{i1}$	$Y_i+\tau_{i2}$

From the TT and CC randomizations, we obtain $t_{i2}-t_{i1}$ and $\tau_{i2}-\tau_{i1}$ respectively. Using the assumption that $\sum_{j=1}^{2} t_{ij}=0$ and $\sum_{j=1}^{2}\tau_{ij}=0$, we obtain $2t_{i1}$ and $2\tau_{i1}$ from which we get the combination $\left(\bar{t}+\bar{\tau}\right)$ where $\bar{t}=\frac{1}{n}\sum_{i=1}^{n} t_{i1}$ and $\bar{\tau}=\frac{1}{n}\sum_{i=1}^{n}\tau_{i1}$ are the observed mean time effects obtained from those who stayed on TT and CC respectively. Thus, we denote an estimate of $\overline{t+\tau}$ by $\overline{\hat{t+\tau}}$ where $\overline{\hat{t+\tau}}=\left(\bar{t}+\bar{\tau}\right)$.

Define $\hat{S}_{t+\tau}^2$ as the finite population variance of $t+\tau$.

Assuming t and τ are independent, we can estimate $S^2_{t+\tau}$ by $\hat{S}^2_{t+\tau}$. Plugging-in the estimated bias, we have

$$\hat{Bias} = \frac{2n-1}{2n}\left(\hat{S}^2_{t+\tau} + \left(\overline{\hat{t+\tau}} \right)^2 \right)$$

where $\overline{\hat{t+\tau}} = \left(\bar{t} + \bar{\tau} \right)$.

Suppose we designate the estimated true individual treatment effect variability as $\hat{S}^2_D$, $\hat{S}^2_D$ is given as

$$\hat{S}^2_D = \frac{2n-1}{2n} S^2_d - \hat{Bias}$$

Remark 2.3

Despite the added information from the TT – CC design, it is worthwhile noting that estimation of S^2_D, denoted $\hat{S}^2_D$, was possible because of the important assumption of independence between t and τ. Without this assumption only bounds for S^2_D can be estimated (as was mentioned in Gadbury, 2001, though Gadbury did not produce the bounds nor were the TT CC sequences considered.

2.2: A Two Period TC – CT – TT – CC Design with Binary Responses

This is similar to the topic in Section 2.2 except for binary responses. Some related work was done by Gadbury et al., (2004) for matched-pairs, but exchangeability assumptions that were relevant for matched-pairs do not necessarily hold when subjects are matched to themselves over

time periods. Thus, the details in Gadbury et al., (2004) will be modified to redefine a 'successful' response to treatment and to deal with time effects as opposed to matching criteria in matched-pairs. Chapter 4 provides the detailed analyses for this two period TC – CT – TT – CC design with binary responses. In the next chapter, we analyze a two treatment, three period crossover design. This design facilitates the estimation of the individual treatment effect variability, a task that could not be achieved with the two periods TC – CT crossover design considered in chapter 2.

CHAPTER 3 - Chapter Three: A Three Period, Two Treatment Design with Quantitative

3. 0. A Three Period, Two Treatment Design with Quantitative Responses

In classic jargon, this design falls under the general classification referred to as "n – of – 1 trial" or Repeated Period Cross-Over design (Senn, 2001). These types of design are particularly useful for patients with chronic diseases – like hypertension, cancer, diabetes, alzheimer, arthritis, asthma and so on – although it has been known to be useful other purposes like examining the short term choice of drugs for osteoarthritis (Yelland et al, 2006). In addition, repeated period cross-over designs are necessary for cases where the physician doubts the effectiveness of a certain drug on a patient. Generally, the main advantage of repeated period cross-over design is that patients act as their own control.

Assume two treatments denoted T and C where one could be the control. Also assume we have n_k subjects assigned to the k^{th} sequence. Define $N = \sum n_k$. Let the i^{th} subject and the j^{th} period be such that $i = 1, 2, 3, ...N$ and $j = 1, 2, 3$. Ratkowsky et al. (1993) compared the efficiencies of various 2 treatments and 3 periods design for estimating a mean treatment effect. In the pictures below, we present a few of the designs compared.

Design 1:

Sequence	Period 1	2	3
1	*T*	*C*	*C*
2	*C*	*T*	*C*
3	*C*	*T*	*T*
4	*T*	*C*	*T*

Design 2:

Sequence	Period 1	2	3
1	*T*	*C*	*C*
2	*C*	*T*	*T*
3	*C*	*C*	*T*
4	*T*	*T*	*C*

Design 3:

Sequence	Period 1	2	3
1	*T*	*T*	*C*
2	*T*	*C*	*T*
3	*C*	*T*	*T*
4	*C*	*C*	*T*
5	*C*	*T*	*C*
6	*T*	*C*	*C*

They concluded that the most efficient of the three designs – in terms of variability and computational difficulties – was *Design 2*. Using all three designs, we performed some superficial analyses of individual treatment effect variability using the potential outcome method. Among the three, *Design 3* had the advantage of separating time effects from true individual treatment effects. So, *Design 3* is used for analyses in this chapter.

The following random assignment of treatment is used, where 6 sequences are grouped into 2 squares as shown below.

Square 1

Sequence	*Period* 1	2	3
1	*T*	*T*	*C*
2	*T*	*C*	*T*
3	*C*	*T*	*T*

Square 2

Sequence	*Period* 1	2	3
1	*C*	*C*	*T*
2	*C*	*T*	*C*
3	*T*	*C*	*C*

(R-1)

Notice that, for every subject, there are $k = 2^3 - 2 = 6$ different possible assignments of the treatments in the three periods. The treatment options, *TTT* and *CCC* do not concretely capture the cross over design as subjects do not change treatment (parallel design). Initially, we will restrict the randomization of treatments to *Square 1*. *Square 2* is a mirror image of square 1 with *T* and *C* flipped. The analysis will be done under two situations: Unequal and equal number of subjects per sequence.

3.1: Unequal Number of Subjects per Sequence

Assume the subjects are independently and randomly assigned to the sequences. Note that this allows for a possible unequal number of subjects per sequence. Let an indicator random variable δ_{ij} be a sequence assignment variable for the i^{th} subject, j^{th} period with $i = 1, 2, 3, ..., N$ and $j = 1, 2, 3$.

For square 1, define

$$\delta_{ij} = \begin{cases} 1, & \text{if subject } i \text{ receives } C \text{ is in period } j \\ 0, & \text{otherwise} \end{cases} \qquad (1.1)$$

Thus, $\delta_{ij} \sim Binomial(1, 1/3)$. Thus, $P(\delta_{ij} = 1) = 1/3$ for all j and

$$\begin{aligned} CTT \;\} \; \delta_{i1} \; or \; \delta_{i1} = 1,\ \delta_{i2} = 0,\ \delta_{i3} = 0 \\ TCT \;\} \; \delta_{i2} \; or \; \delta_{i2} = 1,\ \delta_{i1} = 0,\ \delta_{i3} = 0 \\ TTC \;\} \; \delta_{i3} \; or \; \delta_{i3} = 1,\ \delta_{i1} = 0,\ \delta_{i2} = 0 \end{aligned}$$

For each i and $j \neq j'$, we have that, $E_\delta(\delta_{ij}) = 1/3$, $E_\delta(\delta_{ij}\delta_{ij'}) = 0$

$$E_{\delta}\left(\delta_{ij}\delta_{i'j'}\right) = P\left(\delta_{ij}=1,\ \delta_{i'j'}=1\right)$$
$$= P\left(\delta_{ij}=1\right)P\left(\delta_{i'j'}=1\right)$$
$$= \begin{cases} 1/9, & \text{if } j = j' \\ 1/9, & \text{if } j \neq j' \end{cases} \quad (i \neq i')$$

and $E_{\delta}\left(\delta_{ij}{}^{2}\right) = E_{\delta}\left(\delta_{ij}\right)$. Note, $E_{\delta}(...)$ denoted the expectation with respect to the finite population randomization. The total number of subjects in the sequences constitutes the size of the population.

We assume no carryover effects. Let t_{ij} and τ_{ij} denote the unobservable time effects due to treatment *T* and *C* respectively. $t' = (t_{i1}, t_{i2}, t_{i3})$ and $\tau' = (\tau_{i1}, \tau_{i2}, \tau_{i3})$ are periodic effect parameters. That is, these parameters quantify the unobservable effects from period to period when the same subject is given the same treatment. Let $X_i + t_{ij}$ and $Y_i + \tau_{ij}$ be the observed responses to T and C respectively. The potential outcomes at time periods 1, 2 and 3 are

$$\underbrace{X_i + t_{i1},\ Y_i + \tau_{i1}}_{P1};\ \underbrace{X_i + t_{i2},\ Y_i + \tau_{i2}}_{P2};\ \underbrace{X_i + t_{i3},\ Y_i + \tau_{i3}}_{P3} \tag{1.2}$$

where *P1, P3* and *P3* denote the periods . X_i and Y_i denote the average response (to treatments *T* and *C* respectively) over the three periods. For the i^{th} subject we have, $\sum_{j=1}^{3} t_{ij} = 0$ and $\sum_{j=1}^{3} \tau_{ij} = 0$.

These potential treatment outcomes are established under the condition that only one measurement of the subject's response at a particular period is observable. The true individual treatment effect on the i^{th} subject in the j^{th} period is defined

$$D_{ij} = X_i - Y_i + t_{ij} - \tau_{ij} \text{ .....(\#)}$$

Define, $D_i = \frac{1}{3}\sum_{j=1}^{3} D_{ij}$. Thus, $D_i = X_i - Y_i + \sum_{j=1}^{3} t_{ij} - \sum_{j=1}^{3} \tau_{ij} = X_i - Y_i$ (since $\sum_{j=1}^{3} t_{ij} = 0$ and $\sum_{j=1}^{3} \tau_{ij} = 0$)

Note D_i is not observable for any $i = 1, 2, ..., N$. The "true" (overall) mean effect of treatment $\bar{D}$ is given as $\bar{D} = \bar{X} - \bar{Y}$. That is,

$$\bar{D} = \left(\frac{1}{N}\sum_{i=1}^{N} X_i - \frac{1}{N}\sum_{i=1}^{N} Y_i \right)$$

Remark 3.1

$\bar{D}$ is the true finite population mean treatment effect of the N subjects in the study. The finite population variance of the true individual treatment effects (denoted S_D^2) is $S_D^2 = \frac{1}{N}\sum_{i=1}^{N} \left(D_i - \bar{D} \right)^2$.

Remark 3.2

S_D^2 represents the overall individual treatment response variability or overall subject-by-treatment interaction. That is, the variability of subjects' responses on the same treatment that

depends on the treatment administered or the extent to which the difference between treatments T and C depend on the subject.

Let d_i be the observed treatment effect for the i^{th} subject taken over the *Square 1* randomization (R-1). We have that,

$$d_i = \left(\frac{X_i + t_{i1} + X_i + t_{i2}}{2} - \left(Y_i + \tau_{i3} \right) \right) \delta_{i3} \qquad \text{(for TTC)}$$

$$+ \left(\frac{X_i + t_{i2} + X_i + t_{i3}}{2} - \left(Y_i + \tau_{i1} \right) \right) \delta_{i1} \qquad \text{(for CTT)}$$

$$+ \left(\frac{X_i + t_{i1} + X_i + t_{i3}}{2} - \left(Y_i + \tau_{i2} \right) \right) \delta_{i2} \qquad \text{(for TCT)}$$

$$\Rightarrow \qquad d_i = \left(X_i - Y_i \right) \sum_{j=1}^{3} \delta_{ij} + \left(\frac{t_{i2} + t_{i3}}{2} \right) \delta_{i1} + \left(\frac{t_{i1} + t_{i3}}{2} \right) \delta_{i2} + \left(\frac{t_{i1} + t_{i2}}{2} \right) \delta_{i3} - \tau_{i1} \delta_{i1} - \tau_{i2} \delta_{i2} - \tau_{i3} \delta_{i3}$$

Thus, $$d_i = \left(X_i - Y_i \right) \sum_{j=1}^{3} \delta_{ij} \; - \sum_{j=1}^{3} \alpha_{ij} \delta_{ij}$$

$$d_i = \left(X_i - Y_i \right) \; - \sum_{j=1}^{3} \alpha_{ij} \delta_{ij} \qquad (1.3)$$

where $\alpha_{ij} = \left(\frac{t_{ij}}{2} + \tau_{ij} \right)$ and $\sum_{j=1}^{3} \delta_{ij} = 1$. Now, let $\bar{d}$ be the mean observed effect over all subjects. We have that, $\bar{d} = \frac{1}{N} \sum_{i=1}^{N} d_i$. We also assume that the observed individual response variability, denoted S_d^2, is $S_d^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(d_i - \bar{d} \right)^2$.

Remark 3.3

S_d^2 is the total observed variability that results from subjects given different treatments at different periods of time. Hence, S_d^2 may be seen as the sum of subject-by-treatment interaction and the variability within the subject over time.

Proposition 3.1

The mean observed treatment effect is an unbiased estimate of the true mean effect. That is,

$$E_\delta\left(\bar{d}\right) = \bar{D} \tag{1.4}$$

Proof:

$$E_\delta\left(\bar{d}\right) = \frac{1}{N}\sum_{i=1}^{N} E\left(d_i\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left[\left(X_i - Y_i\right) - \sum_{j=1}^{3}\alpha_{ij}\delta_{ij}\right]$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left[\left(X_i - Y_i\right) - \sum_{j=1}^{3}\alpha_{ij}E\left(\delta_{ij}\right)\right]$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left[\left(X_i - Y_i\right) - \frac{1}{3}\sum_{j=1}^{3}\alpha_{ij}\right]$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(X_i - Y_j\right) \qquad \left(\text{since } \sum_{j=1}^{3} t_{ij} = 0 \text{ and } \sum_{j=1}^{3}\tau_{ij} = 0\right)$$

$$= \bar{D} \qquad \blacksquare$$

Proposition 3.2

For each subject, the observe treatment effect variability is not an unbiased estimator of the true individual treatment variability. That is,

$$E_{\delta}\left(\frac{N-1}{N}S_d^2\right) = S_d^2 + Bias_{IND}$$

Hence, S_d^2 is a biased estimate of S_D^2, where $Bias_{IND} = \frac{N-1}{3N^2}\sum_{i=1}^{N}\sum_{j=1}^{3}\alpha_{ij}^2$, $\alpha_{ij} = \frac{t_{ij}}{2} + \tau_{ij}$ and $Bias_{IND}$ represents the bias for the design where subjects are independently assigned to sequences.

Proof: *See Appendix B*

Remark 3.4

$Bias_{IND}$ involves only time effect parameters and may be thought to quantify variability of treatment responses within subject. That is, the variability that results when the same subject is given the same treatment at different periods of time.

Corollary 3.1

Define $t_{ij} = -2\tau_{ij}$ then, from (1.3), $d_i = D_i$. That is, the observed treatment effect, d_i, is same as the true treatment effect, D_i, but D_{ij}, defined in Equation (#), is not constant across periods because, under this condition, $D_{i1} = X_i - Y_i - 3\tau_{i1}$, $D_{i2} = X_i - Y_i - 3\tau_{i2}$,

$D_{i3} = X_i - Y_i - 3\tau_{i3}$ and each of these depend on the time effect parameter τ_{ij}. Note, again, that $D_i = \frac{1}{3}\sum_{j=1}^{3} D_{ij}$.

A stronger condition exists when t_{ij} and τ_{ij} are all equal to zero. In this case, $d_i = D_i$, $S_d^2 = S_D^2$ and D_i is the same across periods since, in this case, $D_{i1} = D_{i2} = D_{i3} = X_i - Y_i$.

Next, we turn our attention to the second situation in the analyses of individual treatment effect variability for quantitative responses: Equal number of subjects per sequence.

3.2: Equal Number of Subjects Assigned to Sequence

In the last section, we dealt with a possible case of unequal number of subjects per sequence due to the independent assignment of subjects to sequences. Now, assume that the randomization must result in equal number of subjects per sequence. Suppose we assign *n* subjects to each sequence, we would have a total of $N = 3n$ subjects assigned to the three sequences under consideration.

$$\text{Total Randomization} = \underbrace{\binom{N}{n}}_{\#\text{ for seq1}} \underbrace{\binom{N-n}{n}}_{\#\text{ for seq2}} \underbrace{\binom{n}{n}}_{\#\text{ for seq3}}$$

Furthermore, $\delta_{ij} \sim Bernoulli(1/3)$. Thus, $P(\delta_{ij} = 1) = 1/3$ for all j and

$$N = 3n \begin{cases} \rightarrow CTT \;\} \; \delta_{i1} \; or \; \delta_1 = 1,\ \delta_2 = 0,\ \delta_3 = 0 \\ \rightarrow TCT \;\} \; \delta_{i2} \; or \; \delta_2 = 1,\ \delta_1 = 0,\ \delta_3 = 0 \\ \rightarrow TTC \;\} \; \delta_{i3} \; or \; \delta_3 = 1,\ \delta_1 = 0,\ \delta_2 = 0 \end{cases}$$

For the i^{th} subject, $E_\delta(\delta_{ij}) = 1/3$, $E_\delta(\delta_{ij}^2) = E_\delta(\delta_{ij})$ and $E_\delta(\delta_{ij}\delta_{ij'}) = 0$ (for $j \neq j'$).

Now, for $(i \neq i')$, $E_\delta(\delta_{ij}\delta_{i'j'}) = P(\delta_{ij} = 1,\ \delta_{i'j'} = 1)$

$$= P(\delta_{ij} = 1 \mid \delta_{i'j'} = 1)P(\delta_{i'j'} = 1)$$

$$= \begin{cases} \left(\frac{n-1}{N-1}\right)\frac{1}{3} & \text{for } j = j' \\ \left(\frac{n}{N-1}\right)\frac{1}{3} & \text{for } j \neq j' \end{cases}$$

Proposition 3.3

For each subject, the observe treatment effect variability is not an unbiased estimator of the true individual treatment variability. That is,

$$E_\delta\left(\frac{N-1}{N}S_d^2\right) = S_D^2 + Bias_{DEP} \qquad (1.5)$$

Thus, S_d^2 is not an unbiased estimate of S_D^2 where

$$Bias_{DEP} = \frac{1}{3N^2}\left((N-1)\sum_{i=1}^{N}\sum_{j=1}^{3}\alpha_{ij}^2 - \underbrace{\frac{n-1}{N-1}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3}\alpha_{ij}\alpha_{i'j}}_{i\neq i'} - \underbrace{\frac{n}{N-1}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3}\sum_{j'=1}^{3}\alpha_{ij}\alpha_{i'j'}}_{(i\neq i')\ and\ (j\neq j')}\right)$$

where $\alpha_{ij} = \left(\frac{t_{ij}}{2} + \tau_{ij}\right)$ and $Bias_{DEP}$ represents the bias for the design where the assignment of the next subject to a sequence depends on the previous subject's assignment.

Proof: *See Appendix C*

Corollary 3.2

If $t_{ij} = -2\tau_{ij}$, the conditions in corollary 3.1 also apply here. In general, the observed treatment effect variability is biased for the true treatment effect variability.

3.3: Parameter Estimation

To proceed further with the estimation of the individual treatment effect variability, we use a population of time effects. We make the following assumptions. Let $t' = (t_{i1}, t_{i2}, t_{i3})$ and $\tau' = (\tau_{i1}, \tau_{i2}, \tau_{i3})$, for $i = 1, 2, ..., n$ and $j = 1, 2, 3$. Assume $\begin{pmatrix} t_{i1} \\ t_{i2} \end{pmatrix}$ and $\begin{pmatrix} \tau_{i1} \\ \tau_{i2} \end{pmatrix}$ are independent and identically distributed *(i.i.d.)* $\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \begin{pmatrix} 1 & \rho_t \\ \rho_t & 1 \end{pmatrix} \sigma_t^2 \right)$ and $\left(\begin{pmatrix} \mu_3 \\ \mu_4 \end{pmatrix} \begin{pmatrix} 1 & \rho_\tau \\ \rho_\tau & 1 \end{pmatrix} \sigma_\tau^2 \right)$, respectively. Also assume, $\begin{pmatrix} t_{i1} \\ t_{i2} \end{pmatrix}$ and $\begin{pmatrix} \tau_{i1} \\ \tau_{i2} \end{pmatrix}$ are jointly independent. We note that $t_{i3} = -(t_{i1} + t_{i2})$ and $\tau_{i3} = -(\tau_{i1} + \tau_{i2})$. Estimates of the distribution parameters will be derived.

The variables t_{i1}, t_{i2}, t_{i3} or $\tau_{i1}, \tau_{i2}, \tau_{i3}$ cannot be observed separately. However, certain combinations of $t' = (t_{i1}, t_{i2}, t_{i3})$ or $\tau' = (\tau_{i1}, \tau_{i2}, \tau_{i3})$ can be used as estimates of the effect parameters. That is, from the combination TTC, we can observe $(t_{i1} - t_{i2})$. Similarly, from the data in sequences TCT and CTT, we can observe $(t_{i1} - t_{i3})$ and $(t_{i2} - t_{i3})$, respectively. Upon substituting for t_{i3}, they simplify to $(t_{i1} - t_{i2})$, $(2t_{i1} + t_{i2})$ and $(t_{i1} + 2t_{i2})$ respectively. The matrix

$M = \begin{bmatrix} t_{i1} - t_{i2} \\ t_{i1} + 2t_{i2} \\ 2t_{i1} + t_{i2} \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} t_{i1} \\ t_{i2} \end{bmatrix}$ contains only two linear combinations that are linearly independent so one can make use of any two of $(t_{i1} - t_{i2})$, $(t_{i1} + 2t_{i2})$ and $(2t_{i1} + t_{i2})$. Correspondingly, from the data in *Square 2* of (R-1), we can observe $(\tau_{i1} - \tau_{i2})$, $(\tau_{i1} + 2\tau_{i2})$ and $(2\tau_{i1} + \tau_{i2})$. Henceforth, we will omit the "i" in expressions like $(t_{i1} - t_{i2})$, $(\tau_{i1} - \tau_{i2})$ etc. Thus, we have that,

$$E(t_1 - t_2) = \mu_1 - \mu_2 \qquad (1.6)$$

$$E(t_1 + 2t_2) = \mu_1 + 2\mu_2 \qquad (1.7)$$

$$E(2t_1 + t_2) = 2\mu_1 + \mu_2 \qquad (1.8)$$

where the expectation is taken with respect to the population of time effects given above. Using the above equations, we propose the estimates, $\hat{\mu}_1$ and $\hat{\mu}_2$, corresponding to the population means, μ_1 and μ_2, as $\hat{\mu}_1 = \frac{1}{3}\left[2(\overline{t_1 - t_2}) + (\overline{t_1 + 2t_2})\right]$ and $\hat{\mu}_2 = \frac{1}{3}\left[(\overline{t_1 + 2t_2}) - (\overline{t_1 - t_2})\right]$. Similarly we have, $\hat{\mu}_3 = \frac{1}{3}\left[2\left(\overline{\tau_1 - \tau_2}\right) + (\overline{\tau_1 + 2\tau_2})\right]$ and $\hat{\mu}_4 = \frac{1}{3}\left[(\overline{\tau_1 + 2\tau_2}) - (\overline{\tau_1 - \tau_2})\right]$, where $\overline{t_1 - t_2}$ and $\overline{t_1 + 2t_2}$, are the means of the observed differences between responses from subjects who received treatment T in sequences TTC and CTT, respectively. Similarly, $\overline{\tau_1 - \tau_2}$ and $\overline{\tau_1 + 2\tau_2}$ are the means for those who received treatment C in sequences CCT and TCC, respectively.

Let $S_1^2,\ S_2^2,\ S_3^2,\ S_4^2$ be the sample variances of $t_1 - t_2,\ t_1 + 2t_2,\ \tau_1 - \tau_2$ and $\tau_1 + 2\tau_2$ respectively. We observe that, $E(S_1^2) = Var(t_1 - t_2)$ and $E(S_2^2) = Var(t_1 + 2t_2)$. Similarly, $E(S_3^2) = Var(\tau_1 - \tau_2)$ and $E(S_4^2) = Var(\tau_1 + 2\tau_2)$.

Proposition 3.4

Let $u_i = t_{i1} - t_{i2}$ or simply $u = t_1 - t_2$ and $v = t_1 + 2t_2$. Assume $u_1, u_2, \ldots$ and $v_1, v_2, \ldots$ are *i.i.d* with finite fourth moments. Define $\hat{\sigma}_t^2 = \frac{1}{9}\left[2S_1^2 + S_2^2\right]$ and $\hat{\sigma}_\tau^2 = \frac{1}{9}\left[2S_3^2 + S_4^2\right]$. Then, $\hat{\sigma}_t^2$ and $\hat{\sigma}_\tau^2$ are unbiased and consistent estimates of σ_t^2 and σ_τ^2 respectively.

Proof:

We will show the proof for the formulas involving t. Those with τ follow in the same manner.

$$E(S_1^2) = Var(t_1 - t_2)$$

$$= 2\sigma_t^2 - 2Cov(t_1, t_2) \qquad (1.9)$$

$$E(S_2^2) = Var(t_1 + 2t_2)$$

$$= 5\sigma_t^2 + 4Cov(t_1, t_2) \qquad (1.10)$$

Hence, by elimination,

$$\sigma_t^2 = \frac{1}{9}\left[2E(S_1^2) + E(S_2^2)\right].$$

Similarly, we obtain $\sigma_\tau^2 = \frac{1}{9}\left[2S_3^2 + S_4^2\right]$. Define $\hat{\sigma}_t^2 = \frac{1}{9}\left[2S_1^2 + S_2^2\right]$ and $\hat{\sigma}_\tau^2 = \frac{1}{9}\left[2S_3^2 + S_4^2\right]$, we have that,

$$
\begin{aligned}
E\left(\hat{\sigma}_t^2\right) &= \frac{1}{9}\left[2E\left(S_1^2\right)+E\left(S_2^2\right)\right] \\
&= \frac{1}{9}\left[2Var(t_1-t_2)+Var(t_1+2t_2)\right] \\
&= \frac{1}{9}\left[2\left(2\sigma_t^2-2Cov(t_1,t_2)\right)+\left(5\sigma_t^2+4Cov(t_1,t_2)\right)\right] \\
&= \sigma_t^2
\end{aligned}
$$

Hence, $\hat{\sigma}_t^2$ is an unbiased estimate of σ_t^2. Similarly, we can show that $\hat{\sigma}_\tau^2$ is unbiased for σ_τ^2.

Consistency: Since the fourth moments exists, by using two applications of the weak law of large numbers and the continuous mapping theorem, we have that,

$$S_1^2 = \frac{N}{N-1}\left[\frac{1}{N}\sum_{i=1}^{N}u_i^2-\bar{u}_N^2\right]\xrightarrow{P}1\left(E\left(u_1^2\right)-E\left(u_1\right)^2\right)=\text{var}\left(u_1\right).$$

Thus, $S_1^2 \xrightarrow{P} Var\left(t_1-t_2\right)$.

Similarly, $S_2^2 \xrightarrow{P} Var\left(t_1+2t_2\right)$. Hence, $\left(S_1^2,S_2^2\right)\xrightarrow{P}\left(Var\left(t_1-t_2\right),Var\left(t_1+2t_2\right)\right)$.

Now, let $\hat{\sigma}_t^2 = g\left(S_1^2,S_2^2\right)$. Also, let, $g:\mathbb{R}^2\to\mathbb{R}^2$ be continuous at every point on a set G such that, $P\left(\left(S_1^2,S_2^2\right)\in G\right)=1$. Further applications of continuous mapping theorem and Slutsky theorem leads to

$\hat{\sigma}_t^2 = g\left(S_1^2,S_2^2\right)\xrightarrow{P}\frac{1}{9}\left[2Var\left(t_1-t_2\right)+Var\left(t_1+2t_2\right)\right]=\sigma_t^2$. That is $\hat{\sigma}_t^2\xrightarrow{P}\sigma_t^2$. Hence, $\hat{\sigma}_t^2$ is a consistent estimator of σ_t^2. Similar proof for $\hat{\sigma}_\tau^2$ can be established. ■

Corollary 3.3

From proposition 3.4, we may define estimates of the true correlation values ρ_t and ρ_τ as

$\hat{\rho}_t = \frac{1}{2}\left(\frac{2S_2^2 - 5S_1^2}{S_2^2 + 2S_1^2}\right)$ and $\hat{\rho}_\tau = \frac{1}{2}\left(\frac{2S_4^2 - 5S_3^2}{S_4^2 + 2S_3^2}\right)$. We propose that $\hat{\rho}_t$ and $\hat{\rho}_\tau$ are consistent estimates of ρ_t and ρ_τ respectively.

Proof:

Using equations (1.8) and (1.9), we have that

$$Cov(t_1, t_2) = \sigma_t^2 - \frac{1}{2}E(S_1^2)$$

$$= \frac{1}{18}\left(2E(S_2^2) - 5E(S_1^2)\right)$$

after substituting σ_t^2 from above. Hence,

$$\rho_t = \frac{Cov(t_1, t_2)}{\sqrt{Var(t_1)Var(t_2)}} = \frac{1}{2}\left[\frac{2E(S_2^2) - 5E(S_1^2)}{E(S_2^2) + 2E(S_1^2)}\right]$$

Define $\hat{\rho}_t = \frac{1}{2}\left(\frac{2S_2^2 - 5S_1^2}{S_2^2 + 2S_1^2}\right)$ and $\hat{\rho}_\tau = \frac{1}{2}\left(\frac{2S_4^2 - 5S_3^2}{S_4^2 + 2S_3^2}\right)$. From *Proposition 3.4*, $S_2^2 \xrightarrow{P} Var(t_1 + 2t_2)$. Hence, $\left(S_1^2, S_2^2\right) \xrightarrow{P} \left(Var(t_1 - t_2), Var(t_1 + 2t_2)\right)$. Assume $\hat{\rho}_t = h\left(S_1^2, S_2^2\right)$. Let, $h : \mathbb{R}^2 \to \mathbb{R}^2$ be continuous at every point on a set H such that, $P\left(\left(S_1^2, S_2^2\right) \in \mathrm{H}\right) = 1$. Further applications of continuous mapping theorem and Slutsky theorem, gives

$$\hat{\rho}_t = h\left(S_1^2, S_2^2\right) = \frac{1}{2}\left(\frac{2S_2^2 - 5S_1^2}{S_2^2 + 2S_1^2}\right) \xrightarrow{P} \frac{1}{2}\left(\frac{2Var(t_1 + 2t_2) - 5Var(t_1 - t_2)}{Var(t_1 + 2t_2) + 2Var(t_1 - t_2)}\right) = \rho_t$$

Thus, $\hat{\rho}_t \xrightarrow{P} \rho_t$ and $\hat{\rho}_t$ is a consistent estimator of ρ_t. The proof for $\hat{\rho}_\tau$ follows in a similar manner. ■

3.4: Expected Bias Estimation

Now find estimates for the *bias* factor developed in the previous sections are found under various situations.

3.4.1: Bias Estimation for the Case of Unequal Subjects per Sequence

First consider the case for the bias under the independent assignment of subjects which was given as

$$Bias_{IND} = \frac{N-1}{3N^2}\sum_{i=1}^{N}\sum_{j=1}^{3}\alpha_{ij}^{\ 2}$$

where $\alpha_{ij} = \left(\frac{t_{ij}}{2} + \tau_{ij}\right)$ and $i = 1, 2, 3, ..., N$ is the total number of subjects in the sequences.

Proposition 3.5

Given the $\begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$ and $\begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix}$ are independent and identically distributed $\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \begin{pmatrix} 1 & \rho_t \\ \rho_t & 1 \end{pmatrix}\sigma_t^2\right)$ and $\left(\begin{pmatrix} \mu_3 \\ \mu_4 \end{pmatrix} \begin{pmatrix} 1 & \rho_\tau \\ \rho_\tau & 1 \end{pmatrix}\sigma_\tau^2\right)$, respectively, and assuming $\begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$ and $\begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix}$ are jointly independent, we propose that,

$$E(Bias_{IND}) = \frac{N-1}{6N} Q$$

where expectation is taken over the distribution of $\begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$ and $\begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix}$ and

$$Q = (2+\rho_t)\sigma_t^2 + 4(2+\rho_\tau)\sigma_\tau^2 + (\mu_1+\mu_2)^2 - \mu_1\mu_2 + 4(\mu_3+\mu_4)^2 - 4\mu_3\mu_4 + 2\big((\mu_1+\mu_2)(\mu_3+\mu_4) + \mu_1\mu_4 + \mu_2\mu_3\big)$$

Proof:

$$E\left(Bias_{IND}\right) = \frac{N-1}{3N^2}\sum_{i=1}^{N}\sum_{j=1}^{3} E\left(\frac{t_{ij}}{2} + \tau_{ij}\right)^2 .$$

But, $$\sum_{j=1}^{3} E\left(\frac{t_j}{2} + \tau_j\right)^2 = \sum_{j=1}^{3} E\left(\frac{t_j^2}{4}\right) + \sum_{j=1}^{3} E\left(\tau_j^2\right) + \sum_{j=1}^{3} E\left(t_j\tau_j\right)$$

$$= \sum_{j=1}^{2} E\left(\frac{t_j^2}{4}\right) + E\left(\frac{t_3^2}{4}\right) + \sum_{j=1}^{2} E\left(\tau_j^2\right) + E\left(\tau_3^2\right) + \sum_{j=1}^{2} E\left(t_j\tau_j\right) + E\left(t_3\tau_3\right)$$

Now, $\sum_{j=1}^{3} t_j = 0 \Rightarrow t_3 = -(t_1+t_2)$ and $\sum_{j=1}^{3} \tau_j = 0 \Rightarrow \tau_3 = -(\tau_1+\tau_2)$. Thus,

$$\sum_{j=1}^{3} E\left(\frac{t_j}{2} + \tau_j\right)^2 = \sum_{j=1}^{2} E\left(\frac{t_j^2}{4}\right) + \frac{1}{4}E(t_1+t_2)^2 + \sum_{j=1}^{2} E\left(\tau_j^2\right) + E(\tau_1+\tau_2)^2 + \sum_{j=1}^{2} E\left(t_j\tau_j\right) + E\big((t_1+t_2)(\tau_1+\tau_2)\big)$$

$$= \frac{1}{4}\sum_{j=1}^{2} E\left(t_j^2\right) + \frac{1}{4}E\left(t_1^2\right) + \frac{1}{4}E\left(t_2^2\right) + \frac{1}{2}E\left(t_1t_2\right) + \sum_{j=1}^{2} E\left(\tau_j^2\right) + E\left(\tau_1+\tau_2\right)^2$$
$$+ \sum_{j=1}^{2} E\left(t_j\tau_j\right) + E\left(t_1\tau_1\right) + E\left(t_1\tau_2\right) + E\left(t_2\tau_1\right) + E\left(t_2\tau_2\right)$$

From Section 3.3, $E\left(t_1^2\right)=\sigma_t^2+\mu_1^2$ and $E\left(t_1t_2\right)=Cov\left(t_1,t_2\right)+\mu_1\mu_2=\rho_t\sigma_t^2+\mu_1\mu_2$. In addition, $E\left(t_1\tau_1\right)=Cov\left(t_1,\tau_1\right)+E\left(t_1\right)E\left(\tau_1\right)=\mu_1\mu_3$ (τ 's and t 's jointly independent) - this relationship applies to other expectations in the formula above. Thus,

$$\begin{aligned}\sum_{j=1}^{3}E\left(\frac{t_j}{2}+\tau_j\right)^2 &= \frac{2}{4}\left(\sigma_t^2+\mu_1^2\right)+\frac{2}{4}\left(\sigma_t^2+\mu_2^2\right)+\frac{1}{2}\left(\rho_t\sigma_t^2+\mu_1\mu_2\right)\\ &\quad+2\left(\sigma_\tau^2+\mu_3^2\right)+2\left(\sigma_\tau^2+\mu_4^2\right)+2\left(\rho_\tau\sigma_\tau^2+\mu_3\mu_4\right)\\ &\quad+2\left(\mu_1\mu_3+\mu_2\mu_4\right)+\mu_1\mu_4+\mu_2\mu_3\\ &=\frac{1}{2}\Big[\left(2+\rho_t\right)\sigma_t^2+4\left(2+\rho_\tau\right)\sigma_\tau^2+\left(\mu_1+\mu_2\right)^2-\mu_1\mu_2+4\left(\mu_3+\mu_4\right)^2\\ &\quad-4\mu_3\mu_4+2\left(\left(\mu_1+\mu_2\right)\left(\mu_3+\mu_4\right)+\mu_1\mu_4+\mu_2\mu_3\right)\Big]\\ &=\frac{1}{2}Q \qquad (1.11)\end{aligned}$$

where,

$$\begin{aligned}Q=&\left(2+\rho_t\right)\sigma_t^2+4\left(2+\rho_\tau\right)\sigma_\tau^2+\left(\mu_1+\mu_2\right)^2-\mu_1\mu_2+4\left(\mu_3+\mu_4\right)^2\\ &-4\mu_3\mu_4+2\left(\left(\mu_1+\mu_2\right)\left(\mu_3+\mu_4\right)+\mu_1\mu_4+\mu_2\mu_3\right)\end{aligned}$$

We then have,

$$E\left(Bias_{IND}\right)=\frac{N-1}{3N^2}\sum_{i=1}^{N}\sum_{j=1}^{3}E\left(\frac{t_{ij}}{2}+\tau_{ij}\right)^2$$

$$\Rightarrow E\left(Bias_{IND}\right)=\frac{N-1}{3N^2}\sum_{i=1}^{N}\frac{1}{2}Q$$

$$=\frac{N-1}{6N}Q \qquad \blacksquare$$

Next, the analyses are continued with the bias estimation for the case where the randomization should result in equal number of subjects per sequence.

3.4.2: Bias Estimation for the Case of Equal Number of Subjects per Sequence

Consider finding the estimate of the expectation of the second bias. From *proposition 3.3,* the bias is given as

$$Bias_{DEP} = \frac{1}{3N^2}\left((N-1)\sum_{i=1}^{N}\sum_{j=1}^{3}\alpha_{ij}^2 - \underbrace{\frac{n-1}{N-1}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3}\alpha_{ij}\alpha_{i'j'}}_{i\neq i'} - \underbrace{\frac{n}{N-1}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3}\sum_{j'=1}^{3}\alpha_{ij}\alpha_{i'j'}}_{(i\neq i')\ and\ (j\neq j')} \right)$$

where $\alpha_{ij} = \left(\frac{t_{ij}}{2} + \tau_{ij} \right)$ and $N = 3n$ is the total number of subjects in the three sequences with n subjects each.

Proposition 3.6

Given the $\begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$ and $\begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix}$ are independent and identically distributed $\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \begin{pmatrix} 1 & \rho_t \\ \rho_t & 1 \end{pmatrix} \sigma_t^2 \right)$ and $\left(\begin{pmatrix} \mu_3 \\ \mu_4 \end{pmatrix} \begin{pmatrix} 1 & \rho_\tau \\ \rho_\tau & 1 \end{pmatrix} \sigma_\tau^2 \right)$, respectively, and assuming $\begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$ and $\begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix}$ are jointly independent, we propose that,

$$E(Bias_{DEP}) = \frac{1}{6}Q$$

where expectation is taken over the distribution of $\begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$ and $\begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix}$ and

$$Q = (2+\rho_t)\sigma_t^2 + 4(2+\rho_\tau)\sigma_\tau^2 + (\mu_1+\mu_2)^2 - \mu_1\mu_2 + 4(\mu_3+\mu_4)^2 - 4\mu_3\mu_4 + 2\big((\mu_1+\mu_2)(\mu_3+\mu_4) + \mu_1\mu_4 + \mu_2\mu_3\big)$$

Proof:

Let's define the following quantities as

$$U = \sum_{i=1}^{N}\sum_{j=1}^{3}\alpha_{ij}^2,\ V = \underbrace{\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3}\alpha_{ij}\alpha_{i'j}}_{i\neq i'} \text{ and } W = \underbrace{\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3}\sum_{j'=1}^{3}\alpha_{ij}\alpha_{i'j'}}_{(i\neq i')\ and\ (j\neq j')}$$

$$E(Bias_{DEP}) = \frac{1}{3N^2}\left((N-1)E(U) - \frac{n-1}{N-1}E(V) - \frac{n}{N-1}E(W)\right) \tag{1.12}$$

where expectation is taken over the distribution of $\begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$ and $\begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix}$. Next, we find the expectation of each quantity in the bias statement. Hence,

$$E(U) = \sum_{i=1}^{N}\sum_{j=1}^{3}E\left(\frac{t_{ij}}{2}+\tau_{ij}\right)^2 = \frac{N}{2}Q \tag{1.13}$$

$$E(V) = \underbrace{\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3}E\left(\alpha_{ij}\alpha_{i'j}\right)}_{i\neq i'}$$

$$= \underbrace{\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3}E\left(\frac{t_{ij}}{2}+\tau_{ij}\right)^2}_{i\neq i'} \qquad \text{(since } i \text{ and } i' \text{ have the same distribution)}$$

$$= \underbrace{\sum_{i=1}^{N}\sum_{i'=1}^{N}\frac{1}{2}Q}_{i\neq i'},$$

$$= \frac{N(N-1)}{2} Q \tag{1.14}$$

Finally,

$$E(W) = \underbrace{E \sum_{i=1}^{N} \sum_{i'=1}^{N} \sum_{j=1}^{3} \sum_{j'=1}^{3} \alpha_{ij} \alpha_{i'j'}}_{(i \neq i') \text{ and } (j \neq j')}$$

$$= \underbrace{E \sum_{i=1}^{N} \sum_{i'=1}^{N} \left(\underbrace{\sum_{j=1}^{3} \sum_{j'=1}^{3} \alpha_{ij} \alpha_{i'j'}}_{(j \neq j')} \right)}_{(i \neq i')}$$

$$= \underbrace{E \sum_{i=1}^{N} \sum_{i'=1}^{N} \left(\left\{ \sum_{j=1}^{3} \alpha_{ij} \right\}^2 - \sum_{j=1}^{3} \alpha_{ij} \alpha_{i'j} \right)}_{(i \neq i')} \qquad \left(\underbrace{\text{since } \sum_{j=1}^{3} \sum_{j'=1}^{3} \alpha_{ij} \alpha_{i'j'}}_{(j \neq j')} = \left\{ \sum_{j=1}^{3} \alpha_{ij} \right\}^2 - \sum_{j=1}^{3} \alpha_{ij} \alpha_{i'j} \right)$$

$$= \underbrace{-E \sum_{i=1}^{N} \sum_{i'=1}^{N} \sum_{j=1}^{3} \alpha_{ij} \alpha_{i'j}}_{(i \neq i')} \qquad \left(\text{since } \sum_{j=1}^{3} \alpha_{ij} = 0 \right)$$

$$= -E(V) = -\frac{N(N-1)}{2} Q \tag{1.15}$$

Putting (1.13), (1.14) and (1.15) into (1.12) gives

$$E(Bias_{DEP}) = \frac{1}{3N^2} \left(\frac{N(N-1)}{2} Q - \frac{n-1}{N-1} \frac{N(N-1)}{2} Q + \frac{n}{N-1} \frac{N(N-1)}{2} Q \right) = \frac{1}{6} Q \qquad \blacksquare$$

Remark 3.6

From the estimated expected bias formulas above, we observe that $E(Bias_{IND}) \le E(Bias_{DEP})$ for all n values. In addition,

$\underset{N\to\infty}{Lim} E(Bias_{IND}) = E(Bias_{DEP})$ because $\underset{N\to\infty}{Lim}\left(\frac{N-1}{6N}Q\right) = \frac{1}{6}Q$ since $\frac{N-1}{6N} \to \frac{1}{6}$ as $N \to \infty$.

Earlier, we saw that, the true individual treatment effect variability, S_D^2, cannot be measured. Furthermore, the observed treatment effect variability is not unbiased for S_D^2. However, having established the formulas for the expected bias in both situations, in the next section, we propose an estimate the true individual treatment effect variability.

3.5: Estimate of the True Individual Treatment Response Variability

Consider the case when the subject assignment to treatment is independent. We had that, $E_\delta\left(\frac{N-1}{N}S_d^2\right) = S_D^2 + Bias_{IND}$. Thus, $S_D^2 = \frac{N-1}{N}E_\delta\left(S_d^2\right) - Bias_{IND}$. An estimate of S_D^2, denoted $\hat{S}_D^2$, is given as

$$\hat{S}_D^2 = \frac{N-1}{N}S_d^2 - \frac{N-1}{6N}\hat{Q}$$

where S_d^2 is the observed individual treatment response variance and

$$\hat{Q} = \left(2+\hat{\rho}_t\right)\hat{\sigma}_t^2 + 4\left(2+\hat{\rho}_\tau\right)\hat{\sigma}_\tau^2 + \left(\hat{\mu}_1+\hat{\mu}_2\right)^2 - \hat{\mu}_1\hat{\mu}_2 + 4\left(\hat{\mu}_3+\hat{\mu}_4\right)^2 - 4\hat{\mu}_3\hat{\mu}_4 + 2\left(\left(\hat{\mu}_1+\hat{\mu}_2\right)\left(\hat{\mu}_3+\hat{\mu}_4\right) + \hat{\mu}_1\hat{\mu}_4 + \hat{\mu}_2\hat{\mu}_3\right)$$

Remark 3.7

An estimate for the case of equal number of subjects per sequence is given as

$$\hat{S}_D^2 = \frac{N-1}{N} S_d{}^2 - \frac{1}{6}\hat{Q}$$

where

$$\hat{Q} = (2+\hat{\rho}_t)\hat{\sigma}_t^2 + 4(2+\hat{\rho}_\tau)\hat{\sigma}_\tau^2 + (\hat{\mu}_1+\hat{\mu}_2)^2 - \hat{\mu}_1\hat{\mu}_2 + 4(\hat{\mu}_3+\hat{\mu}_4)^2 - 4\hat{\mu}_3\hat{\mu}_4 + 2\left((\hat{\mu}_1+\hat{\mu}_2)(\hat{\mu}_3+\hat{\mu}_4) + \hat{\mu}_1\hat{\mu}_4 + \hat{\mu}_2\hat{\mu}_3\right)$$

In the next, we compare the estimated values with the actual parameters in an illustrated example. The illustrated example puts a normal distribution to the second population type stated above. That is, time effect parameters are given a bivariate normal distribution.

3.5.1: Illustrative Example 1: Estimating the Bias Term

As an example to compare the actual parameter values with the estimated values, we simulated two bivariate normal data of $\begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$ and $\begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix}$. Assume

$\begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 5.667 \\ -7.333 \end{pmatrix} \begin{pmatrix} 31.333 & -6.667 \\ -6.667 & 31.333 \end{pmatrix}\right)$ giving $\rho_t = -0.213$ and

$\begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 0.667 \\ -1.333 \end{pmatrix} \begin{pmatrix} 7.333 & -6.167 \\ -6.167 & 7.333 \end{pmatrix}\right)$ producing $\rho_\tau = -0.841$. The parameter values were taken from a previous simulation work on randomization and they are considered known. 1000 estimates (using the formulas above) were calculated from simulated data sets consisting of 300 subjects per sequence. Table 3.1 summarizes the findings.

Table 3.1**:** *Comparing True Values and Values from Estimators using 1000 simulated data of size 300*

Population Parameter	True Pop. Value	*Estimated Mean	*Estimated Std dev.
Mean of t_1	5.667	5.663	0.387
Mean of t_2	-7.333	-7.335	0.272
Mean of τ_1	0.667	0.668	0.205
Mean of τ_2	-1.333	-1.334	0.113
Variance of t	31.333	31.340	1.820
Variance of τ	7.333	7.333	0.507
Covariance (t_1, t_2)	-6.667	-6.646	2.134
Covariance (τ_1, τ_2)	-6.167	-6.162	0.630
Correlation (t_1, t_2)	-0.213	-0.211	0.066
Correlation (τ_1, τ_2)	-0.841	-0.839	0.039
Expected bias – Indep.	28.191	28.270	1.602
Expected bias – Dep.	28.163	28.300	1.604

*Estimated mean is the mean of 1000 estimates obtained from simulated data with 300 subjects per sequence.
*Estimated Stdev. is standard deviation of 1000 estimates obtained from simulated data.

Notice the closeness between the estimates and the actual values. In addition, the standard errors of the estimates are small. The graphs below further explore the estimated bias (for the case of equal subjects per sequence) with increasing sample sizes. Increasing the sample size reduces the difference.

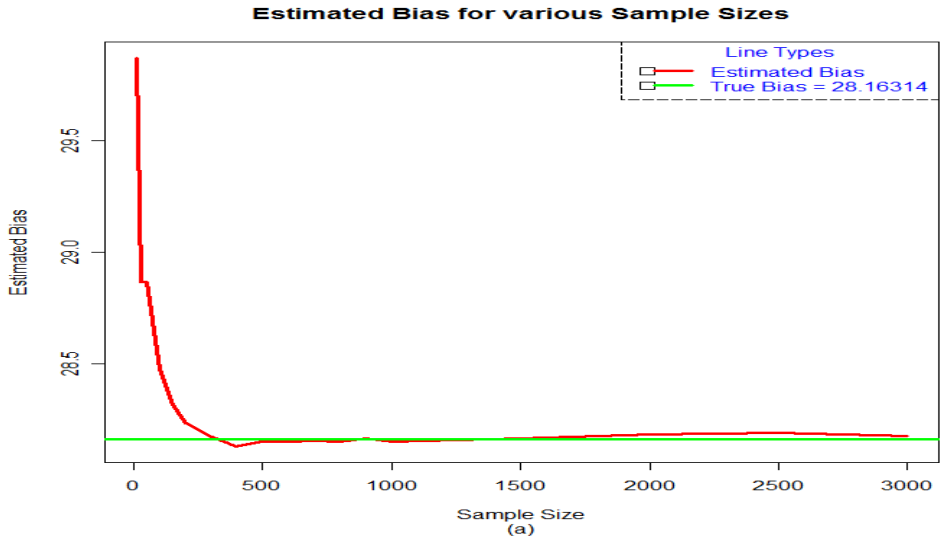


Figure 3.1: *Graph of estimated bias for various sample sizes*

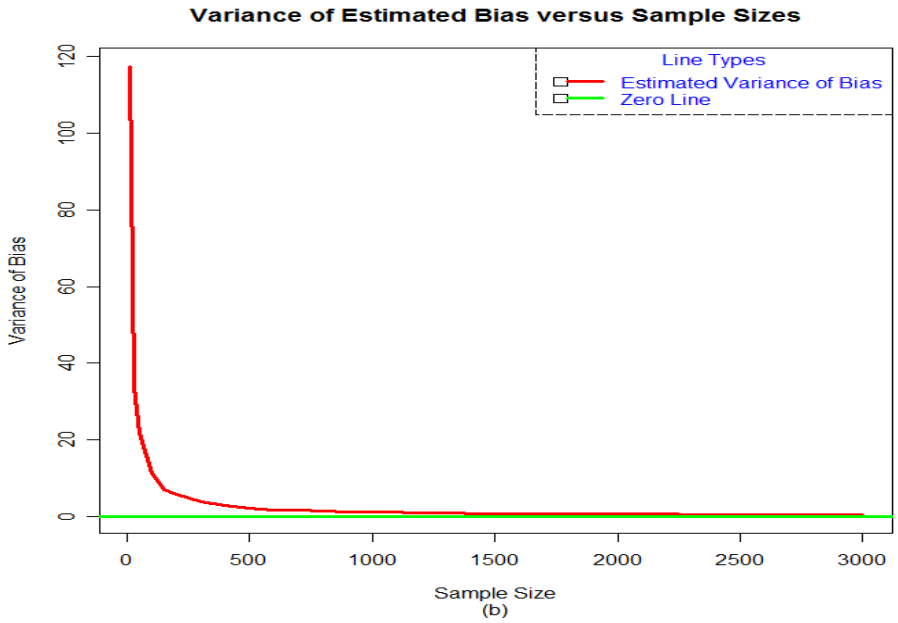


Figure 3.2: *Graph showing the variance of the estimated bias for various sample sizes*

We note that, for increasing sample sizes, the estimated bias approaches the true bias value (a). In addition, the variance of the estimated bias approaches zero with increasing sample size (b). This illustrates consistency of the bias estimator.

Recall that, we analyzed *Squares 1* and *2* sequences separately. In the following Section 3.6, we combine the two analyses and propose an estimate of the overall true individual treatment effect variability, which, as expected, consists of the estimates from *Squares 1* and *2*. A note on symbols used here: Estimates with subscripts "1" indicate that they were derived using *Square 1* sequences; likewise, those with subscripts "2" were are derived using *Square 2* sequences. The estimates from the combined sequences have subscripts "12."

3.6: Generalization to all Six Sequences

It is important to note that, up to this point, we focused on just the three sequences of *Square 1*. Let $N_1, N_2, \bar{D}_1, \bar{D}_2, S^2_{D_1}$ and $S^2_{D_2}$ denote: the sample sizes, the true finite population mean effects and variances of the effects from *Squares 1* and *2,* respectively. Let $S^2_{D_{12}}$ denote the true individual treatment effect variability from the two *Squares*, we have that, $\bar{D}_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} D_i$, $\bar{D}_2 = \frac{1}{N_2}\sum_{i=1}^{N_2} D_i$. We define $S^2_{D_{12}} = \frac{1}{N_1+N_2}\left[\sum_{i=1}^{N_1+N_2}\left(D_i - \bar{D}_{12}\right)^2\right]$ where

$\bar{D}_{12} = \frac{1}{N_1+N_2}\sum_{i=1}^{N_1+N_2} D_i = \frac{1}{N_1+N_2}\left(N_1\bar{D}_1 + N_2\bar{D}_2\right)$. Hence, we have

$$S^2_{D_{12}} = \frac{1}{N_1+N_2}\left[N_1 S^2_{D_1} + N_2 S^2_{D_2} + N_1\bar{D}_1^2 + N_2\bar{D}_2^2 - \frac{1}{N_1+N_2}\left(N_1\bar{D}_1 + N_2\bar{D}_2\right)^2\right]$$

We state the plug-in estimate of $S^2_{D_{12}}$ as,

$$\hat{S}^2_{D12} = \frac{1}{N_1 + N_2}\left[N_1\hat{S}^2_{D_1} + N_2\hat{S}^2_{D_2} + N_1\bar{d}_1^{\,2} + N_2\bar{d}_2^{\,2} - \frac{1}{N_1 + N_2}\left(N_1\bar{d}_1 + N_2\bar{d}_2\right)^2 \right] \quad (1.16)$$

If $N_1 = N_2 = N_{12}$ then $\bar{D}_{12} = \frac{1}{2N_{12}}\sum_{i=1}^{2N_{12}} D_i = \frac{1}{2}\left(\bar{D}_1 + \bar{D}_2\right)$, $\bar{d}_{12} = \frac{1}{2}\left(\bar{d}_1 + \bar{d}_2\right)$

$S^2_{D12} = \frac{1}{2}\left[\left(S^2_{D_1} + S^2_{D_2}\right) + \frac{1}{2}\left(\bar{D}_1 - \bar{D}_2\right)^2\right]$ and $\hat{S}^2_{D12} = \frac{1}{2}\left[\left(\hat{S}^2_{D_1} + \hat{S}^2_{D_2}\right) + \frac{1}{2}\left(\bar{d}_1 - \bar{d}_2\right)^2\right]$

3.7: Illustrative Example 2: Simulated Blood Pressure Dataset

The following example is based on equal number of subjects per sequence. Blood pressure (a.k.a. arterial pressure) is the force of circulating blood on the walls of blood vessel. Blood pressure is one of the four vital signs sensitive to periodic changes and large individual variations. The other three are body temperature, pulse or heart rate and respiratory rate. Blood pressure can be systolic or diastolic in nature. For healthy, resting human adults, normal blood pressure ranges from a systolic level less than120mmHg (millimeter mercury) to a diastolic level less than 80mmHg, often written as 120/80mmHg. In this example, we simulate blood pressure (BP in mmHg) data for 1800 patients who received two treatments T and C for high blood pressure related disease at different time periods. The BP values are assumed to come from a normal distribution. 300 subjects as randomly assigned to each of the six sequences. The Table 3.2 below shows the result of the randomization reported in the potential outcome format. All values are in millimeters mercury (mmHg). The light grey shaded cells show the observed values following assignment to the three sequences in *Square 1*.

This example brings into focus our third population: A super – population distribution of treatment effects. Initially, we generate bivariate normal blood pressure responses (X,Y) to two treatments T and C. In addition, we simulated a bivariate normal time effects under three different time periods, the third being determined by the other two. Both treatment responses were simulated for each of the three periods. Hence, it was possible to compute the actual parameter values and compare with estimates derived using the observed data in *Squares 1* and *2*. Section 3.7.1 below details the simulated example.

3.7.1: Detailed Distributional Specifications for Blood Pressure Date

In this section, we provide the details of the distributional assumptions that produced the blood pressure data, summarized on Table 3.2. The distributions used here illustrate the list of distributions mentioned in Section 1.4.

First, we assumed a pair of blood pressure treatment response variables (X,Y) have an independent and identically distributed bivariate normal distribution with means $\mu_X = 100mmHg$ and $\mu_Y = 90mmHg$, variances $\sigma_X^2 = 9mmHg^2$ and $\sigma_Y^2 = 5mmHg^2$, respectively. Let $Cov(X,Y) = 2$. That is, $\begin{pmatrix} X \\ Y \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 100 \\ 90 \end{pmatrix} \begin{pmatrix} 9 & 2 \\ 2 & 5 \end{pmatrix}\right)$. From the distribution of (X,Y), we define a "super-population" distribution of individual treatment effects $D = X - Y$ with parameters $\mu_D = E(X-Y)$ and $\sigma_D^2 = Var(X-Y)$. That is, $D \overset{i.i.d}{\sim} N(10,10)$.

Second, from the super-population, we draw a finite random sample of 1800 values with finite population mean $\bar{D} = 9.88 mmHg$ and finite population variance $S_D^2 = 10.62 mmHg^2$. We note that, the quantitative section of this dissertation in centered on estimating S_D^2.

Third, we let the time effects be distributed as such: $\begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 5 \\ 7 \end{pmatrix} \begin{pmatrix} 8 & 1 \\ 1 & 8 \end{pmatrix}\right)$ and $\begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix} \begin{pmatrix} 5 & 1 \\ 1 & 5 \end{pmatrix}\right)$. Using the values of (X, Y) and the time effects, we found the potential outcomes for 1800 subjects and the observed outcomes when 300 subjects are assigned to each of the six sequences in *Squares* 1 and 2. The resulting data for *Squares* 1 is shown on Table 3.2 below.

Table 3.2: *Simulated data of Blood Pressure (BP in mmHg) for 900 subjects with 300 per sequence. The light- grey shaded cells are observed data values for corresponding to the sequences.*

			Time Period 1		Time Period 2		Time Period 3	
			$x+t_1$	$y+\tau_1$	$x+t_2$	$y+\tau_2$	$x+t_3$	$y+\tau_3$
Square	Seq.	Subject						
		1	109	92	103	89	85	79
		2	97	89	101	87	88	88
		3	98	90	105	90	86	90
		*	*	*	*	*	*	*
	TTC	*	*	*	*	*	*	*
		298	103	87	100	89	101	83
		299	99	93	108	91	104	92
Square 1		300	106	92	109	92	99	83
		301	101	89	104	91	85	92
		302	106	89	110	90	90	83
		303	110	93	110	91	83	89
		*	*	*	*	*	*	*
	TCT	*	*	*	*	*	*	*
		598	98	94	105	93	86	83
		599	107	85	105	84	96	92
		600	108	89	113	94	82	90
		601	110	90	113	93	89	88
		602	103	85	108	88	94	90
		603	106	91	107	95	83	77
		*	*	*	*	*	*	*
	CTT	*	*	*	*	*	*	*
		898	106	90	107	95	87	81
		899	107	92	108	88	85	84
		900	103	88	101	88	98	94

Using the potential outcome method, the true finite population average is $\bar{D}_1 = 9.796\text{mmHg}$ with point estimate of $\bar{d}_1 = 9.82\text{mmHg}$ for *Square 1* data. For the randomizations in *Square 2*, the true finite population mean is $\bar{D}_2 = 9.958\text{mmHg}$ with estimate given as $\bar{d}_2 = 9.871\text{mmHg}$. For the combined dataset of six sequences, the true finite population average is $\bar{D}_{12} = 9.88\text{mmHg}$ with point estimate of $\bar{d}_{12} = 9.85\text{mmHg}$. These

estimates give the average increase in blood pressure due to treatment T relative to treatment C. These are the finite population mean estimates upon which decisions about the treatment efficacy are sometimes based with less consideration of the individual treatment response effect variability. Furthermore, the true standard deviation of the individual treatment effect for the BP data and its estimate are $S_{D_{12}} = 3.258\text{mmHg}$ and $\hat{S}_{D_{12}} = 3.692\text{mmHg}$, respectively. The estimated coefficient of variation is 0.375 compared to the actual value of 0.330 for the final population. The coefficient of variation represents the proportion of the mean treatment effect to the standard deviation of the effects.

The overall actual finite population mean effect, $\bar{D}_{12} = 9.88\text{mmHg}$, is positive and the standard deviation is a 37.482% "fraction" of the mean. If it is important that the treatment produces a positive effect for most subjects, then the fraction should be small; otherwise the fraction could be bigger in favor of applying the treatment to a large population (Longford, 1999).

3.8: A Probability of Negative treatment effect

Let P_{-} denote the probability that a subject will experience a "negative" individual treatment effect. Suppose the focus is on the effect of treatment T say, then, negative treatment effect means $D = X - Y < \lambda$, where $\lambda \geq 0$ is a threshold value. That is, for those individual, C is more effective relative to T. Assume that D has a normal distribution with mean μ_D and variance σ_D^2. Note that the distribution of D is determined by the distribution of response variables X and Y which, in this case, are assumed normal.

Estimate P_- as

$$\begin{aligned} P_- &= P(D < \lambda) \\ &= P\left(Z_D < \frac{\lambda - \mu_D}{\sigma_D} \right) \\ &= \Phi\left(\frac{\lambda - \mu_D}{\sigma_D} \right) \end{aligned}$$

where $\Phi\left(\frac{\lambda - \mu_D}{\sigma_D} \right)$ is the cumulative standard normal distribution function evaluated at $\frac{\lambda - \mu_D}{\sigma_D}$. Assuming the finite population is large and representative enough, we use the finite population mean, $\bar{d}_{12}$, and variance, $\hat{S}_{D_{12}}$, as estimates of μ_D and σ_D^2, respectively. That is, $\hat{\mu}_D = \bar{d}_{12}$ and $\hat{\sigma}_D = \hat{S}_{D_{12}}$. Thus, we estimate P_- by $\hat{P}_- = \Phi\left(\frac{\lambda - \bar{d}_{12}}{\hat{S}_{D_{12}}} \right)$.

3.8.1: Illustrative Example 3: Probability of Negative Effect

Continuing with the analysis of the blood pressure data with $\lambda = 0,$ suppose we wish to estimate the probability that the true effect of T is less effective in treating hypertension than that of C, that is, $P(X < Y) = P(D < 0)$. Using the potential treatment method, the estimated probability is given as

$$\hat{P}_- = P(\overset{\Lambda}{D} < 0) = \Phi\left(\frac{0 - 9.85}{3.692} \right) = \Phi(-2.67) = 0.0038, \quad \text{where} \quad \lambda = 0,\ \hat{\mu}_D = \bar{d}_{12} = 9.85 \text{ and}$$

$\hat{\sigma}_D = \hat{S}_{D_{12}} = 3.692$. Thus, a randomly selected individual has an estimated probability of 0.0038

of being better off on treatment C relative to treatment T. C is estimated to be more effective than T for at most 0.38% of the population. If there is a tolerance or threshold probability level, we may then decide whether of not T is superior to C.

3.9: Repeated Measure Analyses and Potential Outcome Method

The Grizzle (1965, 1974) model for two treatments - two periods crossover design could be extended to higher design. Here we extend it to the 2 treatments, 3 periods crossover design without carryover. With more than one subject per sequence, the general model for the treatment response variable Y with random subjects within sequence specification – modified form of *Cross-Over Experiment* by Ratkowski et al. (1993, page 60) and Jones and Kenward (1989) – can be written as

$$y_{ijk} = \mu + \eta_{i(k)} + \pi_j + \theta_t + (S\theta)_{it} + \varepsilon_{ijk} \tag{1.17}$$

with

$$E\left(y_{ijk}\right) = \mu_{ij} = \mu + \pi_j + \theta_t$$

where,

y_{ijk} is the observed response for the i^{th} subject in the j^{th} period of the k^{th} sequence

μ_{ij} is the true mean response for the i^{th} subject in the j^{th} period

μ = an overall mean effect

$\eta_{i(k)}$ = the random effect due to the i^{th} subject in the k^{th} sequence; $k = 1,2,3,...,6$;

$i = 1,2,3,...,n_k$, n_k being the number of subjects per sequence

$\pi_j =$ the period effect, $j = 1, 2, 3$

$\theta_t =$ the effect of treatment t

$(\eta\theta)_{it} =$ the random effect of interaction between the i^{th} subject and the t^{th} treatment.

$\varepsilon_{ijk} =$ the random experimental error effect of the i^{th} subject in period j of sequence k^{th}.

Assume $\eta_{i(k)} \overset{iid}{\sim} N\left(0, \sigma_s^2\right)$, $(\eta\theta)_{it} \overset{iid}{\sim} N\left(0, \sigma_{s\theta}^2\right)$ and $\varepsilon_{ijk} \overset{iid}{\sim} N\left(0, \sigma_w^2\right)$. Observed values at the different periods and sequences *(in Square 1)* are listed on Table 3.3 below

Table 3.3: *Observed data for GLM model and the potential outcome values for the sequences in Square 1*

	Time Period 1		Time Period 2		Time Period 3	
Sequences	GLM	POT. OUT	GLM	POT. OUT	GLM	POT. OUT
TTC	y_{i11}	$(X_i + t_{i1})$	y_{i21}	$(X_i + t_{i2})$	y_{i31}	$(Y_i + \tau_{i3})$
TCT	y_{i12}	$(X_i + t_{i1})$	y_{i22}	$(Y_i + \tau_{i2})$	y_{i32}	$(X_i + t_{i3})$
CTT	y_{i13}	$(Y_i + \tau_{i1})$	y_{i23}	$(X_i + t_{i2})$	y_{i33}	$(X_i + t_{i3})$

Using the model (1.17) we have that,

Sequence TTC:

$$y_{i11} = \mu + \eta_{i(1)} + \pi_1 + \theta_1 + (\eta\theta)_{i1} + \varepsilon_{i11}$$

$$y_{i21} = \mu + \eta_{i(1)} + \pi_2 + \theta_1 + (\eta\theta)_{i1} + \varepsilon_{i21}$$

$$y_{i31} = \mu + \eta_{i(1)} + \pi_3 + \theta_2 + (\eta\theta)_{i2} + \varepsilon_{i31}$$

Sequence TCT:

$$y_{i12} = \mu + \eta_{i(2)} + \pi_1 + \theta_1 + (\eta\theta)_{i1} + \varepsilon_{i12}$$

$$y_{i22} = \mu + \eta_{i(2)} + \pi_2 + \theta_2 + (\eta\theta)_{i2} + \varepsilon_{i22}$$

$$y_{i32} = \mu + \eta_{i(2)} + \pi_3 + \theta_1 + (\eta\theta)_{i1} + \varepsilon_{i32}$$

Sequence CTT:

$$y_{i13} = \mu + \eta_{i(3)} + \pi_1 + \theta_2 + (\eta\theta)_{i2} + \varepsilon_{i13}$$

$$y_{i23} = \mu + \eta_{i(3)} + \pi_2 + \theta_1 + (\eta\theta)_{i1} + \varepsilon_{i23}$$

$$y_{i33} = \mu + \eta_{i(3)} + \pi_3 + \theta_1 + (\eta\theta)_{i1} + \varepsilon_{i33}$$

Proposition 3.7

Given the model (1.17) and the potential outcomes observed data, the difference between the least square mean of the treatment effects is an unbiased estimate of the mean of the observed treatment effect (hence, an unbiased estimate of the true mean treatment effect). Let n_1, n_2 and n_3 be the number of subjects in sequences TTC, TCT and CTT respectively. We assume $n_1 = n_2 = n_3 = n$. Define $\hat{\theta}_1$ and $\hat{\theta}_2$ as

$$\hat{\theta}_1 = \frac{1}{3}\left[\frac{\sum_{i=1}^{n_1}(y_{i11} + y_{i21})}{2n_1} + \frac{\sum_{i=1}^{n_2}(y_{i12} + y_{i32})}{2n_2} + \frac{\sum_{i=1}^{n_3}(y_{i23} + y_{i33})}{2n_3}\right] \tag{1.18}$$

and

$$\hat{\theta}_2 = \frac{1}{3}\left[\frac{\sum_{i=1}^{n_1} y_{i31}}{n_1} + \frac{\sum_{i=1}^{n_2} y_{i22}}{n_2} + \frac{\sum_{i=1}^{n_3} y_{i13}}{n_3}\right] \tag{1.19}$$

then,

$$\hat{\theta}_1 - \hat{\theta}_2 = \bar{d}$$

where $\bar{d}$ is the observed mean treatment effect from the potential outcome method.

Proof:

$$\hat{\theta}_1 = \frac{1}{3}\left[\frac{\sum_{i=1}^{n_1}(y_{i11} + y_{i21})}{2n_1} + \frac{\sum_{i=1}^{n_2}(y_{i12} + y_{i32})}{2n_2} + \frac{\sum_{i=1}^{n_3}(y_{i23} + y_{i33})}{2n_3}\right]$$

$$\Rightarrow \hat{\theta}_1 = \frac{1}{3}\left[\frac{\sum_{i=1}^{n_1}\left((X_i + t_{i1}) + (X_i + t_{i2})\right)}{2n_1} + \frac{\sum_{i'=1}^{n_2}\left((X_i + t_{i1}) + (X_i + t_{i3})\right)}{2n_2} + \frac{\sum_{i''=1}^{n_3}\left((X_i + t_{i2}) + (X_i + t_{i3})\right)}{2n_3}\right]$$

$$\Rightarrow \hat{\theta}_1 = \frac{1}{3}\left[\frac{\sum_{i=1}^{n_1}(2X_i + t_{i2} + t_{i1})}{2n_1} + \frac{\sum_{i=1}^{n_2}(2X_i + t_{i1} + t_{i3})}{2n_2} + \frac{\sum_{i=1}^{n_3}(2X_i + t_{i2} + t_{i3})}{2n_3}\right].$$

Similarly,

$$\hat{\theta}_2 = \frac{1}{3}\left[\frac{\sum_{i=1}^{n_1}(Y_i + \tau_{i3})}{n_1} + \frac{\sum_{i=1}^{n_2}(Y_i + \tau_{i2})}{n_2} + \frac{\sum_{i=1}^{n_3}(Y_i + \tau_{i1})}{n_3}\right].$$

Hence,

$$\hat{\theta}_1-\hat{\theta}_2=\frac{1}{3}\left[\frac{1}{n_1}\sum_{i=1}^{n_1}\left(\frac{2X_i+t_{i2}+t_{i1}}{2}-\left(Y_i+\tau_{i3}\right)\right)+\frac{1}{n_2}\sum_{i=1}^{n_2}\left(\frac{2X_i+t_{i1}+t_{i3}}{2}-\left(Y_i+\tau_{i2}\right)\right)+\frac{1}{n_3}\sum_{i=1}^{n_3}\left(\frac{2X_i+t_{i2}+t_{i3}}{2}-\left(Y_i+\tau_{i1}\right)\right)\right]$$

$$\hat{\theta}_1-\hat{\theta}_2=\frac{1}{3}\left[\frac{1}{n_1}\sum_{i=1}^{n_1}\left(\left(X_i-Y_i\right)+\frac{t_{i2}+t_{i1}}{2}-\tau_{i3}\right)+\frac{1}{n_2}\sum_{i=1}^{n_2}\left(\left(X_i-Y_i\right)+\frac{t_{i1}+t_{i3}}{2}-\tau_{i2}\right)+\frac{1}{n_3}\sum_{i=1}^{n_3}\left(\left(X_i-Y_i\right)+\frac{t_{i2}+t_{i3}}{2}-\tau_{i1}\right)\right]$$

Now, $n_1=n_2=n_3=n$

$$\hat{\theta}_1-\hat{\theta}_2=\frac{1}{3}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\left(X_i-Y_i\right)+\frac{t_{i2}+t_{i1}}{2}-\tau_{i3}\right)\delta_{i3}+\frac{1}{n}\sum_{i=1}^{n}\left(\left(X_i-Y_i\right)+\frac{t_{i1}+t_{i3}}{2}-\tau_{i2}\right)\delta_{i2}+\frac{1}{n}\sum_{i=1}^{n}\left(\left(X_i-Y_i\right)+\frac{t_{i2}+t_{i3}}{2}-\tau_{i1}\right)\delta_{i1}\right]$$

$$\Rightarrow\hat{\theta}_1-\hat{\theta}_2=\frac{1}{n}\left(\sum_{i=1}^{n}\left(X_i-Y_i\right)-\sum_{i=1}^{n}\sum_{j=1}^{3}\left(\frac{t_{ij}}{2}+\tau_{ij}\right)\delta_{ij}\right).$$

And from (1.3),

$$\hat{\theta}_1-\hat{\theta}_2=\frac{1}{n}\sum_{i=1}^{n}d_i=\bar{d}$$ ■

Remark 3.5

A similar correspondence between estimates of time effects in a repeated measure framework and the potential outcome framework was also noted and shown in a numerical illustration.

3.9.1: Illustrative Example 4: Comparing Potential Outcome with Repeated Measures Estimates

We continue with *Example 2* with application of the generalized mixed model (1.17) with repeated measures. We obtain $\bar{d}_1 = 9.82\text{mmHg}$, $\bar{d}_2 = 9.87\text{mmHg}$ with standard errors of 0.1722 and 0.2245 for the randomization in *Squares 1* and *2* respectively. Furthermore, for the combined data, we have $\bar{d}_{12} = 9.88\text{mmHg}$ with a standard error of 0.1581. Thus, as stated in *Proposition 3.7*, the GLM estimates are equal to the potential outcomes estimates. That is, we may think of the PROC GLM or PROC MIXED outputs as estimates of the true population mean treatment effects. Furthermore, from the GLM output, the p-value is less than 0.0001, hence, there is evidence that $\mu_X > \mu_Y$.

We also compared the estimates of the linear combination of the time effects parameters as explained in *Corollary 3.5*. Using the potential outcome method on subjects in the first and third sequences of *Square 1,* we have estimates: $t_1 \overset{\wedge}{-} t_2 = -2.127$, $t_1 \overset{\wedge}{-} t_3 = 16.650$ and $t_2 \overset{\wedge}{-} t_3 = 2t_2 \overset{\wedge}{+} t_1 = 18.777$. Similarly, for *Square 2* we have estimates: $\tau_1 \overset{\wedge}{-} \tau_2 = -1.043$, $\tau_1 \overset{\wedge}{-} \tau_3 = 3.693$ and $\tau_2 \overset{\wedge}{-} \tau_3 = 4.737$.

Using all the sequences in square 1, the estimates of the period contrasts produced by GLM are $\hat{\pi}_1 - \hat{\pi}_2 = -2.124$, $\hat{\pi}_1 - \hat{\pi}_3 = 16.654$ and $\hat{\pi}_2 - \hat{\pi}_3 = 18.779$ with standard error 0.2812 in all cases. For *Square 2*, SAS Proc GLM gives estimates: $\hat{\pi}_1 - \hat{\pi}_2 = -0.932$, $\hat{\pi}_1 - \hat{\pi}_3 = 3.916$ and $\hat{\pi}_2 - \hat{\pi}_3 = 4.848$, all, with standard error 0.23358.

In this chapter and chapter 2, we estimated the true individual treatment effect variability and the probability of negative effects for quantitative response variable using the method of potential outcomes. The analyses employed the two treatments, two periods and the two treatments, three periods randomizations. Furthermore, we compared the potential outcome estimates with the usual repeated measures estimates gotten using GLM. In the following chapter, we extend our analyses to qualitative response, precisely, the binary response variables. We limit the analyses to two treatments, two periods TC – CT and the TC – CT – TT – CC designs. For these designs, we will estimate the average treatment effect and the probability of negative effect – a component that implicitly reflects the individual treatment effect variability. Earlier, we saw that some work had been done these designs by Gadbury et al. (2004). However, their analyses assumed "exchangeability." Initially, we relax the exchangeability assumption and later consider it. We show that, when the exchangeability assumption is factored into our analyses, the "without – exchangeability" analyses boils down to the analyses presented in the paper Gadbury et al. (2004).

CHAPTER 4 - Analysis Using Binary Data

4.0: Two Treatments, Two Periods with Binary Outcome

In this chapter, we will base our analyses on two treatments, two periods with a binary response. The focus will be on the designs: TC – CT and TC – CT – TT – CC. The first part of the analyses will deal with the TC – CT crossover design for which we will estimate the true average treatment effect and construct bounds for an inestimable "probability of negative effect." The second part involves the design, TC – CT – TT – CC. That is, some subjects will stay on the same treatment over the two periods. Observed responses from these subjects will enable us to construct tighter bounds for the probability of a negative effect. The expression of "probability of negative effect" carries a connotation of an "unexplained individual treatment variation" (Gadbury et al, 2004). An example will be used for illustrations.

4.1: A Two Treatment Design with Binary Response: One Time Point

Let X and Y represent the response to treatments T and C respectively. We let "1" denote "success" and "0" denote "failure." In addition, let (X,Y) be a set of bivariate discrete potential outcomes from an infinite population of outcomes. The joint discrete probability distribution of (X,Y), $P(X = x,\ Y = y)$, is given as on Table 4.1 below (Gadbury et al., 2004).

Table 4.1: *Joint Probability distribution of* (x, y)

(x, y)	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$	Total
$P(X = x, Y = y)$	ω_1	ω_2	ω_3	ω_4	$\sum_{i=1}^{4} \omega_i = 1$

where (x, y) are the observed outcomes of (X, Y) and $P(X = x, Y = y) = \omega_i$, $(i = 1,2,3,4)$, is the true probability of (x, y) for an individual at a specific time. Since only one of either X or Y is measured at a specific time (the fundamental problem of causal inference), $\omega_i \in [0,1]$ cannot be directly estimated for $i = 1,2,3,4$ separately.

As before, define $D = X - Y$ as the true treatment (causal) effect. That is, D expresses the actual effect of T relative to C and note that D is not observable. Let $P(D)$ denote the probability of D. Note that $P(D)$ is a discrete probability distribution. Possible values of D and the associated probabilities are listed in Table 4.2 and 4.3 below

Table 4.2: *Possible values of* D

(x, y)	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
$D = X - Y$	0	-1	1	0

For example, $D = -1$ means treatment T leads to an unfavorable effect(failure) relative to treatment C at a particular time.

Table 4.3: *Probability of observing* D

D	-1	0	1
$P(D)$	ω_2	$\omega_1 + \omega_4$	ω_3

Remark 4.1.1

Let $p_1 = E(X)$ and $p_2 = E(Y)$ be the mean of the marginal distributions of X and Y, and note that these are estimable. Then from Table 4.3, $p_1 = P(X = 1) = \omega_3 + \omega_4$ and $p_2 = E(Y = 1) = \omega_2 + \omega_4$. Denote the true average individual treatment effect of T relative to C as $E(D)$. Then,

$$E(D) = \omega_3 - \omega_2 = p_1 - p_2 .$$

Note that $E(D)$ represents the true mean treatment effect of T relative to C at a particular time.

For example, $E(D) = 0.6$ could mean several things

1). $\omega_3 = 0.60$ *and* $\omega_2 = 0$: That is, 60% of the patients will succeed on T but fail on C and the remaining 40% will either succeed on both T and C or fail on both T and C.

2). $\omega_3 = 0.65$ *and* $\omega_2 = 0.05$: That is, 65% of the patients will succeed on T and fail on C, 5% will succeed on C and fail on T and 30% will either succeed on both T and C or fail on both T and C.

3). $\omega_3 = 0.80$ *and* $\omega_2 = 0.20$: 80% of the patients will succeed on T and fail on C, 20% will succeed on C and fail on T and 0% will either succeed on both T and C or fail on both T and C.

So, if the average treatment effect probability equals 0.60 does not immediately imply C is completely ineffective as (1) may suggest. Notice that (3) indicates 20% responded well on C and fail on T.

Remark 4.1.2

The variance of the true individual treatment effect is given as

$$\begin{aligned} Var(D) &= E(D^2) - E(D)^2 \\ &= \omega_2 + \omega_3 - (\omega_3 - \omega_2)^2 \\ &= \omega_2(1-\omega_2) + \omega_3(1-\omega_3) + 2\omega_2\omega_3 \end{aligned}$$

$Var(D)$ cannot be estimated because D cannot be observed. In this chapter, less focus will be placed on $Var(D)$ although we will establish bounds for it. Furthermore, we move the analysis of $Var(D)$ to the appendix section of this dissertation. Instead, we will focus on estimating the probability of negative individual treatment effect, $P(D<0)$ – a component that results from the variability of the individual effect – and the average individual treatment effect of T relative to C. In addition, the variance of a discrete distribution, usually, is a function of the mean. These make $Var(D)$ difficult to interpret.

In the next section, we present the two treatments, two periods TC – CT crossover design for binary outcomes. Results will be outlined with and without the assumptions of "exchangeability."

4.2: Two Treatments Two Periods: Potential Outcomes and True Probabilities

The prior section established the potential outcomes framework for a particular time point. This section considers two time points periods, so potential outcome variables are in four dimensions with (X_1, Y_1) for time point *Period 1* and (X_2, Y_2) for time point *Period 2*. Thus,

there is a true individual treatment effects for both periods, given by $D_1 = X_1 - Y_1$ and $D_2 = X_2 - Y_2$. It will be assumed that the bivariate marginal distribution for each time period will be given as in Table 4.4. However, in the 4 – dimensional joint distribution, exchangeability of bivariate outcomes may not hold due to time effects. Initially, it is assumed that exchangeability does not hold. As stated above, suppose there are two treatments T and C with binary response, Table 4.4 below shows a constructed distribution of all possible potential outcomes for a population.

Table 4.4: *Potential outcomes framework and probabilities of two treatments, two periods crossover design*

		Period 2				
	For an Individual	$(x_2, y_2) = (0,0)^{T\ C}$	$(x_2, y_2) = (0,1)^{T\ C}$	$(x_2, y_2) = (1,0)^{T\ C}$	$(x_2, y_2) = (1,1)^{T\ C}$	Marginal Total
Per.1	$(x_1, y_1) = (0,0)^{T\ C}$	β_{11}	β_{12}	β_{13}	β_{14}	ω_1
	$(x_1, y_1) = (0,1)^{T\ C}$	β_{21}	β_{22}	β_{23}	β_{24}	ω_2
	$(x_1, y_1) = (1,0)^{T\ C}$	β_{31}	β_{32}	β_{33}	β_{34}	ω_3
	$(x_1, y_1) = (1,1)^{T\ C}$	β_{41}	β_{42}	β_{43}	β_{44}	ω_4
	Marginal Total	ω_1	ω_2	ω_3	ω_4	$\sum \omega = 1$

From Table 4.4, we note that marginal distributions are assumed equal. That is, $P(X_1 = x_i, Y_1 = y_i) = \omega_i = P(X_2 = x_i, Y_2 = y_i)$ for $i = 1,2,3,4$. Furthermore, $\beta_{ij} = P\left[(X_1, Y_1) = (x_i, y_i), (X_2, Y_2) = (x_j, y_j)\right]$ for $i, j = 1,2,3,4$ is the true joint probability of (X_1, Y_1) outcomes in *Period 1* and (X_2, Y_2) outcomes in *Period 2* (as on Table 4.4). Note here that, "i" goes with the row outcomes in *Period 1* while "j" is associated with the columns

outcomes in *Period 2*. So, β_{ij} is the actual probability of the i^{th} outcome in *Period 1* and the j^{th} outcome in *Period 2*, as given on Table 4.4. In addition, $\beta_{ij} \in [0,1]$ is inestimable for $i, j = 1,2,3,4$, since we cannot observed both (X_1, Y_1) and (X_2, Y_2), simultaneously. For example $\beta_{11} = P[(X_1, Y_1) = (0,0), (X_2, Y_2) = (0,0)]$ is the true probability that both treatments are ineffective in periods 1 and 2 and $\beta_{23} = P[(X_1, Y_1) = (0,1), (X_2, Y_2) = (1,0)]$ is the true probability of succeeding on C and failing on T in *Period 1* and succeeding on T and on failing C in *Period 2*. These probabilities cannot be estimated.

Remark 4.2.1

Additional remarks about Table 4.4:

1). The true probability of potential outcomes at one period, conditioned on the outcomes at another period are the same, regardless of which period is conditioned, as long as the outcomes are the same at each period. That is,

$$P[(X_2, Y_2) = (x_i, y_i) \mid (X_1, Y_1) = (x_i, y_i)] = \frac{\beta_{ii}}{\omega_i} = P[(X_1, Y_1) = (x_i, y_i) \mid (X_2, Y_2) = (x_i, y_i)],$$

for $i = 1,2,3,4$.

2). Exchangeability occurs when

$$\beta_{ij} = P[(X_1, Y_1) = (x_i, y_i), (X_2, Y_2) = (x_j, y_j)] = P[(X_2, Y_2) = (x_i, y_i), (X_1, Y_1) = (x_j, y_j)] = \beta_{ji}$$

We will reserve the detailed look at the notion of exchangeability for the later part of the dissertation. However, it is worthwhile noting that, for $i, j = 1,2,3,4$:

$$P\left[\left(X_2,Y_2\right)=\left(x_j,y_j\right)\;\middle|\;\left(X_1,Y_1\right)=\left(x_i,y_i\right)\right]\neq P\left[\left(X_1,Y_1\right)=\left(x_i,y_i\right)\;\middle|\;\left(X_2,Y_2\right)=\left(x_j,y_j\right)\right],$$

unless $i=j$.

This difference in probabilities suggests a possible time period effect or a dependency of outcomes in one period on outcomes in the other. Notice that with the assumption of exchangeability, the conditional probability effect is same regardless of the time period the treatment is administered.

Marginal Probability

The distribution of Table 4.4 has been constructed such that, even in the absence of exchangeability, the row and column probabilities for an individual in a given period sum to ω_i as presented on the Table 4.4. That is, $\sum_{j=1}^{4}\beta_{ij}=\omega_i=\sum_{i=1}^{4}\beta_{ij}$ and $\sum_{i=1}^{4}\omega_i=1$.

Remark 4.2.2

Having defined the distribution on Table 4.4, in the following sections, we use this distribution to study our observed data from the TC – CT (Section 4.3) and the TC – CT – TT – CC (Section 4.6) designs. We will express the probabilities of observing a particular outcome as a function of the actual probabilities on Table 4.4. Estimates of these estimable probabilities of observing an outcome will be found. In the first part of the analyses, we assume exchangeability condition does not hold (Section 4.5, 4.7). Later, we assume that it does hold and show that the analyses become that presented in Gadbury et al. (2004).

4.3: Potential Outcomes and Probabilities of Observed Outcomes for the TC – CT Design

Consider the two treatments, two periods crossover design. The observed data will be of the form

		Period 1	2
Sequence	1	T	C
	2	C	T

The following schematic diagram illustrates the possible observed outcomes under the TC – CT design:

Sequence	Period 1	Period 2
T C	$X_1 = 1$	$Y_2 = 1$
	$X_1 = 0$	$Y_2 = 0$
C T	$Y_1 = 1$	$X_2 = 1$
	$Y_1 = 0$	$X_2 = 0$

The options below illustrate the possible probabilities ϕ_{ki} $(i = 1,2,...,4; k = 1,2)$ for the i^{th} observed outcome in the k^{th} sequence expressed in terms of the true probability values β_{ij} on Table 4.4. Hereafter, we will label these probabilities of observed data, "Estimable Probabilities."

TC – Sequence (1):

1.)

Estimable Probability	Potential Outcomes				
	Period 1		Period 2		
$\phi_{11} = P(X_1 = 0\ ,\ Y_2 = 0)$	$(X_1 = 0\ ,\ Y_1 = y_1)$		$(X_2 = x_2\ ,\ Y_2 = 0)$		Probability
	0	0	0	0	β_{11}
	0	0	1	0	β_{13}
	0	1	0	0	β_{21}
	0	1	1	0	β_{23}

Thus, $\phi_{11} = {}_{TC}P_{00} = P(X_1 = 0\ ,\ Y_2 = 0) = \beta_{11} + \beta_{13} + \beta_{21} + \beta_{23}$

2.)

Estimable Probability	Potential Outcomes				
	Period 1		Period 2		
$\phi_{12} = P(X_1 = 0\ ,\ Y_2 = 1)$	$(X_1 = 0\ ,\ Y_1 = y_1)$		$(X_2 = x_2\ ,\ Y_2 = 1)$		Probability
	0	0	0	1	β_{12}
	0	0	1	1	β_{14}
	0	1	0	1	β_{22}
	0	1	1	1	β_{24}

$\phi_{12} = {}_{TC}P_{01} = P(X_1 = 0\ ,\ Y_2 = 1) = \beta_{12} + \beta_{14} + \beta_{22} + \beta_{24}$

3.)

Estimable Probability	Potential Outcomes				
	Period 1		Period 2		
$\phi_{13} = P(X_1 = 1\ ,\ Y_2 = 0)$	$(X_1 = 1\ ,\ Y_1 = y_1)$		$(X_2 = x_2\ ,\ Y_2 = 0)$		Probability
	1	0	0	0	β_{31}
	1	0	1	0	β_{33}
	1	1	0	0	β_{41}
	1	1	1	0	β_{43}

$\phi_{13} = {}_{TC}P_{10} = P(X_1 = 1\ ,\ Y_2 = 0) = \beta_{31} + \beta_{33} + \beta_{41} + \beta_{43}$

4.)

Estimable Probability	Potential Outcomes				
	Period 1		Period 2		
$\phi_{14} = P(X_1 = 1\ ,\ Y_2 = 1)$	$(X_1 = 1\ ,\ Y_1 = y_1)$		$(X_2 = x_2\ ,\ Y_2 = 1)$		Probability
	1	0	0	1	β_{32}
	1	0	1	1	β_{34}
	1	1	0	1	β_{42}
	1	1	1	1	β_{44}

$\phi_{14} = {}_{TC}P_{11} = P(X_1 = 1\ ,\ Y_2 = 1) = \beta_{32} + \beta_{34} + \beta_{42} + \beta_{44}$

Similarly, for sequence CT,

CT – Sequence (2):

5.)

Estimable Probability	Potential Outcomes				
	Period 1		Period 2		
$\phi_{21} = P(X_2 = 0\ ,\ Y_1 = 0)$	$(X_1 = x_1\ ,\ Y_1 = 0)$		$(X_2 = 0\ ,\ Y_2 = y_2)$		Probability
	0	0	0	0	β_{11}
	0	0	0	1	β_{12}
	1	0	0	0	β_{31}
	1	0	0	1	β_{32}

$\phi_{21} = {}_{CT}P_{00} = P(X_2 = 0\ ,\ Y_1 = 0) = \beta_{11} + \beta_{12} + \beta_{31} + \beta_{32}$

6.)

Estimable Probability	Potential Outcomes				
	Period 1		Period 2		
$\phi_{22} = P(\ X_2 = 0\ ,\ Y_1 = 1)$	$(X_1 = x_1\ ,\ Y_1 = 1)$		$(X_2 = 0\ ,\ Y_2 = y_2)$		Probability
	0	1	0	0	β_{21}
	0	1	0	1	β_{22}
	1	1	0	0	β_{41}
	1	1	0	1	β_{42}

$\phi_{22} = {}_{CT}P_{10} = P(X_2 = 0\ ,\ Y_1 = 1) = \beta_{21} + \beta_{22} + \beta_{41} + \beta_{42}$

7.)

Estimable Probability	Potential Outcomes				
	Period 1		Period 2		
$\phi_{23} = P(\mathrm{X}_2 = 1\ ,\ Y_1 = 0)$	$(X_1 = x_1\ ,\ Y_1 = 0)$		$(X_2 = 1\ ,\ Y_2 = y_2)$		Probability
	0	0	1	0	β_{13}
	0	0	1	1	β_{14}
	1	0	1	0	β_{33}
	1	0	1	1	β_{34}

$$\phi_{23} = {}_{CT}P_{01} = P(\mathrm{X}_2 = 1\ ,\ Y_1 = 0) = \beta_{13} + \beta_{14} + \beta_{33} + \beta_{34}$$

8.)

Estimable Probability	Potential Outcomes				
	Period 1		Period 2		
$\phi_{24} = P(\mathrm{X}_2 = 1\ ,\ Y_1 = 1)$	$(X_1 = x_1\ ,\ Y_1 = 1)$		$(X_2 = 1\ ,\ Y_2 = y_2)$		Probability
	0	1	1	0	β_{23}
	0	1	1	1	β_{24}
	1	1	1	0	β_{43}
	1	1	1	1	β_{44}

$$\phi_{24} = {}_{CT}P_{11} = P(\mathrm{X}_2 = 1\ ,\ Y_1 = 1) = \beta_{23} + \beta_{24} + \beta_{43} + \beta_{44}$$

In summary, the estimable probabilities of the four possible observed outcomes from each sequence are

Sequence (1) TC:

$$\phi_{11} = {}_{TC}P_{00} = P(X_1 = 0\ ,\ Y_2 = 0) = \beta_{11} + \beta_{13} + \beta_{21} + \beta_{23} \qquad (1.20)$$

$$\phi_{12} = {}_{TC}P_{01} = P(X_1 = 0\ ,\ Y_2 = 1) = \beta_{12} + \beta_{14} + \beta_{22} + \beta_{24} \qquad (1.21)$$

$$\phi_{13} = {}_{TC}P_{10} = P(X_1 = 1\ ,\ Y_2 = 0) = \beta_{31} + \beta_{33} + \beta_{41} + \beta_{43} \qquad (1.22)$$

$$\phi_{14} = {}_{TC}P_{11} = P(X_1 = 1\ ,\ Y_2 = 1) = \beta_{32} + \beta_{34} + \beta_{42} + \beta_{44} \qquad (1.23)$$

Sequence (2) CT:

$$\phi_{21} = {}_{CT}P_{00} = P\left(X_2 = 0\ ,\ Y_1 = 0\right) = \beta_{11} + \beta_{12} + \beta_{31} + \beta_{32} \qquad (1.24)$$

$$\phi_{22} = {}_{CT}P_{10} = P\left(\mathrm{X}_2 = 0\ ,\ Y_1 = 1\right) = \beta_{21} + \beta_{22} + \beta_{41} + \beta_{42} \qquad (1.25)$$

$$\phi_{23} = {}_{CT}P_{01} = P\left(\mathrm{X}_2 = 1\ ,\ Y_1 = 0\right) = \beta_{13} + \beta_{14} + \beta_{33} + \beta_{34} \qquad (1.26)$$

$$\phi_{24} = {}_{CT}P_{11} = P\left(\mathrm{X}_2 = 1\ ,\ Y_1 = 1\right) = \beta_{23} + \beta_{24} + \beta_{43} + \beta_{44} \qquad (1.27)$$

Remark 4.3.1

1.) We note that, $\sum_{i=1}^{4}\phi_{ki} = 1$ for $k = 1,2$, $\phi_{ki} \in [0,1]$ since $\sum_{i=1}^{4}\phi_{ki} = \sum_{i=1}^{4}\sum_{j=1}^{4}\beta_{ij} = \sum_{i=1}^{4}\omega_i = 1$.

2.) If arbitrary labeling of the time periods was possible, then, $\beta_{ij} = \beta_{ji}$ for $i \neq j$ and $i, j = 1,2,3,4$. This would imply exchangeability holds and thus, we can equate probabilities from the two sequences. For example, consider the outcome $\left(X_1 = 1\ ,\ Y_2 = 0\right)$ in (1.22) above and $\left(\mathrm{X}_2 = 1\ ,\ Y_1 = 0\right)$ in (1.26), if exchangeability was possible then $\beta_{13} = \beta_{31}$, $\beta_{34} = \beta_{43}$ and hence, $\phi_{13} = {}_{TC}P_{10} = {}_{CT}P_{01} = \phi_{23}$. This would suggest no sequence effect for the subjects with this outcome. But the reverse is not sufficiently true. That is, if ${}_{TC}P_{10} = {}_{CT}P_{01}$ does not directly imply exchangeability. That is,

$$\left(\beta_{13} = \beta_{31}\right) \text{ and } \left(\beta_{34} = \beta_{43}\right) \not\Rightarrow \left(\phi_{13} = {}_{TC}P_{10} = {}_{CT}P_{01} = \phi_{23}\right)$$

This relationship is true for all other combinations of probabilities involving the two sequences. The proof will be outlined later.

4.3.1: Analysis of the Estimable Probabilities, ϕ_{ki}, of the Observed Outcomes

Here is a more detailed look at the probabilities, ϕ_{ki} $(i = 1,2,...,4; k = 1,2)$, for the i^{th} observed outcome in the k^{th} sequence expressed in terms of the true or actual probability values β_{ij} on Table 4.4. The probabilities are given in equations (1.20) through (1.27) above. Below, ϕ_{ki} are expressed in terms of the marginal probabilities, ω_i and the actual joint probabilities β_{ij}, for $i = 1,2,...,4;\ j = 1,2$ and $k = 1,2$. Inferences will then be drawn from the established relationships after some algebraic manipulations. For example,

$$\begin{aligned}\phi_{11} &= \beta_{11} + \beta_{13} + \beta_{21} + \beta_{23} \\ &= \omega_1 + (\beta_{21} - \beta_{12}) + (\beta_{23} - \beta_{14}) \\ &= \omega_1 - \Delta_1\end{aligned}$$

from the fact that $\omega_1 = \beta_{11} + \beta_{12} + \beta_{13} + \beta_{14} \Rightarrow (\beta_{11} + \beta_{13}) = \omega_1 - (\beta_{12} + \beta_{14})$. Note that $\Delta_1 = (\beta_{12} - \beta_{21}) + (\beta_{14} - \beta_{23})$. This derivation involving ω_1 was established using the *Period 1* marginal. We can easily extend it to ω_2, ω_3 and ω_4. Due to the assumption of equality in marginals, similar equations, for the two sequences TC and CT, can be found using *Period 2* marginals (that is, column totals in Table 4.4). This is addressed in the following two columns:

Using time Period 1 marginals in sequence TC		Using time Period 2 marginals in sequence TC	
$\phi_{11} = \omega_1 - \Delta_1$	(1.28)	$\phi_{11} = \omega_1 - \Delta_3$	(1.29)
$\phi_{12} = \omega_2 + \Delta_1$	(1.30)	$\phi_{12} = \omega_2 + \Delta_4$	(1.31)
$\phi_{13} = \omega_3 + \Delta_2$	(1.32)	$\phi_{13} = \omega_3 + \Delta_3$	(1.33)
$\phi_{14} = \omega_4 - \Delta_2$	(1.34)	$\phi_{14} = \omega_4 - \Delta_4$	(1.35)

Using time Period 1 marginals in sequence CT		Using time Period 2 marginals in sequence CT	
$\phi_{21} = \omega_1 - \Delta_5$	(1.36)	$\phi_{21} = \omega_1 - \Delta_7$	(1.37)
$\phi_{22} = \omega_2 + \Delta_6$	(1.38)	$\phi_{22} = \omega_2 + \Delta_7$	(1.39)
$\phi_{23} = \omega_3 + \Delta_5$	(1.40)	$\phi_{23} = \omega_3 + \Delta_8$	(1.41)
$\phi_{24} = \omega_4 - \Delta_6$	(1.42)	$\phi_{24} = \omega_4 - \Delta_8$	(1.43)

where,

$\Delta_1 = (\beta_{12} - \beta_{21}) + (\beta_{14} - \beta_{23})$, $\Delta_2 = (\beta_{43} - \beta_{34}) + (\beta_{41} - \beta_{32})$, $\Delta_3 = (\beta_{31} - \beta_{13}) + (\beta_{41} - \beta_{23})$,

$\Delta_4 = (\beta_{24} - \beta_{42}) + (\beta_{14} - \beta_{32})$, $\Delta_5 = (\beta_{13} - \beta_{31}) + (\beta_{14} - \beta_{32})$, $\Delta_6 = (\beta_{42} - \beta_{24}) + (\beta_{41} - \beta_{23})$,

$\Delta_7 = (\beta_{21} - \beta_{12}) + (\beta_{41} - \beta_{32})$, $\Delta_8 = (\beta_{34} - \beta_{43}) + (\beta_{14} - \beta_{23})$.

Remark 4.3.1.1

From Equations (1.28) through (1.35), there are some noticeable equalities between the equations to be considered. For instance, Equations (1.28) and (1.29) imply $\Delta_1 - \Delta_3 = 0$. Similarly, we note that, $\Delta_1 - \Delta_4 = 0, \Delta_2 - \Delta_3 = 0$ and $\Delta_2 - \Delta_4 = 0$. For example,

$\Delta_1 - \Delta_3 = (\beta_{12} - \beta_{21} + \beta_{14} - \beta_{23}) - (\beta_{31} - \beta_{13} + \beta_{41} - \beta_{23}) = (\omega_1 - \beta_{11}) - (\omega_1 - \beta_{11}) = 0$.

Furthermore, it can also be showed that, $\Delta_1 - \Delta_4 = (\omega_2 - \beta_{22}) - (\omega_2 - \beta_{22}) = 0$, $\Delta_2 - \Delta_3 = (\omega_3 - \beta_{33}) - (\omega_3 - \beta_{33}) = 0$ and $\Delta_2 - \Delta_4 = (\omega_4 - \beta_{44}) - (\omega_4 - \beta_{44}) = 0$. Hence, we note that, $\Delta_1 = \Delta_2 = \Delta_3 = \Delta_4 = \Delta_{TC}$ and, $\Delta_5 = \Delta_6 = \Delta_7 = \Delta_8 = \Delta_{CT}$. Thus, as anticipated, any one of the two time period 1 and 2 marginals provides sufficient information about the actual or

true probability, ω_i, from the probabilities of observed outcomes, ϕ_{ki}. Thus, (1.28) through (1.43) simplify to (1.48) through (1.51) given below.

For sequence TC:

$$\phi_{11} = \omega_1 - \Delta_{TC} \qquad (1.44)$$

$$\phi_{12} = \omega_2 + \Delta_{TC} \qquad (1.45)$$

$$\phi_{13} = \omega_3 + \Delta_{TC} \qquad (1.46)$$

$$\phi_{14} = \omega_4 - \Delta_{TC} \qquad (1.47)$$

For sequence CT:

$$\phi_{21} = \omega_1 - \Delta_{CT} \qquad (1.48)$$

$$\phi_{22} = \omega_2 + \Delta_{CT} \qquad (1.49)$$

$$\phi_{23} = \omega_3 + \Delta_{CT} \qquad (1.50)$$

$$\phi_{24} = \omega_4 - \Delta_{CT} \qquad (1.51)$$

From various combinations of (1.44) to (1.47) the Δ_{TC} cancels out and we have, for sequence TC: $\phi_{11} + \phi_{12} = \omega_1 + \omega_2$, $\phi_{11} + \phi_{13} = \omega_1 + \omega_3$, $\phi_{14} + \phi_{12} = \omega_4 + \omega_2$, $\phi_{13} + \phi_{14} = \omega_3 + \omega_4$ and for sequence CT: $\phi_{21} + \phi_{22} = \omega_1 + \omega_2$, $\phi_{21} + \phi_{23} = \omega_1 + \omega_3$, $\phi_{24} + \phi_{22} = \omega_4 + \omega_2$, $\phi_{24} + \phi_{23} = \omega_4 + \omega_3$. Thus, linear combinations of ω_i are estimable, although, separately ω_i cannot be estimated even after applying the TC – CT randomization. This point is further reinforced by an attempt to calculate ω_i using the matrix manipulations below. Thus,

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_4 \end{pmatrix} = \begin{bmatrix} \phi_{11} + \phi_{12} \\ \phi_{11} + \phi_{13} \\ \phi_{12} + \phi_{14} \\ \phi_{13} + \phi_{14} \end{bmatrix} \Rightarrow A_{TC} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_4 \end{pmatrix} = \begin{bmatrix} \phi_{11} + \phi_{12} \\ \phi_{11} + \phi_{13} \\ \phi_{12} + \phi_{14} \\ \phi_{13} + \phi_{14} \end{bmatrix}$$

where

$A_{TC} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} = A_{CT}$. We note that the matrix A_{TC} (and hence, A_{CT}) is not a full-rank matrix since the sum 2nd and 3rd columns minus the first equals the 4th column. In addition, the Eigen values of A_{TC} (and hence, A_{CT}) are 2.000, 1.414, 0.000 and -1.414. Hence, no unique solution for ω_i exist. In the next section, we derived estimates for the probabilities in Equations (1.44) to (1.51).

4.4: Estimations Using Data of Observed Counts from a TC – CT Crossover Design

In a typical TC – CT crossover design, the various possible observed outcomes and counts in *Period 1* and 2 can be classified into $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$, where (a,b) indicates response "a" in *Period 1* and "b" in *Period 2*. In addition, for sequence TC, $(a,b) = (x_1, y_2)$. That is, a represents response to T in *Period 1* and b denotes response to C in *Period 2*. Meanwhile for sequence CT, $(a,b) = (y_1, x_2)$. Thus, a symbolizes response to C in *Period 1* and b signifies response to T in *Period 2*.

Let $n_{1\bullet}$ and $n_{2\bullet}$ denote the number of subjects assigned to sequence TC and CT respectively. Also, let n_{11}, n_{12}, n_{13}, n_{14}, n_{21}, n_{22}, n_{23}, n_{24} denote the number of subjects with response such that the first subscript correspond to the sequence and the second subscript enumerate the four outcomes. For example, n_{22} denotes the number of subject that succeeded on treatment C in time *Period 1* and failed on treatment T in time *Period 2*, meanwhile, n_{23} denotes the number of subject that succeeded on treatment T in time *Period 2* and failed on treatment C in time *Period 1*. These counts are illustrated on Table 4.5 below.

Table 4.5: *Standard table of observed counts of responses to treatments in each period in a 2 treatments, 2 periods crossover design with focus on sequence*

Sequence	$(0\ ,\ 0)$	$(0\ ,\ 1)$	$(1\ ,\ 0)$	$(1\ ,\ 1)$	Marginal Total
$1(TC)$	n_{11}	n_{12}	n_{13}	n_{14}	$n_{1\bullet}$
$2(CT)$	n_{21}	n_{23}	n_{22}	n_{24}	$n_{2\bullet}$

Notice the interchange between n_{22} and n_{23}. Here, n_{22} is the number of (y_1, x_2) outcomes where $(y_1 = 1, x_2 = 0)$ and n_{23} is the number of (y_1, x_2) outcomes with $(1,0)$. This interchange was made to match the arrangement of outcomes $(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$.

Remark 4.4.0

Note that $\sum_{i=1}^{4} n_{ki} = n_{k\bullet}$ for $i = 1,2,3,4;\ k = 1,2$. Similarly, $\sum_{k=1}^{2}\sum_{i=1}^{4} n_{ki} = n$, where n denoted the total sample size in both sequences. We note that period effect favors one treatment over the other when $n_{12} + n_{13}$ is substantially different from $n_{22} + n_{23}$ (Jones and Kenward,

1989, p. 93). Inferentially, period effect, in favor of one treatment, is present when $\phi_{12}+\phi_{13}$ differs substantially differs from $\phi_{22}+\phi_{23}$. We have that,

$$\begin{aligned}\phi_{12}+\phi_{13} &= P\left(X_1=0\ ,\ Y_2=1\right)+P\left(X_1=1\ ,\ Y_2=0\right)\\ &=\left(\beta_{12}+\beta_{14}+\beta_{22}+\beta_{24}\right)+\left(\beta_{31}+\beta_{33}+\beta_{41}+\beta_{43}\right)\end{aligned}$$

$$\begin{aligned}\phi_{22}+\phi_{23} &= P\left(X_2=0\ ,\ Y_1=1\right)+P\left(X_2=1\ ,\ Y_1=0\right)\\ &=\left(\beta_{21}+\beta_{41}+\beta_{22}+\beta_{42}\right)+\left(\beta_{13}+\beta_{33}+\beta_{14}+\beta_{34}\right)\end{aligned}$$

In the presence of exchangeability, $\phi_{12}+\phi_{13}=\phi_{22}+\phi_{23}$ and there is no sequence or period effect. That is, time periods are randomly labeled. In the coming section, we explore methods of estimating the probabilities of the observed outcomes ϕ_{ki} for $i=1,2,3,4;k=1,2$.

4.4.1: Maximum Likelihood Estimation of Probabilities Using Observed Count Data

In the previous analysis involving the TC – CT crossover design, we got the estimable probabilities , ϕ_{ki} , for $i=1,2,3,4;k=1,2$.

We assume $\left(n_{k1},n_{k2},n_{k3},n_{k4}\right)\sim multinomial\left(\phi_{k1},\phi_{k2},\phi_{k3},\phi_{k4}\right)$. The likelihood function of the observed data is given as

$$L\left(n_{k1},n_{k2},n_{k3},n_{k4}\mid\phi_{k1},\phi_{k2},\phi_{k3},\phi_{k4}\right)\propto\prod_{k=1}^{2}\prod_{i=1}^{4}\phi_{ki}^{n_{ki}}\quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\left(*\right)$$

where $\sum_{i=1}^{4} n_{ki} = n_{k\bullet}$, for $i = 1,2,3,4;\ k = 1,2$ and subject to the constraints: $\sum_{i=1}^{4} \phi_{1i} = 1$ $\sum_{i=1}^{4} \phi_{2i} = 1$, and $\phi_{12} = \phi_{21} + \phi_{22} - \phi_{11}$, $\phi_{23} = \phi_{11} + \phi_{13} - \phi_{21}$ - a set of constraints that result from the following combinations of probabilities, ϕ_{ki}, in design TC – CT:

$\phi_{11} + \phi_{12} = \phi_{21} + \phi_{22}$, $\phi_{13} + \phi_{14} = \phi_{23} + \phi_{24}$, $\phi_{11} + \phi_{13} = \phi_{21} + \phi_{23}$ and $\phi_{12} + \phi_{14} = \phi_{22} + \phi_{24}$

The likelihood expression in $(*)$ contains 4 distinct parameters leading to 4 nonlinear equations. Solving these equations requires an iterative numerical method subject to 4 constraints. These 4 nonlinear equations could still be reparameterized with no constraints (as will be done in an illustrative example), but then, we would still require an iterative method of solution. Various iterative numerical procedures are available for use. Here, we will use the *optim* package in *R* for evaluation (www.r-project.org).

Assume the maximum likelihood estimates of the probabilities ϕ_{ki} have been found and denote them, $\hat{\phi}_{ki}$, where $i = 1,2,3,4; k = 1,2$. We state the following proposition:

Proposition 4.4.1

The average individual effect of treatment T relative to treatment C, $E(D)$, is given as $E(D) = 0.5\left[(\phi_{13} + \phi_{23}) - (\phi_{12} + \phi_{22})\right]$ and is estimated by

$$\hat{E}(D) = 0.5\left[\left(\hat{\phi}_{13} + \hat{\phi}_{23}\right) - \left(\hat{\phi}_{12} + \hat{\phi}_{22}\right)\right]$$

Proof:

Recall that $E(D) = \omega_3 - \omega_2$. From equations (1.45) and (1.46) in sequence TC: $\omega_3 - \omega_2 = (\phi_{13} - \phi_{12})$. From equations (1.49) and (1.50) in sequence CT: $\omega_3 - \omega_2 = (\phi_{23} - \phi_{22})$. Thus, combining the effect from the two sequences, we have that,

$$E(D) = \omega_3 - \omega_2 = 0.5\left[(\phi_{13} - \phi_{12}) + (\phi_{23} - \phi_{22})\right].$$

$$\Rightarrow \hat{E}(D) = 0.5\left[\left(\hat{\phi}_{13} - \hat{\phi}_{12}\right) + \left(\hat{\phi}_{23} - \hat{\phi}_{22}\right)\right] \quad \blacksquare$$

Remark 4.4.1

$\hat{E}(D)$ is an estimate of $E(D)$- the actual average individual treatment effect of T versus C - in a $TC - CT$ crossover design. $E(D)$ expresses the average difference in response for the $(1,0)$ and $(0,1)$ outcomes at a particular time point. We now turn our focus to the "probability of negative effect."

4.4.2: Probability of a Negative Effect

Consider the probability of fairing well on treatment C relative to treatment T. In this case, $D = X - Y < 0$. From Table 6, the probability of negative effect is given as ω_2. That is, $P(D = -1) = \omega_2$. From Table 4.4, we have that, $P(D = -1) = \omega_2 = \beta_{21} + \beta_{22} + \beta_{23} + \beta_{24}$ (using *Period 1* probabilities) or $P(D = -1) = \omega_2 = \beta_{12} + \beta_{22} + \beta_{32} + \beta_{42}$ (using *Period 2* probabilities), where, $\omega_2 = P(X = 0,\ Y = 1)$. The expression of probability of negative effect

carries a connotation of an "unexplained individual treatment variation" (Gadbury et al, 2004, p. 173). Furthermore, it may also provide information on the "possible magnitude of a treatment by covariate interaction" (Gadbury et al, 2004, p. 173), and treatment by period interaction, thus shedding more light on the extent of the unexplained individual treatment effect variability.

Proposition 4.4.2

For the TC – CT crossover design, the combined probability of a negative effect, $\phi_{12} = {}_{TC}P_{01} = P(X_1 = 0\ ,\ Y_2 = 1)$ and $\phi_{22} = {}_{CT}P_{10} = P(X_2 = 0\ ,\ Y_1 = 1)$ is not an unbiased estimate of the actual probability of negative effect given as $\omega_2 = P(X = 0, Y = 1)$.

Proof:

From equations (1.45) and (1.50), we have that

$$\omega_2 = 0.5\left(\phi_{12} + \phi_{22} - \left(\Delta_{TC} + \Delta_{CT}\right)\right) \tag{1.52}$$

Thus, the probability of observing $(0,1)$ is not unbiased for the true probability of $(0,1)$ denoted ω_2.

Remark 4.4.2

As a consequence of the proposition 4.4.3, we will establish bounds, $[L_1, U_1]$, for the true probability of negative effect, $\omega_2 = P(X = 0, Y = 1)$, since it is impossible to estimate $\left(\Delta_{TC} + \Delta_{CT}\right)$ using the TC – CT design. This statement holds even with the inclusion of exchangeability because, then, $\Delta_{TC} = \Delta_{CT}$ - as will be shown later - and $\left(\Delta_{TC} + \Delta_{CT}\right)$ does not cancel out. The next proposition builds on this remark.

Proposition 4.4.3

The bounds for ω_2 using the TC – CT crossover design are L_1 and U_1 such that

$$L_1 \le \omega_2 \le U_1$$

where $L_1 = \max\left\{0,\ 0.5\left[\left(\phi_{12}+\phi_{22}\right)-\left(\phi_{13}+\phi_{23}\right)\right]\right\}$ and

$U_1 = \min\left\{0.5\left[\left(\phi_{11}+\phi_{21}\right)+\left(\phi_{12}+\phi_{22}\right)\right],\ 0.5\left[\left(\phi_{12}+\phi_{22}\right)+\left(\phi_{14}+\phi_{24}\right)\right]\right\}$

Proof:

From Equations (1.44) of sequence TC and (1.48) of sequence CT, we have, $\phi_{11}+\phi_{12} = \omega_1+\omega_2$ and $\phi_{21}+\phi_{23} = \omega_1+\omega_2$ respectively. This leads us to the equation, $\omega_1+\omega_2 = 0.5\left(\phi_{11}+\phi_{12}+\phi_{21}+\phi_{23}\right)$. Similarly, from Equation (1.46): $\phi_{14}+\phi_{12} = \omega_4+\omega_2$, and Equation (1.50): $\phi_{24}+\phi_{23} = \omega_4+\omega_2$, we have $\omega_4+\omega_2 = 0.5\left(\phi_{14}+\phi_{12}+\phi_{24}+\phi_{23}\right)$.

Hence,

$\omega_2 \le U = \min\left\{0.5\left[\left(\phi_{11}+\phi_{21}\right)+\left(\phi_{12}+\phi_{22}\right)\right],\ 0.5\left[\left(\phi_{12}+\phi_{22}\right)+\left(\phi_{14}+\phi_{24}\right)\right]\right\}$.

Furthermore, we had that $\omega_3-\omega_2 = 0.5\left[\left(\phi_{13}+\phi_{23}\right)-\left(\phi_{12}+\phi_{22}\right)\right]$. This implies $\omega_2 = \omega_3+0.5\left[\left(\phi_{12}-\phi_{13}\right)+\left(\phi_{22}-\phi_{23}\right)\right]$ and thus,

$\omega_2 \ge L = \max\left\{0,\ 0.5\left[\left(\phi_{12}+\phi_{22}\right)-\left(\phi_{13}+\phi_{23}\right)\right]\right\}$ ■

Proposition 4.4.4

L_1 and U_1 are identifiable and have *M.L.E.s* given as $\hat{L}_1$ and $\hat{U}_1$ where

$$\hat{L}_1 = \max\left\{0,\ 0.5\left[\left(\hat{\phi}_{12}+\hat{\phi}_{22}\right)-\left(\hat{\phi}_{13}+\hat{\phi}_{23}\right)\right]\right\} \text{ and}$$

$$\hat{U}_1 = \min\left\{0.5\left[\left(\hat{\phi}_{11}+\hat{\phi}_{21}\right)+\left(\hat{\phi}_{12}+\hat{\phi}_{22}\right)\right],\ 0.5\left[\left(\hat{\phi}_{12}+\hat{\phi}_{22}\right)+\left(\hat{\phi}_{14}+\hat{\phi}_{24}\right)\right]\right\}$$

where $\hat{\phi}_{ki}$, $(i=1,2,3,4 \text{ and } k=1,2)$ are *MLE* estimates of the probabilities ϕ_{ki}. Furthermore, there exists a distribution for which L_1 and U_1 are attained.

Proof:

There exists a unique *M.L.E* for each $\hat{\phi}_{ki}$, $i=1,2,3,4;\ k=1,2$. Thus, identifiability follows.

We also note that,

$$U_1 = \min\left\{0.5\left[\left(\phi_{11}+\phi_{21}\right)+\left(\phi_{12}+\phi_{22}\right)\right],\ 0.5\left[\left(\phi_{12}+\phi_{22}\right)+\left(\phi_{14}+\phi_{24}\right)\right]\right\}$$

$$= \begin{cases} 0.5\left[\left(\phi_{11}+\phi_{21}\right)+\left(\phi_{12}+\phi_{22}\right)\right] & \textit{if } \left(\phi_{11}+\phi_{21}\right) \le \left(\phi_{14}+\phi_{24}\right) \\ 0.5\left[\left(\phi_{14}+\phi_{24}\right)+\left(\phi_{12}+\phi_{22}\right)\right] & \textit{if } \left(\phi_{11}+\phi_{21}\right) \ge \left(\phi_{14}+\phi_{24}\right) \end{cases}$$

These bounds are attained when the distribution of Table 4.4 leads to probability of observed outcomes shown on the Tables 4.6 and 4.7 below.

Table 4.6: *A distribution of probabilities of observed outcomes for which the upper bound is be attained*

For an Indiv.	$\overset{T\ C}{(0,0)}$	$\overset{T\ C}{(0,1)}$	$\overset{T\ C}{(1,0)}$	$\overset{T\ C}{(1,1)}$
$\overset{T\ C}{(0,0)}$	0	0	0	0
$\overset{T\ C}{(0,1)}$	0	$0.5\theta_2$	0	0
$\overset{T\ C}{(1,0)}$	0	0	$0.5\theta_3$	0
$\overset{T\ C}{(1,1)}$	0	0	0	$0.5\theta_4$

or

Table 4.7: *A distribution of probabilities of observed outcomes for which the lower bound is be attained*

For an Indiv.	$\overset{T\ C}{(0,0)}$	$\overset{T\ C}{(0,1)}$	$\overset{T\ C}{(1,0)}$	$\overset{T\ C}{(1,1)}$
$\overset{T\ C}{(0,0)}$	$0.5\theta_1$	0	0	0
$\overset{T\ C}{(0,1)}$	0	$0.5\theta_2$	0	0
$\overset{T\ C}{(1,0)}$	0	0	$0.5\theta_3$	0
$\overset{T\ C}{(1,1)}$	0	0	0	0

respectively, where $\theta_1 = (\phi_{11} + \phi_{21})$, $\theta_2 = (\phi_{12} + \phi_{22})$, $\theta_3 = (\phi_{13} + \phi_{23})$, $\theta_4 = (\phi_{14} + \phi_{24})$.

Similarly,

$$L_1 = \max\left\{0,\ 0.5\left[(\phi_{12} + \phi_{22}) - (\phi_{13} + \phi_{23})\right]\right\}$$

$$= \begin{cases} 0 & if\ \ (\phi_{12} + \phi_{22}) \le (\phi_{13} + \phi_{23}) \\ 0.5\left[(\phi_{12} + \phi_{22}) - (\phi_{13} + \phi_{23})\right] & if\ \ (\phi_{12} + \phi_{22}) \ge (\phi_{13} + \phi_{23}) \end{cases}$$

and the bounds are attained with a distribution of the forms,

Table 4.8: *A distribution of probabilities of observed outcomes for which the lower bound is be attained.*

For an Indiv.	$\overset{T\ C}{(0,0)}$	$\overset{T\ C}{(0,1)}$	$\overset{T\ C}{(1,0)}$	$\overset{T\ C}{(1,1)}$
$\overset{T\ C}{(0,0)}$	$0.5(\theta_1+\theta_2)$	0	0	0
$\overset{T\ C}{(0,1)}$	0	$0.5\theta_2$	0	0
$\overset{T\ C}{(1,0)}$	0	0	$0.5(\theta_3-\theta_2)$	0
$\overset{T\ C}{(1,1)}$	0	0	0	$0.5\theta_4$

or

Table 4.9: *A distribution of probabilities of observed outcomes for which the lower bound is be attained*

For an Indiv.	$\overset{T\ C}{(0,0)}$	$\overset{T\ C}{(0,1)}$	$\overset{T\ C}{(1,0)}$	$\overset{T\ C}{(1,1)}$
$\overset{T\ C}{(0,0)}$	$0.5\theta_4$	0	0	0
$\overset{T\ C}{(0,1)}$	0	$0.5\theta_3$	0	0
$\overset{T\ C}{(1,0)}$	0	0	$0.5(\theta_2-\theta_3)$	0
$\overset{T\ C}{(1,1)}$	0	0	0	$0.5(\theta_1+\theta_3)$

respectively. ■

We continue the analyses with the examination of the concept of exchangeability. We define certain criteria which will be useful, subsequently. Later in the analyses, we impose the exchangeability assumption and investigate the impact on the conclusions already reached.

4.5: Exchangeability

The notion of exchangeability can be explained in terms of the true probabilities β_{ij} and estimable probabilities, ϕ_{ki}, from the observed data, where $i, j = 1,2,3,4;\ k = 1,2$.

First, exchangeability occurs when $\beta_{ij} = \beta_{ji}$. Until now, we have assumed that exchangeability does not hold, that is, $\beta_{ij} \neq \beta_{ji}$ for $i \neq j$ and $i, j = 1,2,3,4$ in Table 4.4.

As a consequence of $\beta_{ij} = \beta_{ji}$, we have exchangeability in the observed outcomes. That is, $\phi_{ki} = \phi_{k'i}$ for $i = 1,2,3,4$ and $k, k' = 1,2 (k \neq k')$. That is, if $\beta_{ij} = \beta_{ji}$, then $\phi_{11} = \phi_{21}$, $\phi_{12} = \phi_{22}$, $\phi_{13} = \phi_{23}$ and $\phi_{14} = \phi_{24}$.

In general, exchangeability may not be reasonable in a two – time period design. This is especially true in the presence of carryover and/or correlation between the outcomes in the first and second periods for a given treatment. The absence of exchangeability emphasizes the presence of time period effect. However, Gadbury et al. (2004) showed that exchangeability is reasonable with matched pairs because it is reasonable to assume subjects 1 and 2 within a pair are randomly labeled. But applying this assumption to a two treatment, two periods crossover design would suggest the periods are randomly labeled.

4.5.1: Definition of Some Criteria

Define the following criteria:

1). *periods - TC - perfect match* if (x_1, y_1) and (x_2, y_2) are such that $x_1 = x_2$ and $y_1 = y_2$. From Table 4.4, the various combinations of outcomes are:

$\{(x_1, y_1),(x_2, y_2)\} = \{(0,0),(0,0)\}$, $\{(x_1, y_1),(x_2, y_2)\} = \{(0,1),(0,1)\}$,

$\{(x_1, y_1),(x_2, y_2)\} = \{(1,0),(1,0)\}$ and $\{(x_1, y_1),(x_2, y_2)\} = \{(1,1),(1,1)\}$.

The actual probabilities associated with these combinations of outcomes are β_{11}, β_{22}, β_{33}, β_{44}, respectively. These are the diagonal probabilities in Table 4.4.

2). *periods - TC - perfect mismatch* if (x_1, y_1) and (x_2, y_2) are such that $x_1 \neq x_2$ and $y_1 \neq y_2$. Combinations of outcomes in this domain include:

$\{(x_1, y_1),(x_2, y_2)\} = \{(0,1),(1,0)\}$, $\{(1,0),(0,1)\}$, $\{(0,0),(1,1)\}$ and $\{(1,1),(0,0)\}$. These constitute the outcomes with cross-diagonal $(\nearrow)$ probabilities in Table 4.4.

3). *periods - T - match* if (x_1, y_1) and (x_2, y_2) are such that $x_1 = x_2$ and $y_1 \neq y_2$. The following combinations of outcomes under this definition include:

$\{(x_1, y_1),(x_2, y_2)\} = \{(0,1),(0,0)\}$, $\{(x_1, y_1),(x_2, y_2)\} = \{(0,0),(0,1)\}$,

$\{(x_1, y_1),(x_2, y_2)\} = \{(1,0),(1,1)\}$ and $\{(x_1, y_1),(x_2, y_2)\} = \{(1,1),(1,0)\}$.

4). *periods - C - match* if (x_1, y_1) and (x_2, y_2) are such that $x_1 \neq x_2$ and $y_1 = y_2$. The following combinations of outcomes fall under this category:

$\{(x_1, y_1),(x_2, y_2)\} = \{(0,0),(1,0)\}$, $\{(x_1, y_1),(x_2, y_2)\} = \{(1,0),(0,0)\}$,

$\{(x_1, y_1),(x_2, y_2)\} = \{(0,1),(1,1)\}$ and $\{(x_1, y_1),(x_2, y_2)\} = \{(1,1),(0,1)\}$.

Proposition 4.5.1

Consider the true probabilities given on Table 4.4.

If exchangeability holds (i.e. $\beta_{ij} = \beta_{ji}$), then $\phi_{ki} = \phi_{k'i}$ for $i, j = 1,2,3,4$ and $k, k' = 1,2 (k \neq k')$.

Assume that the "*periods - TC - perfect match*" and "*periods - TC - perfect mismatch*" probabilities are zero. If $\phi_{ki} = \phi_{k'i}$ then exchangeability holds and we obtain $\beta_{ij} = \beta_{ji}$.

Proof:

We will show the proof for one probability expression, $\phi_{13} = \phi_{23}$. The others follow likewise.

If $\beta_{ij} = \beta_{ji}$ (for $i \neq j$ and $i, j = 1,2,3,4$), then from Table 4.4 and equations (1.19) to (1.26), we will have, for the first case,

$$\begin{aligned}
\phi_{13} = {}_{TC}P_{10} &= P(X_1 = 1,\ Y_2 = 0) \\
&= \beta_{31} + \beta_{33} + \beta_{41} + \beta_{43} \\
&= \beta_{13} + \beta_{33} + \beta_{14} + \beta_{34} \quad (\text{since } \beta_{ij} = \beta_{ji}) \\
&= {}_{CT}P_{10} \\
&= P(X_2 = 1,\ Y_1 = 0) \\
&= \phi_{23}
\end{aligned}$$

This is true for all other combinations of probabilities. Hence, $\phi_{12} = {}_{TC}P_{01} = {}_{CT}P_{10} = \phi_{22}$, $\phi_{14} = {}_{TC}P_{11} = {}_{CT}P_{11} = \phi_{24}$ and $\phi_{11} = {}_{TC}P_{00} = {}_{CT}P_{00} = \phi_{21}$.

If we assume "*periods - TC - perfect match*" and "*periods - TC - perfect mismatch*" probabilities are zero, then the diagonal and cross-diagonal $(\nearrow)$ probabilities are zero. That is, $\beta_{11} = 0,\ \beta_{22} = 0, \beta_{33} = 0,\ \beta_{44} = 0,\ \beta_{14} = 0,\ \beta_{23} = 0,\ \beta_{32} = 0,\ \beta_{41} = 0$. Thus, Table 4.4 becomes Table 4.8 shown below.

Table 4.10: *Probabilities of potential Outcomes for two treatments, two periods* crossover design *showing zero values for periods - TC - perfect match and periods - TC - perfect mismatch*

		Period 2				
	For an Individual	$\overset{T}{(0}\ \overset{C}{0)}$	$\overset{T}{(0}\ \overset{C}{1)}$	$\overset{T}{(1}\ \overset{C}{0)}$	$\overset{T}{(1}\ \overset{C}{1)}$	Marginal Total
Period 1	$\overset{T}{(0}\ \overset{C}{0)}$	0	β_{12}	β_{13}	0	ω_1
	$\overset{T}{(0}\ \overset{C}{1)}$	β_{21}	0	0	β_{24}	ω_2
	$\overset{T}{(1}\ \overset{C}{0)}$	β_{31}	0	0	β_{34}	ω_3
	$\overset{T}{(1}\ \overset{C}{1)}$	0	β_{42}	β_{43}	0	ω_4
	Marginal Total	ω_1	ω_2	ω_3	ω_4	$\sum_{i=1}^{4}\omega_i = 1$

Equating the marginals, ω_i, on Table 14, we have the following equations

$$\beta_{12} + \beta_{13} = \beta_{21} + \beta_{31} \tag{1.53}$$

$$\beta_{21} + \beta_{24} = \beta_{12} + \beta_{42} \tag{1.54}$$

$$\beta_{31} + \beta_{34} = \beta_{13} + \beta_{43} \tag{1.55}$$

$$\beta_{42} + \beta_{43} = \beta_{24} + \beta_{34} \tag{1.56}$$

Now, from (1.19) and (1.25), we had that

$$\phi_{13} = {}_{TC}P_{10} = P(X_1 = 1,\ Y_2 = 0) = \beta_{31} + \beta_{33} + \beta_{41} + \beta_{43}$$
$$= \beta_{31} + \beta_{43} \quad (\text{ since } \beta_{33} = 0,\ \beta_{41} = 0)$$

$$\phi_{23} = {}_{CT}P_{01} = P(X_2 = 1,\ Y_1 = 0) = \beta_{13} + \beta_{14} + \beta_{33} + \beta_{34}$$
$$= \beta_{13} + \beta_{34}$$

So, if we assume $\phi_{13} = {}_{TC}P_{10} = {}_{CT}P_{10} = \theta_{23}$, then,

$$\beta_{31} + \beta_{43} = \beta_{13} + \beta_{34} \tag{1.57}$$

But from Equation (1.56), $\beta_{34} = \beta_{13} + \beta_{43} - \beta_{31}$. Substituting into (1.58) and simplifying gives

$$2\beta_{31} = 2\beta_{13} \Rightarrow \beta_{13} = \beta_{31} \tag{1.58}$$

We can repeat the process for other probability expressions and obtain $\beta_{12} = \beta_{21}$, $\beta_{14} = \beta_{41}$, $\beta_{24} = \beta_{42}$ and $\beta_{34} = \beta_{43}$. Thus, $\beta_{ij} = \beta_{ji}$ for $i, j = 1,2,3,4$. ■

From the above proposition, we note that, exchangeability in the observed probability is a necessary but not a sufficient condition for exchangeability in the true probability. That is, exchangeability in the observed outcome is necessary for the actual exchangeability. But by itself, observed exchangeability is not sufficient. In symbols, $\left(\beta_{ij} = \beta_{ji}\right) \overset{\Rightarrow}{\not\Leftarrow} \left(\phi_{ki} = \phi_{k'i}\right)$.

Remark 4.5.1

So far, we have looked at analyses of the TC – CT crossover design. We were able to express the probabilities of the observed outcomes, ϕ_{ki}, in terms of the true or actual probabilities. Furthermore, we estimated the probabilities, ϕ_{ki}, using observed count data. In addition, we constructed bounds for the probability of negative effect, ω_2, denoted $\left[L_1, U_1\right]$. We note that, using the equation $\omega_2 = 0.5\left(\phi_{12} + \phi_{22} - \left(\Delta_{TC} + \Delta_{CT}\right)\right)$, a new – and hopefully, tighter – bounds, $\left[L_2, U_2\right]$, for ω_2 can be found by first constructing bounds for $\left(\Delta_{TC} + \Delta_{CT}\right)$ using information gained from additional analyses of the TT – CC design. Such bounds for

$(\Delta_{TC} + \Delta_{CT})$ will also be useful in the construction of bounds for the true variance of the individual treatment effect, $Var(D)$ (see *Appendix 4*). Hereafter, we proceed with the analysis of an additional design: TC – CT – TT – CC. Notice this is just an extension of the TC – CT crossover.

4.6: Potential Outcomes and Probabilities of Observed Outcomes for the TC – CT – TT – CC Design

Suppose some patients are allowed to stay on the same treatment over the two periods resulting in the TT – CC. The new TC – CT – TT – CC design is an extension of the TC – CT crossover design. The randomization is of the form

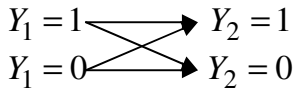
Sequence	*Period* 1	2
1	T	C
2	C	T

Sequence	*Period* 1	2
1	T	T
2	C	C

Consider the following schematic diagram illustrates the observed possible outcomes under the TC – CT design:

Sequence	Period 1	Period 2
T C (1)	$X_1 = 1$	$Y_2 = 1$
	$X_1 = 0$	$Y_2 = 0$
C T (2)	$Y_1 = 1$	$X_2 = 1$
	$Y_1 = 0$	$X_2 = 0$
T T (3)	$X_1 = 1$	$X_2 = 1$
	$X_1 = 0$	$X_2 = 0$

C C (4)



Earlier, we analyzed the probabilities of observed outcomes obtained using the TC – CT design. Below, we present the possible options for these probabilities for the TT – CC design.

TT – Sequence (3):

Estimable Probability	Potential Outcomes				
	Period 1		Period 2		
$\phi_{31} = P(X_1 = 0\ ,\ X_2 = 0)$	$(X_1 = 0\ ,\ Y_1 = y_1)$		$(X_2 = 0,\ Y_2 = y_2)$		Probability
	0	0	0	0	β_{11}
	0	0	0	1	β_{12}
	0	1	0	0	β_{21}
	0	1	0	1	β_{22}

That is, $\phi_{31} = {}_{TT}P_{01} = P(X_1 = 0\ ,\ X_2 = 0) = \beta_{11} + \beta_{12} + \beta_{21} + \beta_{22}$

Similar breakdown results in the following probabilities for the respective observed outcomes:

Sequence (3) TT:

$$\phi_{31} = {}_{TT}P_{01} = P(X_1 = 0\ ,\ X_2 = 0) = \beta_{11} + \beta_{12} + \beta_{21} + \beta_{22} \quad (1.59)$$

$$\phi_{32} = {}_{TT}P_{01} = P(X_1 = 0\ ,\ X_2 = 1) = \beta_{13} + \beta_{14} + \beta_{23} + \beta_{24} \quad (1.60)$$

$$\phi_{33} = {}_{TT}P_{10} = P(X_1 = 1\ ,\ X_2 = 0) = \beta_{31} + \beta_{41} + \beta_{32} + \beta_{42} \quad (1.61)$$

$$\phi_{34} = {}_{TT}P_{11} = P(X_1 = 1\ ,\ X_2 = 1) = \beta_{33} + \beta_{34} + \beta_{43} + \beta_{44} \quad (1.62)$$

Sequence (4) CC:

$$\phi_{41} = {}_{CC}P_{00} = P(Y_1 = 0\ ,\ Y_2 = 0) = \beta_{11} + \beta_{13} + \beta_{31} + \beta_{33} \quad (1.63)$$

$$\phi_{42} = {}_{CT}P_{01} = P(Y_1 = 0\ ,\ Y_2 = 1) = \beta_{12} + \beta_{32} + \beta_{14} + \beta_{34} \quad (1.64)$$

$$\phi_{43} = {}_{CC}P_{10} = P(Y_1 = 1\ ,\ Y_2 = 0) = \beta_{21} + \beta_{23} + \beta_{41} + \beta_{43} \quad (1.65)$$

$$\phi_{44} = {}_{CC}P_{11} = P(Y_1 = 1\ ,\ Y_2 = 1) = \beta_{22} + \beta_{24} + \beta_{42} + \beta_{44} \tag{1.66}$$

From Equation (1.59) through (1.66) we infer the certain combinations, we express the following combination of probabilities: $\phi_{31} + \phi_{32} = \omega_1 + \omega_2$, $\phi_{31} + \phi_{33} = \omega_1 + \omega_3$, $\phi_{34} + \phi_{33} = \omega_3 + \omega_4$, $\phi_{34} + \phi_{32} = \omega_2 + \omega_4$, $\phi_{41} + \phi_{42} = \omega_1 + \omega_2$, $\phi_{41} + \phi_{43} = \omega_1 + \omega_3$, $\phi_{44} + \phi_{43} = \omega_3 + \omega_4$, $\phi_{44} + \phi_{42} = \omega_3 + \omega_4$. The following proposition gives bounds for the $\Delta_{TC} + \Delta_{CT}$ involving estimable probabilities in the TC – CT – TT – CC design.

Proposition 4.6.1

Under the framework established above, $\left|\Delta_{TC} + \Delta_{CT}\right| \le \min\left\{1-(\phi_{31} + \phi_{34}), 1-(\phi_{41} + \phi_{44})\right\}$, where ϕ_{12}, ϕ_{22}, ϕ_{31}, ϕ_{34}, ϕ_{41} and ϕ_{44} are as given above.

Proof:

Recall that $\phi_{31} = P(X_1 = 0\ ,\ X_2 = 0)$, $\phi_{34} = P(X_1 = 1\ ,\ X_2 = 1)$, $\phi_{41} = P(Y_1 = 0\ ,\ Y_2 = 0)$ and $\phi_{44} = P(Y_1 = 1\ ,\ Y_2 = 1)$. In addition, we saw that,

$\Delta_1 = (\beta_{12} - \beta_{21}) + (\beta_{14} - \beta_{23})$, $\Delta_2 = (\beta_{43} - \beta_{34}) + (\beta_{41} - \beta_{32})$, $\Delta_3 = (\beta_{31} - \beta_{13}) + (\beta_{41} - \beta_{23})$, $\Delta_4 = (\beta_{24} - \beta_{42}) + (\beta_{14} - \beta_{32})$, $\Delta_5 = (\beta_{13} - \beta_{31}) + (\beta_{14} - \beta_{32})$, $\Delta_6 = (\beta_{42} - \beta_{24}) + (\beta_{41} - \beta_{23})$, $\Delta_7 = (\beta_{21} - \beta_{12}) + (\beta_{41} - \beta_{32})$, $\Delta_8 = (\beta_{34} - \beta_{43}) + (\beta_{14} - \beta_{23})$. Furthermore, we showed that $\Delta_1 = \Delta_2 = \Delta_3 = \Delta_4 = \Delta_{TC}$ and $\Delta_5 = \Delta_6 = \Delta_7 = \Delta_8 = \Delta_{CT}$.

Consider the Δ_1 and Δ_7 combination.

$$\begin{aligned}\Delta_{TC} + \Delta_{CT} &= \Delta_1 + \Delta_7 \\ &= \beta_{14} - \beta_{23} + \beta_{41} - \beta_{32} \\ &= (\beta_{14} + \beta_{41}) - (\beta_{23} + \beta_{32})\end{aligned}$$

Note that, the expression, $(\beta_{14}+\beta_{41})-(\beta_{23}+\beta_{32})$ involve *perfect – TC – mismatch* probabilities, β_{14}, β_{41}, β_{23} and β_{32} corresponding to $\{(x_1,y_1),(x_2,y_2)\}=\{(0,0),(1,1)\}$, $\{(x_1,y_1),(x_2,y_2)\}=\{(1,1),(0,0)\}$, $\{(x_1,y_1),(x_2,y_2)\}=\{(0,1),(1,0)\}$, and $\{(x_1,y_1),(x_2,y_2)\}=\{(1,0),(0,1)\}$, respectively. Furthermore, using the principle of triangle inequality, we have $|\Delta_{TC}+\Delta_{CT}|\le(\beta_{14}+\beta_{41})+(\beta_{23}+\beta_{32})$. But,

$$\begin{aligned}(\beta_{14}+\beta_{41})+(\beta_{23}+\beta_{32}) &= (\omega_1+\omega_2+\omega_3+\omega_4)-(\beta_{11}+\beta_{12}+\beta_{13})-(\beta_{21}+\beta_{22}+\beta_{24})\\ &\quad -(\beta_{31}+\beta_{33}+\beta_{34})-(\beta_{42}+\beta_{43}+\beta_{44})\\ &\le 1-\left[(\beta_{11}+\beta_{12}+\beta_{21}+\beta_{22})+(\beta_{33}+\beta_{34}+\beta_{43}+\beta_{44})\right]\\ &= 1-\left[P(X_1=0\ ,\ X_2=0)+P(X_1=1\ ,\ X_2=1)\right]\\ &= 1-(\phi_{31}+\phi_{34})\end{aligned}$$

Thus, $(\beta_{14}+\beta_{41})+(\beta_{23}+\beta_{32})\le 1-(\phi_{31}+\phi_{34})$. Similarly,

$$\begin{aligned}(\beta_{14}+\beta_{41})+(\beta_{23}+\beta_{32}) &\le 1-\left[(\beta_{11}+\beta_{13}+\beta_{31}+\beta_{33})+(\beta_{22}+\beta_{24}+\beta_{42}+\beta_{44})\right]\\ &= 1-\left[P(Y_1=0\ ,\ Y_2=0)+P(Y_1=1\ ,Y_2=1)\right]\\ &= 1-(\phi_{41}+\phi_{44})\end{aligned}$$

Hence, $(\beta_{14}+\beta_{41})+(\beta_{23}+\beta_{32})\le 1-(\phi_{41}+\phi_{44})$. Combining gives

$$|\Delta_{TC}+\Delta_{CT}|\le\min\{1-(\phi_{31}+\phi_{34}),1-(\phi_{41}+\phi_{44})\} \qquad (1.67)\ \blacksquare$$

Remark 4.6.1

The proof above uses the $\Delta_{TC}=\Delta_1$ and $\Delta_{CT}=\Delta_7$ combination although $\Delta_1=\Delta_2=\Delta_3=\Delta_4=\Delta_{TC}$ and that $\Delta_5=\Delta_6=\Delta_7=\Delta_8=\Delta_{CT}$. Other possible combinations like Δ_2 and Δ_8; Δ_3 and Δ_5; Δ_4 and Δ_6 will produce the same results with varied degree of

analysis. Equation (1.67) gives bounds for $\Delta_{TC}+\Delta_{CT}$. These bounds will now be used to construct, hopefully tighter, bounds for the probability of negative effect ω_2 and (in *Appendix D*) the true variance individual treatment effects, $Var(D)$

Proposition 4.6.2

Given that $|\Delta_{TC}+\Delta_{CT}| \le \min\{1-(\phi_{31}+\phi_{34}),1-(\phi_{41}+\phi_{44})\}$, a refined set of bounds for ω_2 is

$$L_2 \le \omega_2 \le U_2$$

where, $L_2 = \max\{0, 0.5[(\phi_{12}+\phi_{22}) - \min\{1-(\phi_{31}+\phi_{34}),1-(\phi_{41}+\phi_{44})\}]\}$ and

$U_2 = 0.5[(\phi_{12}+\phi_{22}) + \min\{1-(\phi_{31}+\phi_{34}),1-(\phi_{41}+\phi_{44})\}]$.

Proof:

$$|\Delta_{TC}+\Delta_{CT}| \le \min\{1-(\phi_{31}+\phi_{34}),1-(\phi_{41}+\phi_{44})\}$$
$$\Rightarrow -\min\{1-(\phi_{31}+\phi_{34}),1-(\phi_{41}+\phi_{44})\} \le (\Delta_{TC}+\Delta_{CT}) \le \min\{1-(\phi_{31}+\phi_{34}),1-(\phi_{41}+\phi_{44})\}$$

But we showed that $\omega_2 = 0.5(\phi_{12}+\phi_{22}-(\Delta_{TC}+\Delta_{CT}))$. Thus,

$$0.5[(\phi_{12}+\phi_{22}) - \min\{1-(\phi_{31}+\phi_{34}),1-(\phi_{41}+\phi_{44})\}]$$
$$\le 0.5(\phi_{12}+\phi_{22}-(\Delta_{TC}+\Delta_{CT})) \le$$
$$0.5[(\phi_{12}+\phi_{22}) + \min\{1-(\phi_{31}+\phi_{34}),1-(\phi_{41}+\phi_{44})\}]$$

$$\Rightarrow L_2 \le \omega_2 \le U_2,$$

where, $L_2 = \max\{0, 0.5[(\phi_{12}+\phi_{22}) - \min\{1-(\phi_{31}+\phi_{34}),1-(\phi_{41}+\phi_{44})\}]\}$ and

$U_2 = 0.5[(\phi_{12}+\phi_{22}) + \min\{1-(\phi_{31}+\phi_{34}),1-(\phi_{41}+\phi_{44})\}]$ ■

Proposition 4.6.3 (See *Appendix D*)

Consider an extended data of observed counts on Table 4.11 shown below.

Table 4.11: Standard table of observed counts of responses to treatments in each period in a 2 treatments, 2 periods crossover design with focus on sequence

Sequence	$(0\ ,\ 0)$	$(0\ ,\ 1)$	$(1\ ,\ 0)$	$(1\ ,\ 1)$	Marginal Total
$1(TC)$	n_{11}	n_{12}	n_{13}	n_{14}	$n_{1\bullet}$
$2(CT)$	n_{21}	n_{23}	n_{22}	n_{24}	$n_{2\bullet}$
$3(TT)$	n_{31}	n_{32}	n_{33}	n_{34}	$n_{3\bullet}$
$4(CC)$	n_{41}	n_{42}	n_{43}	n_{44}	$n_{4\bullet}$

Estimating ϕ_{ki} for the TC – CT – TT – CC design follows the same pattern involving iterative numerical evaluations. Consider the likelihood function for the TC – CT – TT – CC as developed below.

Assume $(n_{k1}, n_{k2}, n_{k3}, n_{k4}) \sim multinomial(\phi_{k1}, \phi_{k2}, \phi_{k3}, \phi_{k4})$ for $i, k = 1,2,3,4$. The likelihood function is

$$L(n_{k1}, n_{k2}, n_{k3}, n_{k4} \mid \phi_{k1}, \phi_{k2}, \phi_{k3}, \phi_{k4}) \propto \prod_{k=1}^{4}\prod_{i=1}^{4} \phi_{ki}^{n_{ki}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (**)$$

where $\sum_{i=1}^{4} n_{ki} = n_{k\bullet}$, for $i = 1,2,3,4;\ k = 1,2$ and subject to the two constraints:

$\sum_{i=1}^{4} \phi_{1i} = 1$, $\sum_{i=1}^{4} \phi_{2i} = 1$, $\sum_{i=1}^{4} \phi_{3i} = 1$ and $\sum_{i=1}^{4} \phi_{4i} = 1$. Other constraints are $\phi_{12} = \phi_{21} + \phi_{22} - \phi_{11}$, $\phi_{23} = \phi_{11} + \phi_{13} - \phi_{21}$, $\phi_{32} = \phi_{11} + \phi_{12} - \phi_{31}$ and $\phi_{42} = \phi_{21} + \phi_{23} - \phi_{41}$ - restrictions that result from the following combinations of probabilities in observed outcomes for design TC – CT – TT – CC.

Thus, the likelihood function $(**)$ contains 8 distinct parameters leading to solving a system of 8 nonlinear equations with 8 constraints. Further we may reparameterize into a system of 8 nonlinear equations without constraints. However, like before, their solutions would require an iterative numerical method. Assuming estimates of ϕ_{ki}, denoted $\hat{\phi}_{ki}$, can be found, we proceed with stating estimates for the bounds. An estimate of the bounds, $[L_2, U_2]$, for ω_2 is given as $\left[\hat{L}_2, \hat{U}_2\right]$, where, $\hat{L}_2 = \max\left(0, 0.5\left[\left(\hat{\phi}_{12} + \hat{\phi}_{22}\right) - \min\left\{1-\left(\hat{\phi}_{31} + \hat{\phi}_{34}\right), 1-\left(\hat{\phi}_{41} + \hat{\phi}_{44}\right)\right\}\right]\right)$

and $\hat{U}_2 = 0.5\left[\left(\hat{\phi}_{12} + \hat{\phi}_{22}\right) + \min\left\{1-\left(\hat{\phi}_{31} + \hat{\phi}_{34}\right), 1-\left(\hat{\phi}_{41} + \hat{\phi}_{44}\right)\right\}\right]$.

4.7: Analyses with the Assumption of Exchangeability

So far, our analyses have been based on the fact that exchangeability does not hold. In the following sections, we assume exchangeability and investigate the impact on the analyses this far. The analyses involving exchangeability has been dealt with in Gadbury et al. (2004). In the following sections, we will explain the transition from the analyses without exchangeability and restate (where necessary and without proofs) the key results from the Gadbury et al. (2004) paper. As a consequence of exchangeability, equations (1.48) through (1.51) become

$$\phi_1 = \omega_1 - \Delta \qquad (1.68)$$

$$\phi_2 = \omega_2 + \Delta \qquad (1.69)$$

$$\phi_3 = \omega_3 + \Delta \qquad (1.70)$$

$$\phi_4 = \omega_4 - \Delta \qquad (1.71)$$

where $\phi_i = \phi_{ki} = \phi_{k'i}$ for $i = 1,2,3,4;\ k \neq k' = 1,2$ - for instance $\phi_1 = \phi_{11} = \phi_{21}, \phi_2 = \phi_{12} = \phi_{22}$, $\phi_3 = \phi_{11} = \phi_{23}$ - and thus, $\Delta = \Delta_{TC} = \Delta_{CT} = \beta_{14} - \beta_{23}$. From equations (1.69) through (1.72), we have the following adaptations.

4.7.1: Bounds for the Probability of Negative Effect with Exchangeability

Earlier, we saw that $\omega_2 \in [L_1, U_1]$ where $L_1 = \max\{0,\ 0.5[(\phi_{12} + \phi_{22}) - (\phi_{13} + \phi_{23})]\}$ and $U_1 = \min\{0.5[(\phi_{11} + \phi_{21}) + (\phi_{12} + \phi_{22})],\ 0.5[(\phi_{12} + \phi_{22}) + (\phi_{14} + \phi_{24})]\}$. With the assumption of exchangeability, $\phi_{11} = \phi_{21} = \theta_1$, $\phi_{12} = \phi_{22} = \theta_2$, $\phi_{13} = \phi_{23} = \theta_3$ and $\phi_{14} = \phi_{24} = \theta_4$ where the $\theta_i, i = 1,2,3,4$ are used as in Gadbury et al. (2004). Note that $\theta_1 = P(X_1 = 0\ ,\ Y_2 = 0)$, $\theta_2 = P(X_1 = 0\ ,\ Y_2 = 1), \theta_3 = P(X_1 = 1\ ,\ Y_2 = 0)$ and $\theta_4 = P(X_1 = 1\ ,\ Y_2 = 1)$ for the TC – CT crossover design and $\phi_{31} = P(X_1 = 0\ ,\ X_2 = 0)$, $\phi_{34} = P(X_1 = 1\ ,\ X_2 = 1)$, $\phi_{41} = P(Y_1 = 0\ ,\ Y_2 = 0)$ and $\phi_{44} = P(Y_1 = 1\ ,\ Y_2 = 1)$ for the additional TT – CC design. Thus, L_1 and U_1 become $[L_1^E, U_1^E]$ where

$L_1 = \max\{0,\ 0.5[2\theta_2 - 2\theta_3]\} = \max\{0,\ \theta_2 - \theta_3\} = L_1^E$ and

$U_1 = \min\{0.5(2\theta_1 + 2\theta_2),\ 0.5(2\theta_2 + 2\theta_4)\} = \min\{\theta_1 + \theta_2,\ \theta_2 + \theta_4\} = U_2^E$.

Thus, $L_1^E = \max\{0,\ \theta_2 - \theta_3\}$ and $U_2^E = \min\{\theta_1 + \theta_2,\ \theta_2 + \theta_4\}$. These are same bounds specified in Gadbury et al. (2004).

4.7.2: Refined Bound for the Probability of Negative Effect

Previously, we derived (1.53) that, $\omega_2 = 0.5\left(\phi_{12} + \phi_{22} - \left(\Delta_{TC} + \Delta_{CT}\right)\right)$. Imposing the exchangeability assumption leads to $\omega_2 = \theta_2 - \Delta$ where $\theta_2 = \phi_{12} = \phi_{22}$ and $\Delta = \Delta_{TC} = \Delta_{CT} = \beta_{14} - \beta_{23}$. This constitutes part of equation (4) in Gadbury et al. (2004) where $\Delta = \delta = \beta_{14} - \beta_{23}$. Furthermore, the assumption of exchangeability results in symmetry of probabilities on Table 4.4. Let $p_1 = E(X)$ and $p_2 = E(Y)$ be the mean of the marginal distributions of X and Y. Using symmetry, it can be proven that, $p_1 - \phi_{31} = 1 - p_1 - \phi_{34}$ and $p_2 - \phi_{41} = 1 - p_2 - \phi_{44}$. Applying these equations to the bounds $[L_2, U_2]$ gives bounds $\left[L_2^E, U_2^E\right]$ where, $L_2^E = \theta_2 - \min\{p_1 - \theta_3, p_2 - \theta_4\}$ and $U_2^E = \theta_2 + \min\{p_1 - \theta_3, p_2 - \theta_4\}$. These are the same bounds given in (*Proposition 3* of) Gadbury et al. (2004).

Having developed the theory, the next sections follow with a simulated illustrative example. We will illustrate the results outlined above on a simulated count data. We first state a joint probability distribution similar to that given on Table 4.4. Then, using Table 4.4, we simulate the observed count data, from which our maximum likelihood estimates are found using the *optim* procedure in *R* (www.r-project.com). Because the actual probabilities are known, we are able to find the true probability quantities and compare these with their respective estimates.

4.7.4: Illustrative Example 5: Simulated Observed Count Data

In the following example, we first state a joint probability distribution, β_{ij} (for $i, j = 1,2,3,4$), similar to that given on Table 4.4. Table 4.12 below gives the actual joint probability distribution of the potential outcomes for response variable X and Y.

Table 4.12*: Joint probabilities distribution obtained from the simulated example involving 600 subjects with focus on period 1 marginals using the TC – CT design without the exchangeability assumption*

		Period 2				
	For an Individual	$\overset{T\ C}{(x_2, y_2) = (0,0)}$	$\overset{T\ C}{(x_2, y_2) = (0,1)}$	$\overset{T\ C}{(x_2, y_2) = (1,0)}$	$\overset{T\ C}{(x_2, y_2) = (1,1)}$	Marginal Total
Per. 1	$\overset{T\ C}{(x_1, y_1) = (0,0)}$	0.016	0.065	0.032	0.048	0.161
	$\overset{T\ C}{(x_1, y_1) = (0,1)}$	0.032	0.016	0.113	0.081	0.242
	$\overset{T\ C}{(x_1, y_1) = (1,0)}$	0.081	0.032	0.048	0.113	0.274
	$\overset{T\ C}{(x_1, y_1) = (1,1)}$	0.032	0.129	0.081	0.081	0.323
	Marginal Total	0.161	0.242	0.274	0.323	1.000

Clearly, exchangeability condition does not hold. The marginal probabilities are $(\omega_1, \omega_2, \omega_3, \omega_4) = (0.161, 0.242, 0.274, 0.323)$. The parameter values for the unmatched design are $p_1 = E(X) = 0.597$ and $p_2 = E(Y) = 0.565$. Hence, the true average treatment T effect relative to treatment C is $E(D) = 0.032$. A sample of 600 observations was generated from Table 4.12. Furthermore, from these 600 observations, $n_1 = 200$ and $n_2 = 200$ were randomly selected to receive treatment (T, C) and (C, T), in that order, respectively. Responses are either "1" – for success – or "0" – for failure. The values of the parameters under the TC – CT design

are $\phi_{11} = 0.193$, $\phi_{12} = 0.210$, $\phi_{13} = 0.242$ and $\phi_{14} = 0.355$ for the TC sequence and $\phi_{21} = 0.194$, $\phi_{22} = 0.209$, $\phi_{23} = 0.241$ and $\phi_{24} = 0.356$ for the CT sequence. We also have $\Delta_{TC} + \Delta_{CT} = -0.065$. The observed count data was also recorded. Table 4.13 below shows observed count data corresponding to the given outcomes.

Table 4.13: *Standard table of observed counts in a TC - CT crossover design for the simulated example*

Sequence	$(0\ ,\ 0)$	$(0\ ,\ 1)$	$(1\ ,\ 0)$	$(1\ ,\ 1)$	Marginal Total
$1(TC)$	42	39	59	60	200
$2(CT)$	39	52	50	59	200

Form Table 4.13, the maximum likelihood estimates of the probabilities, $\hat{\phi}_{ki}$, were calculated using the numerical iterative method, *optim* in *R*. The result is presented on Table 4.14 below.

Table 4.14: *Estimates of probabilities of the observed data* $\hat{\phi}_{ki}$ *for the TC - CT crossover design*

Sequence	$(0\ ,\ 0)$	$(0\ ,\ 1)$	$(1\ ,\ 0)$	$(1\ ,\ 1)$	Marginal Total
$1(TC)$	0.2240	0.2295	0.2676	0.2789	1.0000
$2(CT)$	0.2118	0.2798	0.2417	0.2667	1.0000

The standard errors for these estimated probabilities are given on Table 4.15 below.

Table 4.15: *Standard error of estimates* $\hat{\phi}_{ki}$ *for the TC – CT design*

Sequence	$(0\ ,\ 0)$	$(0\ ,\ 1)$	$(1\ ,\ 0)$	$(1\ ,\ 1)$
$1(TC)$	0.0246`	0.0257	0.0250	0.0251
$2(CT)$	0.0246	0.0260	0.0245	0.0249

Using Table 4.14, the estimated average treatment effect, $\hat{E(D)}$, was found to be 0.0382. That is, on average, 3.82% subjects succeeded on treatment T relative to success on C. The actual value of the inestimable probability of negative effect, ω_2, is 0.242. In the theoretical analysis, we showed that this number is not measurable. Thus, we estimate bounds. The estimate of the lower bound for ω_2 is $\hat{L}_1 = 0$ and the estimate of the upper bound is $\hat{U}_1 = 0.4535$. That is, $\omega_2 \in [0, 0.4535]$. Thus, between 0% and 45.4% of the subjects succeeded on C and failed on T.

Furthermore, from the 600 observations, $n_3 = 100$ and $n_4 = 100$ were assigned to sequence TT and CC respectively. For the TT – CC design, the values of the parameters are $\phi_{31} = 0.129$, $\phi_{32} = 0.274$, $\phi_{33} = 0.274$ and $\phi_{34} = 0.323$ for the TT sequence and $\phi_{41} = 0.177$, $\phi_{42} = 0.258$, $\phi_{43} = 0.258$ and $\phi_{44} = 0.307$ for the CC sequence. Table 4.16, shows the added observed count data.

Table 4.16: *Standard table of observed counts in the a TC – CT – TT – CC design*

Sequence	$(0\ ,\ 0)$	$(0\ ,\ 1)$	$(1\ ,\ 0)$	$(1\ ,\ 1)$	Marginal Total
$3(TT)$	12	25	27	36	100
$4(CC)$	20	30	20	30	100

The multinomial parameter estimates using the observed count data on Table 4.16 are given as on Table 4.17 below

Table 4.17: *Estimates of probabilities of the observed data* $\hat{\phi}_{ki}$ *for the* TC – CT – TT – CC design

Sequence	$(0\ ,\ 0)$	$(0\ ,\ 1)$	$(1\ ,\ 0)$	$(1\ ,\ 1)$	Marginal Total
$1(TC)$	0.2086	0.2167	0.2823	0.2925	1.0000
$2(CT)$	0.2113	0.2796	0.2139	0.2952	1.0000
$3(TT)$	0.1461	0.2792	0.3281	0.2467	1.0000
$4(CC)$	0.2259	0.2650	0.1334	0.3757	1.0000

The standard errors for these estimated probabilities given on Table 4.17 are as shown on Table 4.18 below

Table 4.18: *Standard errors of estimates* $\hat{\phi}_{ki}$ *for the TC – CT –* TT – CC *design*

Sequence	$(0\ ,\ 0)$	$(0\ ,\ 1)$	$(1\ ,\ 0)$	$(1\ ,\ 1)$
$1(TC)$	0.0229	0.0233	0.0239	0.0246
$2(CT)$	0.0230	0.0249	0.02224	0.0243
$3(TT)$	0.0346	0.0365	0.0378	0.0351
$4(CC)$	0.0369	0.0366	0.0276	0.0335

The estimated refined lower bound for ω_2 is $\hat{L}_2 = 0.0161$ and the estimate of the upper bound is $\hat{U}_2 = 0.4145$. This constitutes a 12.14% reduction on the previous bounds $\left[\hat{L}_1, \hat{U}_1\right] = [0, 0.4535]$ for the probability of negative effect, ω_2, established under the TC – CT crossover design. Thus, with a sample size of 600, we got a tighter bound upon addition of the information from subjects who stayed on the same treatments over the two periods.

We did not consider the case when exchangeability holds because an example was outlined in Gadbury et al. (2004).

CHAPTER 5 - Summary and Future Challenges

5.1: Summary

This dissertation was centered on using the potential outcomes method to estimate the individual treatment effect variability and a probability of a negative treatment effect in multiple time point settings. The assumptions were: no carryover effects, no covariate effects, no missing data, and a particular assumption about independence of time effects for different potential outcome variables. For a quantitative response, we analyzed a two-treatment, two-periods and a two-treatment, three-periods crossover design. We saw that estimation of the individual treatment effect variable was not possible with the two-treatment, two-periods crossover design unless we add the information provided by subjects under an added parallel design. Under the two-treatment, three-periods design, we proposed an estimate for the finite population treatment effect mean and variance. Furthermore, we estimated a parameter for the probability of negative effect. A simulated blood pressure data was use for illustration.

In the qualitative section, a binary, "0 – 1," response variable was analyzed. Using a given joint probability distribution of potential outcomes, we expressed the probability of the observed outcomes under a two treatment, two periods crossover design. Maximum likelihood estimates based on observed outcomes were found using an iterative numerical method. Using these estimates, we proposed bounds for an inestimable probability of a negative effect. Tighter bounds were obtained with information from subjects that received the same treatments over the two periods. Finally, for illustration, we used a simulated example of count data.

5.2: Future Challenges

We note that, the analyses assumed no carryover effect, no covariate effect and no missing data. It will be interesting to see how the results are affected when there is carryover effect. In addition, covariates like gender, age and others may be factored in. For a brief consideration, let's assume carryover effects are present. The next section introduces the analyses of a two treatment, two periods TC – CT crossover design when carryover effects are added.

5.2.1: Potential Outcome Framework with Carryover Effect – Quantitative Response Variable

In Section 1.3.2, we presented the potential outcome framework when there is no carryover effect. Such a framework was defined as

Subject	Period 1		Period 2	
1	$X_1 + t_{11}$	$Y_1 + \tau_{11}$	$X_1 + t_{12}$	$Y_1 + \tau_{12}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$2n$	$X_{2n} + t_{2n1}$	$Y_{2n} + \tau_{2n1}$	$X_{2n} + t_{2n2}$	$Y_{2n} + \tau_{2n2}$

with potential outcomes $\left(X_i + t_{i1},\ Y_i + \tau_{i1}\right)$ for period 1 and $\left(X_i + t_{i2},\ Y_i + \tau_{i2}\right)$ for period 2 $\left(i = 1, 2, ..., 2n\right)$.

Now, assume there is the effect carryover. The first question arises on how to factor the carryover effect into the potential outcome framework. For instance, for the i^{th} subject, let $\xi_{i1,j-1}$ and $\xi_{i2,j-1}$ denote the carryover effects of treatments C to T and T to C, respectively, administered in the j^{th} period. We assumed $\xi_{i1,0} = 0$ and $\xi_{i2,0} = 0$. In a design involving more

than two periods or with the added TT and CC designs, we may further define the carryover effect from treatment T to T and treatment C to C.

For a two treatment, two periods design, a possible potential outcomes framework is

Subject	Period 1		Period 2	
1	$X_1 + t_{11}$	$Y_1 + \tau_{11}$	$X_1 + t_{12} + \xi_{i1,1}$	$Y_1 + \tau_{12} + \xi_{i2,1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$2n$	$X_{2n} + t_{2n1}$	$Y_{2n} + \tau_{2n1}$	$X_{2n} + t_{2n2} + \xi_{i1,1}$	$Y_{2n} + \tau_{2n2} + \xi_{i2,1}$

That is, the potential outcomes are $\left(X_i + t_{i1},\ Y_i + \tau_{i1}\right)$ for period 1 and $\left(X_i + t_{i2} + \xi_{i1,1},\ Y_i + \tau_{i2} + \xi_{i2,1}\right)$ for period 2 $\left(i = 1, 2, ..., 2n\right)$.

Using this framework for potential outcome in a TC – CT randomization, the observed outcomes framework become

			Period 1	Period 2
Sequence	T	C	$X_i + t_{i1}$	$Y_i + \tau_{i2} + \xi_{i2,1}$
	C	T	$Y_i + \tau_{i1}$	$X_i + t_{i2} + \xi_{i1,1}$

The observed treatment effect is then defined accordingly. Detailed development on this topic and more is left for further research.

In the situation with binary outcomes, carryover effect could imply the marginals, at the two time periods, are not the same as assumed on Table 4.4.

5.2.2. Use of covariates

As noted in Section 2.1 for quantitative outcomes in the TC, CT, TT, CC design, the variance of individual effects could be estimated assuming independence of t and τ. If these are not independent, then only bounds for the variance can be estimated. Producing these bounds and their estimates would be interesting for future research. When covariates (i.e., variables not affected by treatment such as baseline measurements) are available, they may be used to tighten bounds. In such cases, a large lower bound for the variance of individual effects may then be evidence of substantive individual treatment effect heterogeneity.

5.1.3 Missing Data

Missing data may occur in a multiple time point trial when individuals drop out for various reasons. It has been assumed herein that complete data are available, that is, individuals complete the trial. This is equivalent to data missing completely at random. Missing data due to drop out might not be missing at random and this fact would add complexity in even obtaining unbiased estimates of the mean treatment effect. How such missing data would affect the variance of individual effects (or bounds for this variance) could be another avenue of future investigation.

REFERENCES

Appel, J. L., Moore, J. T., Obarzanek. E., Vollmer, M. W., Svetkey, P. L., Sacks, M. F., Bray, A. G., Vogt, M. T., Cutler, A. J., Most-Windhauser, M. M., Lin. P.-H., and Karanja, N. (April 1997). "A Clinical Trial of the Effects of Dietary Patterns on Blood Pressure." *The New England Journal of Medicine*, Vol. 336, no. 16.

Broota, D. K. (1989). "Experimental Design in Behavioral Research John." *Wiley and Sons Ltd*. ISBN 0-470-21641-7.

Creticos, P. S., Adams, W. P., Petty, B. G., Lewis, L. D., MBBCh, Singh, G. J. P., Khattignavong, A. P., Molzon, J. A., Martinez, N. M. Lietman, S. P., Williams, L. R.(2002), "A Methacholine Challenge Dose-Response Study for Development of a Pharmacodynamic Bioequivalence Methodology for Albuterol Metered-Dose Inhalers." *J. Allergy Clin. Immun.*, 110, pp. 713–720.

Dodd, M.G. and Murphy, T.M. (1995). "Accuracy and Precision of Techniques for Counting Great Blue Heron Nests." *J. Wildlife Management*, 59, pp. 667–673.

Egan, M. B. (2003). "Reproducibility of BP Responses to Changes in Dietary Salt: Compelling Evidence for Universal Sodium Restriction? (Editorial Commentary)." *The American Heart Assocition, Inc*., 42, 457.

Gadbury, G. L. (2001). "Randomization Inference and Bias of Standard Errors." *The American Statistician,* Vol. 55, No. 4.

Gadbury, G. L., Iyer, H., Albert, M. J. (2004). "Individual Treatment Effects in Randomized Trials with Binary Outcomes." *Journal of Statistical Planning and Inference*, 121, 163 – 174.

Gadbury, L. G. (2004). "Subject-Treatment Interaction." *Encyclopedia of Biopharmaceutical Statistics*, DOI:10.108/E-EBS 120022255

Gadbury, L. G. and Iyer, H. (2000). "Unit-Treatment Interaction and Its Practical Consequences." *Biometrics*, Vol. 56, No. 3, pp. 882-885.

Gadbury, L. G. and Schreuder, T. H. (2003). "Cause-Effect Relationship in Analytical Surveys: An Illustration of Statistical Issues." *Environmental Monitoring and Assessment* 83, 205 – 227.

Glymour C. (1986). "Statistics and Causal Inference: Comment: Statistics and Metaphysis." Journal of American Statistical Association, Vol. 81, No. 396, 964-966.

Graybill, A. F. (1976). "Theory and Applications of Linear Model." *Duxbury Classic Series*. ISBN 0-534-38019-0

Graybill, A. F. (1983). "Matrices with Applications in Statistics." 2nd Ed. *Duxbury Classic Series.* ISBN 0-5434-40131-7.

Hamer, A.J., Strachan, J.R., Black, M.M., Ibbotson, C. and R.A. Elson (1995). "A new method of Comparative Bone Strength Measurement." *Journal of Medical Engineering and Technology*, 19, pp. 1–5.

Hendricks, W.A. and Robey, W.K.(1936). "The Sampling Distribution of the Coefficient of Variation." *Annals of Mathematical Statistics*, **7**, pp. 129–132.

Holland, P. W. (1986). "Statistics and causal inference". *Journal of the American Statistical Association* Vol. 81, No. 396, 945–970.

Jones, B. and Kenward, G. M. (1989). "Design and Analysis of Cross-Over Trials." *Chapman and Hall Ltd*, ISBN 0 412 30000 1

Kravitz, L. R. , Duan, H. N. and Baslow, J. (2004). "Evidence-Based Medicine, Heterogeneity of Treatment Effects and The Trouble with Averages." *Milbank Quarterly*, Vol. 82(4), pp.661-687

Kuehl, O. R. (1994). "Design of Experiments: Statistical Principles of Research Design and Analysis." 2nd Ed. *Duxbury Express*. ISBN 0-534-36834-4.

Liu, W., Pang, Wan-K. and Huang, Wei-K. (2006). **"**Exact Confidence Bounds for the Coefficient of Variation of a Normally Distributed Population." *International Journal of Statistics and Systems (IJSS)*, Vol.1 No.1, pp. 81-86.

Little, R.J.A. and D.B. Rubin (2002). "Statistical analysis with missing data." *New York: Wiley*

Longford, N. (1999). "Selection Bias and Treatment Heterogeneity in Clinical Trials." *Statistics in Medicine*, 18, 1467-1474.

March, L., Irwig. L., Schwarz, J., Simpson, J., Chock, C., Brooks, P. (1994). "n of 1 trials comparing a non-steroidal anti-inflammatory drug with paracetamol in osteoarthritis." *British Medical Journal (BMJ),* Vol. 309, 1041 – 1046.

Marshall, A., (1997). "Getting the Right Drug into the Right Patient", *Nature Biotechnology*, Vol. 15.

McKay (1931). "The Distribution of the Estimated Coefficient of Variation." *Journal of the Royal Statistical Society*, Vol. 94, No. 4 pp. 564-567.

Miller, E.G. and Karson, M.J.(1977). "Testing the Equality of Two Coefficients of Variation." *American. Statistical. Assoc.: Proceedings of Business and Economics Section*, pp. 278–283.

Neyman, J. (1923), "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles (with discussion)." *Section 9 (In Polish). Statistical Science*, 5, 465–480.

Obarzanek, E., Proschan, A. M., Vollmer, M. W., Moore J. T., Sacks M. F., Appel J. L., Svetkey P. L., Most-Windhauser M. M., Cutler A. J. (2003). "Individual Blood Pressure Responses to Changes in Salt Intake: Result From the DASH-Sodium Trial." *The American Heart Association, Inc*., 42, 459.

Ratkowsky, A. D., Evans, A. M. and Alldredge, J. R. (1993). "Cross-Over Experiments-Design, Analysis and Applications." *Dekker*, Vol. 135

Rubin, B. D. ((2005). "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions." *Journal of the American Statistical Association*, Vol. 100, No. 469.

Rubin, B. D. (1980). "Discussion of "Randomization Analysis of Experimental Data: The Fisher Randomization Test," by D. Basu", *Journal of the American Statistical Association*, Vol. 75, pp. 591-593.

Rubin, D. B. (1974). "Estimating causal effects of treatments in randomized and nonrandomized studies." *J. Educational Psychology* 66 688–701.

Rubin, D. B. (1990). "Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies." *Statistical Science* 5, 472-480.

Saw, L. C. S. (1992). "ANOVA Sums of Squares as Quadratic Forms." *The American Statistician*, Vol. 46, pp. 288-290

Sekhon, S. J. (2007). "The Neyman – Rubin Model of Causal Inference and Estimation via Matching Methods" *The Oxford Handbook of Political Methodology*, Chap. 11.

Senn, S. (2001). "Individual Response to Treatment: Is It a Valid Assumption?" *British Medical Journal (BMJ)*, Vol. 329.

Senn, S. (2001). "Individual Therapy: New Dawn or False Dawn?" *Drug Information Journal*, Vol. 35, pp. 1479 – 1494.

Smith, T. M. F., Sugden R. A. (1988). "Sampling and assignment mechanisms in experiments, surveys and observational studies" *Review of the International Statistical Institute,* Vol.56, No. 2, 165-180.

Taubes, G. (May 2000). "A DASH of Data in the Salt Debate." *Science* 26: Vol. 288. No. 5470, p. 1391.

Vangel, M.G. (1996). "Confidence Intervals for a Normal Coefficient of variation." *American Statistician*, 50, pp. 21–26.

Verrill, S. (2003), "Confidence Interval for Normal and Lognormal Coefficient of Variation." *Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory*, Res. Pap. FPL-RP-609, p. 13.

Yelland, M., Nikles, C. J., McNairn, N., Mar, D. C., Schluter, P. (2006). "Do n-of-1 trials of celecoxib compared to sustained-release paracetamol change patient use of drugs for osteoarthritis?" Bond University.

Appendix A - Proof of Proposition 2.2

Proposition 2.2

$$E\left(\frac{2n-1}{2n}S_d^2\right) = S_D^2 + Bias$$

where $Bias = \frac{2n-1}{2n}\left(S_{t+\tau}^2 + \left(\overline{t+\tau}\right)^2\right)$

$S_{t+\tau}^2$ and $\overline{t+\tau}$ are the population variance and population mean of the sum of time effect terms $t+\tau$.

Proof:

We observe that

$$E_\gamma(S_d^2) = E_\gamma\left[\frac{1}{2n-1}\sum_{i=1}^{2n}(d_i-\bar{d})^2\right]$$

$$= E_\gamma\left[\frac{1}{2n-1}\left(\sum_{i=1}^{2n}d_i - 2n\bar{d}^2\right)\right]$$

$$= \frac{1}{2n-1}\left(\sum_{i=1}^{2n}E_\gamma(d_i^2) - \frac{1}{2n}E_\gamma\left(\sum_{i=1}^{2n}d_i\right)^2\right)$$

$$= \frac{1}{2n-1}\left(\sum_{i=1}^{2n}E_\gamma(d_i^2) - \frac{1}{2n}E_\gamma\left(\sum_{i=1}^{2n}d_i^2 + \underbrace{\sum_{i=1}^{2n}\sum_{i'=1}^{2n}d_i d_{i'}}_{i\neq i'}\right)\right)$$

$$= \frac{1}{2n-1}\left(\frac{2n-1}{2n}\sum_{i=1}^{2n} E_\gamma(d_i^2) - \frac{1}{2n}\underbrace{\sum_{i=1}^{2n}\sum_{i'=1}^{2n} E_\gamma(d_i d_{i'})}_{i\neq i'} \right)$$

$$= \frac{1}{2n}\left(\sum_{i=1}^{2n} E_\gamma(d_i^2) - \frac{1}{2n-1}\underbrace{\sum_{i=1}^{2n}\sum_{i'=1}^{2n} E_\gamma(d_i d_{i'})}_{i\neq i'} \right)$$

But,

$$\begin{aligned} E_\gamma\left(d_i^2\right) &= \left(X_i - Y_i\right)^2 - 2\left(X_i - Y_i\right)E_\gamma(t_{i1}+\tau_{i1})(2\gamma_{i1}-1) + E_\gamma(t_{i1}+\tau_{i1})^2(2\gamma_{i1}-1)^2 \\ &= \left(X_i - Y_i\right)^2 + \left(t_{i1}+\tau_{i1}\right)^2\left(4E_\gamma\left(\gamma_{i1}^2\right) - 4E_\gamma(\gamma_{i1}) + 1\right)^2 \\ &= \left(X_i - Y_i\right)^2 + \left(t_{i1}+\tau_{i1}\right)^2 \end{aligned}$$

since $E_\gamma\left(\gamma_{i1}^2\right) = E_\gamma(\gamma_{i1}) = 1/2$ means $\left(4E_\gamma\left(\gamma_{i1}^2\right) - 4E_\gamma(\gamma_{i1}) + 1\right) = 0$.

$$\begin{aligned} E_\gamma\left(d_i d_{i'}\right) &= \left(X_i - Y_i\right)\left(X_{i'} - Y_{i'}\right) + \left(t_{i1} - \tau_{i1}\right)\left(t_{i'1} - \tau_{i'1}\right)E_\gamma\left(2\gamma_{i1}-1\right)\left(2\gamma_{i'1}-1\right) \\ &= \left(X_i - Y_i\right)\left(X_{i'} - Y_{i'}\right) + \left(t_{i1} - \tau_{i1}\right)\left(t_{i'1} - \tau_{i'1}\right)E_\gamma\left(4\gamma_{i1}\gamma_{i'1} - 2\gamma_{i1} - 2\gamma_{i'1} + 1\right) \\ &= \left(X_i - Y_i\right)\left(X_{i'} - Y_{i'}\right) \end{aligned}$$

since $E_\gamma\left(\gamma_{i1}\gamma_{i'1}\right) = 1/4$ and $E_\gamma\left(\gamma_{i1}\right) = 1/2$ means $E_\gamma\left(4\gamma_{i1}\gamma_{i'1} - 2\gamma_{i1} - 2\gamma_{i'1} + 1\right) = 0$. Thus,

$$E_\gamma(S_d^2)=\frac{1}{2n}\left(\sum_{i=1}^{2n}\left(\left(X_i-Y_i\right)^2+\left(t_{i1}+\tau_{i1}\right)^2\right)-\frac{1}{2n-1}\sum_{i=1}^{2n}\sum_{i'=1}^{2n}\left(X_i-Y_i\right)\left(X_{i'}-Y_{i'}\right)\right)$$

$$=\frac{1}{2n}\left(\sum_{i=1}^{2n}\left(D_i^2+\left(t_{i1}+\tau_{i1}\right)^2\right)-\frac{1}{2n-1}\sum_{i=1}^{2n}\sum_{i'=1}^{2n}D_iD_{i'}\right)$$

$$=\frac{1}{2n-1}\left(\sum_{i=1}^{2n}D_i^2-2n\bar{D}_i^2\right)+\frac{1}{2n}\left(\sum_{i=1}^{2n}\left(t_{i1}+\tau_{i1}\right)^2\right)$$

$$=\frac{1}{2n-1}\left(\sum_{i=1}^{2n}D_i^2-2n\bar{D}_i^2\right)+\frac{1}{2n}\left(\sum_{i=1}^{2n}\left(t_{i1}+\tau_{i1}\right)^2-2n\left(\overline{t+\tau}\right)^2\right)+\left(\overline{t+\tau}\right)^2$$

$$=\frac{2n}{2n-1}S_D^2+S_{t+\tau}^2+\left(\overline{t+\tau}\right)^2$$

Thus,

$$E\left(\frac{2n-1}{2n}S_d^2\right)=S_D^2+Bias$$

where $Bias=\frac{2n-1}{2n}\left(S_{t+\tau}^2+\left(\overline{t+\tau}\right)^2\right)$ ■

It is worthwhile noting that, if we assume that the randomization must result in equal number of subjects in each sequence, we still obtain the same result above because, in this case,

$$E\left(\gamma_{ij}\gamma_{i'j}\right)=\begin{cases}\frac{1}{2}\left(\frac{n-1}{2n-1}\right) & \text{if } j=j'\\ \frac{1}{2}\left(\frac{n}{2n-1}\right) & \text{if } j\neq j'\end{cases}, \qquad i\neq i'\ \left(i=1,2,...2n\ ;j=1,2\right)$$ and

$$E_\gamma\left(\left(t_{i1}-\tau_{ij}\right)\left(t_{i'1}-\tau_{i'j'}\right)\left(2\gamma_{ij}-1\right)\left(2\gamma_{i'j'}-1\right)\right)=\left(t_{i1}-\tau_{ij}\right)\left(t_{i'1}-\tau_{i'j'}\right)E_\gamma\left(4\gamma_{ij}\gamma_{i'j'}-2\gamma_{ij}-2\gamma_{i'j'}+1\right)$$

$$=\left(t_{i1}-\tau_{ij}\right)\left(t_{i'1}-\tau_{i'j'}\right)\left(-\frac{1}{2n-1}+\frac{1}{2n-1}\right)=0$$

$$E_\gamma\left(d_id_{i'}\right)=\left(X_i-Y_i\right)\left(X_{i'}-Y_{i'}\right)+\left(t_{i1}-\tau_{i1}\right)\left(t_{i'1}-\tau_{i'1}\right)E_\gamma\left(2\gamma_{i1}-1\right)\left(2\gamma_{i'1}-1\right)$$

$$=\left(X_i-Y_i\right)\left(X_{i'}-Y_{i'}\right)$$

Appendix B - Proof of Proposition 3.2

Proposition 3.2

$$E_\delta\left(\frac{N-1}{N}S_d^2\right) = S_D^2 + Bias_{IND}$$

where

$$Bias_{IND} = \frac{N-1}{3N^2}\sum_{i=1}^{N}\sum_{j=1}^{3}\alpha_{ij}^2 \text{ and } \alpha_{ij} = \left(\frac{t_{ij}}{2} + \tau_{ij}\right)$$

Proof:

We observe that

$$E_\delta(d_i) = (X_i - Y_i) - 2(X_i - Y_i)\sum_{j=1}^{3}\alpha_{ij}E(\delta_j)$$

$$= (X_i - Y_i) - \frac{2}{3}(X_i - Y_i)\sum_{j=1}^{3}\alpha_{ij} = (X_i - Y_i)$$

Also,

$$d_i^2 = (X_i - Y_i)^2 - 2(X_i - Y_i)\sum_{j=1}^{3}\alpha_{ij}\delta_{ij} + \left(\sum_{j=1}^{3}\alpha_{ij}\delta_{ij}\right)^2$$

$$d_i^2 = (X_i - Y_i)^2 - 2(X_i - Y_i)\sum_{j=1}^{3}\alpha_{ij}\delta_j + \sum_{j=1}^{3}\alpha_{ij}{}^2\delta_j{}^2 + \underbrace{\sum_{j=1}^{3}\sum_{j'=1}^{3}\alpha_{ij}\alpha_{ij'}\delta_j\delta_{j'}}_{j\neq j'}$$

$$E_\delta(d_i^2) = (X_i - Y_i)^2 - 2(X_i - Y_i)\sum_{j=1}^{3}\alpha_{ij}E(\delta_j) + \sum_{j=1}^{3}\alpha_{ij}^2 \underbrace{E(\delta_j^2)}_{=E(\delta_j)} + \underbrace{\sum_{j=1}^{3}\sum_{j'=1}^{3}\alpha_{ij}\alpha_{ij'}E(\delta_j\delta_{j'})}_{j\neq j'}$$

But, $\delta_j \sim \text{Binomial}(1,1/3)$. Thus, $P(\delta_{ij} = 1) = 1/3$ for all j and

$$\begin{aligned} CTT \} \ \delta_{i1} \ or \ \delta_1 = 1,\ \delta_2 = 0,\ \delta_3 = 0 \\ TCT \} \ \delta_{i2} \ or \ \delta_2 = 1,\ \delta_1 = 0,\ \delta_3 = 0 \\ TTC \} \ \delta_{i3} \ or \ \delta_3 = 1,\ \delta_1 = 0,\ \delta_2 = 0 \end{aligned}$$

$\Rightarrow E_\delta(\delta_j) = 1/3,\ E_\delta(\delta_j\delta_{j'}) = P(\delta_j = 1,\ \delta_{j'} = 1) = P(\delta_j = 1)P(\delta_{j'} = 1) = 0$ (for each i and $j \neq j'$)

and $E_\delta(\delta_j^2) = E_\delta(\delta_j)$

Thus,

$$E_\delta(d_i^2) = (X_i - Y_i)^2 - \frac{2}{3}(X_i - Y_i)\sum_{j=1}^{3}\alpha_{ij} + \frac{1}{3}\sum_{j=1}^{3}\alpha_{ij}^2$$

$$= (X_i - Y_i)^2 + \frac{1}{3}\sum_{j=1}^{3}\alpha_{ij}^2 \qquad \left(\text{since } \sum_{j=1}^{3}\alpha_{ij} = 0\right)$$

Now, we have

$$E_\delta(S_d^2) = E_\delta\left[\frac{1}{N-1}\sum_{i=1}^{N}(d_i - \bar{d})^2\right]$$

$$= E_\delta\left[\frac{1}{N-1}\left(\sum_{i=1}^{N}d_i - N\bar{d}^2\right)\right]$$

$$= \frac{1}{N-1}\left(\sum_{i=1}^{N}E_\delta(d_i^2) - \frac{1}{N}E_\delta\left(\sum_{i=1}^{N}d_i\right)^2\right)$$

$$= \frac{1}{N-1}\left(\sum_{i=1}^{N} E_\delta(d_i^2) - \frac{1}{N} E_\delta\left(\sum_{i=1}^{N} d_i^2 + \underbrace{\sum_{i=1}^{N}\sum_{i'=1}^{N} d_i d_{i'}}_{i \neq i'} \right) \right)$$

$$= \frac{1}{N-1}\left(\frac{N-1}{N} \sum_{i=1}^{N} E_\delta(d_i^2) - \frac{1}{N} \underbrace{\sum_{i=1}^{N}\sum_{i'=1}^{N} E_\delta(d_i d_{i'})}_{i \neq i'} \right)$$

$$= \frac{1}{N}\left(\sum_{i=1}^{N} E_\delta(d_i^2) - \frac{1}{N-1} \underbrace{\sum_{i=1}^{N}\sum_{i'=1}^{N} E_\delta(d_i d_{i'})}_{i \neq i'} \right) = \frac{1}{N}\left(\sum_{i=1}^{N} E_\delta({d_i}^2) - \frac{1}{N-1} \underbrace{\sum_{i=1}^{N}\sum_{i'=1}^{N} E_\delta(d_i) E_\delta(d_{i'})}_{i \neq i'} \right)$$

Under independence, $E_\delta(d_i d_{i'}) = E_\delta(d_i) E_\delta(d_{i'})$.

$$E_\delta(S_d^2) = \frac{1}{N}\left(\sum_{i=1}^{N} (X_i - Y_i)^2 + \frac{1}{3} \sum_{i=1}^{N}\sum_{j}^{3} {\alpha_{ij}}^2 - \frac{1}{N-1} \underbrace{\sum_{i=1}^{N}\sum_{i'=1}^{N} (X_i - Y_i)(X_j - Y_{j'})}_{i \neq i'} \right)$$

$$= \frac{1}{N}\left(\sum_{i=1}^{N} (X_i - Y_i)^2 + \frac{1}{3} \sum_{i=1}^{N}\sum_{j}^{3} {\alpha_{ij}}^2 - \frac{1}{N-1}\left(\left\{ \sum_{i=1}^{N} (X_i - Y_i) \right\}^2 - \sum_{i=1}^{N} (X_i - Y_i)^2 \right) \right)$$

because $\left(\underbrace{\sum_{i=1}^{N}\sum_{i'=1}^{N} (X_i - Y_i)(X_j - Y_{j'})}_{i \neq i'} \right) = \left(\sum_{i=1}^{N} (X_i - Y_i) \right)^2 - \sum_{i=1}^{N} (X_i - Y_i)^2$.

Thus,

$$E_\delta(S_d^2) = \frac{1}{N}\left(\sum_{i=1}^{N} D_i^2 + \frac{1}{3} \sum_{i=1}^{N}\sum_{j}^{3} \alpha_{ij}^2 - \frac{1}{N-1}\left(\left\{ N\bar{D} \right\}^2 - \sum_{i=1}^{N} D_i^2 \right) \right)$$

$$= \frac{1}{N-1}\left(\sum_{i=1}^{N} D_i^2 - N\bar{D}^2 + \frac{N-1}{3N} \sum_{i=1}^{N} \sum_{j}^{3} \alpha_{ij}^2 \right)$$

$$= \frac{1}{N-1}\left(NS_D^2 + \frac{N-1}{3N} \sum_{i=1}^{N} \sum_{j}^{3} \left(\frac{t_{ij}}{2} + \tau_{ij} \right)^2 \right)$$

Thus, $E_\delta\left(\frac{N-1}{N} S_d^2\right) = S_D^2 + \frac{N-1}{3N^2} \sum_{i=1}^{N} \sum_{j=1}^{3} \left(\frac{t_{ij}}{2} + \tau_{ij} \right)^2$

$$= S_D^2 + Bias_{IND}$$ ■

Appendix C - Proof of Proposition 3.3

Proposition 3.3

$$E_\delta\left(\frac{N-1}{N}S_d^2\right) = S_D^2 + Bias_{DEP}$$

where

$$Bias_{DEP} = \frac{1}{3N^2}\left[(N-1)\sum_{i=1}^{N}\sum_{j=1}^{3}\alpha_{ij}^2 - \frac{n-1}{N-1}\underbrace{\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3}\alpha_{ij}\alpha_{i'j}}_{i\neq i'} - \frac{n}{N-1}\underbrace{\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3}\alpha_{ij}\alpha_{i'j'}}_{(i\neq i')\ and\ (j\neq j')}\right]$$

where, $\alpha_{ij} = \left(\frac{t_{ij}}{2} + \tau_{ij}\right)$

Proof:

From Appendix B

$$E_\delta(S_d^2) = E_\delta\left[\frac{1}{N-1}\sum_{i=1}^{N}(d_i - \bar{d})^2\right]$$

$$= \frac{1}{N}\left(\sum_{i=1}^{N}E_\delta(d_i^2) - \frac{1}{N-1}\underbrace{\sum_{i=1}^{N}\sum_{i'=1}^{N}E_\delta(d_i d_{i'})}_{i\neq i'}\right) \text{----------(**)}$$

We have that,

$$E_\delta(\underbrace{d_i d_{i'}}_{i\neq i'}) = E_\delta\left[\left(\left(X_i - Y_i\right) - \sum_{j=1}^{3}\alpha_{ij}\delta_{ij}\right)\left(\left(X_{i'} - Y_{i'}\right) - \sum_{j'=1}^{3}\alpha_{i'j'}\delta_{i'j'}\right)\right] \qquad (i \neq i')$$

$$= E_\delta\left[\left(\left(X_i - Y_i\right)\left(X_{i'} - Y_{i'}\right) - \left(X_{i'} - Y_{i'}\right)\sum_{j=1}^{3}\alpha_{ij}\delta_{ij} - \left(X_i - Y_i\right)\sum_{j'=1}^{3}\alpha_{i'j'}\delta_{i'j'} + \sum_{j=1}^{3}\alpha_{ij}\delta_{ij}\sum_{j'=1}^{3}\alpha_{i'j'}\delta_{i'j'}\right)\right]$$

$$= \left(\left(X_i - Y_i\right)\left(X_{i'} - Y_{i'}\right) - \left(X_{i'} - Y_{i'}\right)\sum_{j=1}^{3}\alpha_{ij}E_\delta(\delta_{ij}) - \left(X_i - Y_i\right)\sum_{j'=1}^{3}\alpha_{i'j'}E_\delta(\delta_{i'j'}) + \sum_{j=1}^{3}\sum_{j'=1}^{3}\alpha_{ij}\alpha_{i'j'}E_\delta(\delta_{ij}\delta_{i'j'})\right)$$

$$= \left(\left(X_i - Y_i\right)\left(X_{i'} - Y_{i'}\right) - \underbrace{\frac{1}{3}\left(X_{i'} - Y_{i'}\right)\sum_{j=1}^{3}\alpha_{ij}}_{=0} - \underbrace{\frac{1}{3}\left(X_i - Y_i\right)\sum_{i=1}^{n}\alpha_{i'j'}}_{=0} + \sum_{j=1}^{3}\sum_{j'=1}^{3}\alpha_{ij}\alpha_{i'j'}E_\delta(\delta_{ij}\delta_{i'j'})\right) \qquad (i \neq i')$$

$$= \left(\left(X_i - Y_i\right)\left(X_{i'} - Y_{i'}\right) + \sum_{j=1}^{3}\sum_{j'=1}^{3}\alpha_{ij}\alpha_{i'j'}E_\delta(\delta_{ij}\delta_{i'j'})\right) \qquad (i \neq i')$$

Furthermore, $\sum_{j=1}^{3}\sum_{j'=1}^{3}\alpha_{ij}\alpha_{i'j'}E_\delta(\delta_{ij}\delta_{i'j'}) = \begin{cases} \dfrac{n-1}{3(N-1)}\displaystyle\sum_{j=1}^{3}\sum_{j'=1}^{3}\alpha_{ij}\alpha_{i'j'} & (j = j') \\ \dfrac{n}{3(N-1)}\displaystyle\sum_{j=1}^{3}\sum_{j'=1}^{3}\alpha_{ij}\alpha_{i'j'} & (j \neq j') \end{cases} \qquad (i \neq i')$

Hence from (**),

$$E_\delta(S_d{}^2) = \frac{1}{N}\left(\sum_{i=1}^{N}E_\delta\left(d_i{}^2\right) - \frac{1}{N-1}\underbrace{\sum_{i=1}^{N}\sum_{i'=1}^{N}E_\delta(d_i d_{i'})}_{i \neq i'}\right)$$

$$= \frac{1}{N}\left(\sum_{i=1}^{N}\left(X_i - Y_i\right)^2 + \frac{1}{3}\sum_{i=1}^{N}\sum_{j=1}^{3}\alpha_{ij}^2 - \underbrace{\frac{n-1}{3(N-1)^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3}\alpha_{ij}\alpha_{ij}}_{i \neq i' \text{ and } (j=j')} - \underbrace{\frac{1}{N-1}\sum_{i=1}^{N}\sum_{i'=1}^{N}\left(X_i - Y_i\right)\left(X_{i'} - Y_{i'}\right)}_{i \neq i'} - \underbrace{\frac{n}{3(N-1)^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3}\sum_{j'=1}^{3}\alpha_{ij}\alpha_{ij'}}_{(i \neq i') \text{ and } (j \neq j')}\right)$$

$$= \frac{1}{N}\left(\sum_{i=1}^{N} D_i^2 - \frac{1}{N-1} \underbrace{\sum_{i=1}^{N}\sum_{i'=1}^{N} D_i D_{i'}}_{i \neq i'} + \frac{1}{3}\sum_{i=1}^{N}\sum_{j=1}^{3} \alpha_{ij}^2 - \underbrace{\frac{n-1}{3(N-1)^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3} \alpha_{ij}\alpha_{i'j}}_{i \neq i'} - \underbrace{\frac{n}{3(N-1)^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3}\sum_{j'=1}^{3} \alpha_{ij}\alpha_{i'j'}}_{(i \neq i') \text{ and } (j \neq j')} \right)$$

$$= \frac{1}{N}\left(\sum_{i=1}^{N} D_i^2 - \frac{1}{N-1}\left(\left\{ \sum_{i=1}^{N} D_i \right\}^2 - \sum_{i=1}^{N} D_i^2 \right) + \frac{1}{3}\sum_{i=1}^{N}\sum_{j=1}^{3} \alpha_{ij}^2 - \underbrace{\frac{n-1}{3(N-1)^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3} \alpha_{ij}\alpha_{i'j}}_{i \neq i'} - \underbrace{\frac{n}{3(N-1)^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3}\sum_{j'=1}^{3} \alpha_{ij}\alpha_{i'j'}}_{(i \neq i') \text{ and } (j \neq j')} \right)$$

$$= \frac{1}{N}\left(\frac{N}{N-1}\sum_{i=1}^{N} D_i^2 - \frac{N^2}{N-1}\bar{D}^2 + \frac{1}{3}\sum_{i=1}^{N}\sum_{j=1}^{3} \alpha_{ij}^2 - \underbrace{\frac{n-1}{3(N-1)^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3} \alpha_{ij}\alpha_{i'j}}_{i \neq i'} - \underbrace{\frac{n}{3(N-1)^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3}\sum_{j'=1}^{3} \alpha_{ij}\alpha_{i'j'}}_{(i \neq i') \text{ and } (j \neq j')} \right)$$

$$= \frac{1}{N-1}\left(\sum_{i=1}^{N} D_i^2 - N\bar{D}^2 + \frac{(N-1)}{3N}\sum_{i=1}^{N}\sum_{j=1}^{3} \alpha_{ij}^2 - \underbrace{\frac{n-1}{3N(N-1)}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3} \alpha_{ij}\alpha_{i'j}}_{i \neq i'} - \underbrace{\frac{n}{3N(N-1)}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3}\sum_{j'=1}^{3} \alpha_{ij}\alpha_{i'j'}}_{(i \neq i') \text{ and } (j \neq j')} \right)$$

Hence,

$$E_\delta(S_d^2) = \frac{N}{N-1}S_D^2 + \frac{1}{3N}\sum_{i=1}^{N}\sum_{j=1}^{3} \alpha_{ij}^2 - \underbrace{\frac{n-1}{3N(N-1)^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3} \alpha_{ij}\alpha_{i'j}}_{i \neq i'} - \underbrace{\frac{n}{3N(N-1)^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3}\sum_{j'=1}^{3} \alpha_{ij}\alpha_{i'j'}}_{(i \neq i') \text{ and } (j \neq j')}$$

$$E_\delta\left(\frac{N-1}{N} S_d^2 \right) = S_D^2 + \frac{N-1}{3N^2}\sum_{i=1}^{N}\sum_{j=1}^{3} \alpha_{ij}^2 - \underbrace{\frac{n-1}{3N^2(N-1)}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3} \alpha_{ij}\alpha_{i'j}}_{i \neq i'} - \underbrace{\frac{n}{3N^2(N-1)}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3}\sum_{j'=1}^{3} \alpha_{ij}\alpha_{i'j'}}_{(i \neq i') \text{ and } (j \neq j')}$$

$$= S_D^2 + \frac{1}{3N^2}\left((N-1)\sum_{i=1}^{N}\sum_{j=1}^{3} \alpha_{ij}^2 - \underbrace{\frac{n-1}{N-1}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3} \alpha_{ij}\alpha_{i'j}}_{i \neq i'} - \underbrace{\frac{n}{N-1}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3}\sum_{j'=1}^{3} \alpha_{ij}\alpha_{i'j'}}_{(i \neq i') \text{ and } (j \neq j')} \right)$$

$$= S_D^2 + Bias_{DEP}$$

where, $Bias_{DEP} = \frac{1}{3N^2}\left((N-1)\sum_{i=1}^{N}\sum_{j=1}^{3}\alpha_{ij}^2 - \underbrace{\frac{n-1}{N-1}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3}\alpha_{ij}\alpha_{i'j}}_{i\neq i'} - \underbrace{\frac{n}{N-1}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{3}\sum_{j'=1}^{3}\alpha_{ij}\alpha_{i'j'}}_{(i\neq i')\ and\ (j\neq j')} \right)$ ■

Appendix D - Bounds for the True Variance of Individual Treatment Effects

4.4.1: Bounds for the Variance of Individual Treatment Effects

Previously, we had that the variance of the true individual treatment effect is given as

$$\begin{aligned} Var(D) &= E(D^2) - E(D)^2 \\ &= (\omega_2 + \omega_3) - (\omega_3 - \omega_2)^2 \\ &= \omega_2(1-\omega_2) + \omega_3(1-\omega_3) + 2\omega_2\omega_3 \end{aligned}$$

Using the TC – CT design, $Var(D)$ cannot be estimated since $(\omega_2 + \omega_3)$ cannot be estimated. That is,

$$Var(D) = (\omega_2 + \omega_3) - (E(D))^2$$

However, if we denote an estimate of $Var(D)$ as $\overset{\Lambda}{Var}(D)$, then, we would have,

$$\overset{\Lambda}{Var}(D) = 0.5\left(\hat{\phi}_{13} + \hat{\phi}_{23} + \hat{\phi}_{12} + \hat{\phi}_{22}\right) - (\Delta_{TC} + \Delta_{CT}) - \left(\overset{\Lambda}{E}(D)\right)^2$$

where $\overset{\Lambda}{\omega_2 + \omega_3} = 0.5\left(\hat{\phi}_{13} + \hat{\phi}_{23} + \hat{\phi}_{12} + \hat{\phi}_{23}\right) - (\Delta_{TC} + \Delta_{CT})$ is derived from equations (1.45), (1.46), (1.49) and (1.50). $\overset{\Lambda}{Var}(D)$ cannot be measured using the TC – CT design even with the assumption of exchangeability since $(\Delta_{TC} + \Delta_{CT})$ still persists. However, considering the fact that, an estimate of $\omega_2 + \omega_3$ can be expressed as:

$$\overset{\Lambda}{\omega_2 + \omega_3} = 0.5\left(\hat{\phi}_{13} + \hat{\phi}_{23} + \hat{\phi}_{12} + \hat{\phi}_{23}\right) - (\Delta_{TC} + \Delta_{CT}) \qquad (1.72)$$

Additional information from the TT – CC design can be used to establish bounds for $(\Delta_{TC} + \Delta_{CT})$ and hence for $Var(D)$. The next theorem uses the knowledge from the additional TT – CC design to bound $(\Delta_{TC} + \Delta_{CT})$ and subsequently, $\overset{\Lambda}{Var}(D)$.

Proposition 4.6.3

Given that $|\Delta_{TC} + \Delta_{CT}| \leq \min\{1-(\phi_{31}+\phi_{34}), 1-(\phi_{41}+\phi_{44})\}$, the true variance of the individual treatment effects, $Var(D) \in [L^*, U^*]$

where, $L^* = \max\{0, 0.5[(\phi_{12}+\phi_{22}+\phi_{13}+\phi_{23}) - \min\{1-(\phi_{31}+\phi_{34}), 1-(\phi_{41}+\phi_{44})\}]\}$ and

$U^* = 0.5(\phi_{12}+\phi_{22}+\phi_{13}+\phi_{23}) + \min\{1-(\phi_{31}+\phi_{34}), 1-(\phi_{41}+\phi_{44})\}$

Proof:

Previously, we established that

$$Var(D) = (\omega_2 + \omega_3) - (E(D))^2$$
$$= 0.5(\phi_{13}+\phi_{23}+\phi_{12}+\phi_{22}) - (\Delta_{TC}+\Delta_{CT}) - \left(\overset{\Lambda}{E}(D)\right)^2$$

$(E(D))^2$ is always positive. Thus, $0 \leq Var(D) \leq 0.5(\phi_{13}+\phi_{23}+\phi_{12}+\phi_{22}) - (\Delta_{TC}+\Delta_{CT})$

Using the fact that $|\Delta_{TC} + \Delta_{CT}| \leq \min\{1-(\phi_{31}+\phi_{34}), 1-(\phi_{41}+\phi_{44})\}$, it follows that,

$$\Rightarrow -\min\{1-(\phi_{31}+\phi_{34}), 1-(\phi_{41}+\phi_{44})\} \leq (\Delta_{TC}+\Delta_{CT}) \leq \min\{1-(\phi_{31}+\phi_{34}), 1-(\phi_{41}+\phi_{44})\}$$

.

Thus, $Var(D) \in [L^*, U^*]$

where, $L^* = \max\left\{0, 0.5\left(\phi_{12}+\phi_{22}+\phi_{13}+\phi_{23}\right) - \min\left\{1-\left(\phi_{31}+\phi_{34}\right), 1-\left(\phi_{41}+\phi_{44}\right)\right\}\right\}$ and

$U^* = 0.5\left(\phi_{12}+\phi_{22}+\phi_{13}+\phi_{23}\right) + \min\left\{1-\left(\phi_{31}+\phi_{34}\right), 1-\left(\phi_{41}+\phi_{44}\right)\right\}$ ■

4.7.3: Bound for the Variance of the Individual Treatment Effect with Exchangeability

Using the exchangeability assumption, the bounds for the true variance of the individual treatment effect $\left[L^*, U^*\right]$ given become $\left[L_E^*, U_E^*\right]$ where,

Thus, $Var(D) \in \left[L^*, U^*\right]$

$$L^* = \max\left\{0, 0.5\left(\phi_{12}+\phi_{22}+\phi_{13}+\phi_{23}\right) - \min\left\{1-\left(\phi_{31}+\phi_{34}\right), 1-\left(\phi_{41}+\phi_{44}\right)\right\}\right\}$$

where, $= \max\left\{0, \theta_2+\theta_3 - \min\left\{p_1-\theta_3, p_2-\theta_4\right\}\right\}$ and

$= L_E^*$

$$U^* = 0.5\left(\phi_{12}+\phi_{22}+\phi_{13}+\phi_{23}\right) + \min\left\{1-\left(\phi_{31}+\phi_{34}\right), 1-\left(\phi_{41}+\phi_{44}\right)\right\}$$
$$= \theta_2+\theta_3+\min\left\{p_1-\theta_3, p_2-\theta_4\right\}$$
$$= U_E^*$$

Thus, $L_E^* = \max\left\{0, \theta_2+\theta_3-\min\left\{p_1-\theta_3, p_2-\theta_4\right\}\right\}$

and $U_E^* = \theta_2+\theta_3+\min\left\{p_1-\theta_3, p_2-\theta_4\right\}$,

where, $\theta_1 = P\left(X_1=0\ ,\ Y_2=0\right)$, $\theta_2 = P\left(X_1=0\ ,\ Y_2=1\right), \theta_3 = P\left(X_1=1\ ,\ Y_2=0\right)$ and

$\theta_4 = P\left(X_1=1\ ,\ Y_2=1\right)$.