METHODS IN RANDOMIZATION BASED ANCOVA FOR NOVEL CROSSOVER DESIGNS AND SENSITIVITY ANALYSIS FOR MISSING DATA

Siying Li

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics in the Gillings School of Global Public Health.

Chapel Hill
2017

Approved by:

Gary G. Koch

Amy H. Herring

Anastasia Ivanova

Charles Poole

John S. Preisser

© 2017
Siying Li
ALL RIGHTS RESERVED

ABSTRACT

Siying Li: Methods in Randomization Based ANCOVA for Novel Crossover Designs and Sensitivity Analysis for Missing Data
(Under the direction of Gary G. Koch)

In clinical trials, statistical inference is preferably conducted with less stringent assumptions. This dissertation proposes a non-parametric method for dichotomous and ordinal missing data, and it proposes a structure for the hypothesis testing and estimation for innovative crossover designs.

When data missing not at random (MNAR) arise from randomized multi-visit, multi-center clinical trials, sensitivity analyses to address possibly informative missing are needed. We propose a closed form point and variance matrix estimation for dichotomized missing data by probabilistically redistributing missing counts, adjusting for a stratification factor and/or baseline covariables. The parameter estimates are computed via weighted least squares asymptotic regression through randomization based methods. We further extend the methods to sensitivity analyses for ordinal endpoints.

A novel crossover design, the sequential parallel comparison design (SPCD), where information from placebo responders in the second period are excluded, serves as a design for studies with high placebo response. Estimators for sources of comparison in the traditional SPCD design, as well as other sources of information that are available, are constructed with methods based on the randomization distribution of the observed population using the nonparametric

mean and variance estimates under the null hypothesis, which control Type I error well in hypothesis testing. Baseline imbalance is adjusted by randomization-based ANCOVA. Simulations are performed to study the statistical properties of the proposed methods, which are compared to those of a repeated measures model proposed by Doros et al. (2013).

Point and confidence interval estimation is also addressed by assuming the study population comes from a simple random sample of an almost infinite population. A consistent covariance matrix estimator is constructed and properties of the proposed estimators are studied with simulations, particularly for coverage of confidence intervals. The nominal coverage level is achieved with a t distribution for the approximation to the asymptotic distribution when the sample size is not sufficiently large.

The methodologies are extended to the two-way enrichment design (TED) introduced by Ivanova and Tamura (2011), and to a related bilateral design that applies the four sequence group design to two sides of the same subject instead of two periods.

ACKNOWLEDGEMENTS

I want to take the opportunity to express my thanks to Dr. Gary Koch, who supported me both financially and mentally throughout my graduate school years. His role as mentor of life and his advice on both research and life (and Carolina basketball and football), in the office, at breakfast in Bob Evans and Whole Foods, and over the phone whenever I needed, will always be a memorable experience. My gratitude also goes to Dr. Amy Herring, for coaching me on research projects and serving as my role model as a passionate public health female researcher. Their being my pillar of support and always backing me up has encouraged me through both my good and struggling times. I also want to thank Dr. Preisser and Dr. Ivanova for being coauthors of manuscripts and for the constructive feedbacks they have provided, and Dr. Poole for being on my committee.

I would like to thank my grandparents and parents, who have always set the bar high since I was a kid and believed in my capability in accomplishing goals. I could not express my gratitude enough to my husband, who always support me unconditionally, with great patience along my way. I am also grateful for my fellow students at UNC, SPH, and the biostat department, and my friends at the RTP running club, for making this venture within and outside campus even more meaningful.

TABLE OF CONTENTS

LIST OF TABLES ix

LIST OF FIGURES xi

CHAPTER 1 INTRODUCTION 1

1.1 Handling Random Imbalance of Baseline Covariables 1

1.2 Handling Missing Data 4

1.2.1 Missing Data Mechanism 5

1.2.2 Approaches for Handling Dropouts by Parametric Model 9

1.2.3 Approaches for Handling Dropouts by Imputation 11

1.2.4 Sensitivity Analyses 14

1.3 Crossover Studies 15

1.3.1 Traditional Crossover Designs 15

1.3.2 Innovative Two-Period Crossover Designs 17

1.4 Summary 23

REFERENCES 25

CHAPTER 2: SENSITIVITY ANALYSIS OF FAVORABLE PROPORTION FOR MISSING DICHOTOMOUS DATA IN MULTI-VISIT RANDOMIZED CLINICAL TRIAL 28

2.1 Introduction 28

2.2 Methods 29

2.2.1 Notation 30

2.2.2 Data Structure 30

2.2.3 Favorable Probability Estimation 31

2.2.4 Sensitivity Analysis 32
2.2.5 Covariance Matrix Estimation 34
2.2.6 Treatment Comparison 36
2.3 Example 39
2.4 Discussion 42
REFERENCES 43
CHAPTER 3: SENSITIVITY ANALYSIS IN TREATMENT COMPARISON FOR MISSING ORDINAL DATA IN MULTIE-VISIT RANDOMIZED CLINICAL TRIAL 44
3.1 Introduction 44
3.2 Methods 44
3.2.1 Notation 44
3.2.2 Data Structure 45
3.2.3 Multinomial Probability Estimation 46
3.2.4 Sensitivity Analysis 46
3.2.5 Treatment Comparison 49
3.3 Example 54
REFERENCES 57
CHAPTER 4: RANDOMIZATION-BASED ANCOVA FOR HYPOTHESES TESTING IN THE SEQUENTIAL PARALLEL COMPARISON DESIGN (SPCD) 58
4.1 Introduction 58
4.2 Methods 61
4.3 Simulation Study 69
4.3.1 Simulation Setup 69
4.3.2 Simulation Results 71
4.4 Example 77

REFERENCES 80

CHAPTER 5: RANDOMIZATION-BASED ANCOVA FOR POINT AND CONFIDENCE INTERVAL ESTIMATION IN SEQUENTIAL PARALLEL COMPARISON DESIGN (SPCD) 82

5.1 Introduction 82

5.2 Methods 82

5.3 Simulation Study 87

5.3.1 Simulation Setup 87

5.3.2 Simulation Results 89

REFERENCES 98

CHAPTER 6: RANDOMIZATION-BASED ANCOVA FOR INFERENCE IN TWO-WAY ENRICHMENT DESIGN 99

6.1 Introduction 99

6.2 Methods 99

6.2.1 Estimates for Treatment Comparisons 100

6.2.2 Randomization-Based Covariance Adjusted Estimators 102

REFERENCES 105

CHAPTER 7: RANDOMIZATION-BASED ANCOVA FOR INFERENCE IN BILATERAL DESIGN 106

7.1 Introduction 106

7.2 Methods 106

7.2.1 Estimates for Treatment Comparisons 107

7.2.2 Randomization-Based Covariance Adjusted Estimators 110

REFERENCES 112

LIST OF TABLES

Table 1.1 2×2 Design Parameters 16

Table 2.1 Data Structure 30

Table 2.2 Data Structure for Missing Counts Redistribution 32

Table 2.3 Missing Percentages of Assessment Visits 39

Table 2.4 Results of Sensitivity Analyses for Treatment Comparison Estimators 41

Table 3.1 Data Structure 45

Table 3.2 Data Structure for Missing Counts Redistribution 47

Table 3.3 Missing Counts of Assessment Visits 54

Table 3.4 Results of Sensitivity Analyses 56

Table 4.1 Specifications for Assessments of Statistical Power 70

Table 4.2 Results from 50, 000 replicate simulations for the test statistics of $H0: \Delta 1 = \Delta 4 = 0$ under $\Delta 1 = \Delta 2 = \Delta 3 = \Delta 4 = 0$ 73

Table 4.3 Results from 50, 000 simulations for the test statistics of $H0: \Delta 1 = \Delta 3 = \Delta 4 = 0$ under $\Delta 1 = \Delta 2 = \Delta 3 = \Delta 4 = 0$ 74

Table 4.4 Results from 50, 000 replicate simulations for the test statistics of $H0: \Delta 1 = \Delta 3 = \Delta 3 = \Delta 4 = 0$ under $\Delta 1 = \Delta 2 = \Delta 3 = \Delta 4 = 0$ 75

Table 4.5 Results from 50,000 replicate simulations for power of test statistics under the alternative specified in Table 4.1 76

Table 4.6 Mean and Standard Deviation of Outcome 77

Table 4.7 Estimates of $\Delta 1$ to $\Delta 4$ 78

Table 4.8 Estimates of Weighted Statistics 79

Table 5.1 Specifications for Assessments of Statistical Power 88

Table 5.2 Results from 50, 000 replicate simulations for the test statistics of $H0: \Delta 1 = \Delta 4 = 0$ under $\Delta 1 = \Delta 2 = \Delta 3 = \Delta 4 = 0$ 91

Table 5.3 Results from 50, 000 replicate simulations for the test statistics of $H0: \Delta 1 = \Delta 3 = \Delta 4 = 0$ under $\Delta 1 = \Delta 2 = \Delta 3 = \Delta 4 = 0$ 92

Table 5.4 Results from 50, 000 replicate simulations for the test statistics of H0: $\Delta 1 = \Delta 3 = \Delta 3 = \Delta 4 = 0$ under $\Delta 1 = \Delta 2 = \Delta 3 = \Delta 4 = 0$ 93

Table 5.5 Results from 50,000 replicate simulations for the coverage of confidence intervals and power of test statistics under the alternative specified in Table 5.1 94

Table 5.6 Mean and Standard Deviation of Outcome 95

Table 5.7 Estimates of $\Delta 1$ to $\Delta 4$ 96

Table 5.8 Estimates of Weighted Statistics 97

LIST OF FIGURES

Figure 1.1 Randomized Withdrawal Design 19

Figure 1.2 Placebo Lead-In Design 19

Figure 1.3 P:T, T:P, and T:T Sequence Design 20

Figure 1.4 Randomized Delayed-Start Design 21

Figure 1.5 SPCD Design 22

Figure 1.6 TED Design 23

CHAPTER 1 INTRODUCTION

In clinical trials, statistical inference is preferably conducted with less model assumptions. This dissertation proposes a nonparametric method to handle dichotomous and ordinal missing data, and proposes a structure for hypothesis testing and estimation in innovative crossover designs.

1.1 Handling Random Imbalance of Baseline Covariables

In the statistical analysis plan of a clinical trial, the statistical methods to determine if there is a significant treatment effect need to be stated in an a priori way before the clinical trials are actually carried out. Oftentimes, assumptions have to be made for certain statistical models to be valid, and they are difficult to test before data analysis. Therefore, methods requiring fewer assumptions are more desirable than those complicated ones, especially in the regulatory setting.

This consideration led to the development of nonparametric randomization based analysis of covariance (ANCOVA). In randomized clinical trials, the covariable imbalances (if any) between treatment and control groups are due to random chance, since the treatment assignment is random.

The details about nonparametric randomization based ANCOVA for analyzing randomized clinical trials can be found in (Koch et al., 1998b; LaVange, Durham and Koch, 2005). Briefly, differences between treatment groups with respect to outcome variables and covariables are analyzed simultaneously using weighted least squares (WLS), restricting the covariables differences to be zero. As mentioned before, in a randomized clinical trial, the expected value of such differences for covariables would in fact be equal to zero.

Let y_{gi} be the outcome of subject i in group g, and let $\boldsymbol{x}_{gi} = \left(x_{gi1}, \ldots, x_{gim}\right)'$ be the pre-specified vector of *m* covariables, and let $\boldsymbol{f}_{gi}$ be the response-covariable (m+1) dimensional vector $(y_{gi}, \boldsymbol{x}'_{gi})'$. Then the sample mean of the outcome and covariables of treatment g is $\bar{y}_g = \frac{1}{n_g}\sum_{i=1}^{n_g} y_{gi}$ and $\bar{\boldsymbol{x}}_g = \frac{1}{n_g}\sum_{i=1}^{n_g} \boldsymbol{x}_{gi}$. And $\bar{\boldsymbol{f}}_g = \frac{1}{n_g}\sum_{i=1}^{n_g} \boldsymbol{f}_{gi} = (\bar{y}_g, \bar{\boldsymbol{x}}'_g)'$ is the sample mean of the response-covariables of subjects in group g and $\bar{\boldsymbol{f}} = \frac{1}{n}\sum_{g=1}^{2}\sum_{i=1}^{n_g} \boldsymbol{f}_{gi}$ is the sample mean of all subjects in the trial. Let $\boldsymbol{d} = (d_y, \boldsymbol{d}'_{\boldsymbol{x}})'$ be the vector of differences in means, where $d_y = (\bar{y}_1 - \bar{y}_2)$ and $d_x = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)$.

There are two ways to estimate the variance of the difference $\boldsymbol{d}$. One is through the randomization distribution of $\boldsymbol{d}$ for the finite population selected for the clinical trial assuming the strong null hypothesis $H_0: y_{1i} = y_{2i} = y_{*i}$ that each patient would have the same outcome regardless of the assigned treatment. Under this null hypothesis, the covariance matrix for the difference $\boldsymbol{d}$ is expressed as

$$\boldsymbol{V_0} = \frac{n}{n_1 n_2 (n-1)} \sum_{g=1}^{2} \sum_{i=1}^{n_g} [\boldsymbol{f}_{gi} - \bar{\boldsymbol{f}}][\boldsymbol{f}_{gi} - \bar{\boldsymbol{f}}]' \qquad (1.1).$$

Since $\boldsymbol{V_0}$ is the covariance matrix for the randomization distribution of $\boldsymbol{d}$, permuting all possible randomized assignments to the two treatments for the patients in the clinical trial, it is a matrix of known constant values (rather than random variables), with a conditional nature that the response of this finite population is known.

Alternatively, under the assumption that the patients in the clinical trial are a simple random sample of a very large population, and thus are representative of this large population, an unbiased estimator for the unconditional covariance matrix of the difference $\boldsymbol{d}$ as shown in (1.2).

$$\boldsymbol{V_S} = \sum_{g=1}^{2} \frac{1}{n_i(n_i-1)} \sum_{i=1}^{n_g} [\boldsymbol{f}_{gi} - \bar{\boldsymbol{f}}_g][\boldsymbol{f}_{gi} - \bar{\boldsymbol{f}}_g]' \qquad (1.2).$$

In this case, the covariance matrix $\boldsymbol{V_S}$ is a random matrix instead of a constant matrix, in a sense that the randomness comes from the variability of the simple random sample of patients and the random assignment of treatment groups, regardless of H_0.

Applying the non-parametric analysis of covariance to $\boldsymbol{d}$, it has the form of a linear regression as below,

$$\boldsymbol{d} = \begin{bmatrix} d_y \\ \boldsymbol{d_x} \end{bmatrix} \triangleq \boldsymbol{Z}b = \begin{bmatrix} 1 \\ \mathbf{0}_p \end{bmatrix} b \qquad (1.3)$$

where $\triangleq$ denotes "is estimated by", $\mathbf{0}_p$ denotes a p dimensional vector, $\boldsymbol{Z} = [1\ \mathbf{0}_p']'$, and b is the adjusted mean difference for the response, i.e., the adjusted version of d_y.

Applying WLS, determination of b can be obtained by,

$$b = (\boldsymbol{Z'V^{-1}Z})^{-1}\boldsymbol{Z'V^{-1}f} = d_y - \boldsymbol{V'_{yx}V_{xx}^{-1}d_x} \qquad (1.4)$$

where $\boldsymbol{V} = \begin{bmatrix} \boldsymbol{V_{yy}}\ \boldsymbol{V'_{yx}} \\ \boldsymbol{V_{yx}}\ \boldsymbol{V_{xx}} \end{bmatrix}$ and $\mathbf{V}$ can be either $\boldsymbol{V}_0$ or $\boldsymbol{V}_S$ defined above.

An estimator for the covariance of b is expressed as,

$$V_b = (\boldsymbol{Z'V^{-1}Z})^{-1} = V_{yy} - \boldsymbol{V'_{yx}V_{xx}^{-1}V_{yx}} \qquad (1.5)$$

When $\boldsymbol{V_0}$ is used in place of $\boldsymbol{V}$, $\boldsymbol{V_b}$ is an exact variance of the randomization distribution of the adjusted treatment difference b; and when $\boldsymbol{V}_S$ is used, V_b is a random matrix and a consistent estimator of the covariance matrix of b.

Since the variance V_b of the adjusted mean difference b is smaller than its counterpart V_{yy} of the variance of the unadjusted mean difference d_y, the test based on the adjusted difference b

is more powerful than that based on d_y, and the confidence interval of b is narrower than that of d_y. The variance reduction of b relative to d_y is based on the correlation between the response and the covariable, and the stronger this correlation, the more variance reduction produced (Koch et al., 1998).

The nonparametric ANCOVA can be extended to multivariate response variables, and other types of data including dichotomized, ordinal, and time to event data (Tangen and Koch, 1999).

1.2 Handling Missing Data

In public health studies, repeated measurements of the same subject over time are useful in a number of different contexts, including, but not limited to, reliable estimation by several measurements close in time, testing for a change over time in an experimental study, or comparisons for a difference between treatment groups over time.

In a clinical trial, missing data were planned to be collected but are not present in the database. No matter how well designed and conducted a trial is, some missing data is almost always unavoidable. The consequences of missing data can be wide-ranging in that they might lead to a perceived or real reduction in trial quality and validity, and a reduction in the statistical power of the study.

When missing data are unavoidable and exist in the collected data, assumptions about the missing data can be made. Often, dropout is due to some specific reasons, related (i.e. adverse events, or lack of efficacy) or unrelated (i.e., move out of the neighborhood) to the treatment. Investigators are urged to collect as much information about the reasons of withdrawal as possible when missing data are unpreventable.

The validity of many statistical models that can handle missing data relies heavily on the assumptions for the missing data. For example, generalized estimating equations (GEE) must be carried out along with the missing completely at random (MCAR) assumption and the mixed models for repeated measures with a missing at random (MAR) assumption. These assumptions might not be realistic in real life, and possibly not even verifiable.

With the withdrawal reasons, assumptions of missingness could be checked; or when the assumptions could not be verified, sensitivity analyses could be performed under different scenarios to test against the robustness of a study result.

Besides the assumptions of the missingness, the handling of missing data is complicated by the form of the study outcome, for example, non-normality of data, such as dichotomous data, ordinal data, or skewed continuous data.

1.2.1 Missing Data Mechanism

We now review the mechanisms that lead to missing data, and in particular the question of whether the variables that are missing are related to the underlying values of variables that are observed or not observed in the dataset. It is crucial to understand the missing data mechanism before any analyses are carried out since the properties of missing data methods rely heavily on the nature of the dependencies in these mechanisms.

The role of these mechanisms was largely ignored in the analysis of missing values until these concepts were formalized in the theory of Rubin (1976), through treating the missing data indicators as random variables and constructing a joint distribution among values of interest and the missing indicators. The following notation and terminology is based on the standard missing data framework of Carpenter and Kenward (2012), which is developed from the original paper of Rubin (1976) but is modified to fit the modern statistics literature on missing data.

Notation

In the context of a longitudinal trial, we assume that measurements are obtained at *J* visits at times $j = 1, \cdots, J$ for independent subjects i, $i = 1, \cdots, n$.

Let $\boldsymbol{Y} = (y_{ij})$ denote an $(n \times J)$ rectangular data matrix of the measurements without missing values, with the i-th row $\boldsymbol{y}_i = (y_{i1}, \cdots, y_{iJ})$ being the complete data vector of outcomes for subject i.

Additionally, let $\boldsymbol{X}_i$ be the design matrix of covariates for subject i.

Let $\boldsymbol{r}_i = (r_{i1}, \cdots, r_{iJ})$ be the missing data indicator vector. Specifically, let r_{ij}=1 if y_{ij} is observed and r_{ij}=0 otherwise.

Given the missing data indicator $\boldsymbol{r}_i$, we can partition $\boldsymbol{y}_i$ into $(\boldsymbol{y}_i^O, \boldsymbol{y}_i^M)$, with $\boldsymbol{y}_i^O$ being the observed measurements in $\boldsymbol{y}_i$ and $\boldsymbol{y}_i^M$ being the missing measurements.

The joint distribution of the data and the missing indicator can be formulated as follows and factored into two parts:

$$f(\boldsymbol{y}_i^O, \boldsymbol{y}_i^M, \boldsymbol{r}_i | \boldsymbol{X}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{y}_i^O, \boldsymbol{y}_i^M | \boldsymbol{X}_i, \boldsymbol{\theta}) f(\boldsymbol{r}_i | \boldsymbol{y}_i^O, \boldsymbol{y}_i^M, \boldsymbol{X}_i, \boldsymbol{\psi}) \qquad (1.6)$$

where $\boldsymbol{\theta}$ denotes the parameter vectors for the data and $\boldsymbol{\psi}$ denote the parameter vectors for the missing data mechanism. The first factor on the right-side is the marginal density of the measurements, and the second factor is the conditional density of the missingness on the measurements.

Missing Completely at Random (MCAR)

Under MCAR, the missingness is assumed to be unrelated to either the observed information or the missing, i.e.,

$$f(\boldsymbol{r}_i|\boldsymbol{y}_i^O, \boldsymbol{y}_i^M, \boldsymbol{X}_i, \boldsymbol{\psi}) = f(\boldsymbol{r}_i|\boldsymbol{\psi}) \quad (1.7)$$

Note that this assumption doesn't mean that the missingness itself is random, but rather that this distribution does not depend on the data values.

Therefore, the joint distribution simplifies to

$$f(\boldsymbol{y}_i^O, \boldsymbol{y}_i^M, \boldsymbol{r}_i|\boldsymbol{X}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{y}_i^O, \boldsymbol{y}_i^M|\boldsymbol{X}_i, \boldsymbol{\theta}) f(\boldsymbol{r}_i|\boldsymbol{\psi}) \quad (1.8)$$

which indicates the measurement and the missingness are independent.

The missing data $\boldsymbol{y}_i^M$ can be now integrated out from the joint distribution, and so the joint distribution of the observed measurement and the missing indicator becomes

$$f(\boldsymbol{y}_i^O, \boldsymbol{r}_i|\boldsymbol{X}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{y}_i^O|\boldsymbol{X}_i, \boldsymbol{\theta}) f(\boldsymbol{r}_i|\boldsymbol{\psi}) \quad (1.9)$$

And thus estimation of $\boldsymbol{\theta}$ can be solely based on the observed information $\boldsymbol{y}_i^O$ and does not depend on the nuisance parameter $\boldsymbol{\psi}$.

Missing at Random (MAR)

An assumption less restrictive than MCAR is that missingness depends only on the components that are observed, i.e., $\boldsymbol{y}_i^O$, and not on the components that are missing, i.e., $\boldsymbol{y}_i^M$.

Under MAR, conditional on the observed data, the missingness is independent of the missing measurements, which is,

$$f(\boldsymbol{r}_i|\boldsymbol{y}_i^O, \boldsymbol{y}_i^M, \boldsymbol{X}_i, \boldsymbol{\psi}) = f(\boldsymbol{r}_i|\boldsymbol{y}_i^O, \boldsymbol{X}_i, \boldsymbol{\psi}) \quad (1.10)$$

Therefore, the full data density becomes

$$f(\boldsymbol{y}_i^O, \boldsymbol{y}_i^M, \boldsymbol{r}_i|\boldsymbol{X}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{y}_i^O, \boldsymbol{y}_i^M|\boldsymbol{X}_i, \boldsymbol{\theta}) f(\boldsymbol{r}_i|\boldsymbol{y}_i^O, \boldsymbol{X}_i, \boldsymbol{\psi}) \quad (1.11)$$

The joint distribution of the observed measurements and the missing indicators can again integrate out the missing data and become

$$f(\boldsymbol{y}_i^O, \boldsymbol{r}_i|\boldsymbol{X}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{y}_i^O|\boldsymbol{X}_i, \boldsymbol{\theta}) f(\boldsymbol{r}_i|\boldsymbol{y}_i^O, \boldsymbol{X}_i, \boldsymbol{\psi}) \quad (1.12)$$

With MAR assumption, the model $f(\boldsymbol{r}_i|\boldsymbol{y}_i^O, \boldsymbol{X}_i, \boldsymbol{\psi})$ does not need to be specified to obtain valid likelihood based inferences, and only the model $f(\boldsymbol{y}_i^O, \boldsymbol{y}_i^M|\boldsymbol{X}_i, \boldsymbol{\theta})$ is needed.

MCAR and MAR are often referred to as ignorable missing. The ignorability refers to the fact that once $f(\boldsymbol{r}_i|\boldsymbol{y}_i, \boldsymbol{X}_i)$ not depending on $\boldsymbol{y}_i^M$ can be established, $f(\boldsymbol{r}_i|\boldsymbol{y}_i, \boldsymbol{X}_i)$ can be ignored and a valid likelihood based inference can be obtained given that we model $f(\boldsymbol{y}_i^O, \boldsymbol{y}_i^M|\boldsymbol{X}_i, \boldsymbol{\theta})$ correctly.

Not Missing at Random (NMAR)

If the measurements are NMAR, which means $\boldsymbol{r}_i$ depends on $\boldsymbol{y}_i^M$, the joint distribution can no longer have $\boldsymbol{y}_i^M$ integrated out. No simplification of the joint distribution is possible.

Under the MNAR assumption, the probability of an observation being missing depends on the underlying missing value, and the joint distribution has to be written as in (1.13),

$$f(\boldsymbol{y}_i^O, \boldsymbol{r}_i|\boldsymbol{X}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) = \int f(\boldsymbol{y}_i^O, \boldsymbol{y}_i^M|\boldsymbol{X}_i, \boldsymbol{\theta}) f(\boldsymbol{r}_i|\boldsymbol{y}_i^O, \boldsymbol{y}_i^M, \boldsymbol{X}_i, \boldsymbol{\psi}) d\boldsymbol{y}_i^M \quad (1.13)$$

and inferences could only be made by making further assumptions (Molenberghs and Kenward, 2007).

Monotone versus Non-Monotone Missingness

If the data are arranged as one record per subject, with each record containing outcomes of all visits of a subject, the monotone missing pattern applies when variables can be arranged so that missing values are always occurring as one block at the end of data records; in the case of a non-monotone missing pattern, missing values cannot to be arranged in this way and may happen anywhere in a study record.

In the clinical trials setting, monotone missingness happens when a study subject withdraws from the trial prematurely and doesn't come back to the study, which is commonly referred to as dropout or loss to follow up in longitudinal studies; while non-monotone missing is the case when a subject misses one or more intermediate visits but does come back to provide subsequent measurements (O'Kelly and Ratitch, 2014). A dataset is considered as monotone missing only when all the subjects in the study have a monotone pattern, but it is considered as non-monotone missing if there is intermediate missingness in at least one subject.

1.2.2 Approaches for Handling Dropouts by Parametric Model

Complete Case Analysis

One approach to handling missing data is to have analyses that exclude all data from any subject who drops out. This method is referred to as a complete-case analysis, which is performed by excluding any subjects that missed any intended measurement. It is emphasized that this method is very problematic and is rarely an acceptable approach in most occasions (Fitzmaurice, Laird and Ware, 2012). It will yield unbiased estimates of the mean response trends only when the dropout can be assumed as MCAR. When dropout is MCAR, the study completers are a random subsample of the original sample from the population. However, even in occasions where the MCAR assumptions might hold, a complete-case analysis is not an appealing one since it leads to reduction in the number of subjects and hence results in reduction in statistical power.

Generalizing Estimating Equations (GEE)

GEE is a semiparametric method that models a known function of the marginal expectation of a clustered dependent variable via a linear or non-linear link function for a linear function of one or more explanatory variables (Liang and Zeger, 1986). It is based on the concept

of estimating equations and the use of a non-linear link function for the marginal model of the correlated response can facilitate the analyses of continuous or discrete responses.

The correlation of the clustered dependent variable can be specified via a working correlation matrix and the consistency of parameter estimates do not rely on correct specification of the correlation. The dependent variable doesn't need to have the same number of elements across clusters and thus in the longitudinal data context, missing data is allowed. However, if the data is MAR, as GEE methods only require a model for the mean response but do not specify the multivariate joint distribution for the response vector, the standard GEE methods do not provide valid estimates of the regression parameters (Fitzmaurice et al., 2012).

An adaption of GEE methods for the MAR assumption is to model the missingness $f(\boldsymbol{r}_i|\boldsymbol{y}_i, \boldsymbol{X}_i)$ and weighting the analysis by including it in the estimating equations accordingly (Robins, Rotnitzky and Zhao, 1995).

Mixed Model for Repeated Measures (MMRM)

MMRM is a likelihood based method that assumes a multivariate normal distribution for the repeated measurements. The mean structure can take into account the time effect and thus the time effect within subjects at different measurements can be modeled. The correlation within subjects can also be modeled by the covariance structure of the repeated measurements through assumptions of dependence among the different measurements. This approach includes all subjects with at least one observed measurement, and the missing measurements are assumed to have the same distribution as the observed. Since MMRM is likelihood based, it provides valid inference on the model if the data are MAR when the joint distribution of the responses is correctly specified (O'Kelly and Ratitch, 2014).

1.2.3 Approaches for Handling Dropouts by Imputation

Ad-hoc Single Imputation

Imputation replaces the missing values with plausible ones. There are many approaches to do single imputation. The missing values could be filled in from an individual imputation, where these values are coming from the same individual with the actual missing values, or from a group imputation, in which information from the entire sample or a portion of the sample is used to fill in for the missing value of an individual (Fitzmaurice et al., 2012).

Two of the most commonly used individual imputations are, baseline observation carried forward (BOCF) and last observation carried forward (LOCF), where either the baseline value or the last observed value is substituted for the missing values of the study subject. For example, in the LOCF case, if an individual was supposed to have five measurements but only the first three measurements were observed, the last two missing measurements would be filled in with the third value as it was the last observed value before the loss to follow up. The assumption of BOCF or LOCF is conservative in estimating the missing outcome if a subject does benefit from the trial. The imputation using BOCF or LOCF would underestimate the variability of the estimation and result in smaller standard errors estimate. Other single imputation includes the individual mean substitution, the group mean substitution, the individual worst case substitution, or the interpolation of last and next observed values if the missingness is not monotonic.

Multiple Imputation (MI)

Multiple imputation was first introduced by Rubin (1987) to handle missing data in sample surveys, and has been developed to spread to other areas including observational studies and randomized clinical trials. The application of multiple imputation has become very popular in recent years as many analysts become familiar with it, and as many software packages such as

SAS, R, and Stata have included procedures or packages to deal with it, which reduces the computational burden and complexity. Multiple imputation is a more flexible and powerful tool to handle incomplete data than the other parametric methods such as GEE and MMRM, in that it has both an imputation model and an analysis model and these two don't have to be the same, and it is more acceptable in most settings than the single imputation.

Multiple imputation adopts a three-step approach to fill in the incomplete data and analyze the resulting data structure. First, an imputation model is assumed for the missing outcomes, and plausible values for missing observations are imputed with a draw from the imputation model, usually as a posterior distribution of the missing values conditioning on the imputation model covariates and any previous visits is assumed. This process is repeated to reflect uncertainty about the missing values, resulting in the creation of a number of complete datasets. The number of needed imputations depends on the fraction of missing data, and usually a number of K>5 would be sufficient for most applications to obtain acceptable properties (that is, correct confidence interval coverage) (Carpenter and Kenward, 2012). Second, each of these K complete datasets is analyzed with an analysis model, which need not be the same as the imputation model. Finally, the results are combined for overall inference using Rubin's combination rule (Rubin, 1987).

Rubin's combination rule is as follows. Assume the parameter of interest of the complete data analysis is θ and denote $\hat{\theta}_k$ and $\hat{V}_k$ as the point estimate and variance estimate of θ from the k-th imputed dataset, $k = 1, \dots, K$. Then the MI estimate of θ can be expressed as the average of the estimates from the K complete datasets,

$$\hat{\theta}_{MI} = \frac{1}{K}\sum_{k=1}^{K}\hat{\theta}_k \qquad (1.14)$$

The measure of precision for $\hat{\theta}_{MI}$ consists of two parts, the between imputation variance and the within imputation variance. Define

$$\widehat{W} = \frac{1}{K}\sum_{k=1}^{K} \hat{V}_k \qquad (1.15)$$

to be the average within imputation variance, and

$$\hat{B} = \frac{1}{K}\sum_{k=1}^{K} (\hat{\theta}_k - \hat{\theta}_{MI})^2 \qquad (1.16)$$

to be the between imputation variance. Then an estimate of the variance of $\hat{\theta}_{MI}$ is given by

$$\hat{V}_{MI} = \widehat{W} + \left(1 + \frac{1}{K}\right) \hat{B} \qquad (1.17)$$

A few practical considerations occur in the first step of multiple imputation. Imputation can be performed in a variety of different ways, depending on the type of the missing data pattern (monotone or non-monotone), and depending on the type of missing variables (i.e. continuous or categorical). One of the common practical considerations is to impute the non-monotone missing data to monotone. When all response variables are continuous for data with non-monotone missingness, imputation can be done by drawing from a Bayesian posterior for the multivariate normal distribution, with a Markov Chain Monte Carlo (MCMC) simulation. MCMC will impute data partially, filling in only those missing values that have a non-monotone pattern (O'Kelly and Ratitch, 2014). This method is most suitable when all variables included in the imputation model are continuous, however, this approach has also been applied when some variables are categorical; and it is usually the case that the covariates in a model include both continuous and categorical variables. To apply MCMC imputation, nominal categorical values can be dummy-coded as a set of binary variables, while ordinal variables can sometimes be

treated as continuous in this partial imputation step. However, in the clinical trial setting, when multiple clinical centers need to be adjusted for, and when the number of centers is large (>10), the center variable might have to be removed from the multivariate normal model for the partial imputation. This assumption could be reasonable if the non-monotone missingness doesn't vary by center, otherwise the imputed values would not have taken into account the variability introduced by study center.

1.2.4 Sensitivity Analyses

Sensitivity analysis in missing data situations is usually carried out through stressing the assumption of MAR. It is important to examine the sensitivity of statistical inferences when departures from the MAR assumption are in question, because this assumption cannot be verified using the data (O'Kelly and Ratitch, 2014). In this regard, the primary purpose of a sensitivity analysis in a clinical trial is to seek to answer the question that if plausible unfavorable outcomes happen to the withdrawal in the experimental treatment, does the significant results drawn from the primary analysis remain credible or not?

For example, the loss to follow up outcomes that are suspected to be different from what would have happened if remaining in the study, could be made worse by a clinically significant value, if the outcome is continuous, or by an odds ratio if the outcome is categorical. This is often known as the delta adjustment method (National Research Council Panel, 2010), where delta is the clinically important difference, or odds ratio. Delta adjustment could be applied to all the treatment groups, or it might be of more interest to be used to penalize the withdrawals in the experimental treatment (O'Kelly and Ratitch, 2014). Tipping point analysis is the application of a sequence of delta adjustments, by positing a wide range of assumptions from less pessimistic to more pessimistic to explore the influence of missingness on the study conclusion (National

Research Council Panel, 2010). The tipping point in this range of assumptions is the value that overturns the conclusion from being favorable to the experimental treatment, to being not different from the reference group. In terms of hypothesis testing of a treatment effect, the tipping point is the value at which the p value of the test changes from significant to non-significant.

1.3 Crossover Studies

Crossover studies are experimental designs for which each subject is randomly assigned to receive a sequence of treatments during consecutive periods for some response variables. There are many possible designs of crossover studies, depending on the number of treatments to compare, the number of periods of each treatment, and the aim of the trials (Jones and Kenward, 2014).

1.3.1 Traditional Crossover Designs

One of the most well-known crossover designs is the one with two sequence groups for two treatments in two periods. This is also the simplest crossover design, which is known as the 2×2 design, or the two-period two-treatment design. The main advantage of the crossover study is that treatments are compared within subjects, where every subject provides two periods of different treatments and thus removal of the subject effect is enabled by direct comparison within subject (Jones and Kenward, 2014).

Secondly, since every subject provides two response measurements in the two periods, at a fixed sample size, the power of the treatment comparison is improved. In addition, since all the patients would receive the experimental treatment in one period or another, the dropout rate in this design could be minimized, at least for the first study period. For example, Pincus et al. (2001) performed a randomized crossover trial of an experimental drug versus active control in

ambulatory patients with osteoarthritis of the hip or knee and achieved a low dropout rate. Of the 227 enrolled patients, 218 (96.0%) patients completed the first treatment period and 181 (79.7%) completed both treatment periods.

However, the feature of repeated measurements in crossover designs brings disadvantages along with its advantages; for example, the possibility that the effect of an earlier period would be carried into the later period, and the potential risk of more dropouts due to longer study duration compared to a single period trial.

Statistical Methods for 2×2 Design

Tudor and Koch (1994) review nonparametric methods for analyzing the traditional crossover studies comparing two treatments with small sample sizes and the parametric counterparts when sample sizes are sufficiently large. The methods apply to various types of outcome including continuous, dichotomous, ordinal, and censored time-to-event response.

In particular, for a 2x2 crossover design with a univariate continuous outcome, the structure for the inference is as shown in Table 1.1.

Table 1.1 2×2 Design Parameters

Group	Period 1	Period 2
AB	$\mu + \pi_1 + \tau_A$	$\mu + \pi_2 + \tau_B + \lambda_A$
BA	$\mu + \pi_1 + \tau_B$	$\mu + \pi_2 + \tau_A + \lambda_B$

τ_A and τ_B are direct treatment effects of treatment A and treatment B, and π_1 and π_2 are period effects of periods 1 and 2, and λ_A and λ_B are carryover effects of treatment A and treatment B respectively.

With small sample sizes, observed statistics are compared to permutation distributions to provide p values for the hypothesis testing of similarity of treatment effects $H_{0\tau}: \tau_A = \tau_B$ using

data from both periods in a few steps. In the first step, one would test H_0, where $H_0: 2(\tau_B - \tau_A) - (\lambda_B - \lambda_A) = 0$. If the H_0 is contradicted, one moves on to test the equality of carryover effects $H_{0\lambda}: \lambda_A = \lambda_B$; and if this similarity is not contradicted, one could have confidence that the contradiction of H_0 is mainly due to the difference in treatment effects τ_A and τ_B and the equivalence of testing H_0 and $H_{0\tau}$; or if this equality of carryover effects $H_{0\lambda}$ is contradicted, meaning there are different carryover effects by the two treatments, which may partly account for the contradiction of H_0, one would have to move on to compare the difference between treatments only using data from Period 1, which does not depend on the carryover effects.

The above tests in each step could be replaced by the asymptotic tests with approximate distributions when the sample size is sufficiently large (sample size per sequence $\geq$15).

A more comprehensive review of analyses in the traditional 2x2 design and other higher order designs is provided in Jones and Kenward (2014).

1.3.2 Innovative Two-Period Crossover Designs

Innovative crossover designs can have multiple designs embedded within them, which are also in the general class of re-randomization designs. Instead of re-randomization at the beginning of the second period, randomization before the trial could be performed to get the randomized sequences. Without loss of generality, this literature review limits the scope of the discussion to the design with fixed randomization sequences at the initiation of the study without re-randomization later.

With a two-period design comparing test treatment (T) to placebo treatment (P), four sequence groups P:P, P:T, T:P, and T:T could be of interest. Designs with some combination of those sequence groups have provided useful features for the studies of different patient

populations. Other advances of crossover designs with this structure could be made use of with other added design features, such as enrichment.

Enriched Two-Sequence Design

Common crossover designs that use two of four sequence groups are the T:P, and P:T design (the 2×2 design), the T:P and T:T design, and the P:P and P:T design. The latter two-sequence designs are usually used with enrichment features.

The *randomized withdrawal design*, with the T:P and T:T sequence groups, makes use of only the patients who respond to the drug in the first period for continuation to the second period. This design is helpful when there is heterogeneity in the patient population itself to respond to a treatment. For example, Temple (1994) discusses situations where the gold standard randomized, double-blind, placebo-controlled study design with continuous treatment might not be able to provide an optimal study when certain diseases are treated, such as irritable bowel syndrome (IBS), a gastrointestinal disorder, which he suggested might be due to IBS being a "common response to a diverse group of abnormality". The FDA has proposed to conduct clinical trials to include only IBS patients identified by their clinicians as responders to the study treatment (Dunger-Baldauf, Racine, and Koch 2006).

The P:P and P:T design, usually known as the *placebo lead-in design*, has the other two sequence groups of the four as compared to randomized withdrawal design. In this design, only the patient who doesn't respond to placebo in the first period remains in the study. This design is practical in studies such as drugs to treat disease in the central nerve system, where there are many placebo responders. The placebo response rate in antidepressant and antipsychotic trials is reported to increase overtime in meta analyses of trials between 1985 and 2000 (Khan et al., 2005). With only the placebo nonresponders identified in the first period continuing in the

second period to receive either experimental treatment or placebo, the treatment effect is maximized since patients who do not respond to the first period are not expected to become placebo-responders in the second period (Fava et al., 2003).

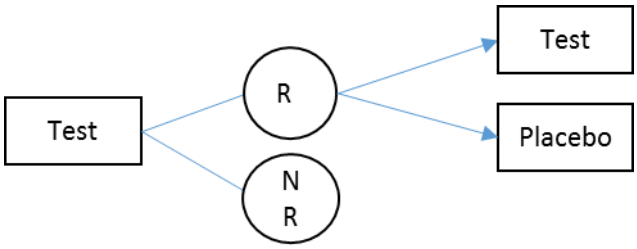


Figure 1.1 Randomized Withdrawal Design

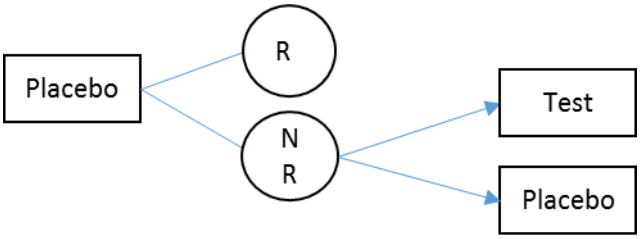


Figure 1.2 Placebo Lead-In Design

Three-Sequence Design

Designs with three or more sequence groups could provide additional benefits to the two-sequence designs mentioned above.

One design in this class to improve completeness of data with this crossover structure was proposed by Koch, Davis, and Anderson (1998). This design has three of the four sequence groups, P:T, T:P, and T:T; as shown in Figure 1.3, it provides T during the second period to patients with P during the first period, and provides continued treatment of T to some fraction of the patients who received T during the first period. This design embeds the comparison that is capable in the 2×2 design and the randomized withdrawal design.

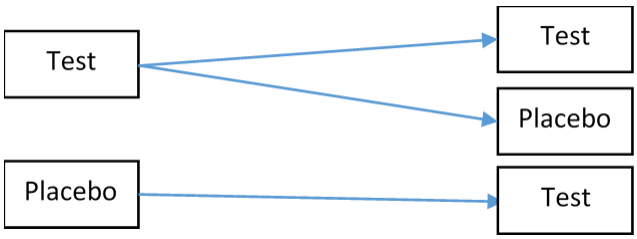


Figure 1.3 P:T, T:P, and T:T Sequence Design

Another design with attractive features also has three sequence groups as P:P, P:T, and T:T, and it is commonly known as the *randomized delayed-start design* . In this design, patients are initially randomized to placebo or test drug in the first period, and patients who are in the placebo group in the first period would receive either placebo or test drug in the second period, while patients who receive test drug in the first period remain on the same treatment. This design is suitable to evaluate treatments for disease with long term progression to distinguish the symptomatic improvement from the true disease modifying effect. The effect of the active treatment in the first period compared to placebo could be due to either the symptomatic improvement or the true effect on modifying the disease, but in the second period when the delayed-start P:T sequence patients receive the active treatment, if the early-start T:T sequence patients show benefits from being in the trial longer than the P:T sequence, it indicates a disease modifying effect of the active treatment. If these two sequence groups are showing similar improvement from baseline, then the active treatment might only reflect a symptomatic relief in the course (Dunger-Baldauf, Racine and Koch, 2006) (Clinical Trial Design in Parkinson's Disease 2013 p3).

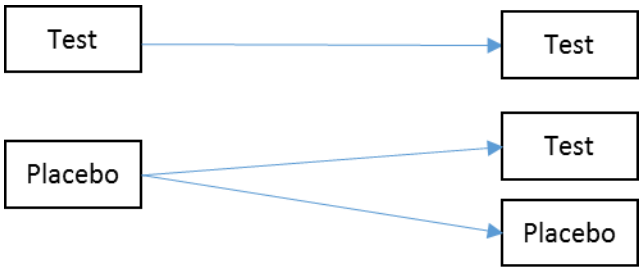


Figure 1.4 Randomized Delayed-Start Design

These two three-sequence designs both have an advantage over the two-sequence randomized withdrawal design in that there is a higher chance in receiving the better treatment in the second period, which might provide a favorable impact on the patient retention and reduce non-compliance with the protocol, at least at the end of the first period (Dunger-Baldauf, 2007).

Enriched Multiple-Sequence Design

Another study design that has the same sequence groups as the randomized delayed-start design is the sequential parallel comparison design (SPCD). SPCD design was proposed by Fava et al. (2013), and it is different from the delayed-start design in that only the placebo non-responders of the first period in the P:P and P:T groups continue into the second period. SPCD serves similar purposes as the two-sequence placebo lead-in design in psychiatric clinical trials with a high placebo response rate, except that it has an additional T:T sequence. And thus besides sharing the high compliance benefit of the three sequence design mentioned above, it also eliminates the potential risk brought by the placebo lead-in design that it is more difficult to identify placebo-responders when it is hard to hide from the clinicians that only placebo is given in the first period (Fava et al., 2003; Ivanova, Qaqish and Schoenfeld, 2011; Doros et al., 2013).

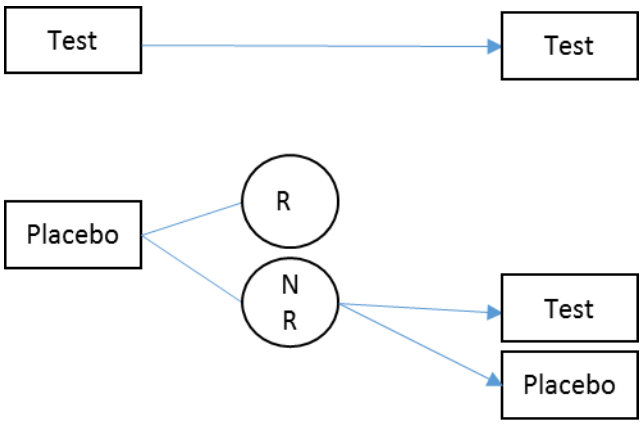


Figure 1.5 SPCD Design

The original paper of Fava et al. (2003) focused on the study outcome as dichotomized data. Other methods have been proposed for the analyses of binary outcomes in the context of SPCD designs by Huang and Tamura (2010), Ivanova, Qaqish and Schoenfeld (2011), and Huang, Tamura, and Boos (2011).

Recent uses of the SPCD design have been extended to continuous or ordinal outcome as it arises more naturally than the dichotomizing of a continuous measurement. Huang and Tamura (2010) considered seemingly unrelated regression (SUR) to account for the correlation between subjects in the two periods of the trial. Chen et al (2011) proposed an ordinary least squares approach and Doros et al. (2013) proposed a repeated measures model that includes all possible outcome data collected in the trial.

The two-way enrichment design (TED), introduced by Ivanova and Tamura (2011), has all four sequence groups, P:P, P:T, T:P, and T:T, but in the second period, only the non-responders to the placebo in the first period of the P:P and P:T sequences, and the responders to the active treatment in the first period of the T:P and T:T sequences remain in the study. And thus the TED design has the advantage of both the placebo lead-in design and the randomized withdrawal designs. This design is suitable to study the maintenance of efficacy of an active treatment

(through the randomized withdrawal design) for a disease with a high placebo response rate. For example, for generalized anxiety disorder (GAD), which is a central nerve system disease with a high placebo response rate, and a chronic disease for which worsening would quickly occur after discontinuation from an active treatment, a trial to evaluate an active treatment versus placebo would benefit from the TED design.

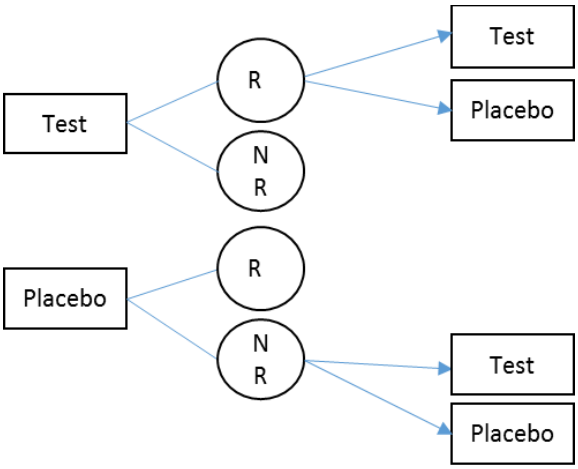


Figure 1.6 TED Design

Bilateral Design

Besides the studies mentioned above, the four sequence group design could also be applied to two sides of the same subject, instead of two periods. For example, Kawaguchi and Koch (2009) studied the two eyes of the same patients with four sequence group, with the treatments assigned to the two eyes instead of the two periods respectively.

1.4 Summary

This literature review covers many aspects of clinical trials. The existing methodology for handling missing data in different missing data scenarios is reviewed and different designs and usage of crossover studies is discussed. Also, this chapter discusses a nonparametric way to handle the random imbalance in covariables, as well as its usefulness in reducing the variability of estimation.

The following chapters are organized as follows. In Chapter 2, we discuss a method for sensitivity analyses of estimation in favorable proportion for missing dichotomous data. Chapter 3 extends the method in Chapter 2 to outcome data with an ordinal nature. Chapter 4 presents a method for statistical inference under the null for SPCD trials and Chapter 5 further studies the point and confidence interval estimation under the alternative of this design. And Chapters 6 and 7 provides an outline and statistical planning for topic in hypothesis testing for TED trials and bilateral design.

REFERENCES

Chen, Y.-F., Yang, Y., Hung, H.M.J. and Wang, S.-J. (2011) Evaluation of performance of some enrichment designs dealing with high placebo response in psychiatric clinical trials. *Contemporary Clinical Trials*, **32**, 592–604.

Doros, G., Pencina, M., Rybin, D., Meisner, A. and Fava, M. (2013) A repeated measures model for analysis of continuous outcomes in sequential parallel comparison design studies. *Statistics in Medicine*, **32**, 2767–2789.

Dunger-Baldauf, C. (2007) Designs with Randomization Following Initial Study Treatment. *Wiley Encyclopedia of Clinical Trials* p. John Wiley & Sons, Inc.

Dunger-Baldauf, C., Racine, A. and Koch, G.G. (2006) Re-treatment studies. *Drug Information Journal*, **40**.

Fava, M., Evins, A.E., Dorer, D.J. and Schoenfeld, D.A. (2003) The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach. *Psychotherapy and Psychosomatics*, **72**, 115–127.

Fitzmaurice, G.M., Laird, N.M. and Ware, J.H. (2012) *Applied Longitudinal Analysis*. John Wiley & Sons.

Huang, X. and Tamura, R.N. (2010) Comparison of Test Statistics for the Sequential Parallel Design. *Statistics in Biopharmaceutical Research*, **2**, 42–50.

Ivanova, A., Qaqish, B. and Schoenfeld, D.A. (2011) Optimality, sample size, and power calculations for the sequential parallel comparison design. *Statistics in Medicine*, **30**, 2793–2803.

Jones, B. and Kenward, M. (2014) Design and Analysis of Cross-Over Trials, Third Edition. URL https://www.crcpress.com/Design-and-Analysis-of-Cross-Over-Trials-Third-Edition/Jones-Kenward/p/book/9781439861424 [accessed 16 July 2016]

Kawaguchi, A., Koch, G.G. and Ramaswamy, R. (2009) Applications of extensions of bivariate rank sum statistics to the crossover design to compare two treatments through four sequence groups. *Biometrics*, **65**, 979–988.

Kenward, M. (2012) *Multiple Imputation and Its Application*. John Wiley & Sons.

Khan, A., Kolts, R.L., Rapaport, M.H., Krishnan, K.R.R., Brodhead, A.E. and Browns, W.A. (2005) Magnitude of placebo response and drug-placebo differences across psychiatric disorders. *Psychological Medicine*, **35**, 743–749.

Koch, G.G., Tangen, C.M., Jung, J.G. and Amara, I.A. (1998) Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. Statistics in Medicine, 17, 1863–1892.

LaVange, L.M., Durham, T.A. and Koch, G.G. (2005) Randomization-based nonparametric methods for the analysis of multicentre trials. *Statistical Methods in Medical Research*, **14**, 281–301.

Lehmacher, W., Wassmer, G. and Reitmeir, P. (1991) Procedures for two-sample comparisons with multiple endpoints controlling the experimentwise error rate. *Biometrics*, **47**, 511–521.

Liang, K.-Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22.

Little, R.J.A. and Rubin, D.B. (2014) *Statistical Analysis with Missing Data*. John Wiley & Sons.

Molenberghs, G. and Kenward, M. (2007) *Missing Data in Clinical Studies*. John Wiley & Sons.

National Research Council Panel. (2010) *The Prevention and Treatment of Missing Data in Clinical Trials*. National Academies Press (US), Washington (DC).

O'Kelly, M. and Ratitch, B. (2014) *Clinical Trials with Missing Data: A Guide for Practitioners*. John Wiley & Sons.

Parkinson's Disease Edition: Early Diagnosis and Comprehensive Management Table of Contents - Clinical Trial Design in Parkinson's Disease. URL http://lmt.projectsinknowledge.com/Activity/index.cfm?showfile=b&jn=2094&sj=2094.02&sc=2094.02.3 [accessed 10 August 2016]

Pincus, T., Koch, G.G., Lei, H., Mangal, B., Sokka, T., Moskowitz, R., et al. (2004) Patient Preference for Placebo, Acetaminophen (paracetamol) or Celecoxib Efficacy Studies (PACES): two randomised, double blind, placebo controlled, crossover clinical trials in patients with knee or hip osteoarthritis. *Annals of the Rheumatic Diseases*, **63**, 931–939.

Robins, J.M., Rotnitzky, A. and Zhao, L.P. (1995) Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data. *Journal of the American Statistical Association*, **90**, 106–121.

Rubin, D.B. (1976) Inference and missing data. *Biometrika*, **63**, 581–592.

Rubin, D.B. (1987) *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons.

Tamura, R.N. and Huang, X. (2007) An examination of the efficiency of the sequential parallel design in psychiatric clinical trials. *Clinical Trials (London, England)*, **4**, 309–317.

Tamura, R.N., Huang, X. and Boos, D.D. (2011) Estimation of treatment effect for the sequential parallel design. *Statistics in Medicine*, **30**, 3496–3506.

Tangen, C.M. and Koch, G.G. (1999) Nonparametric Analysis of Covariance for Hypothesis Testing with Logrank and Wilcoxon Scores and Survival-Rate Estimation in a Randomized Clinical Trial. *Journal of Biopharmaceutical Statistics*, **9**, 307–338.

Temple, R.J. (1994) Special study designs: early escape, enrichment, studies in non-responders. *Communications in Statistics - Theory and Methods*, **23**, 499–531.

Tudor, G. and Koch, G.G. (1994) Review of nonparametric methods for the analysis of crossover studies. *Statistical Methods in Medical Research*, **3**, 345–381.

CHAPTER 2: SENSITIVITY ANALYSIS OF FAVORABLE PROPORTION FOR MISSING DICHOTOMOUS DATA IN MULTI-VISIT RANDOMIZED CLINICAL TRIAL

2.1 Introduction

In clinical trials, the dichotomous endpoint only has two possible outcomes for an observation, either directly or via categorization of an ordinal or continuous observation. However, missing data often occur for one or more visits during a multi-visit study. No matter how well designed and conducted a trial is, some missing data can almost always be expected (O'Kelly and Ratitch, 2014). Oftentimes, missing data are due to some specific reasons, and they can be related to the treatment for a patient (e.g., adverse events, or lack of efficacy) or unrelated (e.g., move from the community for treatment).

When loss to follow-up occurs, investigators are urged to collect as much information as possible for the withdrawal reasons. Given the withdrawal reasons, sensitivity analyses could be performed under different scenarios. In the regulatory setting, a tipping point analysis is usually needed to assess what conditions would overturn the statistical significance of the claimed treatment difference and whether such pivotal conditions are potentially possible in the real trial (National Research Council Panel, 2010).

In the situations for missing data in a clinical trial, the two quantities of primary interest are the treatment comparison estimate under different assumptions and its corresponding variance. They can be obtained through imputation for missing data, or with statistical models with some assumptions for the missingness and covariance structure for the data. In the statistical

analysis plan of the clinical trial, the statistical methods need to be stated a priori for the assessment of the treatment comparison before the clinical trials are actually conducted. Oftentimes, possibly unrealistic assumptions are required for the validity of certain statistical models, and they are difficult to evaluate before data analysis. Therefore, methods requiring fewer assumptions are desired rather than those with complex assumptions, especially in the regulatory setting.

In this paper, we propose a method that mathematically redistributes the missing counts as favorable or unfavorable under different specifications for the missing data, so as to provide resulting estimates for the treatment comparisons in a multi-visit clinical trial and a corresponding covariance matrix. Also, adjustment for covariates is possible through randomization-based analysis of covariance (ANCOVA) so as to provide variance reduction and offset random imbalances. Section 2.2 introduces the data set up and the methodology, and the methods are illustrated with an example in Section 2.3. Chapter 3 provides an extension to an ordinal categorical outcome.

2.2 Methods

Consider a study comparing a test treatment and a control treatment for a favorable outcome or not through assessments at each of several visits. For the dichotomous outcome with possibly missing data, there are three possible responses, favorable, unfavorable, or missing, although the outcome for missing could be expanded to include the applicable reason. While outcome and response are often interchangeable in the literature, here we make a distinction between outcome categories and response categories, where the former could only have two outcomes and the latter includes missing as a category.

2.2.1 Notation

Let y_{ghijk} be the indicator variable for the response of subject i in group g and stratum h at visit j being k, where group $g = 1, 2$ index the test treatment and control treatment respectively; stratum $h = 1, 2, \cdots, H$ index the stratum for the subject; subject $i = 1, 2, \cdots, n_{gh}$; visit $j = 1, 2, \cdots, J$, and response $k = 1, 2, 3$, where 1, 2 and 3 index favorable, unfavorable, and missing response respectively. If there were 3 potential reasons for missing such as lack of efficacy, unacceptable tolerability, and other, then $k = 1, 2, 3, 4, 5$ could apply; but throughout this paper, only one missing category is mainly considered. We define a three-dimensional vector $\boldsymbol{Y}_{ghij*} = (y_{ghij1}, y_{ghij2}, y_{ghij3})'$ to combine the three indicators. For example, $Y_{1111*} = (0,1,0)$ means Subject 1 for the test treatment and Stratum 1 at time point 1 has unfavorable outcome. Accordingly, we further define a data vector that includes all visits as $\boldsymbol{Y}_{ghi**} = \left(\boldsymbol{Y}'_{ghi1*}, \cdots, \boldsymbol{Y}'_{ghiJ*}\right)' = \left(y_{ghi11}, y_{ghi12}, y_{ghi13}, \cdots, y_{ghJ11}, y_{ghJ12}, y_{ghJ13}\right)'$.

2.2.2 Data Structure

For subjects in treatment group g and stratum h, the observed data can be arranged in a contingency table format as in Table 2.1. After including missing as a category, the number of

Table 2.1 Data Structure

Time \ Response	Fav	UnFav	Missing	Total
1	$n_{gh11}(p_{gh11})$	$n_{gh12}(p_{gh12})$	$n_{gh13}(p_{gh13})$	n_{gh}
⋮	⋮	⋮	⋮	n_{gh}
J	$n_{ghJ1}(p_{ghJ1})$	$n_{ghJ2}(p_{ghJ2})$	$n_{ghJ3}(p_{ghJ3})$	n_{gh}

subjects at each visit is fixed as n_{gh}. The cell count n_{ghjk} is $n_{ghjk} = \sum_{i=1}^{n_{gh}} y_{ghijk}$, and cell proportion p_{ghjk} is $p_{ghjk} = \frac{n_{ghjk}}{n_{gh}}$. The counts vector at row j of the table is expressed as

$n_{ghj} = (n_{ghj1}, n_{ghj2}, n_{ghj3})$, which follows a multinomial distribution, $\boldsymbol{n}_{ghj} \sim Multinominal(n_{gh}, \boldsymbol{\pi}_{ghj})$, where $\boldsymbol{\pi}_{ghj} = \left(\pi_{ghj1}, \pi_{ghj2}, \pi_{ghj3}\right)'$ is the marginal multinomial probability vector of response for the j-th visit. From the properties of multinomial distributions, the unbiased estimator of the multinomial probability $\boldsymbol{\pi}_{ghj}$, is the proportion vector at row j, $\boldsymbol{p}_{ghj} = \left(p_{ghj1}, p_{ghj2}, p_{ghj3}\right)'$. Combining across all time points, $\boldsymbol{p}_{gh} = \left(\boldsymbol{p}'_{gh1}, \dots, \boldsymbol{p}'_{ghJ}\right)'$ and $\boldsymbol{\pi}_{gh} = (\boldsymbol{\pi}'_{gh1}, \dots, \boldsymbol{\pi}'_{ghJ})'$ apply to the correlated multinomial distributions for the $\boldsymbol{n}_{ghj}$ for the J visits. The estimated covariance matrix of $\boldsymbol{p}_{gh}$ as the mean of the $\boldsymbol{Y}_{ghi**}$ is shown in (2.1).

$$\boldsymbol{V}_{p_{gh}} = \frac{1}{n_{gh}(n_{gh}-1)} \sum_{i=1}^{n_{gh}} \left(\boldsymbol{Y}_{ghi**} - \boldsymbol{p}_{gh}\right)\left(\boldsymbol{Y}_{ghi**} - \boldsymbol{p}_{gh}\right)' \quad (2.1)$$

2.2.3 Favorable Probability Estimation

Now the estimation of interest is for the probability of favorable outcome. For this purpose, the missing response category is redistributed to the favorable and unfavorable outcomes according to a missing outcome specification.

Let q_{ghj} be the probability estimator that a subject in group g and stratum h would have a favorable outcome at visit j. The redistributed favorable outcome proportion under a missing completely at random (MCAR) specification is shown in (2.2).

$$q_{ghj} = p_{ghj1} + \frac{p_{ghj1}}{p_{ghj1} + p_{ghj2}} p_{ghj3} \quad (2.2)$$

It is also an "observed case" estimate which has the assumption that patients with missing status have the same probability of favorable outcome as those with observed status as shown in (2.3).

$$q_{ghj} = \frac{p_{ghj1}}{p_{ghj1} + p_{ghj2}} \qquad (2.3)$$

2.2.4 Sensitivity Analysis

So far we have assumed the missing responses are MCAR-like. However, in clinical trials, the discontinued patients could have withdrawn from the study due to reasons related to the unobserved outcome, which renders the data non-ignorable missing (NMAR); see Little and Rubin (2014).

For the subsequent discussion in this section, we omit the notation for group g and stratum h for simplicity of presentation, although they can always be included without loss of generality. Now let n_{j1}, n_{j2}, and n_{j3} represent the counts of the favorable, unfavorable, and missing responses at visit j, and $n_{j1} + n_{j2} + n_{j3} = n_j$. We further divide the missing counts n_{j3} into n_{j31} and n_{j32}, where n_{j31} and n_{j32} represent the unobserved counts with a favorable and unfavorable outcome if missing responses were actually observed. Also, p_{j1}, p_{j2}, and p_{j3} are the corresponding proportion estimators. Thus, we have the data structure shown in Table 2.2. Also, Table 2.2 could be expanded to account for counts for two or more reasons for missing responses.

Table 2.2 Data Structure for Missing Counts Redistribution

Visit	Fav	UnFav	Missing	Total
j	n_{j1}	n_{j2}	$n_{j3} = n_{j31} + n_{j32}$	n_j

The odds ratio ratio θ_j comparing the favorable to the unfavorable outcome in the patients with missing status versus observed patients is shown in (2.4); and the solution it implies for n_{j31} is shown in (2.5).

$$\theta_j = \frac{n_{j31}}{n_{j32}} \Big/ \frac{n_{j1}}{n_{j2}} \qquad (2.4)$$

$$n_{j31} = \frac{\theta_j n_{j1} n_{j3}}{\theta_j n_{j1} + n_{j2}} \qquad (2.5)$$

Thus, there can be determination of total favorable outcome counts through an assumed specification of the odds ratio θ_j. Also, separate θ_j's could address two or more reasons for missing responses.

If $\theta_j = 1$, the missing responses are assumed to be MCAR-like; if $\theta_j > 1$ the missing responses are regarded as more likely to have better outcome than those observed; if $\theta_j < 1$, the missing responses are regarded as more likely to have worse outcome, as is often the case for patients who discontinue the test treatment. Through adjusting for different θ_j, different specifications of the possible outcomes for the missing response can be obtained, and thus we call the θ_j the sensitivity parameters for missingness.

The adjusted favorable proportion estimator can be expressed in terms of the observed proportions for the responses and the θ_j as shown in (2.6).

$$q_{j\theta} = \frac{n_{j1} + n_{j31}}{n_j} = \frac{1}{n_j}\left(n_{j1} + \frac{\theta_j n_{j1}}{\theta_j n_{j1} + n_{j2}} n_{j3}\right) = p_{j1} + \frac{\theta_j p_{j1}}{\theta_j p_{j1} + p_{j2}} p_{j3} \qquad (2.6)$$

By construction, the $q_{j\theta}$ are comparable to what might be expected by random multiple imputation of the missing responses via (2.5), but they are alternatively produced from mathematical redistribution as in (2.5). From (2.6), it follows that the adjusted odds of favorable versus unfavorable outcome at time j has the structure shown in (2.7).

$$\frac{q_{j\theta}}{1 - q_{j\theta}} = \frac{n_{j1}\left(\theta_j n_{j1} + n_{j2} + \theta_j n_{j3}\right)}{n_{j2}\left(\theta_j n_{j1} + n_{j2} + n_{j3}\right)}$$

$$= \frac{p_{j1}\left(\theta_j p_{j1} + p_{j2} + \theta_j\left(1 - p_{j1} - p_{j2}\right)\right)}{p_{j2}\left(\theta_j p_{j1} + p_{j2} + \left(1 - p_{j1} - p_{j2}\right)\right)}$$

$$= \frac{p_{j1}\left(p_{j2} + \theta_j(1 - p_{j2})\right)}{p_{j2}\left(\theta_j p_{j1} + (1 - p_{j1})\right)} \qquad (2.7)$$

If θ_j=1, which is the MCAR-like case, $\frac{q_{j\theta}}{1-q_{j\theta}} = \frac{p_{j1}}{p_{j2}}$, with this indicating that the adjusted odds of favorable outcome versus unfavorable is the same as the odds for the observed outcomes.

2.2.5 Covariance Matrix Estimation

Let $\boldsymbol{q_\theta} = (q_{1\theta}, \ldots, q_{J\theta})$ denote the j-dimensional vector of adjusted outcome proportion estimators. In order to use the linear Taylor series methods discussed in Koch et al. (1977), as well as summarized Stokes et al. (2012, Chapter 14), to produce a consistent estimate for the covariance matrix of the adjusted proportion vector $\boldsymbol{q_\theta}$, we express $\boldsymbol{q_\theta}$ in the form of compound functions of the unadjusted proportion vector $\boldsymbol{p} = (\boldsymbol{p}_1', \ldots, \boldsymbol{p}_J')'$ and sensivity parameter θ as shown in (2.8).

$$\boldsymbol{q_\theta} = \boldsymbol{A_3 exp}[\boldsymbol{A_2\, log}(\boldsymbol{A_{1\theta}p})] \qquad (2.8)$$

In (2.8), log() denotes the element-wise vector operation that transforms a vector to the corresponding vector of natural logarithms, and exp[] denotes the element-wise vector operation that transforms a vector to the corresponding vector of exponentiated values, and matrices $\boldsymbol{A_{1\theta}}$, $\boldsymbol{A_2}$, and $\boldsymbol{A_3}$ are shown in (2.9) for which $\boldsymbol{bdiag_J}(\boldsymbol{L_j})$ denotes a diagonal matrix of J blocks, with

$$\boldsymbol{A_{1\theta}} = \boldsymbol{bdiag_J}\begin{pmatrix} 1 & 0 & 0 \\ \theta_j & 0 & 0 \\ \theta_j & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \boldsymbol{A_2} = \boldsymbol{bdiag_J}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 \end{pmatrix}, \boldsymbol{A_3} = \boldsymbol{bdiag_J}(1 \;\; 1) \qquad (2.9)$$

matrix $\boldsymbol{L_j}$ on the j-th diagonal block $j = 1, \ldots, J$. By applying the linear Taylor series methods in Koch et al. (1977), we can obtain a consistent estimator for the covariance matrix of $\boldsymbol{q_\theta}$ as shown in (2.10).

$$\boldsymbol{V}_{q_\theta} = \boldsymbol{B}_\theta \boldsymbol{V}_{\boldsymbol{p}} \boldsymbol{B}'_\theta \qquad (2.10)$$

For (2.10), $\boldsymbol{B}_\theta$ is the elementwise first partial derivative of vector $\boldsymbol{q}_{\boldsymbol{\theta}}$ with respect to vector $\boldsymbol{p}$ and is obtained by applying the chain rule, as shown in (2.11) for which $\boldsymbol{a}_{\boldsymbol{1\theta}} = \boldsymbol{A}_{\boldsymbol{1\theta}}\boldsymbol{p}$,

$$\boldsymbol{B}_{\boldsymbol{\theta}} = \frac{\boldsymbol{\partial q_\theta}}{\boldsymbol{\partial p}} = \frac{\boldsymbol{\partial q_\theta}}{\boldsymbol{\partial a_{3\theta}}} \frac{\boldsymbol{\partial a_{3\theta}}}{\boldsymbol{\partial a_{2\theta}}} \frac{\boldsymbol{\partial a_{2\theta}}}{\boldsymbol{\partial a_{1\theta}}} \frac{\boldsymbol{\partial a_{1\theta}}}{\boldsymbol{\partial p}} = \boldsymbol{A_3 D_{a_{3\theta}} A_2 D_{a_{1\theta}}^{-1} A_{1\theta}} \quad (2.11)$$

$\boldsymbol{a}_{\boldsymbol{2\theta}} = \log(\boldsymbol{a}_{\boldsymbol{1\theta}})$, $\boldsymbol{a}_{\boldsymbol{3\theta}} = \exp[\boldsymbol{A}_{\boldsymbol{2}}\boldsymbol{a}_{\boldsymbol{2\theta}}]$, and $\boldsymbol{q}_{\boldsymbol{\theta}} = \boldsymbol{A}_{\boldsymbol{3}}\boldsymbol{a}_{\boldsymbol{3\theta}}$, and $\boldsymbol{D}_{\boldsymbol{a}_{\boldsymbol{1\theta}}}^{\boldsymbol{-1}}$ is a diagonal matrix with the reciprocals of the elements of the vector $a_{1\theta}$ on the main diagonal and $\boldsymbol{D}_{\boldsymbol{a}_{\boldsymbol{3\theta}}}$ is the diagonal matrix with the elements of the vector $\boldsymbol{a}_{\boldsymbol{3\theta}}$ on the main diagonal.

Now we reconsider group g and stratum h, for which the adjusted favorable proportion estimate is shown in (2.12).

$$q_{ghj\theta} = p_{ghj1} + \frac{\theta_j p_{ghj1}}{\theta_j p_{ghj1} + p_{ghj2}} p_{ghj3} \quad (2.12)$$

The covariance estimate of $\boldsymbol{q}_{\boldsymbol{gh\theta}}$, where $\boldsymbol{q}_{\boldsymbol{gh\theta}} = \left(q_{gh1\theta}, \ldots, q_{ghJ\theta}\right)'$ is shown in (2.13).

$$\boldsymbol{V}_{\boldsymbol{q}_{\boldsymbol{gh\theta}}} = \boldsymbol{B}_{\boldsymbol{gh\theta}} \boldsymbol{V}_{\boldsymbol{p}_{\boldsymbol{gh}}} \boldsymbol{B}'_{\boldsymbol{gh\theta}} \qquad (2.13)$$

The selection of the sensitivity parameter θ_{ghj} could be based on knowledge for the trial being conducted and the nature of disease; see Zhao et al. (2014). In many cases, it can correspond to fractions of the reciprocal for a known odds ratio for the effect of a useful treatment versus placebo.

Typically, the θ_{ghj} are the same for all strata in group g, i.e., $\theta_{ghj} = \theta_{gj}$, and they could vary at different visits j and j'. In addition, one could expect missing responses for patients in the placebo group to have a similar outcome distribution as the observed patients, which corresponds to $\theta_{2hj} = 1$ for any h and any visit j in the placebo group.

2.2.6 Treatment Comparison

Treatment Difference

The treatment difference $\Delta = (\Delta_1, \cdots, \Delta_J)$ between test and placebo visits 1 to j could be estimated using the corresponding adjusted proportion differences, and they can be weighted by the Mantel-Haenszel weights $w_h = \left\{\frac{n_{1h}n_{2h}/(n_{1h}+n_{2h})}{\sum_{h'=1}^{H} n_{1h'}n_{2h'}/(n_{1h'}+n_{2h'})}\right\}$ for the combined strata. The treatment difference estimator at visit j for Δ_j is $d_{j\theta}$ as shown in (2.14).

$$d_{j\theta} = \sum_{h=1}^{H} w_h\left(q_{1hj\theta} - q_{2hj\theta}\right) \qquad (2.14)$$

Letting $\boldsymbol{d_\theta} = (d_{1\theta}, \cdots, d_{J\theta})$, the consistent covariance matrix estimator of $\boldsymbol{d_\theta}$ can be obtained using the covariance matrix estimator $\boldsymbol{V_{q_{gh\theta}}}$ of $\boldsymbol{q_{gh\theta}}$ in (2.13), and it is expressed as (2.15).

$$\boldsymbol{V_{d_\theta}} = \sum_{h=1}^{H} w_h^2 \left(\boldsymbol{V_{q_{1hj\theta}}} + \boldsymbol{V_{q_{2hj\theta}}}\right) \qquad (2.15)$$

Adjusted Treatment Difference via Randomzation-Based ANCOVA

In a randomized clinical trial, baseline covariables are expected to have the same distribution in the randomized groups. Baseline covariables could include the baseline measurement of the outcome, demographic variables, or other variables. However, random imbalances in the baseline covariables could occur as each treatment group is a finite sample of the randomized population, and covariance adjustment for them can offset such imbalances.

Covariance adjustment can also provide variance reduction for treatment comparison estimation when applied in a randomization-based way (Koch et al., 1998; Tangen and Koch, 2001; LaVange et al., 2005). The motivation behind the variance reduction is that the parameter

estimate that is associated with the imbalance would be corrected to offset the direction of the imbalance. The estimation is through randomization-based ANCOVA, which is an approach that applies weighted least squares methods to evaluate differences between treatment groups with respect to outcome variables and covariables simultaneously (Koch et al., 1998).

Here, we introduce some notations for the covariables. Suppose each subject has M baseline covariables $\boldsymbol{x_{ghi}}$, $\boldsymbol{x_{ghi}} = \left(x_{ghi1}, \dots, x_{ghiM}\right)'$, and $\bar{\boldsymbol{x}}_{\boldsymbol{gh}} = \frac{1}{n_{gh}} \sum_{i=1}^{n_{gh}} \boldsymbol{x_{ghi}}$ is the mean vector of the prespecified covariables.

We define $\boldsymbol{f_{ghi}} = \left(\boldsymbol{Y}'_{\boldsymbol{ghi}**}, \boldsymbol{x}'_{\boldsymbol{ghi}}\right)'$ as a $(3 \times J + M)$ dimensional response-covariable vector, and $\bar{\boldsymbol{f}}_{\boldsymbol{gh}} = \frac{1}{n_{gh}} \sum_{i=1}^{n_{gh}} \boldsymbol{f_{ghi}} = \left(\boldsymbol{p}'_{\boldsymbol{gh}}, \bar{\boldsymbol{x}}'_{\boldsymbol{gh}}\right)'$ is the mean of the response-covariable vector.

Further, we transform $\bar{\boldsymbol{f}}_{\boldsymbol{gh}}$ to $\boldsymbol{F_{ghθ}} = \left(\boldsymbol{q}'_{\boldsymbol{ghθ}}, \bar{\boldsymbol{x}}'_{\boldsymbol{gh}}\right)'$, then $\boldsymbol{F_{ghθ}} = \left(\boldsymbol{A_3} \exp\left[\boldsymbol{A_2} \log\left(\boldsymbol{A_{1θ}} \boldsymbol{p_{gh}}\right)\right], \bar{\boldsymbol{x}}'_{\boldsymbol{gh}}\right)'$.

Also, we denote $\boldsymbol{u} = \sum_{h=1}^{H} w_h (\bar{\boldsymbol{x}}_{1h} - \bar{\boldsymbol{x}}_{2h})$ as the covariable difference weighted across the H strata and combine the J treatment differences and M covariable difference to get $\boldsymbol{G_θ} = \left(\widehat{\boldsymbol{d}}'_{\theta}, \boldsymbol{u}'\right)'$.

The consistent covariance matrix estimator $\boldsymbol{V}_{\bar{\boldsymbol{f}}_{\boldsymbol{gh}}}$ of $\bar{\boldsymbol{f}}_{\boldsymbol{gh}}$ can be obtained as in (2.16).

$$\boldsymbol{V}_{\bar{\boldsymbol{f}}_{\boldsymbol{gh}}} = \frac{1}{n_{gh}(n_{gh} - 1)} \sum_{i=1}^{n_{gh}} (\boldsymbol{f_{ghi}} - \bar{\boldsymbol{f}}_{\boldsymbol{gh}})(\boldsymbol{f_{ghi}} - \bar{\boldsymbol{f}}_{\boldsymbol{gh}})' \quad (2.16)$$

And then the covariance matrix $V_{F_{gh\theta}}$ of $F_{gh\theta}$ and $V_{G\theta}$ of G_θ can be obtained as (2.17) and (2.18).

$$\boldsymbol{V}_{\boldsymbol{F_{ghθ}}} = \begin{bmatrix} \boldsymbol{B_{ghθ}} & \boldsymbol{0_{J,M}} \\ \boldsymbol{0_{M,3J}} & \boldsymbol{I_M} \end{bmatrix} \boldsymbol{V}_{\bar{\boldsymbol{f}}_{\boldsymbol{gh}}} \begin{bmatrix} \boldsymbol{B}'_{\boldsymbol{ghθ}} & \boldsymbol{0_{3J,M}} \\ \boldsymbol{0_{M,J}} & \boldsymbol{I_M} \end{bmatrix} \quad (2.17)$$

$$\boldsymbol{V_{G\theta}} = \sum_{h=1}^{H} w_h^2 \left(\boldsymbol{V_{F_{1h\theta}}} + \boldsymbol{V_{F_{2h\theta}}} \right) \qquad (2.18)$$

Then the differences between means for covariables for the two treatment groups are restricted to zero, as is expected by randomization of patients to the two treatments; this constraint can be expressed as shown in (2.19).

$$\boldsymbol{E(u)} = \boldsymbol{0_M} \qquad (2.19)$$

Randomization-based covariable adjustment for the treatment comparison estimator $\boldsymbol{d_\theta}$ with respect to $\boldsymbol{u}$ can be invoked by fitting the linear model in (2.20) by weighted least squares

$$E(\boldsymbol{G_\theta}) \triangleq \begin{bmatrix} I_J \\ 0_{M,J} \end{bmatrix} \boldsymbol{b_\theta} = \boldsymbol{Z b_\theta} = \begin{pmatrix} \boldsymbol{b}_{1\theta} \\ \vdots \\ \boldsymbol{b}_{J\theta} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \qquad (2.20)$$

regression with weights based on $V_{G_\theta}^{-1}$ in (2.18) and with "$\triangleq$" meaning "is estimated by."

The weighted least squares regression for the specification in (2.20) produces the estimator for the covariable-adjusted treatment comparisons $\boldsymbol{b_\theta}$ in (2.21), with a consistent estimator for the covariance matrix of $\boldsymbol{b_\theta}$, as in (2.22).

$$\boldsymbol{b_\theta} = \left(\boldsymbol{Z' V_{G_\theta}^{-1} Z} \right)^{-1} \boldsymbol{Z' V_{G_\theta}^{-1} G_\theta} \qquad (2.21)$$

$$\boldsymbol{V_{b_\theta}} = \left(\boldsymbol{Z' V_{G_\theta}^{-1} Z} \right)^{-1} \qquad (2.22)$$

Hypothesis Testing

For testing H_0: $\boldsymbol{C\Delta} = 0$, the test statistic in (2.23) is applicable where $\widehat{\boldsymbol{\Delta}}$ is the

$$Q_{C\Delta} = \widehat{\boldsymbol{\Delta}}' \boldsymbol{C}' (\boldsymbol{C V_{\widehat{\Delta}} C}')^{-1} \boldsymbol{C} \widehat{\boldsymbol{\Delta}} \quad (2.23)$$

corresponding estimator $\boldsymbol{d_\theta}$ or covariable-adjusted estimator $\boldsymbol{b_\theta}$, and $\boldsymbol{C}$ is the desired full rank contrast matrix with rank $r(\boldsymbol{C})$. Under H_0, $Q_{\boldsymbol{C\Delta}} \sim \chi^2_{r(\boldsymbol{C})}$ when the sample size is sufficiently large, where $\chi^2_{r(\boldsymbol{C})}$ is the central Chi-square distribution with $\boldsymbol{r}(\boldsymbol{C})$ degrees of freedom.

2.3 Example

The proposed method for sensitivity analysis is illustrated with an example of a double-blind, randomized, placebo-controlled, parallel-group study to assess the safety and efficacy of a test medicine for weight loss in obese patients. The sample data consist of 1000 patients in a bootstrap sample from an obesity trial like that discussed in Smith et al. (2010).

One of the co-primary endpoints of this weight loss study is the proportion (%) of patients who achieve ≥5% weight loss from baseline to week 52. The study participants were followed every 4 weeks until the end of the study. The primary assessment was body weight at Week 52, and other important assessment visits were at Week 12, Week 24, and Week 36. These 4 visits are numbered Visits 1 to 4 in chronological order.

Substantial numbers of patients withdrew from the study and didn't return for follow up. In this bootstrap sample, at Week 52, 45.5% and 53.1% of patients had missing responses for the test and placebo group respectively; at all visits, more missingness happened in the placebo group than in the test group, as shown in Table 2.3.

Table 2.3 Missing Percentages of Assessment Visits

Visit / Group	1	2	3	4
Test	10.9%	28.8%	38.5%	45.5%
Placebo	21.8%	34.2%	43.2%	53.1%

Two strata according to gender were considered; and baseline weight, age, and baseline body mass index (BMI) were covariables with adjustment through randomization-based ANCOVA as discussed in Section 2.2.6.

One question of regulatory interest for this example is whether there are 15% or more responders for test treatment than placebo. If missing responses cannot be assumed to be ignorable, how robust the results are if challenged by a sequence of sensitivity parameters θ is a question of interest.

The hypotheses are H_{0j}: $\boldsymbol{I_{j,4}\Delta} \leq 15\%$ versus H_{Aj}: $\boldsymbol{I_{j,4}\Delta} > 15\%$ for $j = 1, 2, 3, 4$ of the 4 visits, where $\boldsymbol{I_{j,4}}$ denote the j^{th} row of the identity matrix $\boldsymbol{I_4}$, and it will be addressed with the direct treatment difference estimator and the covariables-adjusted treatment difference estimator respectively. For sequential testing to address multiplicity, the primary assessment Visit 4 H_{04} would be addressed as a first step with two-sided hypothesis testing at the $\alpha = 0.05$ level, and if significant, H_{03} would be addressed, etc.

The methodology can accommodate different sensitivity parameters for different visits, strata and treatment groups. But for convenience of illustration, we only consider different sensitivity parameters for the test and placebo treatment, and assume $\theta_{ghj} = \theta_g$ for $g = 1,2$ for test medicine and placebo groups, $h = 1, 2$ for females and males, $j = 1, 2, 3, 4$ for visits 1 to 4. For the placebo group, one would typically use $\theta_2 = 1$, which could be a realistic assumption since patients would usually experience similar results as if they remained in the placebo group. The values of $\theta_1 \leq 1$ for the test medicine group could address the assumptions that the post-withdrawal experience of a test drug patient was less favorable than for a patient remaining in the study. As a tipping point analyses, we use a sequence of θ_1 values to see under what assumptions the result of rejection of the null hypothesis would remain unchanged (at the two-sided

significance level of 0.05), although the attention can additionally be given to the point estimate and confidence interval. The estimates of unadjusted and adjusted treatment differences, and their standard errors (SE), and the Chi-Square values and corresponding p-values of the testing for the hypothesis H_0 above are listed in Table 2.4.

Table 2.4 Results of Sensitivity Analyses for Treatment Comparison Estimators

Sensitivity parameter θ	Visit	Unadjusted				Adjusted			
		d	SE	ChiSq	p value	b	SE	ChiSq	p value
$\theta_1 = 1$	1	0.266	0.0309	14.007	0.0002	0.266	0.0306	14.235	0.0002
	2	0.339	0.0358	27.825	<0.0001	0.341	0.0353	29.354	<0.0001
	3	0.331	0.0390	21.466	<0.0001	0.333	0.0384	22.611	<0.0001
	4	0.313	0.0424	14.817	0.0001	0.315	0.0418	15.496	0.0001
$\theta_1 = 0.5$	1	0.248	0.0306	10.146	0.0014	0.247	0.0304	10.231	0.0014
	2	0.294	0.0362	15.880	0.0001	0.296	0.0356	16.797	<0.0001
	3	0.272	0.0398	9.312	0.0023	0.273	0.0391	9.880	0.0017
	4	0.241	0.0434	4.387	0.0362	0.242	0.0427	4.634	0.0313
$\theta_1 = 1/3$	1	0.239	0.0304	8.547	0.0035	0.238	0.0302	8.571	0.0034
	2	0.269	0.0361	10.851	0.0010	0.270	0.0355	11.478	0.0007
	3	0.237	0.0398	4.742	0.0294	0.238	0.0390	5.056	0.0245
	4	0.198	0.0431	1.255	0.2627	0.199	0.0423	1.342	0.2467

Note: $\theta_2 = 1$

When the missingness is MCAR in either the test group or placebo group with $\theta_1 = \theta_2 = 1$, the conclusion that the test treatment had 15% or more responders than the placebo treatment is well supported, at all follow-up visits. As we place more stringent penalties on the missing data in the test treatment while keeping the placebo missingness as MCAR, the estimated treatment difference becomes smaller, particularly for visits 3 and 4 where missing data are much more extensive. At visit 4, when the assumption is made that the test treatment's missing response has only (1/3) the odds for favorable outcome as observed responses, the test treatment has about 20% more responders than the placebo, with standard error at about 4%, and the conclusion of 15% more responders for comparing test to placebo no longer holds. The results of

the estimates of the unadjusted difference d_θ and covariable adjusted difference b_θ are similar. And the standard errors of the adjusted difference b_θ estimator are only slightly smaller than those of the unadjusted d_θ.

2.4 Discussion

For situations where MNAR missing dichotomous response data exist for a randomized clinical trial, this paper discusses how the mathematical re-distribution of missing responses can provide useful sensitivity analyses to address the robustness of treatment comparisons from methods with possibly unrealistic assumptions such as MCAR. The tipping point, which is the sensitivity parameter that turns a significant result into a nonsignificant one, can be examined to see whether it is realistic or not relative to knowledge about the nature of the treatment and disorder being studied. Also, the sensitivity analyses are applicable with Mantel-Haenszel adjustment for strata and/or covariables through randomization-based ANCOVA. An extension of the methods in this paper to an ordered categorical outcome is in Chapter 3.

REFERENCES

Koch, G.G., Tangen, C.M., Jung, J.G. and Amara, I.A. (1998) Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. Statistics in Medicine, 17, 1863–1892.

Koch, G.G., Landis, J.R., Freeman, J.L., Freeman, D.H., Jr., and Lehnen, R.G. (1977). A general methodology for the analysis of experiments with repeated measurements of categorical data. *Biometrics*, 33, 133-158.

LaVange, L.M., Durham, T.A. and Koch, G.G. (2005) Randomization-based nonparametric methods for the analysis of multicentre trials. *Statistical Methods in Medical Research*, **14**, 281–301.

Little, R.J.A. and Rubin, D.B. (2014) *Statistical Analysis with Missing Data*. John Wiley & Sons.

National Research Council Panel. (2010) *The Prevention and Treatment of Missing Data in Clinical Trials*. National Academies Press (US), Washington (DC).

O'Kelly, M. and Ratitch, B. (2014) *Clinical Trials with Missing Data: A Guide for Practitioners*. John Wiley & Sons.

Rubin, D.B. (1976) Inference and missing data. *Biometrika*, **63**, 581–592.

Smith, S.R., Weissman, N.J., Anderson, C.M., Sanchez, M., Chuang, E., Stubbe, S., et al. (2010) Multicenter, Placebo-Controlled Trial of Lorcaserin for Weight Management. *New England Journal of Medicine*, **363**, 245–256.

Tamura, R.N., Huang, X. and Boos, D.D. (2011) Estimation of treatment effect for the sequential parallel design. *Statistics in Medicine*, **30**, 3496–3506.

Tangen, C.M. and Koch, G.G. (2001) Non-parametric analysis of covariance for confirmatory randomized clinical trials to evaluate dose–response relationships. *Statistics in Medicine*, **20**, 2585–2607.

Zhao, Y., Preisser, J.S. and Koch, G.G. (2014) Sensitivity Analysis for Withdrawals in Grouped Time-to-Event Data. *Statistics in Biopharmaceutical Research*, **6**, 41–54.

CHAPTER 3: SENSITIVITY ANALYSIS IN TREATMENT COMPARISON FOR MISSING ORDINAL DATA IN MULTIE-VISIT RANDOMIZED CLINICAL TRIAL

3.1 Introduction

The methodology in Chapter 2 is extended to an ordinal outcome with the comparison between treatments utilizing the Mann Whitney probability estimator for the ordinal nature of the outcome. Section 3.2 will introduce the data set up and the methodology for ordinal data, and they will be illustrated with an example in Section 3.3.

3.2 Methods

3.2.1 Notation

Consider a study comparing a test treatment and a control treatment for an ordinal outcome with K categories and assessments at each of the J visits. For an ordinal outcome with missing data, there are $(K+1)$ possible responses, with these being the K ordered outcome categories plus the missing category. While outcome and response are sometimes interchangeable in the literature, here we make a distinction between outcome categories and response categories, where the former could only have K outcomes and the latter includes missing as a category.

Let y_{gijk} be the indicator variable for the response of subject i in group g at visit j being k, where group $g = 1, 2$ index the test treatment and control treatment respectively; subject $i = 1, 2, \cdots, n_g$; visit $j = 1, 2, \cdots, J$, and response $k = 1, 2, \cdots, K, K+1$, where $1, \ldots, K$ index the 1 to K outcomes and $(K+1)$ indexes missing response respectively. We define a $(K+1)$-

dimensional vector $Y_{gij*} = (y_{gi11}, \cdots, y_{gi1K}, y_{gi1(K+1)})'$ to combine the $(K+1)$ indicators. For example, $Y_{111*} = (0, \cdots, 1, 0)$ means Subject 1 in the test treatment at time point 1 has outcome category as K. Accordingly, we further define a data vector that includes all visits as $Y_{gi**} = \left(Y'_{gi1*}, \cdots, Y'_{giJ*}\right)' = \left(y_{gi11}, \cdots, y_{gi1K}, y_{gi1(K+1)}, \ldots, y_{giJ1}, \cdots, y_{giJK}, y_{giJ(K+1)}\right)'$.

3.2.2 Data Structure

For subjects in treatment group g, the observed data can be arranged in a contingency table format as Table 3.1 below. After including missing as a category, the number of subjects at each visit is fixed as n_g.

Table 3.1 Data Structure

Time \ Response	1	⋯	K	$K+1$	Total
1	$n_{g11}(p_{g11})$	⋯	$n_{g1K}(p_{g1K})$	$n_{g1(K+1)}(p_{g1(K+1)})$	n_g
⋮	⋮		⋮	⋮	n_g
J	$n_{gJ1}(p_{gJ1})$	⋯	$n_{gJK}(p_{gJK})$	$n_{gJ(K+1)}(p_{gJ(K+1)})$	n_g

The counts vector at row j of the table is expressed as $\boldsymbol{n_{gj}} = (n_{gj1}, \cdots, n_{ghjK}, n_{ghj(K+1)})$, which marginally follows a multinomial distribution, $\boldsymbol{n_{gj}} \sim Multinominal(n_g, \boldsymbol{\pi_{gj}})$, where $\boldsymbol{\pi_{gj}}$ is the marginal multinomial probability vector of responses. From the properties of the multinomial distribution, the unbiased estimator of the multinomial probability $\boldsymbol{\pi_{gj}}$, is the proportion vector at row j, $\boldsymbol{p}_{gj} = \left(p_{gj1}, \cdots, p_{gjK}, p_{gj(K+1)}\right)$. Combining across all time points, $\boldsymbol{p_g} = \left(\boldsymbol{p'_{g1}}, \ldots, \boldsymbol{p'_{gJ}}\right)'$, and $\boldsymbol{\pi_g} = \left(\boldsymbol{\pi'_{g1}}, \ldots, \boldsymbol{\pi'_{gJ}}\right)'$. Extensions to allow strata would proceed similarly to the methods discussed in Chapter 2 for dichotomous outcomes. An unbiased estimator of the covariance matrix of the probability estimator $\boldsymbol{p_g}$ containing the marginal proportions of the correlated multinomial distributions in the rows of Table 3.1 is shown in (3.1).

$$\boldsymbol{V}_{\boldsymbol{p}_g} = \frac{1}{n_g(n_g-1)}\sum_{i=1}^{n_g}(\boldsymbol{Y}_{\boldsymbol{gi}**} - \boldsymbol{p}_g)(\boldsymbol{Y}_{\boldsymbol{gi}**} - \boldsymbol{p}_g)' \qquad (3.1)$$

3.2.3 Multinomial Probability Estimation

The objective of the estimation is the distribution of outcomes; and for an ordinal outcome, cumulative probabilities are usually considered. Without loss of generality, assume lower values of the ordinal outcome are more favorable, i.e., $k = 1$ is the most favorable outcome. Under MCAR, the cumulative favorable proportion of the first l categories for visit j is expressed in (3.2).

$$\sum_{k=1}^{l} q_{gjk} = \frac{\sum_{k=1}^{l} p_{gjk}}{\sum_{k=1}^{K} p_{gjk}} = \sum_{k=1}^{l} p_{gjk} + \frac{\sum_{k=1}^{l} p_{gjk}}{\sum_{k=1}^{K} p_{gjk}}\, p_{gj(K+1)} \qquad (3.2)$$

As MCAR, (3.2) implies that the probability of each outcome category in the patients with missing responses is the same for those with responses observed.

3.2.4 Sensitivity Analysis

For the subsequent discussion, we omit the notation for group g for simplicity of presentation. Now let $n_{j1},\cdots, n_{jK}$, and $n_{j(K+1)}$ represent the counts of the K ordinal outcome categories and the missing responses at visit j, and $n_{j1} + \cdots + n_{jK} + n_{j(K+1)} = n_j$. We further divide the missing counts $n_{j(K+1)}$ into the counts of the K outcome categories $n_{j(K+1)1}, \cdots, n_{j(K+1)K}$, and they represent the counts among the patients with missing status for the K outcome categories respectively if their missing status did not occur, and so we have the data structure as Table 3.2.

Table 3.2 Data Structure for Missing Counts Redistribution

Visit	1	⋯	K	Missing	Total
j	n_{j1}	⋯	n_{jK}	$n_{j(K+1)} = n_{j(K+1)1} + \cdots + n_{j(K+1)K}$	n_j

The odds ratio comparing the first l categories to the last $(K-l)$ categories in the patients with missing responses versus those with observed responses as shown in (3.3), which implies (3.4) for the redistribution to the first l categories combined.

$$\theta_{jl} = \frac{\sum_{k=1}^{l} n_{j(K+1)k}}{\sum_{k=l+1}^{K} n_{j(K+1)k}} \Big/ \frac{\sum_{k=1}^{l} n_{jk}}{\sum_{k=l+1}^{K} n_{jk}}, l = 1, \cdots, (K-1). \qquad (3.3)$$

$$\sum_{k=1}^{l} n_{j(K+1)k} = \frac{\theta_{jl} \sum_{k=1}^{l} n_{jk}}{\theta_{jl} \sum_{k=1}^{l} n_{jk} + \sum_{k=l+1}^{K} n_{jk}} n_{j(K+1)} \qquad (3.4)$$

If $\theta_{jl} = 1$ for all $l = 1, \cdots, (K-1)$, an MCAR-like structure applies; if $\theta_{jl} > 1$, the patients with missing status are assumed to be more likely to have better outcome than those observed; if $\theta_{jl} < 1$, the patients with missing status are assumed to be more likely to have worse outcome Furthermore, if $\theta_{j1} = \cdots = \theta_{j(K-1)}, K \geq 2$, a proportional odds assumption is imposed.

Through adjusting for different θ, different specifications of the possible outcomes in the missing observations could be attained, and thus we call θ the sensitivity parameter for missingness. The adjusted cumulative proportion estimator could be expressed in terms of the unadjusted proportions $\boldsymbol{p}$ and θ_{jl} as shown in (3.5).

$$\sum_{k=1}^{l} q_{jk\theta} = \frac{\sum_{k=1}^{l} n_{jk} + \sum_{k=1}^{l} n_{j(K+1)k}}{n_j} = \frac{1}{n_j}\left(\sum_{k=1}^{l} n_{jk} + \frac{\theta_{jl} \sum_{k=1}^{l} n_{jk}}{\theta_{jl} \sum_{k=1}^{l} n_{jk} + \sum_{k=l+1}^{K} n_{jk}} n_{j(K+1)}\right)$$

$$= \sum_{k=1}^{l} p_{jk} + \frac{\theta_{jl} \sum_{k=1}^{l} p_{jk}}{\theta_{jl} \sum_{k=1}^{l} p_{jk} + \sum_{k=l+1}^{K} p_{jk}} p_{j(K+1)}, \quad l = 1, \cdots, (K-1) \qquad (3.5)$$

$$q_{jK\theta} = 1 - \sum_{k=1}^{(K-1)} q_{jk\theta}$$

Now we consider group g, and we express the adjusted cumulative proportion vector $\boldsymbol{q_{g\theta}} = \left(\boldsymbol{q'_{g1\theta}}, \dots, \boldsymbol{q'_{gJ\theta}}\right)'$ with $\boldsymbol{q_{gj\theta}} = \left(q_{gj1\theta}, \dots, q_{gjK\theta}\right)'$ in the form of compound functions of the unadjusted proportion vector $\boldsymbol{p_g}$ and sensitivity parameters $\boldsymbol{\theta}$ as shown in (3.6) in order to use

$$\boldsymbol{q_{g\theta}} = \boldsymbol{A_4 A_3 exp}\left[\boldsymbol{A_2\, log}\left(\boldsymbol{A_{1g\theta} p_g}\right)\right] + \boldsymbol{c} \qquad (3.6)$$

linear Taylor series methods to produce a consistent estimator for the corresponding matrix. For (3.6), matrices $\boldsymbol{A_{1g\theta}}$, $\boldsymbol{A_2}$, $\boldsymbol{A_3}$, and $\boldsymbol{A_4}$, and $(JK \times 1)$ vector $\boldsymbol{c}$ are shown in (3.7) and (3.8).

$$\boldsymbol{A_{1g\theta}} = \boldsymbol{bdiag_J}\left(\boldsymbol{A_{1gj\theta}}\right)$$

$$\boldsymbol{A_{1gj\theta}} = \begin{bmatrix} \boldsymbol{A_{1gj1\theta}} \\ \boldsymbol{A_{1gj2\theta}} \\ \vdots \\ \boldsymbol{A_{1gj(K-1)\theta}} \end{bmatrix}_{3(K-1)\times(K+1)} \qquad (3.7)$$

$$\boldsymbol{A_{1gjk\theta}} = \begin{pmatrix} \boldsymbol{1'_k}\ \boldsymbol{0'_{K-k}}\ 0 \\ \boldsymbol{\theta_{gjk}}\boldsymbol{1'_k}\ \boldsymbol{1'_{K-k}}\ 0 \\ \boldsymbol{0'_k}\boldsymbol{0'_{K-k}}\ \theta_{gjk} \end{pmatrix}_{3\times(K+1)}, k = 1, \cdots, (K-1).$$

$$\boldsymbol{A_2} = \boldsymbol{bdiag_J}\left\{\boldsymbol{bdiag_{(K-1)}}\begin{pmatrix} 1\ 0 & 0 \\ 1 & -1\ 1 \end{pmatrix}\right\} \qquad (3.8)$$

$$\boldsymbol{A_3} = \boldsymbol{bdiag_J}\left\{\boldsymbol{bdiag_{(K-1)}}(1\ 1)\right\}$$

$$\boldsymbol{A_4} = \boldsymbol{bdiag_J}\begin{pmatrix} 10 & 00 & 00 \\ -11 & 00 & 00 \\ 0-1 & 10 & 00 \\ \vdots\cdots & \cdots\cdots & \cdots\vdots \\ 00 & 00 & -11 \\ 00 & 00 & 0-1 \end{pmatrix}_{K\times(K-1)}$$

$$\boldsymbol{c} = (0, 0, \dots, 0, 1, \dots, \dots, \dots, 0, 0, \dots, 0, 1)'_{(JK\times 1)}$$

By applying linear Taylor series, we can obtain a consistent estimator for the covariance matrix of $\boldsymbol{q_{g\theta}}$ as shown in (3.9) and (3.10) where $\boldsymbol{B}_{1g\theta}$ is the elementwise first partial derivative

$$\boldsymbol{V_{q_{g\theta}}} = \boldsymbol{B}_{1g\theta}\boldsymbol{V_{p_g}}\boldsymbol{B}'_{1g\theta} \qquad (3.9)$$

$$\boldsymbol{B_{1g\theta}} = \frac{\boldsymbol{\partial q_{g\theta}}}{\boldsymbol{\partial p_g}} = \frac{\boldsymbol{\partial q_{g\theta}}}{\boldsymbol{\partial a_{3\theta}}}\frac{\boldsymbol{\partial a_{3\theta}}}{\boldsymbol{\partial a_{2\theta}}}\frac{\boldsymbol{\partial a_{2\theta}}}{\boldsymbol{\partial a_{1\theta}}}\frac{\boldsymbol{\partial a_{1\theta}}}{\boldsymbol{\partial p_g}} = \boldsymbol{A_4 A_3 D_{a_{3\theta}} A_2 D^{-1}_{a_{1\theta}} A_{1g\theta}} \quad (3.10)$$

of vector $\boldsymbol{q_{g\theta}}$ with respect to vector $\boldsymbol{p_g}$ and is obtained by the chain rule, with $\boldsymbol{a_{1g\theta}} = \boldsymbol{A_{1g\theta}p_g}$, $\boldsymbol{a_{2g\theta}} = \mathbf{log}(\boldsymbol{a_{1g\theta}})$, $\boldsymbol{a_{3g\theta}} = \mathbf{exp}[\boldsymbol{A_2 a_{2g\theta}}]$, and $\boldsymbol{q_{g\theta}} = \left(\boldsymbol{A_4 A_3 a_{3g\theta}} + \boldsymbol{c}\right)$, and $\boldsymbol{D^{-1}_{a_{1g\theta}}}$ is a diagonal matrix with the reciprocals of the elements of the vector $\boldsymbol{a}_{1g\theta}$ on the main diagonal and $\boldsymbol{D_{a_{3g\theta}}}$ is the diagonal matrix with the elements of the vector $\boldsymbol{a_{3g\theta}}$ on the main diagonal.

Typically, if we assume θ_{gjk} are the same for all categories k, i.e., $\theta_{gjk} = \theta_{gj}$, and they could vary at different visits j and j', and so a proportional odds assumption is made. In addition, one often could expect missing responses for patients in the placebo group to have a similar outcome distribution as the observed patients, which means $\theta_{2jk} = 1$ for any visit j in the placebo group could be specified.

3.2.5 Treatment Comparison

Mann-Whitney Probability for Treatment Difference

A Mann-Whitney probability estimator (MW estimator) can be used for the treatment comparison for a strictly ordinal outcome. Comparing the outcome of test treatment to placebo, the Mann-Whitney probability estimates the probability that a randomly selected patient with the test treatment has better outcome than a randomly selected patient with placebo. If the two treatments are equally effective, the chance of having a better response for the test treatment would be 0.5.

Denote the Mann-Whitney probability as ξ_j for visit j. The null hypothesis for comparing the test treatment with the placebo is $H_{0j}: \xi_j < 0.5$ versus $H_{Aj}: \xi_j \geq 0.5, , j = 1, \cdots, J$.

Denote T_{gj} as the outcome for a subject in group g at visit j. If a smaller value of the outcome is better, the probability ξ_j of test treatment being better than placebo at visit j is expressed in (3.11).

Denote T_{gj} as the outcome for a subject in group g at visit j. If a smaller value of the outcome is better, the probability ξ_j of test treatment being better than placebo at visit j is expressed in (3.11).

$$\xi_j = P(T_{1j} \leq T_{2j}) = \sum_{k=1}^{K} P(T_{1j} = k)\left[P(T_{2j} \geq k) - \frac{1}{2}P(T_{2j} = k)\right] \quad (3.11)$$

Then the Mann-Whitney probability estimator $r_{j\theta}$ of ξ_j at visit j, for the sensitivity analysis is expressed as shown in (3.12).

$$r_{j\theta} = \sum_{k=1}^{K} q_{1jk\theta}\left(\sum_{l=k}^{K} q_{2jk\theta} - \frac{1}{2}q_{2jk\theta}\right) \quad (3.12)$$

Letting $\boldsymbol{r_\theta} = (r_{1\theta}, \cdots, r_{J\theta})'$, the Mann-Whitney probability estimator could be expressed as the compound function of the proportions $\boldsymbol{q_\theta} = (\boldsymbol{q'_{1\theta}}, \boldsymbol{q'_{2\theta}})'$ as shown in (3.13) where matrices $\boldsymbol{A_5}, \boldsymbol{A_6},$ and $\boldsymbol{A_7}$ are defined in (3.14) where T_K is a $K \times K$ upper triangular matrix with all

$$\boldsymbol{r_\theta} = \boldsymbol{A_7}\,\mathbf{exp}[\boldsymbol{A_6}\,\boldsymbol{log}(\boldsymbol{A_5 q_\theta})] \quad (3.13)$$

$$\boldsymbol{A_5} = \begin{pmatrix} \boldsymbol{I_{JK}} & \boldsymbol{0_{JK,JK}} \\ \boldsymbol{0_{JK,JK}} & \boldsymbol{bdiag_J}(\boldsymbol{T_K} - 0.5\boldsymbol{I_K}) \end{pmatrix}_{2KJ \times 2KJ}, \boldsymbol{A_6} = \left(\boldsymbol{I_{JK}}, \boldsymbol{I_{JK}}\right)_{KJ \times 2KJ}, \boldsymbol{A_7}$$

$$= \boldsymbol{bdiag_J}(\boldsymbol{1'_K})_{J \times KJ} \quad (3.14)$$

elements on or below the diagonal equal to 1.

We apply the chain rule again in (3.15) to obtain the first partial derivative matrix of $\hat{\xi}$

$$\boldsymbol{B_{2\theta}} = \frac{\partial \boldsymbol{r_\theta}}{\partial \boldsymbol{q_\theta}} = \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{a_6}} \frac{\partial \boldsymbol{a_6}}{\partial \boldsymbol{a_5}} \frac{\partial \boldsymbol{a_5}}{\partial \boldsymbol{a_4}} \frac{\partial \boldsymbol{a_4}}{\partial \boldsymbol{q_\theta}} = \boldsymbol{A_7 D_{a_6} A_6\, D_{a_4}^{-1} A_5} \qquad (3.15)$$

with respect to $\boldsymbol{q_\theta}$ with $\boldsymbol{a_{4\theta}} = \boldsymbol{A_5 q_\theta}$, $\boldsymbol{a_{5\theta}} = \mathbf{log}(\boldsymbol{a_{4\theta}})$,$\boldsymbol{a_{6\theta}} = \mathbf{exp}[\boldsymbol{A_6 a_{5\theta}}]$, and $\boldsymbol{r_\theta} = \boldsymbol{A_7 a_{6\theta}}$.

Thus, a consistent estimator for the covariance matrix $\boldsymbol{V_{r_\theta}}$ of $\boldsymbol{r_\theta}$ is obtained on the basis of linear Taylor series approximations, and it is expressed in (3.16) where $\boldsymbol{V_{q_\theta}} = \boldsymbol{bdiag}(\boldsymbol{V_{q_{1\theta}}}, \boldsymbol{V_{q_{2\theta}}})$.

$$\boldsymbol{V_{r_\theta}} = \boldsymbol{B_{2\theta} V_{q_\theta} B'_{2\theta}} \qquad (3.16)$$

Adjusted Treatment Difference via Randomization-Based ANCOVA

In a randomized clinical trial, baseline covariables are expected to have the same distributions in the randomized groups. Baseline covariables could include the baseline measurements of the outcome, demographic variables, or other variables to be taken into account. However, random imbalances in the baseline covariables could occur as each treatment group is a finite sample of the randomized population.

Covariance adjustment can provide variance reduction in estimation for treatment comparisons, together with correction for random imbalances between treatments (Koch et al., 1998b; Tangen and Koch, 2001; LaVange et al., 2005). The motivation behind the variance reduction is that the parameter estimator that is associated with the imbalance is corrected to offset the imbalance in covariables.

The estimation is through nonparametric ANCOVA, which is an approach that applies weighted least squares methods to evaluate comparisons between treatment group with respect to outcome variables and covariables simultaneously (Koch et al., 1998b).

Here, we introduce some notation for the covariables.

Suppose each subject has M baseline covariables $\boldsymbol{x_{gi}}$, $\boldsymbol{x_{ghi}} = \left(x_{gi1}, \ldots, x_{giM}\right)'$, and $\overline{\boldsymbol{x}}_{\boldsymbol{g}} = \frac{1}{n_{gh}} \sum_{i=1}^{n_{gh}} \boldsymbol{x_{gi}}$ is the mean vector of the prespecified covariables.

We define $\boldsymbol{f_{gi}} = \left(\boldsymbol{Y}'_{\boldsymbol{gi}**}, \boldsymbol{x}'_{\boldsymbol{gi}}\right)'$ as a $((K+1) \times J + M)$ dimensional response-covariable vector, and $\bar{\boldsymbol{f}}_{\boldsymbol{g}} = \frac{1}{n_{gh}} \sum_{i=1}^{n_{gh}} \boldsymbol{f_{gi}} = \left(\boldsymbol{p}'_{\boldsymbol{g}}, \overline{\boldsymbol{x}}'_{\boldsymbol{g}}\right)'$ is the mean of the responses-covariables.

We define a $(J+M)$ dimensional vector $\boldsymbol{G_\theta} = (\boldsymbol{r}'_{\boldsymbol{\theta}}, \boldsymbol{u}')'$ to include both the treatment comparisons $\boldsymbol{r_{g\theta}}$ and the covariable differences $\boldsymbol{u} = \overline{\boldsymbol{x}}_{\boldsymbol{1}} - \overline{\boldsymbol{x}}_{\boldsymbol{2}}$.

The consistent covariance matrix estimator $\boldsymbol{V}_{\bar{\boldsymbol{f}}_{\boldsymbol{g}}}$ of $\bar{\boldsymbol{f}}_{\boldsymbol{g}}$ can be obtained nonparametrically as shown in (3.17).

For $\bar{\boldsymbol{f}} = (\boldsymbol{p}'_{\boldsymbol{1}}, \overline{\boldsymbol{x}}_{\boldsymbol{1}}, \boldsymbol{p}'_{\boldsymbol{2}}, \overline{\boldsymbol{x}}_{\boldsymbol{2}})'$, we have $\boldsymbol{V}_{\bar{\boldsymbol{f}}} = \boldsymbol{bdiag}\left(\boldsymbol{V}_{\bar{\boldsymbol{f}}_{\boldsymbol{1}}}, \boldsymbol{V}_{\bar{\boldsymbol{f}}_{\boldsymbol{2}}}\right)$. Also, we let $\tilde{\boldsymbol{f}} = (\boldsymbol{p}'_{\boldsymbol{1}}, \boldsymbol{p}'_{\boldsymbol{2}}, \overline{\boldsymbol{x}}'_{\boldsymbol{1}}, \overline{\boldsymbol{x}}'_{\boldsymbol{2}})'$, and so $\tilde{\boldsymbol{f}} = \boldsymbol{B_f}\bar{\boldsymbol{f}}$ with (3.18), the covariance matrix estimator $\boldsymbol{V_{G_\theta}}$ of $\boldsymbol{G_\theta}$ can be obtained as (3.19) where $\boldsymbol{B_{1\theta}} = \boldsymbol{bdiag}(\boldsymbol{B_{11\theta}}, \boldsymbol{B_{12\theta}})$ and $\boldsymbol{V}_{\tilde{\boldsymbol{f}}} = \boldsymbol{B_f V}_{\bar{\boldsymbol{f}}} \boldsymbol{B}'_{\boldsymbol{f}}$.

$$\boldsymbol{B_f} = \begin{pmatrix} \boldsymbol{I_{J(K+1)}} & \boldsymbol{0_{J(K+1),M}} & \boldsymbol{O_{J(K+1),J(K+1)}} & \boldsymbol{0_{J(K+1),M}} \\ \boldsymbol{O_{J(K+1),J(K+1)}} & \boldsymbol{0_{J(K+1),M}} & \boldsymbol{I_{J(K+1)}} & \boldsymbol{0_{J(K+1),M}} \\ \boldsymbol{0_{M,J(K+1)}} & \boldsymbol{I_M} & \boldsymbol{0_{M,J(K+1)}} & \boldsymbol{O_{M,M}} \\ \boldsymbol{0_{M,J(K+1)}} & \boldsymbol{O_{M,M}} & \boldsymbol{0_{M,J(K+1)}} & \boldsymbol{I_M} \end{pmatrix}_{\boldsymbol{2((K+1)J+M) \times 2((K+1)J+M)}} \tag{3.18}$$

$$\boldsymbol{V_{G_\theta}} = \begin{bmatrix} \boldsymbol{B_{2\theta}} \boldsymbol{B_{1\theta}} & \boldsymbol{0_{J,M}} & \boldsymbol{0_{J,M}} \\ \boldsymbol{0_{2J,M}} & \boldsymbol{I_M} & \boldsymbol{-I_M} \end{bmatrix} \boldsymbol{V}_{\bar{\boldsymbol{f}}} \begin{bmatrix} \boldsymbol{B_{2\theta}} \boldsymbol{B_{1\theta}} & \boldsymbol{0_{J,M}} & \boldsymbol{0_{J,M}} \\ \boldsymbol{0_{2J,M}} & \boldsymbol{I_M} & \boldsymbol{-I_M} \end{bmatrix}' \tag{3.19}$$

Weighted least squares regression can be applied to $\boldsymbol{G_\theta}$ so as to account for the constraints for the expected differences of means for covariables between the two treatment groups to be zero on the basis of randomization of patients to the two treatments; such constraints are expressed in (3.20).

$$\boldsymbol{E(u) = 0_M} \quad (3.20)$$

Randomization-based covariable adjustment for the treatment comparison estimator $\boldsymbol{r_\theta}$ with respect to $\boldsymbol{u}$ can be invoked by fitting the linear model in (3.21) by weighted least squares

$$E(\boldsymbol{G_\theta}) \triangleq \begin{bmatrix} \boldsymbol{I_J} \\ \boldsymbol{0_{MJ}} \end{bmatrix} \boldsymbol{b} = \boldsymbol{Zb} = \begin{pmatrix} \boldsymbol{b}_1 \\ \vdots \\ \boldsymbol{b}_J \\ 0 \\ \vdots \\ 0 \end{pmatrix} \qquad (3.21)$$

regression with weights based from $\boldsymbol{V_{G_\theta}^{-1}}$ in (3.19) and with "$\triangleq$" meaning "is estimated by."

The weighted least squares regression yields the estimator $\boldsymbol{b_\theta}$ in (3.22) for the

$$\boldsymbol{b_\theta = \left(Z' V_{G_\theta}^{-1} Z\right)^{-1} Z' V_{G_\theta}^{-1} G_\theta} \qquad (3.22)$$

covariable-adjusted treatment comparisons, and the consistent estimator for the covariance matrix of $\boldsymbol{b_\theta}$ is $\boldsymbol{V_{b_\theta}}$ in (3.23).

$$\boldsymbol{V_{b_\theta} = \left(Z' V_{G_\theta}^{-1} Z\right)^{-1}} \qquad (3.23)$$

Hypothesis Testing

For the hypothesis H_0: $\boldsymbol{C\xi = 0}$, we have the test statistic in (3.24) where $\hat{\boldsymbol{\xi}}$ is the

$$Q_{C\xi} = \hat{\boldsymbol{\xi}}' \boldsymbol{C}' \left(\boldsymbol{C V_{\hat{\xi}} C}'\right)^{-1} \boldsymbol{C} \hat{\boldsymbol{\xi}} \quad (3.24)$$

corresponding treatment comparison estimator $\boldsymbol{r_\theta}$ or covariable-adjusted estimator $\boldsymbol{b_\theta}$ of the Mann Whitney criteria $\boldsymbol{\xi}$, and $\boldsymbol{C}$ is the specified contrast matrix. Under H_0, $Q_{C\xi} \sim \chi^2_{rank(\boldsymbol{C})}$ when the sample size is sufficiently large, where $\chi^2_{rank(\boldsymbol{C})}$is the central Chi-square distribution with $rank(\boldsymbol{C})$ degrees of freedom, and $rank(\boldsymbol{C})$ is the rank of the full rank contrast matrix $\boldsymbol{C}$.

3.3 Example

The proposed method is illustrated with an example based on an adaptation of data in Stanish et al [1978]. The clinical trial was to evaluate the efficacy of a new drug relative to placebo for skin conditions.

One of the endpoints for this skin condition study is disease condition improvement from baseline at visit 3. The study participants were followed for 3 visits and the ordinal outcome is 1=Rapidly Improving, 2=Slowly Improving, 3=Stable, 4=Slowly Worsening, 5=Rapidly worsening.

The study had three visits. Sample sizes were 88 (test) and 84 (placebo) respectively. The missing data count in each visit is shown in Table 3.3.

Table 3.3 Missing Counts of Assessment Visits

Visit / Group	1	2	3
Test	2	9	15
Placebo	1	7	15

The question of interest is whether patients in the test treatment improved more than those in the placebo treatment. If the loss to follow-up cannot be assumed to be ignorable, the robustness of the results to challening by a sequence of sensitivity parameters θ is of interest.

The hypotheses are H_{0j}: $\boldsymbol{I_{j,3}}\boldsymbol{\xi} \leq 0.5$ versus H_{Aj}: $\boldsymbol{I_{j,3}}\boldsymbol{\xi} > 0.5$ for $j = 1, 2, 3$ of the 3 visits, where $\boldsymbol{I_{j,3}}$ denotes the j^{th} row of the identity matrix $\boldsymbol{I_3}$, and it will be addressed with the direct treatment comparison estimator and the covariable adjusted treatment comparison estimator respectively. For sequential testing, the primary assessment Visit 3 for H_{03} would be addressed as a first step with two-sided $\alpha = 0.05$, and if significant, H_{02} would be addressed, etc.

The methodology can accommodate different sensitivity parameters for different visits and treatment groups and different specifications for the odds among the ordinal categories. But

for convenience of illustration, we only consider different sensitivity parameters for the test and placebo treatment with the proportional odds assumption, and specify $\theta_{ghj} = \theta_g$ for $g = 1,2$ for test and placebo groups and $j = 1, 2, 3$ for visits 1 to 3. For the placebo group, one would typically use $\theta_2 = 1$, which could be a realistic assumption if patients with missing outcomes would usually experience similar results as if they remained in the placebo group. The values of $\theta_1 \leq 1$ for the test group could address the specifications that the post-withdrawal experience of a test drug patient was less favorable than patients with observed outcomes.

As a tipping point analysis, we use a sequence of θ_1 values to see under what specifications the result of rejection of the null hypothesis would remain unchanged (at the two sided significance level of 0.05). The estimates for unadjusted and adjusted treatment comparisons, and their standard errors (SE), and the chi-square values and corresponding p values of the testing for the hypothesis H_0 above are listed in Table 3.4.

When the loss to follow-up is MCAR in either the test group or the placebo group with $\theta_1 = \theta_2 = 1$, the conclusion that the test treatment is better than the placebo treatment is well supported at the primary assessment visit (Visit 3), but not so significant at Visits 2 and 1. As we place more stringent penalties on the missing data for the test treatment while keeping the placebo missingness as MCAR, the estimator for the treatment comparison becomes closer to 0.5. At Visit 3, when the specification is made that the odds for missing outcomes is only (1/3) as good as for observed outcomes for the test treatment, the probability for test being better than placebo is 0.589 with a standard error at about 0.047, and the conclusion that the test treatment is better than the placebo no longer holds.

The results of the estimates for the unadjusted comparisons $\boldsymbol{r_\theta}$ and the covariable adjusted comparisons $\boldsymbol{b_\theta}$ are similar. And the standard errors of the adjusted comparison $\boldsymbol{b_\theta}$ are only slightly smaller than those of the unadjusted $\boldsymbol{r_\theta}$.

Table 3.4 Results of Sensitivity Analyses

Sensitivity parameter θ	Visit	Unadjusted				Adjusted			
		r	SE	Chi-Sq	p value	b	SE	Chi-Sq	p value
$\theta_1 = 1$	1	0.565	0.0429	2.294	0.130	0.562	0.0426	2.121	0.145
	2	0.581	0.0449	3.294	0.070	0.580	0.0448	3.171	0.075
	3	0.614	0.0457	6.249	0.012	0.615	0.0457	6.272	0.012
$\theta_1 = 0.5$	1	0.563	0.0429	2.128	0.145	0.560	0.0427	1.960	0.162
	2	0.570	0.0452	2.409	0.121	0.568	0.0451	2.276	0.131
	3	0.598	0.0464	4.490	0.034	0.598	0.0464	4.473	0.034
$\theta_1 = 1/3$	1	0.561	0.0429	2.038	0.153	0.558	0.0427	1.872	0.171
	2	0.563	0.0453	1.952	0.162	0.561	0.0452	1.818	0.178
	3	0.589	0.0467	3.597	0.058	0.588	0.0467	3.564	0.059

Note: $\theta_2 = 1$

REFERENCES

Koch, G.G., Tangen, C.M., Jung, J-W, and Amara, I.A. (1998). Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. *Statistics in Medicine*, 17, 1863-1892.

LaVange, L.M., Durham, T.A., Koch, G.G. (2005). Randomization-based nonparametric methods for analysis of multicentre trials. *Statistical Methods in Medical Research*, 14, 281-301.

Stanish, W.M, Gillings, D.B., and Koch, G.G. (1978). An application of multivariate ratio methods for the analysis of a longitudinal clinical trial with missing data. *Biometrics*, 34, 305-317.

Tangen, C.M. and Koch, G.G. (2001). Non-parametric analysis of covariance for confirmatory randomized clinical trials to evaluate dose-response relationships. *Statistics in Medicine*, 20, 2585-2607.

CHAPTER 4: RANDOMIZATION-BASED ANCOVA FOR HYPOTHESES TESTING IN THE SEQUENTIAL PARALLEL COMPARISON DESIGN (SPCD)

4.1 Introduction

Crossover designs have been utilized in addition to the gold standard randomized parallel-group placebo-controlled trial for the assessment of efficacy for drugs intended for regulatory submission. There are many possible designs of crossover studies, depending on the number of treatments to compare, the number of periods of each treatment, and the aim of the trials (Jones and Kenward, 2014b).

There are several benefits of crossover designs compared to the traditional parallel-group trial. One of them is that the power of the treatment comparison is improved at a fixed sample size, mainly because every subject provides multiple response measurements of the outcome in the multiple periods. Another feature is that its use could reduce the dropout rates, at least for the first period, because the subjects in the placebo group could expect a treatment no worse than the first period if they continue (Koch, Davis and Anderson, 1998a). For example, Pincus et al. (2001) performed a randomized crossover trial of the experimental drug versus active control in ambulatory patients with osteoarthritis of the hip or knee and achieved a low dropout rate for the first period. Of the 227 enrolled patients, 218 (96.0%) patients provided data for the first treatment period and 181 (79.7%) provided data for both treatment periods.

With a two-period design comparing test treatment (T) to placebo treatment (P), four sequence groups P:P, P:T, T:P, and T:T could be of interest. Designs with some combination of those sequence groups have useful features for studies of different patient populations. Also,

other extensions of crossover designs with this structure could be achieved with other added design features such as enrichment. Designs with a crossover structure are in the general class of re-randomization designs, for which subjects could be re-randomized at the start of the second period. Alternatively, rather than re-randomization at the beginning of the second period, randomization before the first period could be performed to produce the multiple randomized sequences.

One of the popular crossover designs with the enriched feature is the randomized withdrawal design with the T:P and T:T sequence groups, and it focuses on only the patients who respond to the drug in the first period and continue to the second period. This design is helpful when there is heterogeneity in the patient population itself to respond to a treatment (Dunger-Baldauf et al., 2006; Dunger-Baldauf, 2007).

The P:P and P:T design, usually known as the placebo lead-in design, has the other two sequence groups in contrast to the randomized withdrawal design. In this design, only the patients who do not respond to placebo in the first period continue to the second period in the study. This design is practical in studies to treat disorders in the central nerve system, where there are many placebo responders (Fava et al., 2003). The placebo response rate in antidepressant and antipsychotic trials has been reported to increase over time in meta analyses of trials between 1985 and 2000 (Khan et al., 2005). With only the placebo nonresponders identified in the first period continuing in the second period to receive either experimental treatment or placebo, the treatment effect is maximized since patients who do not respond to the first period are not expected to become placebo-responders in the second period (Fava et al., 2003).

The design of interest for this paper is the one with three sequence groups, P:P, P:T, and T:T, and it is sometimes known as the sequential parallel comparison design (SPCD) (Fava et al., 2003). Another popular design that has these three sequence groups is the randomized delayed-start design (RDSD), which is useful to evaluate treatments for disease with long term progression by distinguishing the symptomatic improvement from the true disease modifying effect (Dunger-Baldauf et al., 2006). The SPCD, with the P:P, P:T, and T:T sequences, serves similar purposes as the two-sequence placebo lead-in design in psychiatric clinical trials with high placebo response, except that it has an additional T:T sequence, which can be useful for masking the treatment in the first period, reducing the dropout rate during the first period, and enabling a treatment comparison during the first period.

The original paper by Fava et al. (2003) focused on a dichotomized outcome for a study; and other methods have been proposed for the analysis of dichotomous outcomes in the context of SPCD designs by Huang and Tamura (2010, 2011), and Huang, Tamura, and Boos (2011). Recent uses of the SPCD design have extensions to continuous or ordinal outcomes as they arise more naturally than the binary outcomes with dichotomization of a continuous measurement. Huang and Tamura (2010) considered seemingly unrelated regression (SUR) to account for the correlation between subjects in the two periods of the trial. Chen et al (2011) proposed an ordinary least squares approach, and Doros et al. (2013) proposed a repeated measures model that more extensively includes the outcome data collected in the trial.

In this paper, we consider sources of information for the comparison between a test treatment and placebo that are provided in the traditional SPCD design, i.e., the first period treatment difference in the overall population and the second period treatment difference in the

placebo non-responders; and we also evaluate the potential role of other sources of information that are available in this design and that could be of potential interest.

For the analysis of the SPCD design with scope for outcomes in both placebo non-responders and responders, we propose in Section 4.2 a hypothesis testing method based on the randomization distribution of the observed population using the randomization-based mean and variance estimates under the null hypothesis to control Type I error. Further, with this method, adjustment is possible for covariables at baseline for all patients and at the beginning of the second period for patients in the P:P and P:T groups so as to produce variance reduction and to eliminate random imbalances for the covariables. In this regard, we introduce a randomization-based ANCOVA as opposed to a traditional model-based ANCOVA (Koch et al. 1998). In Section 4.3, we report results from simulation studies for the statistical properties of the proposed methods, and we compare their performance with applicable counterparts from the repeated measures model of Doros et al. (2013). A hypothetical study with the three SPCD sequence groups is provided to illustrate the use of the method.

4.2 Methods

For a randomized clinical trial to compare a test treatment T to placebo P during two periods for patients with a chronic (or recurrent) disorder, such as osteoarthritis (or migraine headache), let $i = 1, 2, 3$ index P:P, P:T, T:T as the sequence groups for the two treatments in the two periods; and let $k = 1, 2, \dots, n$ index the population of patients who are eligible for inclusion in the clinical trial and who are randomly assigned to the three sequence groups. In this regard, let U_{ik} denote a random variable which has the value 1 for the assignment of the k-th patient to the i-th group and the value 0 otherwise. The specification for the $\{U_{ik}\}$ has n_i patients randomly

assigned to the i-th group according to simple random sampling without replacement. Thus, the $\{U_{ik}\}$ have the expected values and covariance structure shown in (4.1).

$$
\begin{aligned}
E\{U_{ik}\} &= Pr\{U_{ik}=1\} = (n_i/n) \text{ for all } i,k \\
Var\{U_{ik}\} &= n_i(n-n_i)/n^2 \text{ for all } i,k \\
Cov\{U_{ik},U_{i'k'}\} &= \{[Pr(U_{ik}=U_{i'k'}=1)] - \Pr(U_{ik}=1)\Pr(U_{i'k'}=1)\} \qquad (4.1) \\
&= 0 - \frac{n_i n_{i'}}{n^2} = -\frac{n_i n_{i'}}{n^2} \text{ for } i \neq i' \text{and } k = k' \\
&= \frac{n_i(n_i-1)}{n(n-1)} - \frac{n_i^2}{n^2} = -\frac{n_i(n-n_i)}{n^2(n-1)} \text{ for } i = i' \text{with } k \neq k' \\
&= \frac{n_i n_{i'}}{n(n-1)} - \frac{n_i n_{i'}}{n^2} = \frac{n_i n_{i'}}{n^2(n-1)} \text{ for } i \neq i' \text{and } k \neq k'
\end{aligned}
$$

Let $\boldsymbol{U}_{*k} = (U_{1k}, U_{2k}, U_{3k})'$ and let $\boldsymbol{n} = (n_1, n_2, n_3)'$. From (4.1), it then follows that the $\boldsymbol{U}_{*k}$ have the expected values and covariance structure shown in (4.2) where $\boldsymbol{D_n}$ is a diagonal

$$
\begin{aligned}
\boldsymbol{E}\{\boldsymbol{U}_{*k}\} &= (\boldsymbol{n}/n) \\
\boldsymbol{Var}\{\boldsymbol{U}_{*k}\} &= (n\boldsymbol{D_n} - \boldsymbol{nn}')/n^2 = \boldsymbol{V_U} \text{ for all } k \qquad (4.2) \\
\boldsymbol{Cov}\{\boldsymbol{U}_{*k}, \boldsymbol{U}_{*k'}\} &= -\frac{\boldsymbol{V_U}}{n-1} \text{for all } k \neq k'
\end{aligned}
$$

matrix with diagonal elements $\boldsymbol{n}$.

Let $j = 0, 1, 2$ index the baseline, period 1, and period 2. Let y_{ijk} denote the constant that corresponds to the observed response for the k-th patient during the j-th period according to a non-negative numeric (or ordinal) scale if their random assignment is to the i-th sequence group. Also, $y_{*jk} = \sum_{i=1}^{3} U_{ik}\, y_{ijk}$ represents the random response of the *k*-th patient during the *j*-th period with the U_{ik} being the basis of its randomness. In this regard, $y_{10k} = y_{20k} = y_{30k} = y_{*0k}$ since the baseline response for the k-th patient is the same regardless of their randomly assigned

group; moreover, $y_{11k} = y_{21k}$ since the *k*-th patient receives placebo in period 1 if randomly assigned to either the P:P group or the P:T group. Let $\boldsymbol{y}_{ik} = (y_{i0k}, y_{i1k}, y_{i2k})'$ with no missing data being assumed for all patients. Under the global null hypothesis H_0 that each patient has the same responses for the three periods regardless of their randomly assigned sequence group (and thereby their randomly assigned treatment), the specification $\boldsymbol{y}_{1k} = \boldsymbol{y}_{2k} = \boldsymbol{y}_{3k} = \boldsymbol{y}_{*k}$ applies.

Let z_{ik} be a dichotomous responder variable for period 1 such that $z_{ik} = 1$ if the k-th patient has favorable response during period 1 in the sense that $(y_{i1k} \leq L)$ if their random assignment is to the i-th group with its corresponding treatment for period 1 versus $z_{ik} = 0$ if $(y_{i1k} > L)$; alternatively the $\{z_{ik}\}$ could be based on change (or percent change) from baseline. Let $f_{ijk} = z_{ik}y_{ijk}$ so as to equal y_{ijk} for period $j = 1, 2$ for responders in period 1 and be equal to 0 for non-responders in period 1; and let $g_{ik} = (y_{i2k} - f_{i2k}) = (1 - z_{ik})y_{i2k}$ so as to equal y_{i2k} for period 2 for period 1 non-responders and to equal 0 for period 1 responders. Let $\boldsymbol{F}_{ik} = (y_{i0k}, y_{i1k}, z_{ik}, f_{i1k}, y_{i2k}, f_{i2k})'$ with the assumption of no missing data for its components. Under H_0, $\boldsymbol{F}_{1k} = \boldsymbol{F}_{2k} = \boldsymbol{F}_{3k} = \boldsymbol{F}_{*k}$ applies as a consequence of the $\boldsymbol{y}_{1k} = \boldsymbol{y}_{2k} = \boldsymbol{y}_{3k} = \boldsymbol{y}_{*k}$. Also, $\boldsymbol{F}_{ik}$ could be expanded to include one or more other covariables $\boldsymbol{x}_{*0k}$ at baseline, such as age (in addition to y_{*0k}); but the presentation is more straightforward and sufficient without this extension because the same considerations apply to both y_{*0k} and $\boldsymbol{x}_{*0k}$.

Let $\overline{\boldsymbol{F}}_i = (\sum_{k=1}^{n} U_{ik}\boldsymbol{F}_{ik}/n_i) = (\bar{y}_{i0}, \bar{y}_{i1}, \bar{z}_i, \bar{f}_{i1}, \bar{y}_{i2}, \bar{f}_{i2})'$ denote the vector of the means for the i-th sequence group. From (4.2), it follows that the randomization distributions of the $\overline{\boldsymbol{F}}_i$ have expected values and covariance structure as shown in (4.3).

$$\boldsymbol{E}\{\overline{\boldsymbol{F}}_i\} = \left(\sum_{k=1}^{n} \boldsymbol{F}_{ik} n_i / n_i\, n\right) = \boldsymbol{\mu}_i$$

$$\boldsymbol{Var}\{\overline{\boldsymbol{F}}_i\} = \frac{1}{n_i^2}\left[\sum_{k=1}^{n} \frac{n_i(n-n_i)}{n^2} \boldsymbol{F}_{ik}\boldsymbol{F}'_{ik} - \sum_{k\neq k'} \frac{n_i(n-n_i)}{n^2(n-1)} \boldsymbol{F}_{ik}\boldsymbol{F}'_{ik'}\right]$$

$$= \frac{(n-n_i)}{n_i n(n-1)}\left[\sum_{k=1}^{n} (\boldsymbol{F}_{ik} - \boldsymbol{\mu}_i)\,(\boldsymbol{F}_{ik} - \boldsymbol{\mu}_i)'\right]$$

$$= \frac{(n-n_i)}{n_i n} \boldsymbol{V}_{\boldsymbol{F}_i} = \left(\frac{1}{n_i} - \frac{1}{n}\right) \boldsymbol{V}_{\boldsymbol{F}_i} \qquad (4.3)$$

$$\boldsymbol{Cov}\{\overline{\boldsymbol{F}}_i, \overline{\boldsymbol{F}}_{i'}\} = \frac{1}{n_i n_{i'}}\left[\sum_{k=1}^{n} \frac{-n_i n_{i'}}{n^2} \boldsymbol{F}_{ik}\boldsymbol{F}'_{i'k} + \sum_{k\neq k'} \frac{n_i n_{i'}}{n^2(n-1)} \boldsymbol{F}_{ik}\boldsymbol{F}'_{i'k'}\right]$$

$$= \frac{-1}{n(n-1)}\left[\sum_{k=1}^{n} (\boldsymbol{F}_{ik} - \boldsymbol{\mu}_i)(\boldsymbol{F}_{i'k} - \boldsymbol{\mu}_{i'})\right] = -\frac{1}{n}\boldsymbol{V}_{\boldsymbol{F}_i,\boldsymbol{F}_{i'}}$$

In (4.3), $\boldsymbol{\mu}_i$ is the mean vector for the $\boldsymbol{F}_{ik}$ in the finite population of the n randomized patients and $\boldsymbol{V}_{\boldsymbol{F}_i}$ is the corresponding finite population covariance matrix. Since $\boldsymbol{F}_{1k} = \boldsymbol{F}_{2k} = \boldsymbol{F}_{3k} = \boldsymbol{F}_{*k} = (y_{*0k}, y_{*1k}, z_{*k}, f_{*1k}, y_{*2k}, f_{*2k})'$ under H_0, it follows that (4.3) simplifies to (4.4) under H_0.

$$\boldsymbol{E}\{\overline{\boldsymbol{F}}_i \mid H_0\} = \left(\sum_{k=1}^{n} \boldsymbol{F}_{*k} n_i / n_i n\right) = \boldsymbol{\mu}_{*0} = \left(\mu_0, \mu_{10}, \mu_{z_0}, \mu_{f_{10}}, \mu_{20}, \mu_{f_{20}}\right)',$$

$$\boldsymbol{Var}\{\overline{\boldsymbol{F}}_i \mid H_0\} = \left(\frac{1}{n_i} - \frac{1}{n}\right)\left[\sum_{k=1}^{n} (\boldsymbol{F}_{*k} - \boldsymbol{\mu}_{*0})(\boldsymbol{F}_{*k} - \boldsymbol{\mu}_{*0})'\right] = \left(\frac{1}{n_i} - \frac{1}{n}\right)\boldsymbol{V}_{F,0}, \qquad (4.4)$$

$$\boldsymbol{Cov}\{\overline{\boldsymbol{F}}_i, \overline{\boldsymbol{F}}_{i'} \mid H_0\} = -\frac{1}{n}\boldsymbol{V}_{F,0}.$$

Thus, for $\overline{\boldsymbol{F}} = (\overline{\boldsymbol{F}}'_1, \overline{\boldsymbol{F}}'_2, \overline{\boldsymbol{F}}'_3)'$ it follows that $\boldsymbol{E}\{\overline{\boldsymbol{F}} \mid H_0\} = \mathbf{1}_3 \otimes \boldsymbol{\mu}_{*0}$ where $\mathbf{1}_3$ is the (3 x 1) vector of 1's and $\otimes$ denotes the right Kronecker product for the multiplication of each element of

the vector on the left by the vector on the right; and the covariance structure of $\overline{\boldsymbol{F}}$ under H_0 is as shown in (4.5).

$$\boldsymbol{V}_{\overline{\boldsymbol{F}},0} = \boldsymbol{Var}\{\overline{\boldsymbol{F}} \mid H_0\} = \begin{bmatrix} \left(\frac{1}{n_1} - \frac{1}{n}\right) & -\frac{1}{n} & -\frac{1}{n} \\ -\frac{1}{n} & \left(\frac{1}{n_2} - \frac{1}{n}\right) & -\frac{1}{n} \\ -\frac{1}{n} & -\frac{1}{n} & \left(\frac{1}{n_3} - \frac{1}{n}\right) \end{bmatrix} \otimes \boldsymbol{V}_{\boldsymbol{F},0} \qquad (4.5)$$

$$= [\boldsymbol{D}_{\boldsymbol{n}}^{-1} - (\mathbf{1}_3\mathbf{1}_3'/n)] \otimes \boldsymbol{V}_{\boldsymbol{F},0}.$$

In this regard, $\boldsymbol{E}\{\overline{\boldsymbol{F}} \mid H_0\}$ and $\boldsymbol{Var}\{\overline{\boldsymbol{F}} \mid H_0\}$ pertain to the randomization distribution of $\overline{\boldsymbol{F}}$ under H_0 for re-randomizations of the finite population of the n randomized patients. Since $\boldsymbol{V}_{\overline{\boldsymbol{F}},0}$ is singular through $\sum_{i=1}^{3} n_i\overline{\boldsymbol{F}}_i/n = \boldsymbol{\mu}_{*0}$ (or $(\boldsymbol{n}'\otimes \boldsymbol{I}_6)\,\boldsymbol{V}_{\overline{\boldsymbol{F}}}\,(\boldsymbol{n}\otimes \boldsymbol{I}_6) = \mathbf{0}_{6,6}$ with $\mathbf{0}_{6,6}$ being a matrix of 0's), assessment of H_0 with $\overline{\boldsymbol{F}}$ is through $\boldsymbol{a} = \boldsymbol{A}\overline{\boldsymbol{F}}$ as shown in (4.6), and the

$$\boldsymbol{a} = \boldsymbol{A}\overline{\boldsymbol{F}} = \begin{bmatrix} -\boldsymbol{I}_6 & \boldsymbol{I}_6 & \mathbf{0}_{6,6} \\ -\boldsymbol{I}_6 & \mathbf{0}_{6,6} & \boldsymbol{I}_6 \end{bmatrix} \overline{\boldsymbol{F}} = \begin{bmatrix} (\overline{\boldsymbol{F}}_2 - \overline{\boldsymbol{F}}_1) \\ (\overline{\boldsymbol{F}}_3 - \overline{\boldsymbol{F}}_1) \end{bmatrix} \qquad (4.6).$$

corresponding covariance structure of $\boldsymbol{a}$ under H_0 is shown in (4.7).

$$\boldsymbol{V}_{\boldsymbol{a},0} = Var\{\boldsymbol{a} \mid H_0\} = \boldsymbol{A}\boldsymbol{V}_{\overline{\boldsymbol{F}},\mathbf{0}}\boldsymbol{A}' = \begin{bmatrix} \left(\frac{1}{n_1} + \frac{1}{n_2}\right) & \frac{1}{n_1} \\ \frac{1}{n_1} & \left(\frac{1}{n_1} + \frac{1}{n_3}\right) \end{bmatrix} \otimes \boldsymbol{V}_{\boldsymbol{F},0} \qquad (4.7)$$

For the assessment of H_0 without adjustment for $y_{10k} = y_{20k} = y_{30k}$, $y_{11k} = y_{21k}$, $z_{1k} = z_{2k}$, and $f_{11k} = f_{21k}$, the potential comparisons of interest are shown in (4.8), and they are linear functions $\boldsymbol{c} = \boldsymbol{C}\boldsymbol{a} = \boldsymbol{C}\boldsymbol{A}\overline{\boldsymbol{F}}$ of $\boldsymbol{a}$ with $\boldsymbol{C}$ shown in (4.9); also, $\mu_{z_0} = (\sum_{k=1}^{n} z_{*k}/n)$, $\mu_{f_{20}} = (\sum_{k=1}^{n} f_{*2k}/n)$, and $\mu_{20} = (\sum_{k=1}^{n} y_{*2k}/n)$. The rationale for c_3 is that $\left(\mu_{z_0}\bar{f}_{i2} - \mu_{f_{20}}\bar{z}_i\right)/\mu_{z_0}^2$ for $i = 1, 2$ is the mean of deviations $\left(\mu_{z_0}f_{i2k} - \mu_{f_{20}}z_{ik}\right)/\mu_{z_0}^2$ which equal 0 when $z_{ik} = 0$, and so its behavior is through the $\left(y_{i2k} - \left(\mu_{f_{20}}/\mu_{z_0}\right)\right)/\mu_{z_0}$ when $z_{ik} = 1$; also, $\left(\mu_{z_0}\bar{f}_{i2} - \mu_{f_{20}}\bar{z}_i\right)/\mu_{z_0}^2 =$

$(\bar{z}_i/\mu_{z_0})[(\bar{f}_{i2}/\bar{z}_i) - (\mu_{f_{20}}/\mu_{z_0})]$, and so c_3 addresses the difference between the means of period 2 responders and their population counterparts $(\mu_{f_{20}}/\mu_{z_0})$ under H_0 in a rescaled linearized way. Similar considerations pertain to c_4.

$$
\begin{aligned}
c_1 &= \left\{\bar{y}_{31} - \frac{n_1\bar{y}_{11} + n_2\bar{y}_{21}}{(n_1+n_2)}\right\} \\
c_2 &= (\bar{y}_{32} - \bar{y}_{12}) \qquad (4.8), \\
c_3 &= \left\{(\bar{f}_{22} - \bar{f}_{12})/\mu_{z_0}\right\} - \left\{\mu_{f_{20}}(\bar{z}_2 - \bar{z}_1)/\mu_{z_0}^2\right\} \\
c_4 &= \left\{(\bar{y}_{22} - \bar{f}_{22} - \bar{y}_{12} + \bar{f}_{12})/(1-\mu_{z_0})\right\} - \left\{(\mu_{20} - \mu_{f_{20}})(\bar{z}_1 - \bar{z}_2)/(1-\mu_{z_0})^2\right\}
\end{aligned}
$$

$$
C = \begin{bmatrix}
0 & -\frac{n_2}{(n_1+n_2)} & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & -\frac{\mu_{f_{20}}}{\mu_{z_0}^2} & 0 & 0 & \frac{1}{\mu_{z_0}} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{(\mu_{20}-\mu_{f_{20}})}{(1-\mu_{z_0})^2} & 0 & \frac{1}{(1-\mu_{z_0})} & -\frac{1}{(1-\mu_{z_0})} & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix} \qquad (4.9)
$$

Under H_0, $\boldsymbol{E}\{\boldsymbol{c} \mid H_0\} = \mathbf{0}_4$; and the covariance structure for $\boldsymbol{c}$ under H_0 is $\boldsymbol{V}_{\boldsymbol{c},0} = \boldsymbol{C}\boldsymbol{V}_{\boldsymbol{a},0}\boldsymbol{C}'$. For $\boldsymbol{c} = (c_1, c_2, c_3, c_4)'$, the comparison c_1 pertains to T versus P in period 1; the comparison c_2 pertains to T:T versus P:P in period 2; the comparison c_3 pertains to P:T versus P:P in period 2 for responders to placebo in period 1; c_4 pertains to P:T versus P:P in period 2 for non-responders to placebo in period 1; and both c_3 and c_4 have linearized adjustments for the corresponding influences of differences in responder proportions for the P:P and P:T groups. Univariate test statistics for H_0, and thereby for the comparison between T and P, can be based on weighted linear combinations $c_w = \sum_{h=1}^{4} w_h c_h$ where $\boldsymbol{w} = (w_1, w_2, w_3, w_4)'$ is a vector of weights such that all $w_h \geq 0$ and $\sum_{h=1}^{4} w_h = 1$. In this regard, all $w_h = 0.25$ corresponds to equal weights and $\boldsymbol{w}_{inv} = \left(\mathbf{1}'\boldsymbol{V}_{\boldsymbol{c},0}^{-1}\mathbf{1}\right)^{-1}\mathbf{1}'\boldsymbol{V}_{\boldsymbol{c},0}^{-1}$ corresponds to inverse covariance matrix weights, with the latter being optimal in the sense of minimum variance and potentially statistical power

under the alternatives to H_0 whereby the c_h express similar non-null differences between T and P. With the weights $\boldsymbol{w}$, the test statistic for the global null hypothesis H_0 at the one-sided significance level α is $T_{w,\boldsymbol{c}} = \boldsymbol{w}'\boldsymbol{c}/\left(\boldsymbol{w}'\boldsymbol{V}_{\boldsymbol{c},0}\boldsymbol{w}\right)^{0.5}$. Under H_0, $T_{w,\boldsymbol{c}}$ approximately has a standard normal distribution with mean 0 and variance 1. If the one-sided p-value for $T_{w,\boldsymbol{c}}$ significantly contradicts H_0 in the sense that $\mathrm{p} < \alpha$ for some specified type 1 error level, such as $\alpha = 0.025$, then the closed testing methods of Lehmacher et al. (1991) can be used to test H_0 for subsets of the c_h through counterparts of $T_{w,\boldsymbol{c}}$ for weighted averages of subsets of the c_h. In this regard, if $\mathrm{p} < \alpha$ for such test statistics for all subsets that include c_h, then statistical significance applies to c_h in its own right for contradicting H_0 with strong control of type 1 error for the corresponding scope of multiple comparisons. Also, in some situations, a subset of the c_h is of more interest for the assessment of H_0 than all components of $\boldsymbol{c}$, and so it can be assessed directly in its own right with particular cases of interest being c_1 and c_4 for the SPCD design, and perhaps c_1, c_2, and c_4 for other versions of the crossover design with P:P, P:T, and T:T sequence groups.

The constraints $\boldsymbol{c}_0 = \boldsymbol{C}_0\boldsymbol{a} = \boldsymbol{C}_0\boldsymbol{A}\overline{\boldsymbol{F}}$ for $\boldsymbol{a}$ with null expected values regardless of whether H_0 applies are shown in (4.10) with $\boldsymbol{C}_0$ in (4.11).

$$\boldsymbol{c}_0 = \begin{bmatrix} (\bar{y}_{20} - \bar{y}_{10}) \\ (\bar{y}_{30} - \bar{y}_{10}) \\ (\bar{y}_{21} - \bar{y}_{11}) \\ (\bar{z}_2 - \bar{z}_1) \\ \left\{\dfrac{\left(\bar{f}_{21} - \bar{f}_{11}\right)}{\mu_{z_0}}\right\} - \left\{\dfrac{\mu_{f_{10}}(\bar{z}_2 - \bar{z}_1)}{\mu_{z_0}^2}\right\} \end{bmatrix} \qquad (4.10)$$

$$\boldsymbol{C}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\dfrac{\mu_{f_{10}}}{\mu_{z_0}^2} & \dfrac{1}{\mu_{z_0}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad (4.11)$$

Let $\boldsymbol{c}_+ = [\boldsymbol{c}', \boldsymbol{c}_0']' = [\boldsymbol{C}', \boldsymbol{C}_0']'\boldsymbol{a} = \boldsymbol{C}_+\boldsymbol{a}$ denote the combined set of comparisons $\boldsymbol{c}$ pertaining to H_0 and constraints $\boldsymbol{c}_0$. Under H_0, the covariance structure for $\boldsymbol{c}_+$ is $\boldsymbol{V}_{\boldsymbol{c}_+,0} = \boldsymbol{C}_+\boldsymbol{V}_{\boldsymbol{a},0}\boldsymbol{C}_+'$. Since $\boldsymbol{E}\{\boldsymbol{c}_0\} = \mathbf{0}_5$ regardless of whether H_0 applies, randomization-based covariance adjustment for $\boldsymbol{c}$ with respect to the constraints $\boldsymbol{c}_0$ is invoked by fitting the linear model shown in (4.12) by weighted least squares with weights based on $\boldsymbol{V}_{\boldsymbol{c}_+,0}^{-1}$, and with "$\triangleq$" meaning

$$\boldsymbol{E}\{\boldsymbol{c}_+\} \triangleq \begin{bmatrix} \boldsymbol{I}_4 \\ \mathbf{0}_{5,4} \end{bmatrix} \boldsymbol{b}_0 = \boldsymbol{X}\boldsymbol{b}_0 \qquad (4.12)$$

"is estimated by." To account for other covariables $\boldsymbol{x}_{*0k}$ at baseline, $\boldsymbol{c}_0$ in (4.10) is expanded to include $(\overline{\boldsymbol{x}}_{20} - \overline{\boldsymbol{x}}_{10})$ and $(\overline{\boldsymbol{x}}_{30} - \overline{\boldsymbol{x}}_{10})$ where $\overline{\boldsymbol{x}}_{i0} = \frac{1}{n_i}\sum_{k=1}^{n} U_{ik}\boldsymbol{x}_{*0k}$.

For the model in (4.12), $\boldsymbol{b}_0 = (b_{01}, b_{02}, b_{03}, b_{04})'$ are covariance adjusted counterparts of $\boldsymbol{c}$, and $\boldsymbol{b_0} = \left(\boldsymbol{X}'\boldsymbol{V}_{\boldsymbol{c}_+,0}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{V}_{\boldsymbol{c}_+,0}^{-1}\boldsymbol{c}_+$; also, the covariance matrix for $\boldsymbol{b_0}$ is $\boldsymbol{V}_{\mathrm{b}_0} = \left(\boldsymbol{X}'\boldsymbol{V}_{\boldsymbol{c}_+,0}^{-1}\boldsymbol{X}\right)^{-1}$. As a consequence of the structure of $\boldsymbol{X}$ in (4.12), $\boldsymbol{b_0}$ and $\boldsymbol{V}_{\boldsymbol{b}_0}$ can be expressed as shown in (4.13)

$$\begin{aligned} \boldsymbol{b_0} &= \boldsymbol{c} - \left(\boldsymbol{C}\boldsymbol{V}_{a,0}\boldsymbol{C}_0'\right)\left(\boldsymbol{C}_0\boldsymbol{V}_{a,0}\boldsymbol{C}_0'\right)^{-1}\boldsymbol{c}_0 \\ \boldsymbol{V}_{\boldsymbol{b_0}} &= \boldsymbol{V}_{\boldsymbol{c},0} - \left(\boldsymbol{C}\boldsymbol{V}_{a,0}\boldsymbol{C}_0'\right)\left(\boldsymbol{C}_0\boldsymbol{V}_{a,0}\boldsymbol{C}_0'\right)^{-1}\left(\boldsymbol{C}_0\boldsymbol{V}_{a,0}\boldsymbol{C}'\right) \qquad (4.13) \\ &= \boldsymbol{C}\left[\boldsymbol{V}_{a,0} - \boldsymbol{V}_{a,0}\boldsymbol{C}_0'\left(\boldsymbol{C}_0\boldsymbol{V}_{a,0}\boldsymbol{C}_0'\right)^{-1}\boldsymbol{C}_0\boldsymbol{V}_{a,0}\right]\boldsymbol{C}' \end{aligned}$$

so as to show the nature of randomization-based covariance adjustment. Covariance adjusted test statistics for H_0 can be based on weighted linear combinations $\boldsymbol{b}_{\boldsymbol{w},\mathbf{0}} = \sum_{h=1}^{4} w_h\boldsymbol{b}_{\boldsymbol{h},\mathbf{0}}$ in ways similar to those discussed for $\boldsymbol{c}$. In this regard, $T_{w,\boldsymbol{b_0}} = \boldsymbol{w}'\boldsymbol{b_0}/\left(\boldsymbol{w}'\boldsymbol{V}_{\boldsymbol{b_0}}\boldsymbol{w}\right)^{0.5}$ approximately has the normal distribution with mean 0 and variance 1 under H_0. Finally, the unadjusted test statistics based on $\boldsymbol{c}$ and their randomization-based covariance adjusted counterparts with respect to $\mathbf{b_0}$ are applicable to transformations of the elements of the $\boldsymbol{y}_{*k}$ that apply under H_0 such as

ranks across subjects in the pooled groups or to dichotomous indicators with the value 1 if some criterion, such as $\left(y_{*jk} < L^* < L\right)$, is satisfied for $j = 1, 2$ and the value 0 if otherwise.

4.3 Simulation Study

4.3.1 Simulation Setup

The Type I errors and powers of the methods discussed in this chapter were evaluated with simulation studies. The responses Y_{ij} of patients in the *i*-th group for the *j*-th period, where $i =$ P:P, P:T, T:T were randomly generated in the manner shown in (4.14) for which Z_{i1} is an

$$\begin{bmatrix} Y_{i0} \\ Y_{i1} \\ Y_{i2} \end{bmatrix} = \begin{bmatrix} e_{i0} \\ e_{i1} \\ e_{i2} \end{bmatrix} + \begin{bmatrix} \xi_0 \\ \xi_{i1} \\ Z_{i1}\xi_{i2} + (1 - Z_{i1})\xi_{i3} \end{bmatrix} \qquad (4.14)$$

indicator for a period 1 responder in the *i*-th group with $Z_{i1} = 1$ if $Y_{i1} \leq L$, $Z_{i1} = 0$ if $Y_{i1} > L$ for L as the specified criterion for a responder or not. In (4.14), the $\boldsymbol{e}_i = (e_{i0}, e_{i1}, e_{i2})$ are independently generated, random errors from the trivariate normal distribution with $\mathbf{0}_3$ as the common mean for all three groups, and with $\wedge$ in (4.15) as the common covariance matrix for all

$$\wedge = \sigma^2 \begin{bmatrix} 1 & \rho_{01} & \rho_{02} \\ \rho_{01} & 1 & \rho_{12} \\ \rho_{02} & \rho_{12} & 1 \end{bmatrix} \qquad (4.15)$$

three groups where σ^2 is the common variance for all three periods and $\rho_{jj'}$ is the correlation for periods j and j'. Additionally, ξ_0 is the common mean at baseline for all three groups; ξ_{i1} is the mean for period 1 for the *i*-th group; ξ_{i2} is an *i*-th group shift parameter that applies to period 2 for period 1 responders; and ξ_{i3} is an *i*-th group shift parameter that applies to period 2 for period 1 non-responders.

For all simulation studies, the specified covariance matrix $\wedge$ had $\sigma^2 = 36$, with the scope of correlations being $\rho_{12} = \rho_{13} = \rho_{23} = 0.3, 0.5$ as exchangeable structures and $\rho_{12} = \rho_{23} = \rho_{13}^{0.5} = 0.5, 0.7$ as autoregressive structures. The specifications for the assessments of type 1 error

were $\xi_0 = 40$, $\xi_{\text{P:P},1} = \xi_{\text{P:T},1} = \xi_{\text{T:T},1} = 35$, $\xi_{\text{P:P},2} = \xi_{\text{P:T},2} = \xi_{\text{T:T},2} = 32$, and $\xi_{\text{P:P},3} = \xi_{\text{P:T},3} = \xi_{\text{T:T},3} = 35$, with $Y_{i1} \leq L = 33$ being the criterion for a period 1 responder in the *i*-th group. Also, with $\pi_i = E\{Z_{i1}\}$ as the probability of responder status for group *i*, it follows from $\xi_{\text{P:P},1} = \xi_{\text{P:T},1} = \xi_{\text{T:T},1} = 35$ that $\pi_{\text{P:P}} = \pi_{\text{P:T}} = \pi_{\text{T:T}} = 0.37$ for the assessments of type 1 error under the global null hypothesis H_0 as specified in Section 4.2.

The specifications of the ξ_{i1}, ξ_{i2}, and ξ_{i3} for the assessments of statistical power are shown in Table 4.1. Also, $\xi_0 = 40$ and $\pi_{\text{P:P}} = \pi_{\text{P:T}} = 0.37$, but $\pi_{\text{T:T}} = 0.47$ so that the difference in period 1 responder rates for the T:T group versus the P:P and P:T groups is about 0.10. From Table 4.1, it follows that $\Delta_1 = \left(\xi_{\text{T:T},1} - \xi_{\text{P:P},1}\right) = -1.5$, $\Delta_3 = \left(\xi_{\text{P:T},2} - \xi_{\text{P:P},2}\right) = -1.0$, and $\Delta_4 = \left(\xi_{\text{P:T},3} - \xi_{\text{P:P},3}\right) = -2.0$. Since $E\{Y_{i2}\} = \{\pi_i \xi_{i2} + (1 - \pi_i)\xi_{i3}\} = \eta_i$, it follows that $\Delta_2 = (\eta_{\text{T:T}} - \eta_{\text{P:P}}) = (32.1 - 33.9) = -1.8$. Thus, $(\Delta_1, \Delta_2, \Delta_3, \Delta_4) = (-1.5, -1.8, -1, -2)$ is the specification that corresponds to the assessments of statistical power with respect to (c_1, c_2, c_3, c_4) and $(b_{10}, b_{20}, b_{30}, b_{40})$; and $(\Delta_1, \Delta_2, \Delta_3, \Delta_4) = (0, 0, 0, 0)$ corresponds to the assessments of type 1 error.

Table 4.1 Specifications for Assessments of Statistical Power

Group	Period 1	Period 2 for Period 1 Responder	Period 2 for Period 1 Non-Responder
P:P	$\xi_{\text{P:P},1} = 35$	$\xi_{\text{P:P},2} = 32$	$\xi_{\text{P:P},3} = 35$
P:T	$\xi_{\text{P:T},1} = 35$	$\xi_{\text{P:T},2} = 31$	$\xi_{\text{P:T},3} = 33$
T:T	$\xi_{\text{T:T},1} = 33.5$	$\xi_{\text{T:T},2} = 30.5$	$\xi_{\text{T:T},3} = 33.5$

The simulations were performed with equal sample sizes $n_* = 40, 80, 160$ patients per group, with the total sample sizes respectively being $3n_* = 120, 240, 480$. For each simulation, the responses of the baseline, period 1, and period 2 are generated via (4.14) with the previously

noted specifications. For testing the hypothesis $H_0: \Delta_w = 0 \; versus \; H_A: \Delta_w \neq 0$ in correspondence to c_w or $b_{0,w}$, the determination of p value for the test statistic $T_{w,c}$ or T_{wb_0} is based on reference of its squared value to the chi-square value $\chi^2_{1,1-\alpha}$ where α is the specified significance level and we chose α to be 0.05. The simulation results are based on 50,000 replicates for the specifications previously stated. The results of Type I error and power from the simulations are means of indicator variables for whether $p \leq \alpha$ applies for testing H_0, and the empirical standard deviation (ESD) is taken as the square root of the variance of the estimated c_w or $b_{w,0}$ across all simulations, and the average of the estimated standard error (ASE) of c_w or $b_{w,0}$ in the simulations is provided.

4.3.2 Simulation Results

The results that only address Δ_1 and Δ_4, as is usually the case of interest in the SPCD design under the null hypothesis, are displayed in Table 4.2. As shown there, under the null hypothesis, the unadjusted and adjusted methods provide unbiased point estimates and reasonable estimates of the standard errors of the estimators, as the ASE and ESD are similar. Type I errors under all scenarios are well-controlled at the nominal 0.05 level regardless of the sample sizes using the unadjusted $\boldsymbol{c}$ or covariable adjusted $\boldsymbol{b_0}$. The MMRM approach of Doros et al. (2013) also had its type 1 error evaluated with the simulation studies, and it provided good control when the sample size per group is 80 and 160; but it had an elevated type one error when the sample size is 40 per group, under each of the different correlation specifications. Additionally, the covariable adjusted method $\boldsymbol{b_0}$ provides variance reduction of 10% to 45%, depending on the correlation levels and sample sizes, as shown in the efficiency column.

Under H_0 for no treatment differences, results that additionally address Δ_2, the treatment difference between P:P and T:T sequences, in addition to the usual Δ_1 and Δ_4 for the SPCD

design, are shown in Table 4.3. These results also show good control of the Type I error. For utilizing all available information from a study, a weighted statistic can address all Δ_1 to Δ_4, with equal weighting or inverse variance weighting, and Table 4.4 shows that it controls the type I error well at the 0.05 significance level.

Under the alternative parameter specification in Table 4.1, the powers with the three methods are presented in Table 4.5. The covariable adjusted method $\boldsymbol{b_0}$ and MMRM approach of Doros et al. (2013) for addressing Δ_1 and Δ_4 have similar power, with both having higher power than the unadjusted method $\boldsymbol{c}$. Under the specification of Δ_2 for Table 4.1, when taking Δ_2 into account in addition to Δ_1 and Δ_4, the powers are slightly better than without addressing Δ_2. When considering all the available sources of information with equal weights, the power decreases due to smaller Δ_3 compared to Δ_1, Δ_2, Δ_4. But when inverse variance weighting is used, the power is similar to that when Δ_3 is not addressed, mainly because the weight corresponding to Δ_3 is smaller than those corresponding to Δ_1, Δ_2, Δ_4.

Table 4.2 Results from 50, 000 replicate simulations for the test statistics of $H_0: \Delta_1 = \Delta_4 = 0$ under $\Delta_1 = \Delta_2 = \Delta_3 = \Delta_4 = 0$

ρ	n_i	Weight	Unadjusted				Adjusted				MMRM				Efficiency	
			Bias	ASE	ESD	Type I error	Bias	ASE	ESD	Type I error	Bias	ASE	ESD	Type I error	a/u	a/m
0.3 (EX)	40	Equal	-0.0031	1.005	1.009	0.0491	-0.0013	0.956	0.956	0.0490	-0.0014	0.973	0.977	0.0520	0.90	0.96
		InvVar	-0.0030	0.944	0.950	0.0499	-0.0009	0.898	0.899	0.0490	0.0085	0.908	0.919	0.0526	0.90	0.96
	80	Equal	-0.0045	0.712	0.716	0.0500	-0.0042	0.680	0.683	0.0505	-0.0042	0.684	0.688	0.0513	0.91	0.98
		InvVar	-0.0029	0.670	0.671	0.0500	-0.0029	0.639	0.640	0.0492	0.0020	0.641	0.645	0.0515	0.91	0.99
	160	Equal	0.0037	0.504	0.503	0.0488	0.0025	0.482	0.482	0.0497	0.0027	0.483	0.482	0.0494	0.92	1.00
		InvVar	0.0030	0.474	0.473	0.0494	0.0019	0.454	0.453	0.0500	0.0042	0.453	0.454	0.0506	0.92	1.00
0.5 (AR)	40	Equal	-0.0027	0.974	0.980	0.0493	-0.0038	0.878	0.883	0.0500	-0.0037	0.912	0.928	0.0542	0.81	0.91
		InvVar	-0.0038	0.929	0.937	0.0495	-0.0052	0.820	0.824	0.0506	0.0021	0.838	0.854	0.0552	0.77	0.93
	80	Equal	-0.0026	0.689	0.693	0.0499	-0.0031	0.624	0.627	0.0516	-0.0021	0.642	0.648	0.0540	0.82	0.93
		InvVar	-0.0015	0.659	0.661	0.0504	-0.0029	0.583	0.583	0.0509	0.0016	0.592	0.597	0.0528	0.78	0.96
	160	Equal	0.0029	0.488	0.486	0.0488	0.0032	0.443	0.442	0.0497	0.0026	0.453	0.454	0.0509	0.83	0.95
		InvVar	0.0029	0.466	0.465	0.0495	0.0037	0.413	0.413	0.0499	0.0052	0.419	0.420	0.0519	0.79	0.97
0.5 (EX)	40	Equal	-0.0032	0.974	0.980	0.0519	-0.0014	0.857	0.864	0.0514	-0.0022	0.865	0.873	0.0538	0.78	0.98
		InvVar	-0.0026	0.929	0.936	0.0510	-0.0012	0.810	0.816	0.0515	0.0061	0.816	0.829	0.0561	0.76	0.97
	80	Equal	0.0011	0.689	0.692	0.0504	0.0008	0.609	0.613	0.0508	0.0016	0.609	0.612	0.0509	0.78	1.00
		InvVar	0.0019	0.659	0.660	0.0504	0.0006	0.576	0.578	0.0502	0.0050	0.576	0.580	0.0519	0.77	0.99
	160	Equal	0.0010	0.488	0.484	0.0476	-0.0002	0.432	0.430	0.0486	0.0001	0.429	0.427	0.0489	0.79	1.02
		InvVar	0.0006	0.466	0.464	0.0474	0.0000	0.409	0.407	0.0488	0.0020	0.407	0.406	0.0496	0.77	1.00
0.7 (AR)	40	Equal	0.0003	0.923	0.928	0.0512	0.0011	0.724	0.725	0.0492	0.0001	0.752	0.758	0.0533	0.61	0.91
		InvVar	-0.0011	0.909	0.917	0.0523	0.0005	0.678	0.679	0.0502	0.0052	0.690	0.700	0.0536	0.55	0.94
	80	Equal	-0.0017	0.654	0.655	0.0486	-0.0016	0.515	0.516	0.0501	-0.0010	0.529	0.533	0.0519	0.62	0.94
		InvVar	-0.0003	0.645	0.644	0.0483	-0.0009	0.482	0.481	0.0491	0.0018	0.488	0.490	0.0508	0.56	0.96
	160	Equal	0.0046	0.462	0.460	0.0483	0.0025	0.365	0.364	0.0501	-0.0009	0.373	0.376	0.0518	0.63	0.94
		InvVar	0.0037	0.457	0.454	0.0486	0.0018	0.342	0.341	0.0498	0.0005	0.345	0.346	0.0516	0.56	0.97

Note: Bias = mean of (estimate-true value); Type I error = rejection rate of null hypothesis (when $Z_w^2 > \chi_{0.95}^2$); ASE=average standard error; ESD=empirical standard deviation. a/u=(adjusted variance) / (unadjusted variance); a/m = (adjusted variance) / (MMRM variance)

Table 4.3 Results from 50, 000 simulations for the test statistics of $H_0: \Delta_1 = \Delta_3 = \Delta_4 = 0$ under $\Delta_1 = \Delta_2 = \Delta_3 = \Delta_4 = 0$

ρ	n_i	Weight	Unadjusted				Adjusted				Efficiency
			Bias	ASE	ESD	Type I error	Bias	ASE	ESD	Type I error	a/u
0.3 (EX)	40	Equal	-0.0013	1.009	1.015	0.0504	0.0009	0.940	0.942	0.0494	0.86
		InvVar	-0.0030	0.931	0.937	0.0503	-0.0006	0.875	0.877	0.0503	0.88
	80	Equal	-0.0016	0.714	0.716	0.0506	-0.0014	0.668	0.670	0.0496	0.87
		InvVar	-0.0019	0.661	0.663	0.0500	-0.0014	0.623	0.624	0.0491	0.89
	160	Equal	0.0022	0.505	0.505	0.0496	0.0009	0.474	0.474	0.0492	0.88
		InvVar	0.0024	0.468	0.467	0.0502	0.0009	0.443	0.443	0.0496	0.90
0.5 (AR)	40	Equal	-0.0053	1.023	1.030	0.0506	-0.0061	0.919	0.922	0.0497	0.80
		InvVar	-0.0035	0.926	0.933	0.0500	-0.0049	0.816	0.821	0.0500	0.77
	80	Equal	-0.0013	0.724	0.725	0.0504	-0.0024	0.653	0.654	0.0490	0.81
		InvVar	-0.0016	0.657	0.660	0.0504	-0.0030	0.582	0.583	0.0510	0.78
	160	Equal	0.0034	0.512	0.510	0.0497	0.0037	0.463	0.462	0.0493	0.82
		InvVar	0.0029	0.466	0.465	0.0495	0.0037	0.413	0.413	0.0496	0.79
0.5 (EX)	40	Equal	-0.0051	1.023	1.031	0.0512	-0.0039	0.859	0.865	0.0513	0.70
		InvVar	-0.0030	0.926	0.933	0.0513	-0.0029	0.799	0.805	0.0518	0.74
	80	Equal	0.0010	0.724	0.726	0.0500	-0.0001	0.610	0.613	0.0500	0.71
		InvVar	0.0018	0.657	0.659	0.0496	0.0002	0.569	0.571	0.0508	0.75
	160	Equal	0.0023	0.512	0.508	0.0477	0.0014	0.433	0.430	0.0470	0.71
		InvVar	0.0006	0.466	0.463	0.0476	0.0006	0.404	0.402	0.0489	0.75
0.7 (AR)	40	Equal	0.0007	1.026	1.033	0.0505	0.0028	0.794	0.795	0.0496	0.59
		InvVar	-0.0014	0.884	0.890	0.0515	-0.0005	0.665	0.667	0.0496	0.56
	80	Equal	0.0002	0.726	0.726	0.0495	-0.0001	0.564	0.563	0.0488	0.60
		InvVar	-0.0016	0.628	0.628	0.0492	-0.0015	0.474	0.474	0.0500	0.57
	160	Equal	0.0030	0.514	0.512	0.0494	0.0009	0.400	0.400	0.0500	0.61
		InvVar	0.0046	0.445	0.442	0.0489	0.0024	0.337	0.335	0.0491	0.57

Note: Bias = mean of (estimate-true value); Type I error = rejection rate of null hypothesis (when $Z_w^2 > \chi_{0.95}^2$); ASE=average standard error; ESD=empirical standard deviation. a/u=(adjusted variance) / (unadjusted variance).

Table 4.4 Results from 50, 000 replicate simulations for the test statistics of $H_0: \Delta_1 = \Delta_3 = \Delta_3 = \Delta_4 = 0$ under $\Delta_1 = \Delta_2 = \Delta_3 = \Delta_4 = 0$

ρ	n_i	Weight	Unadjusted				Adjusted				Efficiency
			Bias	ASE	ESD	Type I	Bias	ASE	ESD	Type I	a/u
0.3 (EX)	40	Equal	-0.0006	0.980	0.985	0.0505	0.0004	0.917	0.920	0.0496	0.87
		InvVar	-0.0026	0.857	0.862	0.0511	-0.0012	0.809	0.811	0.0504	0.89
	80	Equal	-0.0018	0.694	0.694	0.0477	-0.0007	0.653	0.653	0.0491	0.88
		InvVar	-0.0026	0.610	0.610	0.0500	-0.0017	0.578	0.578	0.0493	0.90
	160	Equal	0.0022	0.491	0.490	0.0489	0.0013	0.464	0.464	0.0488	0.89
		InvVar	0.0028	0.432	0.431	0.0499	0.0017	0.411	0.411	0.0497	0.91
0.5 (AR)	40	Equal	-0.0084	0.965	0.968	0.0496	-0.0100	0.881	0.883	0.0505	0.83
		InvVar	-0.0050	0.827	0.833	0.0514	-0.0067	0.738	0.743	0.0517	0.79
	80	Equal	-0.0014	0.683	0.682	0.0491	-0.0014	0.627	0.627	0.0480	0.84
		InvVar	-0.0021	0.588	0.590	0.0509	-0.0025	0.527	0.527	0.0501	0.80
	160	Equal	0.0030	0.483	0.481	0.0482	0.0037	0.445	0.443	0.0489	0.85
		InvVar	0.0025	0.417	0.416	0.0500	0.0036	0.375	0.374	0.0493	0.81
0.5 (EX)	40	Equal	-0.0060	0.964	0.970	0.0501	-0.0050	0.824	0.829	0.0499	0.73
		InvVar	-0.0036	0.827	0.831	0.0505	-0.0033	0.725	0.729	0.0509	0.77
	80	Equal	0.0005	0.683	0.685	0.0509	0.0002	0.587	0.590	0.0501	0.74
		InvVar	0.0014	0.588	0.592	0.0508	0.0006	0.518	0.522	0.0511	0.78
	160	Equal	0.0021	0.483	0.481	0.0492	0.0013	0.416	0.415	0.0494	0.74
		InvVar	0.0004	0.417	0.416	0.0510	0.0001	0.369	0.367	0.0489	0.78
0.7 (AR)	40	Equal	0.0023	0.929	0.935	0.0498	0.0027	0.748	0.750	0.0491	0.64
		InvVar	0.0006	0.761	0.764	0.0513	-0.0002	0.591	0.592	0.0508	0.60
	80	Equal	-0.0020	0.658	0.658	0.0491	-0.0015	0.532	0.532	0.0498	0.65
		InvVar	-0.0035	0.541	0.541	0.0500	-0.0026	0.422	0.422	0.0489	0.61
	160	Equal	0.0018	0.466	0.465	0.0495	0.0006	0.378	0.378	0.0493	0.66
		InvVar	0.0036	0.384	0.383	0.0503	0.0022	0.300	0.300	0.0497	0.61

Note: Bias = mean of (estimate-true value); Type I error = rejection rate of null hypothesis (when $Z_w^2 > \chi_{0.95}^2$); ASE=average standard error; ESD=empirical standard deviation. a/u=(adjusted variance) / (unadjusted variance).

Table 4.5 Results from 50,000 replicate simulations for power of test statistics under the alternative specified in Table 4.1

ρ	n_i	Weight	Δ_1 and Δ_4			Δ_1, Δ_2, and Δ_4		Δ_1, Δ_2, Δ_3, and Δ_4	
			Unadjusted	Adjusted	MMRM	Unadjusted	Adjusted	Unadjusted	Adjusted
0.3 (EX)	40	Equal	0.418	0.438	0.443	0.426	0.460	0.368	0.396
		InvVar	0.426	0.448	0.451	0.434	0.464	0.427	0.459
	80	Equal	0.688	0.721	0.723	0.698	0.745	0.631	0.673
		InvVar	0.699	0.736	0.733	0.712	0.757	0.712	0.755
	160	Equal	0.931	0.948	0.950	0.936	0.959	0.900	0.927
		InvVar	0.938	0.955	0.954	0.943	0.963	0.945	0.964
0.5 (AR)	40	Equal	0.441	0.501	0.491	0.418	0.477	0.379	0.424
		InvVar	0.435	0.510	0.506	0.435	0.506	0.436	0.511
	80	Equal	0.716	0.791	0.775	0.689	0.765	0.644	0.708
		InvVar	0.711	0.801	0.793	0.714	0.801	0.723	0.807
	160	Equal	0.946	0.974	0.969	0.932	0.966	0.909	0.944
		InvVar	0.943	0.977	0.976	0.946	0.977	0.951	0.980
0.5 (EX)	40	Equal	0.443	0.519	0.529	0.416	0.527	0.376	0.472
		InvVar	0.434	0.525	0.535	0.436	0.535	0.437	0.539
	80	Equal	0.716	0.805	0.815	0.686	0.815	0.642	0.763
		InvVar	0.711	0.816	0.820	0.714	0.825	0.722	0.829
	160	Equal	0.948	0.980	0.982	0.933	0.983	0.908	0.968
		InvVar	0.945	0.982	0.983	0.947	0.984	0.952	0.986
0.7 (AR)	40	Equal	0.478	0.655	0.643	0.414	0.588	0.398	0.541
		InvVar	0.445	0.656	0.663	0.462	0.662	0.478	0.680
	80	Equal	0.763	0.917	0.907	0.686	0.873	0.676	0.839
		InvVar	0.727	0.920	0.919	0.750	0.923	0.778	0.935
	160	Equal	0.964	0.997	0.997	0.932	0.992	0.926	0.987
		InvVar	0.950	0.997	0.997	0.960	0.997	0.970	0.998

Note: Power = rejection rate of null hypothesis (when $Z_w^2 > \chi_{0.95}^2$).

4.4 Example

A hypothetical placebo-controlled pre-randomized two-period study that makes use of the P:P, P:T, and T:T sequence groups is created to illustrate an application of the proposed methods and to compare them to those from the MMRM method proposed by Doros et al. (2013). The example has 240 subjects which are equally assigned to the three sequence groups. Assessments of responses occur at the baseline, end of period 1 and period 2, with a lower outcome score being more beneficial. At period 1, a score less than 33 is considered as a responder to the treatment. The means and standard deviations of the responses are provided in Table 4.6. The means of responses at baseline are similar across the three sequence groups, with a slightly larger mean in the T:T group. There are 31%, 34%, and 39% of responders in the three groups respectively, as shown in the $\bar{Z}$ row, with only 6% more responders in the test treatment group than the placebo group at Period 1. And as the trial continues to the second period, the mean responses there for the P:T and T:T groups are almost the same, and the P:P group is slightly worse.

Table 4.6 Mean and Standard Deviation of Outcome

Statistics	P:P	P:T	T:T
$\bar{Y}_0$	40.54	40.45	40.80
SD	5.36	5.57	6.91
$\bar{Y}_1$	35.15	34.92	34.24
SD	6.17	5.72	6.48
$\bar{z}$	0.31	0.34	0.39
$\bar{Y}_2$	33.98	32.44	32.39
SD	5.69	5.04	6.13

The estimates using the proposed unadjusted and adjusted methods as well as the MMRM method are shown in Table 4.7. The standard errors of the estimates increase as the sample sizes decrease, as those pertaining to Δ_1 have the smallest standard errors and those

pertaining to Δ_3 have the largest standard errors (since only 30% to 40% of patients are responders in period 1). The covariable-adjusted method and the MMRM method provide similar estimates for Δ_1, the treatment difference at the end of the first period with a value of 0.95, whereas the unadjusted estimate is somewhat closer to the null. For the period 1 treatment difference only, the p values of 0.289, 0.111, and 0.130 for the unadjusted method, the adjusted method, and the MMRM method, respectively, similarly fail to contradict the null hypothesis.

Table 4.7 Estimates of Δ_1 to Δ_4

Method	Statistics	Δ_1	Δ_2	Δ_3	Δ_4	$\left(\frac{\Delta_1}{SE}\right)^2$	p value
Unadjusted	Etimate	-0.791	-1.591	-0.666	-1.731	1.126	0.289
	SE	0.745	1.033	1.440	1.091		
Adjusted	Estimate	-0.960	-1.600	-0.467	-1.838	2.535	0.111
	SE	0.603	0.951	1.385	1.035		
MMRM	Estimate	-0.951	NA	-0.111	-1.909	2.488	0.130
	SE	0.603	NA	1.546	1.051		

Equal: equal weights; InvVar: Inverse variance weighting.

Importantly, the estimate pertaining to Δ_4 for the treatment difference at the end of the second period for the placebo non-responders in period 1 is twice as large as that pertaining to Δ_1 and at least three times bigger than that pertaining to Δ_3 as the difference for the placebo responders; and this indicates that the placebo non-responders, when given the test treatment, are more informative than the placebo responders. The comparison that addresses Δ_2 for the difference between the T:T group and the P:P group also show better improvement than that for the first period, and this consideration could possibly contribute to the overall treatment comparison if taken into account.

As shown in Table 4.8, all of the statistics provided by the unadjusted method, except for the one accounting for Δ_2 with equal weight, fail to contradict the null hypothesis at the 0.05

level, mainly due to somewhat larger variance; but the adjusted method, with any specification for Δ's, and the MMRM method considering Δ_1 and Δ_4, similarly show significant results at the 0.05 level.

Table 4.8 Estimates of Weighted Statistics

Method	Statistics	Δ_1 and Δ_4		Δ_1, Δ_2 and Δ_4		Δ_1, Δ_2 Δ_3 and Δ_4	
		Equal	InvVar	Equal	InvVar	Equal	InvVar
Unadjusted	Weighted estimate	-1.261	-1.090	-1.371	-1.082	-1.195	-0.967
	SE	0.661	0.615	0.696	0.615	0.669	0.560
	P value	0.056	0.077	0.049	0.079	0.074	0.084
Adjusted	Weighted estimate	-1.399	-1.199	-1.466	-1.167	-1.216	-1.041
	SE	0.599	0.521	0.631	0.520	0.620	0.480
	P value	0.020	0.022	0.020	0.025	0.050	0.030
MMRM	Weighted estimate	-1.430	-1.188				
	SE	0.606	0.523				
	P value	0.025	0.031				

Equal: equal weights; InvVar: Inverse variance weighting.

REFERENCES

Chen, Y.-F., Yang, Y., Hung, H.M.J. and Wang, S.-J. (2011) Evaluation of performance of some enrichment designs dealing with high placebo response in psychiatric clinical trials. *Contemporary Clinical Trials*, **32**, 592–604.

Doros, G., Pencina, M., Rybin, D., Meisner, A. and Fava, M. (2013) A repeated measures model for analysis of continuous outcomes in sequential parallel comparison design studies. *Statistics in Medicine*, **32**, 2767–2789.

Dunger-Baldauf, C. (2007) Designs with Randomization Following Initial Study Treatment. *Wiley Encyclopedia of Clinical Trials* p. John Wiley & Sons, Inc.

Dunger-Baldauf, C., Racine, A. and Koch, G.G. (2006) Re-treatment studies. *Drug Information Journal*, **40**.

Fava, M., Evins, A.E., Dorer, D.J. and Schoenfeld, D.A. (2003) The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach. *Psychotherapy and Psychosomatics*, **72**, 115–127.

Huang, X. and Tamura, R.N. (2010) Comparison of Test Statistics for the Sequential Parallel Design. *Statistics in Biopharmaceutical Research*, **2**, 42–50.

Jones, B. and Kenward, M.G. (2014) *Design and Analysis of Cross-Over Trials, Third Edition*. CRC Press.

Khan, A., Kolts, R.L., Rapaport, M.H., Krishnan, K.R.R., Brodhead, A.E. and Browns, W.A. (2005) Magnitude of placebo response and drug-placebo differences across psychiatric disorders. *Psychological Medicine*, **35**, 743–749.

Koch, G.G., Davis, S.M. and Anderson, R.L. (1998) Methodological advances and plans for improving regulatory success for confirmatory studies. Statistics in Medicine, 17, 1675–1690.

Koch, G.G., Tangen, C.M., Jung, J.G. and Amara, I.A. (1998) Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. Statistics in Medicine, 17, 1863–1892.

Lehmacher, W., Wassmer, G. and Reitmeir, P. (1991) Procedures for two-sample comparisons with multiple endpoints controlling the experimentwise error rate. *Biometrics*, **47**, 511–521.

Pincus, T., Koch, G.G., Lei, H., Mangal, B., Sokka, T., Moskowitz, R., et al. (2004) Patient Preference for Placebo, Acetaminophen (paracetamol) or Celecoxib Efficacy Studies (PACES): two randomised, double blind, placebo controlled, crossover clinical trials in patients with knee or hip osteoarthritis. *Annals of the Rheumatic Diseases*, **63**, 931–939.

Tamura, R.N. and Huang, X. (2007) An examination of the efficiency of the sequential parallel design in psychiatric clinical trials. *Clinical Trials (London, England)*, **4**, 309–317.

Tamura, R.N., Huang, X. and Boos, D.D. (2011) Estimation of treatment effect for the sequential parallel design. *Statistics in Medicine*, **30**, 3496–3506.

Temple, R.J. (1994) Special study designs: early escape, enrichment, studies in non-responders. *Communications in Statistics - Theory and Methods*, **23**, 499–531.

Tudor, G. and Koch, G.G. (1994) Review of nonparametric methods for the analysis of crossover studies. *Statistical Methods in Medical Research*, **3**, 345–381.

CHAPTER 5: RANDOMIZATION-BASED ANCOVA FOR POINT AND CONFIDENCE INTERVAL ESTIMATION IN SEQUENTIAL PARALLEL COMPARISON DESIGN (SPCD)

5.1 Introduction

When the global null hypothesis H_0 for the SPCD design is contradicted by test statistics in Chapter 4 such as $T_{\boldsymbol{w},\boldsymbol{c}}$ or $T_{\boldsymbol{w},b_0}$, or is not expected to apply, randomization-based covariance adjustment can proceed for confidence interval estimation for comparisons between T and P through the expansion of the population for inference to an almost infinitely large population of N patients so that the n randomized patients included in the trial are conceptually representative of this large population in a simple random sampling sense (Koch et al., 1998).

5.2 Methods

In the above setting, the randomization process is comparable to the random assignment of n_i patients to the *i*-th sequence group for $i = 1, 2, 3$ and $(N - n)$ patients to a group without random selection for inclusion in the clinical trial. Accordingly, $\boldsymbol{\mu}_i$ and $\boldsymbol{V}_{\boldsymbol{F}_i}$ in (4.3) become the population mean vector and population covariance matrix for the population of N patients which the n_i patients in the *i*-th group are assumed to represent if all N patients received the i-th sequence of treatments. Also, with n replaced by N in (4.1), (4.2), and (4.3), $\boldsymbol{\mu}_i$ and $\boldsymbol{V}_{\boldsymbol{F}_i}$ have the structure shown in (5.1).

$$\boldsymbol{\mu}_i = \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{F}_{ik}, \quad \boldsymbol{V}_{\boldsymbol{F}_i} = \frac{1}{(N-1)}\sum_{k=1}^{N} (\boldsymbol{F}_{ik} - \boldsymbol{\mu}_i)(\boldsymbol{F}_{ik} - \boldsymbol{\mu}_{i'})' \qquad (5.1)$$

Also, the sample mean estimator for the *i*-th group $\overline{\boldsymbol{F}}_i = (\sum_{k=1}^{N} U_{ik}\boldsymbol{F}_{ik}/n_i)$ applies, although the summation only includes the n_i patients randomized to the *i*-th group.

As $N \to \infty$, the covariance matrix of the sample mean simplifies to (5.2).

$$\boldsymbol{Var}(\overline{\boldsymbol{F}}_i) = \left(\boldsymbol{V}_{\boldsymbol{F}_i}/n_i\right) \qquad (5.2)$$

$$\boldsymbol{Cov}(\overline{\boldsymbol{F}}_i, \overline{\boldsymbol{F}}_{i'}) = 0.$$

An unbiased estimator for $\boldsymbol{Var}(\overline{\boldsymbol{F}}_i)$ is $\widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}_i}$ is as shown in (5.3), since $\boldsymbol{E}\{\widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}_i}\} = \boldsymbol{Var}(\overline{\boldsymbol{F}}_i)$

$$\widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}_i} = \sum_{k=1}^{N} U_{ik}(\boldsymbol{F}_{ik} - \overline{\boldsymbol{F}}_i)(\boldsymbol{F}_{ik} - \overline{\boldsymbol{F}}_i)'/n_i(n_i - 1) \qquad (5.3)$$

$$= \frac{1}{n_i(n_i-1)}\left\{\sum_{k=1}^{N}[U_{ik}(\boldsymbol{F}_{ik} - \boldsymbol{\mu}_i)(\boldsymbol{F}_{ik} - \boldsymbol{\mu}_i)'] - [n_i(\overline{\boldsymbol{F}}_i - \boldsymbol{\mu}_i)(\overline{\boldsymbol{F}}_i - \boldsymbol{\mu}_i)']\right\}$$

can be derived in (5.4), regardless of the large population size N.

$$\boldsymbol{E}\{\widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}_i}\} = \left\{\left[\sum_{k=1}^{N}(\boldsymbol{F}_{ik} - \boldsymbol{\mu}_i)(\boldsymbol{F}_{ik} - \boldsymbol{\mu}_i)'/N(n_i - 1)\right] - \left[\frac{(N-n_i)\boldsymbol{V}_{\boldsymbol{F},i}}{Nn_i(n_i-1)}\right]\right\} \qquad (5.4)$$

$$= \frac{\boldsymbol{V}_{\boldsymbol{F},i}}{N(n_i-1)}\left[(N-1) - \frac{(N-n_i)}{n_i}\right] = \left(\boldsymbol{V}_{\boldsymbol{F},i}/n_i\right) = Var(\overline{\boldsymbol{F}}_i)$$

Thus, $\widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}} = \boldsymbol{Diag}\left(\widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}_1}, \widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}_2}, \widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}_3}\right)$ is the block diagonal estimated covariance matrix for $\overline{\boldsymbol{F}}$ with the $\{\widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}},i}\}$ as its diagonal blocks.

Let $\widetilde{\boldsymbol{F}}_i$ denote the transformation of $\overline{\boldsymbol{F}}_i$ whereby the $\bar{f}_{ij}$ for $j = 1, 2$ are replaced by $\tilde{f}_{ij} = \left(\bar{f}_{ij}/\bar{z}_i\right)$, and $\tilde{g}_{i2} = \left(\bar{y}_{i2} - \bar{f}_{i2}\right)/(1 - \bar{z}_i)$ is also included, as shown in (5.5), with matrices $\boldsymbol{R}_1$, $\boldsymbol{r}$,

$$\widetilde{\boldsymbol{F}}_i = \left(\bar{y}_{i0}, \bar{y}_{i1}, \bar{z}_i, \tilde{f}_{i1}, \bar{y}_{i2}, \tilde{f}_{i2}, \tilde{g}_{i2}\right)' \qquad (5.5)$$

$$= \mathbf{exp}[\boldsymbol{R}_2\, \boldsymbol{log}(\boldsymbol{R}_1\overline{\boldsymbol{F}}_i + \boldsymbol{r})],$$

and $\boldsymbol{R}_2$ as shown in (5.6).

$$\boldsymbol{R}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}, \qquad \boldsymbol{r} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \qquad \boldsymbol{R}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 \end{bmatrix} \qquad (5.6)$$

By linear Taylor series methods as discussed in Koch et al. (1977), a consistent estimate for the covariance matrix of $\widetilde{\boldsymbol{F}}_i$ is $\widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}_i} = \boldsymbol{L}_i \widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}_i} \boldsymbol{L}_i'$ for which $\boldsymbol{L}_i = \boldsymbol{D}_{\widetilde{\boldsymbol{F}}_i} \boldsymbol{R}_2 \boldsymbol{D}^{-1}_{(\boldsymbol{R}_1 \overline{\boldsymbol{F}}_i + \boldsymbol{r})} \boldsymbol{R}_1$ and $(\boldsymbol{R}_1 \overline{\boldsymbol{F}}_i + \boldsymbol{r}) = \left(\bar{y}_{i0}, \bar{y}_{i1}, \bar{z}_i, (1 - \bar{z}_i), \bar{f}_{i1}, \bar{y}_{i2}, \bar{f}_{i2}, \left(\bar{y}_{i2} - \bar{f}_{i2}\right)'\right)$.

Let $\widetilde{\boldsymbol{F}} = \left(\widetilde{\boldsymbol{F}}_1', \widetilde{\boldsymbol{F}}_2', \widetilde{\boldsymbol{F}}_3'\right)'$. A consistent estimate for the covariance matrix of $\widetilde{\boldsymbol{F}}$ is the block diagonal matrix $\widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}} = \boldsymbol{Diag}\left(\widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}_1}, \widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}_2}, \widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}_3}\right)$ with the $\widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}_i}$ as its diagonal blocks.

The difference between sequence groups $\widetilde{\boldsymbol{a}} = \left[\left(\widetilde{\boldsymbol{F}}_2 - \widetilde{\boldsymbol{F}}_1\right)', \left(\widetilde{\boldsymbol{F}}_3 - \widetilde{\boldsymbol{F}}_1\right)'\right]' = \boldsymbol{A}\widetilde{\boldsymbol{F}}$ is constructed as shown in (5.7), and a consistent estimate for the covariance matrix of $\widetilde{\boldsymbol{a}}$

$$\widetilde{\boldsymbol{a}} = \boldsymbol{A}\widetilde{\boldsymbol{F}} = \begin{bmatrix} -\boldsymbol{I}_6 & \boldsymbol{I}_6 & \boldsymbol{0}_{66} \\ -\boldsymbol{I}_6 & \boldsymbol{0}_{66} & \boldsymbol{I}_6 \end{bmatrix} \widetilde{\boldsymbol{F}} \qquad (5.7)$$

is $\widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{a}}} = \boldsymbol{A}\widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}}\boldsymbol{A}'$ as shown in (5.8).

$$\widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{a}}} = \begin{bmatrix} \widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}_1} + \widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}_2} & \widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}_1} \\ \widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}_1} & \widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}_1} + \widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}_3} \end{bmatrix} \qquad (5.8)$$

For the assessment of H_0 without adjustment for $y_{10k} = y_{20k} = y_{30k}$, $y_{11k} = y_{21k}$, $z_{1k} = z_{2k}$, and $f_{11k} = f_{21k}$, the potential comparisons of interest for this crossover design are shown in (5.9), and they are linear functions $\tilde{\boldsymbol{c}} = \widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{a}} = \widetilde{\boldsymbol{C}}\boldsymbol{A}\widetilde{\boldsymbol{F}}$ of $\widetilde{\boldsymbol{a}}$ with matrix $\widetilde{\boldsymbol{C}}$ as shown in (5.10).

$$\tilde{\boldsymbol{c}}_1 = \left\{\bar{y}_{31} - \frac{(n_1\bar{y}_{11} + n_2\bar{y}_{21})}{(n_1 + n_2)}\right\}$$

$$\tilde{\boldsymbol{c}}_2 = (\bar{y}_{32} - \bar{y}_{12})$$

$$\tilde{\boldsymbol{c}}_3 = \left(\tilde{f}_{22} - \tilde{f}_{21}\right) \qquad (5.9)$$

$$\tilde{\boldsymbol{c}}_4 = (\tilde{g}_{22} - \tilde{g}_{21})$$

$$\widetilde{\boldsymbol{C}} = \begin{bmatrix} 0 & -\frac{n_2}{(n_1+n_2)} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad (5.10)$$

A consistent estimate for the covariance matrix of $\tilde{\boldsymbol{c}}$ is $\widehat{\boldsymbol{V}}_{\tilde{\boldsymbol{c}}} = \widetilde{\boldsymbol{C}}\widehat{\boldsymbol{V}}_{\tilde{\boldsymbol{a}}}\widetilde{\boldsymbol{C}}'$. Also, under H_0, $\boldsymbol{E}_A\{\tilde{\boldsymbol{c}} \mid H_0\} = \boldsymbol{0}_4$ where $\boldsymbol{E}_A\{\ \}$ denotes asymptotic expected value with respect to the distribution of $\tilde{\boldsymbol{c}}$ through the assumed random sampling of patients and the invoked randomization for sufficiently large sample size to support its statistical behavior through its linear Taylor series approximation.

Similarly to the previous chapter, for $\tilde{\boldsymbol{c}} = (\tilde{c}_1, \tilde{c}_2, \tilde{c}_3, \tilde{c}_4)'$, the comparison $\tilde{c}_1$ pertains to T versus P in period 1; the comparison $\tilde{c}_2$ pertains to T:T versus P:P in period 2; the comparison $\tilde{c}_3$ pertains to P:T versus P:P in period 2 for responders to placebo in period 1; $\tilde{c}_4$ pertains to P:T versus P:P in period 2 for non-responders to placebo in period 1. Univariate test statistics for H_0 and thereby for the comparison between T and P can be based on weighted linear combinations $\tilde{c}_w = \sum_{h=1}^{4} w_h\tilde{c}_h$ where $\boldsymbol{w} = (w_1, w_2, w_3, w_4)'$ is a vector of weights such that all $w_h \geq 0$ and $\sum_{h=1}^{4} w_h = 1$. With the weights $\boldsymbol{w}$, the test statistic for the global null hypothesis H_0 at the two-sided significance level α is $\tilde{T}_{w,\tilde{c}} = w'\tilde{\boldsymbol{c}}/(\boldsymbol{w}'\boldsymbol{V}_{\tilde{\boldsymbol{c}}}\boldsymbol{w})^{0.5}$. Under H_0, $\tilde{T}_{w,\tilde{c}}$ approximately has a standard normal distribution with mean 0 and variance 1. A confidence interval based

on $\tilde{c}_w$ can be constructed as $\left[\tilde{c}_{\mathrm{w}} - Z_{\frac{\alpha}{2}}\sqrt{V_{\tilde{c}_{\mathrm{w}}}},\ \tilde{c}_{\mathrm{w}} + Z_{\frac{\alpha}{2}}\sqrt{V_{\tilde{c}_{\mathrm{w}}}}\right]$, where $Z_{\frac{\alpha}{2}}$ is the $\left(1-\frac{\alpha}{2}\right)$th percentile of the standard normal distribution and $\boldsymbol{V}_{\tilde{c}_w} = \boldsymbol{w}'\boldsymbol{V}_{\tilde{c}_w}\boldsymbol{w}$.

The constraints $\tilde{\boldsymbol{c}}_0 = \widetilde{\boldsymbol{C}}_0\tilde{\boldsymbol{a}} = \widetilde{\boldsymbol{C}}_0\boldsymbol{A}\widetilde{\boldsymbol{F}}$ for $\tilde{\boldsymbol{a}}$ with $\boldsymbol{E}_A\{\tilde{\boldsymbol{c}}_0\} = 0$ regardless of whether H_0 applies are shown in (5.11).

$$\tilde{\boldsymbol{c}}_0 = \begin{bmatrix}(\bar{y}_{20}-\bar{y}_{10})\\(\bar{y}_{30}-\bar{y}_{10})\\(\bar{y}_{21}-\bar{y}_{11})\\(\bar{z}_2-\bar{z}_1)\\(\tilde{f}_{21}-\tilde{f}_{11})\end{bmatrix} = \begin{bmatrix}1&0&0&0&0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&1&0&0&0&0&0&0\\0&1&0&0&0&0&0&0&0&0&0&0&0&0\\0&0&1&0&0&0&0&0&0&0&0&0&0&0\\0&0&0&1&0&0&0&0&0&0&0&0&0&0\end{bmatrix}\tilde{\boldsymbol{a}}$$

$$= \widetilde{\boldsymbol{C}}_0\tilde{\boldsymbol{a}} \qquad (5.11)$$

Let $\tilde{\boldsymbol{c}}_+ = [\tilde{\boldsymbol{c}}', \tilde{\boldsymbol{c}}_0']' = \left[\widetilde{\boldsymbol{C}}', \widetilde{\boldsymbol{C}}_0'\right]'\tilde{\boldsymbol{a}} = \widetilde{\boldsymbol{C}}_+\tilde{\boldsymbol{a}}$ denote the combined set of comparisons $\tilde{\boldsymbol{c}}$ pertaining to H_0 and constraints $\tilde{\boldsymbol{c}}_0$. The estimated covariance structure for $\tilde{\boldsymbol{c}}_+$ is $\widehat{\boldsymbol{V}}_{\tilde{\boldsymbol{c}}_+} = \widetilde{\boldsymbol{C}}_+\widehat{\boldsymbol{V}}_{\tilde{\boldsymbol{a}}}\widetilde{\boldsymbol{C}}_+'$. Since $\boldsymbol{E}\{\tilde{\boldsymbol{c}}_0\} = \boldsymbol{0}_5$ regardless of whether H_0 applies, randomization-based covariance adjustment for $\tilde{\boldsymbol{c}}$ with respect to the constraints $\tilde{\boldsymbol{c}}_0$ is invoked by fitting the linear model shown in (5.12) by

$$\boldsymbol{E}\{\tilde{\boldsymbol{c}}_+\} \triangleq \begin{bmatrix}\boldsymbol{I}_4\\\boldsymbol{0}_{5,4}\end{bmatrix}\widetilde{\boldsymbol{b}} = \boldsymbol{X}\widetilde{\boldsymbol{b}} \qquad (5.12)$$

weighted least squares with weights based on $\widehat{\boldsymbol{V}}_{\tilde{\boldsymbol{c}}_+}^{-1}$ and with "$\triangleq$" meaning "is estimated by"; and $\widetilde{\boldsymbol{b}} = (\tilde{b}_1, \tilde{b}_2, \tilde{b}_3, \tilde{b}_4)'$ are covariance adjusted counterparts of $\tilde{\boldsymbol{c}}$, and $\widetilde{\boldsymbol{b}} = \left(\boldsymbol{X}'\widehat{\boldsymbol{V}}_{\tilde{\boldsymbol{c}}_+}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\widehat{\boldsymbol{V}}_{\tilde{\boldsymbol{c}}_+}^{-1}\tilde{\boldsymbol{c}}_+$; also, the covariance matrix for $\widetilde{\boldsymbol{b}}$ is $\widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{b}}} = \left(\boldsymbol{X}'\widehat{\boldsymbol{V}}_{\tilde{\boldsymbol{c}}_+}^{-1}\boldsymbol{X}\right)^{-1}$. Covariance adjusted test statistics for H_0 can be based on weighted linear combinations $\widetilde{\boldsymbol{b}}_{\mathrm{w}} = \sum_{h=1}^{4} w_h\tilde{b}_h$ in ways similar to those discussed for $\tilde{\boldsymbol{c}}$.

5.3 Simulation Study

5.3.1 Simulation Setup

The Type I errors and powers of the methods discussed in this chapter were evaluated with simulation studies. The responses Y_{ij} of patients in the *i*-th group for the *j*-th period, where $i =$ P:P, P:T, T:T, were randomly generated in the manner shown in (5.13) for which Z_{i1} is an

$$\begin{bmatrix} Y_{i0} \\ Y_{i1} \\ Y_{i2} \end{bmatrix} = \begin{bmatrix} e_{i0} \\ e_{i1} \\ e_{i2} \end{bmatrix} + \begin{bmatrix} \xi_0 \\ \xi_{i1} \\ Z_{i1}\xi_{i2} + (1 - Z_{i1})\xi_{i3} \end{bmatrix} \qquad (5.13)$$

indicator for a period 1 responder in the *i*-th group with $Z_{i1} = 1$ if $Y_{i1} \leq L$, $Z_{i1} = 0$ if $Y_{i1} > L$ for L as the specified criterion for a responder or not. In (5.13), the $\boldsymbol{e}_i = (e_{i0}, e_{i1}, e_{i2})$ are independently generated, random errors from the trivariate normal distribution with $\mathbf{0}_3$ as the common mean for all three groups, and with $\wedge$ in (5.14) as the common covariance matrix for all

$$\wedge = \sigma^2 \begin{bmatrix} 1 & \rho_{01} & \rho_{02} \\ \rho_{01} & 1 & \rho_{12} \\ \rho_{02} & \rho_{12} & 1 \end{bmatrix} \qquad (5.14)$$

three groups where σ^2 is the common variance for all three periods and $\rho_{jj'}$ is the correlation for periods j and j'. Additionally, ξ_0 is the common mean at baseline for all three groups; ξ_{i1} is the mean for period 1 for the *i*-th group; ξ_{i2} is an *i*-th group shift parameter that applies to period 2 for period 1 responders; and ξ_{i3} is an *i*-th group shift parameter that applies to period 2 for period 1 non-responders.

For all simulation studies, the specified covariance matrix $\wedge$ had $\sigma^2 = 36$, with the scope of correlations being $\rho_{12} = \rho_{13} = \rho_{23} = 0.3, 0.5$ as exchangeable structures and $\rho_{12} = \rho_{23} = \rho_{13}^{0.5} = 0.5, 0.7$ as autoregressive structures. The specifications for the assessments of type 1 error were $\xi_0 = 40$, $\xi_{\text{P:P},1} = \xi_{\text{P:T},1} = \xi_{\text{T:T},1} = 35$, $\xi_{\text{P:P},2} = \xi_{\text{P:T},2} = \xi_{\text{T:T},2} = 32$, and $\xi_{\text{P:P},3} =$

$\xi_{\text{P:T,3}} = \xi_{\text{T:T,3}} = 35$, with $Y_{i1} \leq L = 33$ being the criterion for a period 1 responder in the *i*-th group. Also, with $\pi_i = E\{Z_{i1}\}$ as the probability of responder status for group *i*, it follows from $\xi_{\text{P:P,1}} = \xi_{\text{P:T,1}} = \xi_{\text{T:T,1}} = 35$ that $\pi_{\text{P:P}} = \pi_{\text{P:T}} = \pi_{\text{T:T}} = 0.37$ for the assessments of type 1 error under the global null hypothesis H_0 as specified in Section 4.2.

The specifications of the ξ_{i1}, ξ_{i2}, and ξ_{i3} for the assessments of statistical power are shown in Table 5.1. Also, $\xi_0 = 40$ and $\pi_{\text{P:P}} = \pi_{\text{P:T}} = 0.37$, but $\pi_{\text{T:T}} = 0.47$ so that the difference in period 1 responder rates for the T:T group versus the P:P and P:T groups is about 0.10. From Table 5.1, it follows that $\Delta_1 = \left(\xi_{\text{T:T,1}} - \xi_{\text{P:P,1}}\right) = -1.5$, $\Delta_3 = \left(\xi_{\text{P:T,2}} - \xi_{\text{P:P,2}}\right) = -1.0$, and $\Delta_4 = \left(\xi_{\text{P:T,3}} - \xi_{\text{P:P,3}}\right) = -2.0$. Since $E\{Y_{i2}\} = \{\pi_i\xi_{i2} + (1-\pi_i)\xi_{i3}\} = \eta_i$, it follows that $\Delta_2 = (\eta_{\text{T:T}} - \eta_{\text{P:P}}) = (32.1 - 33.9) = -1.8$. Thus, $(\Delta_1, \Delta_2, \Delta_3, \Delta_4) = (-1.5, -1.8, -1, -2)$ is the specification that corresponds to the assessments of statistical power with respect to (c_1, c_2, c_3, c_4) and (b_1, b_2, b_3, b_4); and $(\Delta_1, \Delta_2, \Delta_3, \Delta_4) = (0, 0, 0, 0)$ corresponds to the assessments of Type 1 error.

Table 5.1 Specifications for Assessments of Statistical Power

Group	Period 1	Period 2 for Period 1 Responder	Period 2 for Period 1 Non-Responder
P:P	$\xi_{\text{P:P,1}} = 35$	$\xi_{\text{P:P,2}} = 32$	$\xi_{\text{P:P,3}} = 35$
P:T	$\xi_{\text{P:T,1}} = 35$	$\xi_{\text{P:T,2}} = 31$	$\xi_{\text{P:T,3}} = 33$
T:T	$\xi_{\text{T:T,1}} = 33.5$	$\xi_{\text{T:T,2}} = 30.5$	$\xi_{\text{T:T,3}} = 33.5$

The simulations were performed with equal sample sizes $n_* = 40, 80, 160$ patients per group, with the total sample sizes respectively being $3n_* = 120, 240, 480$. For each simulation, the responses of the baseline, period 1, and period 2 are generated via (5.13) with the previously noted specifications. For testing the hypothesis $H_0: \Delta_w = 0 \; versus \; H_A: \Delta_w \neq 0$ in correspondence to c_w or $b_{0,w}$, the determination of p value for the test statistic $T_{w,c}$ or T_{wb_0} is

based on reference of its squared value to the chi-square value $\chi^2_{1,1-\alpha}$ where α is the specified significance level and we chose α to be 0.05. The simulation results are based on 50,000 replicates for the specifications previously stated. The results of Type I error and power from the simulations are means of indicator variables for whether $p \leq \alpha$ applies for testing H_0, and the empirical standard deviation (ESD) is taken as the square root of the variance of the estimated c_w or $b_{w,0}$ across all simulations, and the average of the estimated standard error (ASE) of c_w or $b_{w,0}$ in the simulations is provided.

5.3.2 Simulation Results

The results that address Δ_1 and Δ_4 only, which is usually the case of interest in the SPCD design, under the null hypothesis, are displayed in Table 5.2 with equal weighting and inverse variance weighting; and the results under the null that address Δ_1, Δ_2 and Δ_4, and all Δ_1 to Δ_4, are as shown in Tables 5.3 and 5.4 respectively.

For closer to nominal control of Type 1 error at the two-sided 0.05 level and closer to nominal coverage for the two-sided 0.95 confidence interval, the estimator $\hat{V}_{\tilde{c}_+}$ for the variance of $\tilde{c}_+$ has multiplication by $\frac{N-3}{N-m}$, where the subtraction of m corresponds to the number of sample means estimated from the 3 sequence groups combined; for example, $m = 7$ and $m = 12$ for the unadjusted approach and covariate-adjusted approach when the test statistic is based on a weighted mean to address Δ_1 and Δ_4 with ($m = 9, 14$ for addressing $\Delta_1, \Delta_2, \Delta_4$ and $m = 11, 16$ for addressing all Δ_1 to Δ_4). Accordingly, an approximate F distribution with d.f.=$(1, N - m)$ for the p value of hypothesis testing and t-distribution with $\text{d. f.} = (N - n)$ for confidence interval determination is used.

As shown in Table 5.2, under the null hypothesis, all three methods provide unbiased point estimates and reasonably accurate corresponding standard errors, as the ASE and ESD are

similar. When equal weights are used, Type I errors range from 0.0467 to 0.0501 for the unadjusted approach, and from 0.0485 to 0.0533 for the covariate adjusted approach, and 0.0489 to 0.0542 for the MMRM approach of Doros et al. (2013), with the larger Type I errors occurring at the $n = 40$ per group and the somewhat smaller than nominal Type I errors at n=160 per group for all three approaches. Similar results are observed in the inverse variance weighting to the parameters. As shown in the efficiency column, the covariate adjusted and MMRM methods have similar variances for estimation, with these being smaller than that for the unadjusted method; and thus there is better precision and narrower confidence interval estimation. As shown in Table 5.2 vertically, as correlation among the outcomes increases, the variance for the estimation decreases in all three approaches of estimation, with a bigger impact in the covariate-adjusted and MMRM estimators than the unadjusted.

Under H_0 for no treatment differences, results that address Δ_2, the treatment difference between P:P and T:T sequences (in addition to the traditional SPCD design that addresses only Δ_1 and Δ_4) are shown in Table 5.3. Results with inclusion of Δ_2 also have good control of the Type I error. For utilizing all available information from the study, a weighted statistic that addresses all Δ_1 to Δ_4, with equal weighting or inverse variance weighting, is also considered, and there is control for the type I errors at the 0.05 significance level, as shown in Table 5.4.

Under the alternative parameter specification in Table 5.1, the nominal coverage of 95% confidence interval and the power with the three methods, under scenarios of different sample sizes and correlations among outcomes, are presented in Table 5.5. The coverage with equal weighting for addressing Δ_1 and Δ_4 by the unadjusted, adjusted, and MMRM each range from 0.949 to 0.953, and 0.947 to 0.952, and 0.946 to 0.951, respectively, and they all indicate good coverage of the target parameter.

Table 5.2 Results from 50, 000 replicate simulations for the test statistics of $H_0: \Delta_1 = \Delta_4 = 0$ under $\Delta_1 = \Delta_2 = \Delta_3 = \Delta_4 = 0$

ρ	n_i	Weight	Unadjusted				Adjusted				MMRM				Efficiency	
			Bias	ASE	ESD	Type I	Bias	ASE	ESD	Type I	Bias	ASE	ESD	Type I	a/u	a/m
0.3 (EX)	n=40	Equal	-0.0030	1.021	1.014	0.0479	0.0034	0.985	0.992	0.0515	-0.0014	0.973	0.977	0.0520	0.96	1.03
		InvVar	0.0105	0.958	0.961	0.0495	0.0175	0.921	0.941	0.0543	0.0085	0.908	0.919	0.0526	0.96	1.05
	n=80	Equal	-0.0047	0.717	0.718	0.0493	-0.0017	0.690	0.696	0.0523	-0.0042	0.684	0.688	0.0513	0.94	1.02
		InvVar	0.0040	0.674	0.675	0.0501	0.0065	0.647	0.654	0.0518	0.0020	0.641	0.645	0.0515	0.94	1.03
	n=160	Equal	0.0038	0.506	0.503	0.0486	0.0037	0.486	0.487	0.0502	0.0027	0.483	0.482	0.0494	0.93	1.02
		InvVar	0.0065	0.476	0.475	0.0492	0.0063	0.456	0.458	0.0507	0.0042	0.453	0.454	0.0506	0.93	1.02
0.5 (AR)	n=40	Equal	-0.0026	0.989	0.985	0.0478	-0.0007	0.905	0.915	0.0512	-0.0037	0.912	0.928	0.0542	0.86	0.97
		InvVar	0.0096	0.942	0.946	0.0489	0.0088	0.841	0.858	0.0539	0.0021	0.838	0.854	0.0552	0.82	1.01
	n=80	Equal	-0.0026	0.695	0.694	0.0494	-0.0012	0.634	0.637	0.0517	-0.0021	0.642	0.648	0.0540	0.84	0.97
		InvVar	0.0052	0.663	0.664	0.0500	0.0046	0.590	0.595	0.0525	0.0016	0.592	0.597	0.0528	0.80	0.99
	n=160	Equal	0.0029	0.490	0.487	0.0482	0.0038	0.446	0.446	0.0502	0.0026	0.453	0.454	0.0509	0.84	0.97
		InvVar	0.0062	0.468	0.467	0.0490	0.0070	0.416	0.418	0.0513	0.0052	0.419	0.420	0.0519	0.80	0.99
0.5 (EX)	n=40	Equal	-0.0031	0.989	0.985	0.0501	0.0019	0.884	0.897	0.0533	-0.0022	0.865	0.873	0.0538	0.83	1.06
		InvVar	0.0106	0.942	0.945	0.0504	0.0139	0.831	0.851	0.0555	0.0061	0.816	0.829	0.0561	0.81	1.05
	n=80	Equal	0.0010	0.695	0.694	0.0490	0.0028	0.619	0.623	0.0517	0.0016	0.609	0.612	0.0509	0.81	1.04
		InvVar	0.0086	0.663	0.663	0.0498	0.0085	0.583	0.590	0.0523	0.0050	0.576	0.580	0.0519	0.79	1.03
	n=160	Equal	0.0009	0.489	0.485	0.0467	0.0005	0.435	0.433	0.0485	0.0001	0.429	0.427	0.0489	0.80	1.03
		InvVar	0.0040	0.468	0.465	0.0477	0.0036	0.411	0.411	0.0489	0.0020	0.407	0.406	0.0496	0.78	1.02
0.7 (AR)	n=40	Equal	0.0002	0.938	0.933	0.0489	0.0036	0.748	0.751	0.0507	0.0001	0.752	0.758	0.0533	0.65	0.98
		InvVar	0.0113	0.922	0.924	0.0515	0.0104	0.697	0.706	0.0527	0.0052	0.690	0.700	0.0536	0.58	1.02
	n=80	Equal	-0.0017	0.659	0.656	0.0473	0.0002	0.524	0.524	0.0505	-0.0010	0.529	0.533	0.0519	0.64	0.97
		InvVar	0.0061	0.650	0.647	0.0477	0.0046	0.488	0.489	0.0504	0.0018	0.488	0.490	0.0508	0.57	1.00
	n=160	Equal	0.0046	0.464	0.461	0.0482	0.0032	0.368	0.367	0.0503	-0.0009	0.373	0.376	0.0518	0.64	0.96
		InvVar	0.0069	0.458	0.455	0.0486	0.0043	0.344	0.344	0.0499	0.0005	0.345	0.346	0.0516	0.57	0.98

Note: Bias = mean of (estimate-true value); Type I error = rejection rate of null hypothesis (when $Z_w^2 > \chi_{0.95}^2$); ASE=average standard error; ESD=empirical standard deviation. a/u=(adjusted variance) / (unadjusted variance); a/m = (adjusted variance) / (MMRM variance).

Table 5.3 Results from 50, 000 replicate simulations for the test statistics of $H_0: \Delta_1 = \Delta_3 = \Delta_4 = 0$ under $\Delta_1 = \Delta_2 = \Delta_3 = \Delta_4 = 0$

ρ	n_i	Weight	Unadjusted				Adjusted				Efficiency
			Bias	ASE	ESD	Type I	Bias	ASE	ESD	Type I	a/u
0.3 (EX)	n=40	Equal	-0.0012	1.033	1.018	0.0466	0.0065	0.976	0.974	0.0491	0.92
		InvVar	-0.0077	0.948	0.958	0.0516	0.0023	0.902	0.926	0.0563	0.94
	n=80	Equal	-0.0017	0.723	0.717	0.0485	0.0016	0.680	0.681	0.0498	0.90
		InvVar	-0.0035	0.667	0.670	0.0507	0.0008	0.633	0.641	0.0518	0.92
	n=160	Equal	0.0023	0.508	0.505	0.0488	0.0022	0.478	0.479	0.0487	0.90
		InvVar	0.0015	0.471	0.469	0.0501	0.0017	0.446	0.448	0.0511	0.91
0.5 (AR)	n=40	Equal	-0.0052	1.048	1.032	0.0468	-0.0017	0.955	0.953	0.0490	0.85
		InvVar	-0.0004	0.943	0.954	0.0511	0.0031	0.842	0.867	0.0562	0.83
	n=80	Equal	-0.0014	0.733	0.726	0.0478	0.0004	0.665	0.664	0.0487	0.84
		InvVar	-0.0003	0.663	0.667	0.0509	0.0016	0.591	0.598	0.0531	0.80
	n=160	Equal	0.0034	0.515	0.510	0.0486	0.0047	0.467	0.466	0.0491	0.83
		InvVar	0.0035	0.468	0.467	0.0495	0.0055	0.416	0.419	0.0523	0.80
0.5 (EX)	n=40	Equal	-0.0050	1.048	1.034	0.0474	0.0006	0.892	0.896	0.0502	0.75
		InvVar	-0.0004	0.943	0.953	0.0526	0.0012	0.824	0.850	0.0571	0.80
	n=80	Equal	0.0009	0.733	0.727	0.0485	0.0026	0.622	0.622	0.0496	0.73
		InvVar	0.0029	0.664	0.666	0.0502	0.0026	0.578	0.586	0.0528	0.77
	n=160	Equal	0.0022	0.515	0.508	0.0467	0.0023	0.437	0.433	0.0473	0.73
		InvVar	0.0012	0.468	0.466	0.0481	0.0015	0.407	0.407	0.0499	0.76
0.7 (AR)	n=40	Equal	0.0007	1.051	1.036	0.0467	0.0075	0.825	0.821	0.0486	0.63
		InvVar	0.0068	0.901	0.909	0.0520	0.0075	0.688	0.703	0.0553	0.60
	n=80	Equal	0.0002	0.735	0.727	0.0476	0.0025	0.575	0.571	0.0487	0.62
		InvVar	0.0030	0.634	0.635	0.0495	0.0032	0.482	0.486	0.0516	0.59
	n=160	Equal	0.0030	0.517	0.513	0.0485	0.0022	0.403	0.403	0.0504	0.62
		InvVar	0.0067	0.447	0.444	0.0487	0.0044	0.339	0.339	0.0499	0.58

Note: Bias = mean of (estimate-true value); Type I error = rejection rate of null hypothesis (when $Z_w^2 > \chi_{0.95}^2$);
ASE=average standard error; ESD=empirical standard deviation. a/u=(adjusted variance) / (unadjusted variance).

Table 5.4 Results from 50, 000 replicate simulations for the test statistics of $H_0: \Delta_1 = \Delta_3 = \Delta_3 = \Delta_4 = 0$ under $\Delta_1 = \Delta_2 = \Delta_3 = \Delta_4 = 0$

ρ	n_i	Weight	Unadjusted				Adjusted				Efficiency
			Bias	ASE	ESD	Type I	Bias	ASE	ESD	Type I	a/u
0.3 (EX)	n=40	Equal	-0.0011	1.011	0.994	0.0468	0.0036	0.958	0.965	0.0518	0.94
		InvVar	-0.0050	0.878	0.892	0.0532	0.0036	0.839	0.870	0.0586	0.95
	n=80	Equal	-0.0020	0.705	0.698	0.0465	0.0018	0.667	0.668	0.0491	0.92
		InvVar	-0.0030	0.617	0.621	0.0508	0.0017	0.589	0.598	0.0534	0.93
	n=160	Equal	0.0022	0.495	0.492	0.0476	0.0024	0.468	0.469	0.0495	0.91
		InvVar	0.0024	0.435	0.435	0.0499	0.0030	0.415	0.417	0.0511	0.92
0.5 (AR)	n=40	Equal	-0.0088	0.995	0.977	0.0457	-0.0072	0.920	0.926	0.0514	0.90
		InvVar	-0.0080	0.848	0.863	0.0529	-0.0037	0.765	0.796	0.0589	0.85
	n=80	Equal	-0.0011	0.693	0.685	0.0469	0.0009	0.641	0.641	0.0492	0.88
		InvVar	-0.0030	0.596	0.600	0.0514	-0.0001	0.537	0.545	0.0534	0.82
	n=160	Equal	0.0029	0.487	0.482	0.0468	0.0044	0.450	0.448	0.0490	0.87
		InvVar	0.0020	0.420	0.420	0.0498	0.0044	0.378	0.380	0.0515	0.82
0.5 (EX)	n=40	Equal	-0.0062	0.995	0.979	0.0461	-0.0018	0.861	0.871	0.0510	0.79
		InvVar	-0.0061	0.848	0.860	0.0527	0.0005	0.752	0.782	0.0582	0.83
	n=80	Equal	0.0002	0.694	0.689	0.0488	0.0022	0.600	0.603	0.0503	0.77
		InvVar	-0.0001	0.596	0.602	0.0514	0.0026	0.528	0.539	0.0549	0.80
	n=160	Equal	0.0020	0.487	0.482	0.0484	0.0019	0.421	0.419	0.0496	0.76
		InvVar	-0.0002	0.420	0.420	0.0513	0.0010	0.372	0.373	0.0506	0.79
0.7 (AR)	n=40	Equal	0.0021	0.959	0.941	0.0456	0.0062	0.782	0.784	0.0497	0.69
		InvVar	-0.0027	0.780	0.790	0.0527	0.0015	0.614	0.633	0.0582	0.64
	n=80	Equal	-0.0021	0.669	0.660	0.0461	0.0002	0.544	0.544	0.0497	0.68
		InvVar	-0.0049	0.548	0.550	0.0499	-0.0013	0.430	0.435	0.0517	0.63
	n=160	Equal	0.0019	0.469	0.466	0.0483	0.0015	0.382	0.382	0.0492	0.67
		InvVar	0.0027	0.386	0.387	0.0506	0.0026	0.303	0.304	0.0514	0.62

Note: Bias = mean of (estimate-true value); Type I error = rejection rate of null hypothesis (when $Z_w^2 > \chi_{0.95}^2$);
ASE=average standard error; ESD=empirical standard deviation. a/u=(adjusted variance) / (unadjusted variance).

Table 5.5 Results from 50,000 replicate simulations for the coverage of confidence intervals and power of test statistics under the alternative specified in Table 5.1

ρ	n_i	Weight	Δ_1 and Δ_4						Δ_1, Δ_2, and Δ_4				Δ_1, Δ_2, Δ_3, and Δ_4			
			Unadjusted		Adjusted		MMRM		Unadjusted		Adjusted		Unadjusted		Adjusted	
			CR	Power	CR	Power	CR	Power	CR	Power	CR	Power	CR	Power	CR	Power
0.3 (EX)	n=40	Equal	0.952	0.403	0.948	0.425	0.948	0.443	0.954	0.395	0.951	0.431	0.953	0.334	0.948	0.368
		InvVar	0.950	0.410	0.945	0.435	0.946	0.451	0.948	0.426	0.943	0.457	0.944	0.423	0.939	0.454
	n=80	Equal	0.951	0.679	0.948	0.710	0.949	0.723	0.952	0.681	0.950	0.729	0.953	0.605	0.951	0.647
		InvVar	0.950	0.691	0.948	0.724	0.947	0.733	0.949	0.706	0.947	0.748	0.947	0.707	0.945	0.747
	n=160	Equal	0.951	0.929	0.950	0.945	0.951	0.950	0.951	0.933	0.951	0.956	0.952	0.887	0.951	0.916
		InvVar	0.950	0.936	0.949	0.952	0.948	0.954	0.950	0.943	0.949	0.961	0.948	0.943	0.947	0.962
0.5 (AR)	n=40	Equal	0.952	0.422	0.949	0.486	0.946	0.491	0.953	0.385	0.951	0.450	0.954	0.343	0.948	0.395
		InvVar	0.951	0.419	0.946	0.495	0.944	0.506	0.948	0.425	0.943	0.498	0.945	0.433	0.939	0.506
	n=80	Equal	0.951	0.707	0.948	0.782	0.946	0.775	0.952	0.670	0.951	0.750	0.953	0.618	0.951	0.683
		InvVar	0.950	0.702	0.947	0.793	0.946	0.793	0.949	0.708	0.947	0.794	0.947	0.721	0.943	0.803
	n=160	Equal	0.952	0.945	0.950	0.972	0.948	0.969	0.952	0.928	0.951	0.963	0.953	0.899	0.951	0.935
		InvVar	0.950	0.941	0.948	0.975	0.948	0.976	0.950	0.945	0.948	0.976	0.948	0.950	0.947	0.980
0.5 (EX)	n=40	Equal	0.950	0.424	0.947	0.505	0.946	0.529	0.953	0.384	0.950	0.499	0.954	0.342	0.949	0.437
		InvVar	0.949	0.417	0.943	0.510	0.943	0.535	0.947	0.426	0.942	0.526	0.945	0.434	0.940	0.530
	n=80	Equal	0.951	0.708	0.948	0.797	0.949	0.815	0.951	0.667	0.950	0.802	0.951	0.615	0.950	0.737
		InvVar	0.950	0.702	0.947	0.807	0.947	0.820	0.949	0.708	0.947	0.819	0.946	0.721	0.942	0.825
	n=160	Equal	0.951	0.946	0.952	0.979	0.951	0.982	0.951	0.928	0.952	0.981	0.952	0.898	0.951	0.962
		InvVar	0.950	0.944	0.950	0.981	0.949	0.983	0.950	0.946	0.950	0.983	0.948	0.952	0.947	0.985
0.7 (AR)	n=40	Equal	0.951	0.456	0.949	0.637	0.947	0.643	0.953	0.380	0.952	0.558	0.955	0.364	0.950	0.507
		InvVar	0.949	0.427	0.946	0.639	0.946	0.663	0.948	0.449	0.944	0.649	0.946	0.476	0.939	0.671
	n=80	Equal	0.953	0.754	0.950	0.912	0.946	0.907	0.952	0.666	0.952	0.863	0.954	0.651	0.951	0.822
		InvVar	0.952	0.717	0.949	0.914	0.946	0.919	0.951	0.744	0.947	0.918	0.947	0.779	0.946	0.931
	n=160	Equal	0.952	0.963	0.950	0.997	0.949	0.997	0.951	0.927	0.950	0.991	0.952	0.917	0.951	0.984
		InvVar	0.951	0.949	0.949	0.997	0.947	0.997	0.952	0.959	0.950	0.997	0.948	0.971	0.946	0.998

Note: CR = 2 sided coverage rate of the nominal 95% confidence interval; Power = rejection rate of null hypothesis (when $Z_w^2 > \chi_{0.95}^2$).

5.4 Example

A hypothetical placebo-controlled pre-randomized two-period study that makes use of the P:P, P:T, and T:T sequence groups is created to illustrate an application of the proposed methods and to compare them to counterparts from the MMRM method proposed by Doros et al. (2013). The example has 240 subjects which are equally assigned to the three sequence groups. The means and standard deviations of the outcomes are provided in Table 5.6. Assessments of outcomes occur at the baseline, end of period 1 and period 2, with a smaller outcome score being more beneficial. At period 1, a score less than 33 is considered as a responder to the treatment. The means of measurements at baseline $\bar{Y}_0$ are similar across the three sequence groups, with a slightly larger mean in the T:T group. There are 31%, 34%, and 39% of responders in the three groups respectively, as shown in the $\bar{z}$ row, with only 6% more responders in the test treatment group than the placebo group. And as the trial continues to the second period, the mean measurements at the end of this period for the P:T and T:T groups are almost the same, and the P:P group is slightly worse than the P:T or T:T group overall.

Table 5.6 Mean and Standard Deviation of Outcome

Statistics	P:P	P:T	T:T
$\bar{Y}_0$	40.54	40.45	40.80
SD	5.36	5.57	6.91
$\bar{Y}_1$	35.15	34.92	34.24
SD	6.17	5.72	6.48
$\bar{z}$	0.31	0.34	0.39
$\bar{Y}_2$	33.98	32.44	32.39
SD	5.69	5.04	6.13

The estimates using the proposed unadjusted and adjusted methods as well as the MMRM method are shown in Table 5.7. The standard errors of the estimates increase as the sample sizes decrease, as $\widehat{\Delta}_1$ has the smallest standard error and $\widehat{\Delta}_3$ has the largest standard error

(as it only uses about 30% (i.e., the placebo responders) of the sample size in the P:P and P:T groups). The covariable-adjusted method provides similar estimates for $\widehat{\Delta}_1$, the treatment difference at the end of the first period, with a value of 0.95, whereas the unadjusted difference is slightly smaller because it does not correct for the slight random imbalance of the measurement at baseline. For the period 1 treatment difference only, the p values are 0.271 for the unadjusted method, 0.098 for the adjusted method, and 0.115 for the MMRM method, and they fail to contradict the null hypothesis of no treatment difference.

Table 5.7 Estimates of Δ_1 to Δ_4

Method	Statistics	Δ_1	Δ_2	Δ_3	Δ_4	$\left(\frac{\Delta_1}{SE}\right)^2$	p value
Unadjusted	Etimate	-0.791	-1.591	-0.677	-1.664	1.211	0.271
	SE	0.719	1.033	1.504	1.074		
Adjusted	Estimate	-0.955	-1.585	-0.530	-1.749	2.732	0.098
	SE	0.578	0.937	1.468	1.022		
MMRM	Estimate	-0.951	NA	-0.111	-1.909	2.488	0.115
	SE	0.603	NA	1.546	1.051		

Importantly, estimates for addressing Δ_4 for the difference at the end of the second period for the placebo non-responders in period 1 are twice as large as those addressing Δ_1, and at least three times bigger than those addressing Δ_3, the difference for the placebo responders; and these considerations show that the placebo non-responders, when given the test treatment, are more informative than the placebo responders. The comparisons between the T:T and P:P sequences at the second period, which corresponds to Δ_2, also show better improvement than that at the first period, and this could possibly contribute to the overall treatment comparison if taken into consideration.

As shown in Table 5.8, none of the statistics provided by the unadjusted method succeed in contradicting the null hypothesis at the 0.05 level, due to a smaller weighted statistic or a slightly larger variance; but the adjusted method, except for the one also accounting for Δ_3 with equal weight, shows significant results at the 0.05 level, and the MMRM method accounting for Δ_1 and Δ_4 also shows significant p-values.

Table 5.8 Estimates of Weighted Statistics

Method	Statistics	Δ_1 and Δ_4		Δ_1, Δ_2 and Δ_4		Δ_1, Δ_2 Δ_3 and Δ_4	
		Equal	InvVar	Equal	InvVar	Equal	InvVar
Unadjusted	Weighted	-1.227	-1.054	-1.349	-1.063	-1.181	-0.944
	SE/F	0.657	0.607	0.691	0.610	0.697	0.566
	P value	0.063	0.084	0.052	0.083	0.091	0.096
Adjusted	Weighted	-1.352	-1.150	-1.430	-1.132	-1.205	-1.016
	SE/F	0.611	0.524	0.636	0.526	0.662	0.495
	P value	0.028	0.029	0.026	0.032	0.070	0.041
MMRM	Weighted	-1.430	-1.188				
	SE	0.606	0.523				
	P value	0.025	0.031				

SE/F: estimated standard error inflated by the F distribution factor

REFERENCES

Doros, G., Pencina, M., Rybin, D., Meisner, A. and Fava, M. (2013) A repeated measures model for analysis of continuous outcomes in sequential parallel comparison design studies. *Statistics in Medicine*, **32**, 2767–2789.

Huang, X. and Tamura, R.N. (2010) Comparison of Test Statistics for the Sequential Parallel Design. *Statistics in Biopharmaceutical Research*, **2**, 42–50.

Jones, B. and Kenward, M.G. (2014) *Design and Analysis of Cross-Over Trials, Third Edition*. CRC Press.

Kawaguchi, A., Koch, G.G. and Ramaswamy, R. (2009) Applications of extensions of bivariate rank sum statistics to the crossover design to compare two treatments through four sequence groups. *Biometrics*, **65**, 979–988.

Koch, G.G., Landis, J.R., Freeman, J.L., Freeman, D.H., Jr., and Lehnen, R.G. (1977). A general methodology for the analysis of experiments with repeated measurements of categorical data. *Biometrics*, 33, 133-158.

Koch, G.G., Tangen, C.M., Jung, J.G. and Amara, I.A. (1998) Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. *Statistics in Medicine*, **17**, 1863–1892.

Lehmacher, W. (1991) Analysis of the crossover design in the presence of residual effects. *Statistics in Medicine*, **10**, 891–899.

Tamura, R.N. and Huang, X. (2007) An examination of the efficiency of the sequential parallel design in psychiatric clinical trials. *Clinical Trials (London, England)*, **4**, 309–317.

CHAPTER 6: RANDOMIZATION-BASED ANCOVA FOR INFERENCE IN TWO-WAY ENRICHMENT DESIGN

6.1 Introduction

This chapter presents extensions of the methods in Chapter 5 to the two-way enrichment design (TED) discussed by Ivanova and Tamura (2011). For this purpose, let $i = 1, 2, 3, 4$ index P:P, P:T, T:P, T:T as the sequence groups for test treatment T and placebo P in the two periods. For this design, treatment comparisons between T and P during the second period are of particular interest for placebo non-responders during the first period (i.e., P:T vs. P:P) and test treatment responders during the first period (i.e., T:T vs. T:P).

6.2 Methods

Among n patients who are eligible for inclusion in the clinical trial, let n_i denote the number of such patients who are randomly assigned to the i-th sequence group. Also, the $n = \sum_{i=1}^{4} n_i$ patients in the clinical trial are assumed to represent an essentially infinite target population in a sense that is conceptually comparable to a simple random sample with replacement. Let $j = 0, 1, 2$ index the baseline, period 1, and period 2 for the clinical trial; and let Y_{ijk} denote the observed random response during the j-th period for the k-th patient in the i-th sequence group according to a non-negative numeric scale. Let Z_{ik} be a dichotomous responder variable for period 1 such that $Z_{ik} = 1$ if the k-th patient in the i-th group has favorable response during the first period in the sense that $(0 \leq Y_{i1k} \leq L)$ versus $Z_{ik} = 0$ if $(Y_{i1k} > L)$; alternatively, the $\{Z_{ik}\}$ could be based on change (or percent change) from baseline. Let $F_{ijk} =$

$Z_{ik}Y_{ijk}$ so as to equal Y_{ijk} for $j = 1, 2$ for responders in period 1 and to equal 0 for non-responders in period 1, and let $G_{ik} = (Y_{i2k} - F_{i2k}) = (1 - Z_{ik})Y_{i2k}$ so as to equal Y_{i2k} for period 2 for period 1 non-responders and to equal 0 for period 1 responders. Let $\boldsymbol{F}_{ik} = (Y_{i0k}, Y_{i1k}, Z_{ik}, F_{i1k}, Y_{i2k}, F_{i2k})'$ with the assumption of no missing values for its components. Also, the $\boldsymbol{F}_{ik}$ could be expanded to include one or more other covariables $\boldsymbol{X}_{i0k}$ at baseline, such as age (in addition to Y_{i0k}); but the presentation is more straightforward without this extension because the same considerations apply to both Y_{i0k} and $\boldsymbol{X}_{i0k}$.

Let $\overline{\boldsymbol{F}}_i = \left(\sum_{k=1}^{n_i} \boldsymbol{F}_{ik}/n_i\right) = (\bar{Y}_{i0}, \bar{Y}_{i1}, \bar{Z}_i, \bar{F}_{i1}, \bar{Y}_{i2}, \bar{F}_{i2})'$ denote the vector of the means of the $\boldsymbol{F}_{ik}$ for the i-th group; and let $\widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}_i}$ denote the unbiased estimate for its covariance matrix in (6.1).

$$\widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}_i} = \sum_{k=1}^{n_i} (\boldsymbol{F}_{ik} - \overline{\boldsymbol{F}}_i)(\boldsymbol{F}_{ik} - \overline{\boldsymbol{F}}_i)'/n_i(n_i - 1) \qquad (6.1)$$

Let $\overline{\boldsymbol{F}} = (\overline{\boldsymbol{F}}_1', \overline{\boldsymbol{F}}_2', \overline{\boldsymbol{F}}_3', \overline{\boldsymbol{F}}_4')'$ and let $\widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}} = \boldsymbol{Diag}\left(\widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}_i}, \widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}_2}, \widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}_3}, \widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}_4}\right)$ denote its block diagonal covariance matrix so as to account for the statistical independence of the $\{\overline{\boldsymbol{F}}_i\}$ and their corresponding estimated covariance matrices $\widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}_i}$.

6.2.1 Estimates for Treatment Comparisons

Let $\widetilde{\boldsymbol{F}}_i$ denote the transformation in (6.2) whereby the $\bar{F}_{ij}$ for $j = 1, 2$ are replaced by

$$\widetilde{\boldsymbol{F}}_i = \left(\bar{Y}_{i0}, \bar{Y}_{i1}, \bar{Z}_i, \tilde{F}_{i1}, \bar{Y}_{i2}, \tilde{F}_{i2}, \tilde{G}_{i2}\right)' \qquad (6.2)$$

$\tilde{F}_{ij} = \left(\bar{F}_{ij}/\bar{Z}_i\right)$ and $\tilde{G}_{i2} = (\bar{Y}_{i2} - \bar{F}_{i2})/(1 - \bar{Z}_i)$ is also included. In order to apply the linear Taylor series methods discussed in Koch et al. (1977) to produce a consistent estimator $\widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}_i}$ for the covariance matrix of $\widetilde{\boldsymbol{F}}_i$, the transformation of $\overline{\boldsymbol{F}}_i$ to $\widetilde{\boldsymbol{F}}_i$ is expressed as in (6.3) with $\boldsymbol{R}_1$, $\boldsymbol{r}$,

$$\widetilde{\boldsymbol{F}}_i = \boldsymbol{exp}[\boldsymbol{R}_2\, \boldsymbol{log}(\boldsymbol{R}_1\overline{\boldsymbol{F}}_i + \boldsymbol{r})] \qquad (6.3)$$

and $\boldsymbol{R}_2$ as shown in (6.4) and with $\boldsymbol{log}$ (and $\boldsymbol{exp}$) being the operation that transforms a vector to

$$\boldsymbol{R}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}, \boldsymbol{r} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{R}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (6.4)$$

the corresponding vector of natural logarithms (and exponentiated values).

It then follows from Koch et al. (1977) that $\widehat{\boldsymbol{V}}_{\bar{\boldsymbol{F}}_i} = \boldsymbol{L}_i \widehat{\boldsymbol{V}}_{\bar{\boldsymbol{F}}_i} \boldsymbol{L}_i'$ for which $\boldsymbol{L}_i = \boldsymbol{D}_{\bar{\boldsymbol{F}}_i} \boldsymbol{R}_2 \boldsymbol{D}^{-1}_{(\boldsymbol{R}_1 \bar{\boldsymbol{F}}_1 + \boldsymbol{r})} \boldsymbol{R}_1$ and $(\boldsymbol{R}_1 \bar{\boldsymbol{F}}_1 + \boldsymbol{r}) = \left(\bar{Y}_{i0}, \bar{Y}_{i1}, \bar{Z}_i, (1 - \bar{Z}_i), \bar{F}_{i1}, \bar{Y}_{i2}, \bar{F}_{i2}, (\bar{Y}_{i2} - \bar{F}_{i2})\right)'$ is a consistent estimator for the covariance matrix of $\widetilde{\boldsymbol{F}}_i$. Accordingly, for $\widetilde{\boldsymbol{F}} = \left(\widetilde{\boldsymbol{F}}_1', \widetilde{\boldsymbol{F}}_2', \widetilde{\boldsymbol{F}}_3', \widetilde{\boldsymbol{F}}_4'\right)'$, the block diagonal matrix $\widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}} = \boldsymbol{Diag}\left(\widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}_1}, \widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}_2}, \widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}_3}, \widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}_4}\right)$ is a consistent estimator for the corresponding covariance matrix.

For treatment comparisons between T and P, the estimators of interest are shown in (6.5),

$$\begin{aligned} c_1 &= \left\{ \frac{n_3 \bar{Y}_{31} + n_4 \bar{Y}_{41}}{(n_3 + n_4)} - \frac{n_1 \bar{Y}_{11} + n_2 \bar{Y}_{21}}{(n_1 + n_2)} \right\} \\ c_2 &= (\bar{Y}_{42} - \bar{Y}_{12}) \\ c_3 &= \left(\tilde{F}_{22} - \tilde{F}_{21}\right) \\ c_4 &= \left(\tilde{G}_{22} - \tilde{G}_{21}\right) \\ c_5 &= \left(\tilde{F}_{42} - \tilde{F}_{32}\right) \\ c_6 &= \left(\tilde{G}_{42} - \tilde{G}_{32}\right) \end{aligned} \quad (6.5)$$

and they are linear functions $\boldsymbol{c} = \boldsymbol{C}\widetilde{\boldsymbol{F}}$ of $\widetilde{\boldsymbol{F}}$ with the matrix $\boldsymbol{C}$ as shown in (6.6) for which $\boldsymbol{\delta}_{7,u}$ is a

$$\boldsymbol{C} = \begin{bmatrix} -p_1 \boldsymbol{\delta}_{7,2}' & -p_2 \boldsymbol{\delta}_{7,2}' & p_3 \boldsymbol{\delta}_{7,2}' & p_4 \boldsymbol{\delta}_{7,2}' \\ -\boldsymbol{\delta}_{7,5}' & \boldsymbol{0}_7' & \boldsymbol{0}_7' & \boldsymbol{\delta}_{7,5}' \\ -\boldsymbol{\delta}_{7,6}' & \boldsymbol{\delta}_{7,6}' & \boldsymbol{0}_7' & \boldsymbol{0}_7' \\ -\boldsymbol{\delta}_{7,7}' & \boldsymbol{\delta}_{7,7}' & \boldsymbol{0}_7' & \boldsymbol{0}_7' \\ \boldsymbol{0}_7' & \boldsymbol{0}_7' & -\boldsymbol{\delta}_{7,6}' & \boldsymbol{\delta}_{7,6}' \\ \boldsymbol{0}_7' & \boldsymbol{0}_7' & -\boldsymbol{\delta}_{7,7}' & \boldsymbol{\delta}_{7,7}' \end{bmatrix} \quad (6.6)$$

(7×1) vector with 1 in the u-th position and $p_1 = n_1/(n_1 + n_2)$, $p_2 = (1 - p_1)$, $p_3 = n_3/(n_3 + n_4)$, $p_4 = (1 - p_3)$. A consistent estimator for the covariance matrix of $\boldsymbol{c}$ is $\widehat{\boldsymbol{V}}_{\boldsymbol{c}} = \boldsymbol{C}\boldsymbol{V}_{\widetilde{\boldsymbol{F}}}\boldsymbol{C}'$. For $\boldsymbol{c} = (c_1, c_2, c_3, c_4, c_5, c_6)'$, the comparison c_1 pertains to T versus P in the first period; the comparison c_2 pertains to T:T versus P:P in period 2; the comparison c_3 pertains to P:T versus P:P in period 2 for responders to placebo in period 1; c_4 pertains to P:T versus P:P in period 2 for non-responders to placebo in period 1; the comparison c_5 pertains to T:T versus T:P in period 2 for responders to T in period 1; and the comparison c_6 pertains to T:T versus T:P for non-responders to T in period 1. Univariate test statistics for the overall comparison between T and P can be based on weighted linear combinations $c_W = \sum_{h=1}^{6} w_h c_h$ where $\boldsymbol{w} = (w_1, w_2, w_3, w_4, w_5, w_6)'$ is a vector of weights such at all $w_h \geq 0$ and $\sum_{h=1}^{6} w_h = 1$. With the weights $\boldsymbol{w}$, the test statistic for the overall null hypothesis $H_{0\boldsymbol{c}}$ for $\boldsymbol{E}_{\boldsymbol{A}}\{\boldsymbol{c}\} = 0$, where $\boldsymbol{E}_{\boldsymbol{A}}\{\ \}$ denotes asymptotic expected value, is $T_{w,\boldsymbol{c}} = \boldsymbol{w}'\boldsymbol{c}/\left(\boldsymbol{w}'\widehat{\boldsymbol{V}}_{\boldsymbol{c}}\boldsymbol{w}\right)^{0.5}$. Under $H_{0\boldsymbol{c}}$, $T_{w,\boldsymbol{c}}$ approximately has the normal distribution with mean 0 and variance 1. A $(1 - \alpha)$ two-sided confidence interval based on c_w can be constructed as $\left[c_w - Z_{\alpha/2}\sqrt{\hat{v}_{c_w}}, c_w + Z_{\alpha/2}\sqrt{\hat{v}_{c_w}}\right]$ where $Z_{\alpha/2}$ is the $(1 - \alpha/2)$ percentile of the standard normal distribution and $\hat{v}_{c_w} = \boldsymbol{w}'\widehat{\boldsymbol{V}}_{c}\boldsymbol{w}$. Since the comparisons of principal interest for the two-way enrichment design are c_1, c_4, and c_5, a specification of equal weight for them and 0 weight for c_2, c_3, and c_6 is $\boldsymbol{w}_3 = (\,(1/3), 0, 0, (1/3), (1/3), 0)$. Alternatively, the use of other weights for other subsets for $\boldsymbol{c}$ is passible, with the scope including both equal weights and inverse covariance matrix weights.

6.2.2 Randomization-Based Covariance Adjusted Estimators

The constraints $\boldsymbol{c}_0 = \boldsymbol{C}_0\widetilde{\boldsymbol{F}}$ with $\boldsymbol{E}_{\boldsymbol{A}}\{\boldsymbol{c}_0\} = \boldsymbol{0}$ regardless of whether the previously noted overall null hypothesis H_0 applies are shown in (6.7) with the matrix $\boldsymbol{C}_0$ shown in (6.8).

$$\boldsymbol{c}_0 = \begin{bmatrix} (\bar{Y}_{20} - \bar{Y}_{10}) \\ (\bar{Y}_{30} - \bar{Y}_{10}) \\ (\bar{Y}_{40} - \bar{Y}_{10}) \\ (\bar{Y}_{21} - \bar{Y}_{11}) \\ (\bar{Y}_{41} - \bar{Y}_{31}) \\ (\bar{Z}_2 - \bar{Z}_1) \\ (\bar{Z}_4 - \bar{Z}_3) \\ (\tilde{F}_{21} - \tilde{F}_{11}) \\ (\tilde{F}_{41} - \tilde{F}_{31}) \end{bmatrix} \qquad (6.7)$$

$$\boldsymbol{C}_0 = \begin{bmatrix} -\boldsymbol{\delta}'_{7,1} & \boldsymbol{\delta}'_{7,1} & \mathbf{0}'_7 & \mathbf{0}'_7 \\ -\boldsymbol{\delta}'_{7,1} & \mathbf{0}'_7 & \boldsymbol{\delta}'_{7,1} & \mathbf{0}'_7 \\ -\boldsymbol{\delta}'_{7,1} & \mathbf{0}'_7 & \mathbf{0}'_7 & \boldsymbol{\delta}'_{7,1} \\ -\boldsymbol{\delta}'_{7,2} & \boldsymbol{\delta}'_{7,2} & \mathbf{0}'_7 & \mathbf{0}'_7 \\ \mathbf{0}'_7 & \mathbf{0}'_7 & -\boldsymbol{\delta}'_{7,2} & \boldsymbol{\delta}'_{7,2} \\ -\boldsymbol{\delta}'_{7,3} & \boldsymbol{\delta}'_{7,3} & \mathbf{0}'_7 & \mathbf{0}'_7 \\ \mathbf{0}'_7 & \mathbf{0}'_7 & -\boldsymbol{\delta}'_{7,3} & \boldsymbol{\delta}'_{7,3} \\ -\boldsymbol{\delta}'_{7,4} & \boldsymbol{\delta}'_{7,4} & \mathbf{0}'_7 & \mathbf{0}'_7 \\ \mathbf{0}'_7 & \mathbf{0}'_7 & -\boldsymbol{\delta}'_{7,4} & \boldsymbol{\delta}'_{7,4} \end{bmatrix} \qquad (6.8)$$

Let $\boldsymbol{c}_+ = [\boldsymbol{c}', \boldsymbol{c}'_0] = [\boldsymbol{C}', \boldsymbol{C}'_0]' = \boldsymbol{C}_+\tilde{\boldsymbol{F}}$ denote the combined set of comparisons $\boldsymbol{c}$ pertaining to H_0 and constraints $\boldsymbol{c}_0$. The estimated covariance structure for $\boldsymbol{c}_+$ is $\widehat{\boldsymbol{V}}_{\boldsymbol{c}_+} = \boldsymbol{C}_+\widehat{\boldsymbol{V}}_{\tilde{\boldsymbol{F}}}\boldsymbol{C}_+$. Since $\boldsymbol{E}_{\boldsymbol{A}}\{\boldsymbol{c}_0\} = \mathbf{0}$ regardless of whether H_0 applies, randomization-based covariance adjustment for $\boldsymbol{c}$ with respect to the constraints $\boldsymbol{c}_0$ is invoked by fitting the linear model in (6.9) by weighted least

$$\boldsymbol{E}_{\boldsymbol{A}}\{\boldsymbol{c}_+\} = \begin{bmatrix} \boldsymbol{I}_6 \\ \mathbf{0}_{9,6} \end{bmatrix} \boldsymbol{b} = \boldsymbol{A}\boldsymbol{b} \qquad (6.9)$$

squares with weights based on $\widehat{\boldsymbol{V}}_{\boldsymbol{c}_+}$ and with "≙" meaning "is estimated by".

Accordingly, $\boldsymbol{b} = (b_1, b_2, b_3, b_4, b_5, b_6)'$ are covariance adjusted counterparts of $\boldsymbol{c}$. More specifically, $\boldsymbol{b} = \left(\boldsymbol{A}'\widehat{\boldsymbol{V}}_{\boldsymbol{c}_+}^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}'\widehat{\boldsymbol{V}}_{\boldsymbol{c}_+}^{-1}\boldsymbol{c}_+$. Also, a consistent estimator for the covariance matrix of $\boldsymbol{b}$ is $\widehat{\boldsymbol{V}}_{\boldsymbol{b}} = \left(\boldsymbol{A}'\widehat{\boldsymbol{V}}_{\boldsymbol{c}_+}^{-1}\boldsymbol{A}\right)^{-1}$. Covariance adjusted test statistics for H_0 can be based on weighted linear combinations $\boldsymbol{b}_w = \sum_{h=1}^{6} w_h b_h$ in ways similar to those discussed for $\boldsymbol{c}$ in Section 2. In

this regard, test statistics based on $\boldsymbol{b}_w$ can have better power than those based on $\boldsymbol{c}_w$ because of their smaller variance via the structure shown in (6.10) for $\widehat{\boldsymbol{V}}_{\boldsymbol{b}}$.

$$\widehat{\boldsymbol{V}}_{\boldsymbol{b}} = \widehat{\boldsymbol{V}}_{\boldsymbol{c}} - \left(\boldsymbol{C}\widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}}\boldsymbol{C}_0'\right)\left(\boldsymbol{C}_0\widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}}\boldsymbol{C}_0'\right)^{-1}\left(\boldsymbol{C}_0\widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}}\boldsymbol{C}'\right) \qquad (6.10)$$

REFERENCES

Doros, G., Pencina, M., Rybin, D., Meisner, A. and Fava, M. (2013) A repeated measures model for analysis of continuous outcomes in sequential parallel comparison design studies. *Statistics in Medicine*, **32**, 2767–2789.

Huang, X. and Tamura, R.N. (2010) Comparison of Test Statistics for the Sequential Parallel Design. *Statistics in Biopharmaceutical Research*, **2**, 42–50.

Jones, B. and Kenward, M.G. (2014) *Design and Analysis of Cross-Over Trials, Third Edition*. CRC Press.

Kawaguchi, A., Koch, G.G. and Ramaswamy, R. (2009) Applications of extensions of bivariate rank sum statistics to the crossover design to compare two treatments through four sequence groups. *Biometrics*, **65**, 979–988.

Koch, G.G., Landis, J.R., Freeman, J.L., Freeman, D.H., Jr., and Lehnen, R.G. (1977). A general methodology for the analysis of experiments with repeated measurements of categorical data. *Biometrics*, 33, 133-158.

Koch, G.G., Tangen, C.M., Jung, J.G. and Amara, I.A. (1998) Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. *Statistics in Medicine*, **17**, 1863–1892.

Lehmacher, W. (1991) Analysis of the crossover design in the presence of residual effects. *Statistics in Medicine*, **10**, 891–899.

Tamura, R.N. and Huang, X. (2007) An examination of the efficiency of the sequential parallel design in psychiatric clinical trials. *Clinical Trials (London, England)*, **4**, 309–317.

CHAPTER 7: RANDOMIZATION-BASED ANCOVA FOR INFERENCE IN BILATERAL DESIGN

7.1 Introduction

This chapter presents adaptions of the methods in Chapter 6 to the bilateral design in clinical settings such as dermatology whereby patients receive alternative treatments on opposite sides of their bodies. For this purpose, let $i = 1, 2, 3, 4$ index P:P, P:T, T:P, T:T as the groups for test treatment T and placebo P on the left and right sides. For this design, treatment comparisons between T and P on one side are of particular interest for placebo non-responders on the opposite side and for test treatment responders on the opposite side.

7.2 Methods

Among n patients who are eligible for inclusion in the clinical trial, let n_i denote the number of such patients who are randomly assigned to the i-th group. Also, the $n = \sum_{i=1}^{4} n_i$ patients in the clinical trial are assumed to represent an essentially infinite target population in a sense that is conceptually comparable to a simple random sample with replacement. Let $j = 1, 2$ index the left and right sides for a patient; and let Y_{ijk} denote the observed random response for the j-th side of the k-th patient in the i-th group according to a non-negative numerical scale; and let Y_{ijk0} denote the baseline counterpart of Y_{ijk}. Let Z_{ijk} be a dichotomous responder variable for the j-th side of the k-th patient in the i-th group such that $Z_{ijk} = 1$ corresponds to favorable response in the sense that $\left(0 \leq Y_{ijk} \leq L\right)$ versus $Z_{ijk} = 0$ if $\left(Y_{ijk} > 0\right)$; alternatively, the Z_{ijk} could be based on change (or percent change) from baseline. Let $F_{ijk} = Z_{ij'k} Y_{ijk}$ for $j' \neq j$ so as

to equal Y_{ijk} for responders on the j'-th side and to equal 0 for non-responders on the j'-th side; and let $G_{ijk} = \left(Y_{ijk} - F_{ijk}\right) = \left(1 - Z_{ij'k}\right)Y_{ijk}$ so as to equal Y_{ijk} for non-responders on the j'-th side and to equal 0 for responders on the j'-th side. Let $\boldsymbol{F}_{ik} = (Y_{i1k0}, Y_{i1k}, Z_{i1k}, F_{i1k}, Y_{i2k0}, Y_{i2k}, Z_{i2k}, F_{i2k})'$ with the assumption of no missing values for its components. Also, the $\boldsymbol{F}_{ik}$ could be expanded to include one or more covariables $\boldsymbol{X}_{i0k}$ at baseline, such as age (in addition to Y_{ijk0}); but the presentation is more straightforward without this extension because the same considerations apply to both the Y_{ijk0} and $\boldsymbol{X}_{i0k}$.

Let $\overline{\boldsymbol{F}}_i = \left(\sum_{k=1}^{n_i} \boldsymbol{F}_{ik}\right) = (\bar{Y}_{i1*0}, \bar{Y}_{i1}, \bar{Z}_{i1}, \bar{F}_{i1}, \bar{Y}_{i2*0}, \bar{Y}_{i2}, \bar{Z}_{i2}, \bar{F}_{i2})'$ denote the vector of means of the $\boldsymbol{F}_{ik}$ for the i-th group; and let $\widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}_i}$ denote the unbiased estimator for its covariance matrix in (7.1).

$$\widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}_i} = \sum_{k=1}^{n_i} (\boldsymbol{F}_{ik} - \overline{\boldsymbol{F}}_i)(\boldsymbol{F}_{ik} - \overline{\boldsymbol{F}}_i)'/n_i(n_i - 1) \qquad (7.1)$$

Let $\overline{\boldsymbol{F}} = (\overline{\boldsymbol{F}}'_1, \overline{\boldsymbol{F}}'_2, \overline{\boldsymbol{F}}'_3, \overline{\boldsymbol{F}}'_4)'$ and let $\widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}} = \boldsymbol{Diag}\left(\widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}_1}, \widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}_2}, \widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}_3}, \widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}_4}\right)$ denote its block diagonal estimated covariance matrix so as to account for the statistical independence of the $\{\overline{\boldsymbol{F}}_i\}$ and their corresponding estimated covariance matrices $\widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}_i}$.

7.2.1 Estimates for Treatment Comparisons

Let $\widetilde{\boldsymbol{F}}_i$ denote the transformation in (7.2) whereby the $\bar{F}_{ij}$ are replaced by

$$\widetilde{\boldsymbol{F}}_i = \left(\bar{Y}_{i1*0}, \bar{Y}_{i1}, \bar{Z}_{i1}, \tilde{F}_{i1}, \tilde{G}_{i1,}\bar{Y}_{i2*0}, \bar{Y}_{i2}, \bar{Z}_{i2}, \tilde{F}_{i2}, \tilde{G}_{i2}\right)' \qquad (7.2)$$

$\tilde{F}_{ij} = \left(\bar{F}_{ij}/\bar{Z}_{ij'}\right)$ and the $\tilde{G}_{ij} = \left(\bar{Y}_{ij} - \bar{F}_{ij}\right)/\left(1 - \bar{Z}_{ij'}\right)$ are also included. In order to apply the linear Taylor series methods discussed in Koch et al. (1977) to produce a consistent estimator $\widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}_i}$ for the covariance matrix of $\widetilde{\boldsymbol{F}}_i$, the transformation of $\overline{\boldsymbol{F}}_i$ to $\widetilde{\boldsymbol{F}}_i$ is expressed as in (7.3) with

$$\widetilde{\boldsymbol{F}}_i = \boldsymbol{exp}[\boldsymbol{R}_2\, \boldsymbol{log}(\boldsymbol{R}_1\overline{\boldsymbol{F}}_i + \boldsymbol{r})] \qquad (7.3)$$

$\boldsymbol{R}_1$, $\boldsymbol{r}$, and $\boldsymbol{R}_2$ as shown in (7.4) and with $\boldsymbol{log}$ (and $\boldsymbol{exp}$) being the operation that transforms a

$$\boldsymbol{R}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \end{bmatrix}, \boldsymbol{r} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{R}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (7.4)$$

vector to the corresponding vector of natural logarithms (and exponentiated values). It then follows that from Koch et al. (1977) that $\widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}_i} = \boldsymbol{L}_i \widehat{\boldsymbol{V}}_{\overline{\boldsymbol{F}}_i} \boldsymbol{L}_i'$ for which $\boldsymbol{L}_i = \boldsymbol{D}_{\widetilde{\boldsymbol{F}}_i} \boldsymbol{R}_2 \boldsymbol{D}^{-1}_{(\boldsymbol{R}_1 \overline{\boldsymbol{F}}_1 + \boldsymbol{r})} \boldsymbol{R}_1$ and $(\boldsymbol{R}_1 \overline{\boldsymbol{F}}_1 + \boldsymbol{r}) = \big(\bar{Y}_{i1*0}, \bar{Y}_{i1}, \bar{Z}_{i1}, (1 - \bar{Z}_{i1}), \bar{F}_{i1}, (\bar{Y}_{i1} - \bar{F}_{i1}), \bar{Y}_{i2*0}, \bar{Y}_{i2}, \bar{Z}_{i2}, (1 - \bar{Z}_{i2}), \bar{F}_{i2}, (\bar{Y}_{i2} - \bar{F}_{i2}) \big)'$ is a consistent estimator for the covariance matrix of $\widetilde{\boldsymbol{F}}_i$. Accordingly, for $\widetilde{\boldsymbol{F}} = \big(\widetilde{\boldsymbol{F}}_1', \widetilde{\boldsymbol{F}}_2', \widetilde{\boldsymbol{F}}_3', \widetilde{\boldsymbol{F}}_4' \big)'$, the block diagonal matrix $\widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}} = \boldsymbol{Diag}\big(\widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}_1}, \widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}_2}, \widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}_3}, \widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}_4} \big)$ is a consistent estimator for the corresponding covariance matrix.

For treatment comparisons between T and P, the estimators of interest are shown in (7.5),

$$\boldsymbol{c} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \\ c_7 \\ c_8 \end{bmatrix} = \begin{bmatrix} \big(\tilde{F}_{31} - \tilde{F}_{11} \big) \\ \big(\tilde{G}_{31} - \tilde{G}_{11} \big) \\ \big(\tilde{F}_{41} - \tilde{F}_{21} \big) \\ \big(\tilde{G}_{41} - \tilde{G}_{21} \big) \\ \big(\tilde{F}_{22} - \tilde{F}_{12} \big) \\ \big(\tilde{G}_{22} - \tilde{G}_{12} \big) \\ \big(\tilde{F}_{42} - \tilde{F}_{32} \big) \\ \big(\tilde{G}_{42} - \tilde{G}_{32} \big) \end{bmatrix} \quad (7.5)$$

and they are linear functions $\boldsymbol{c} = \boldsymbol{C}\widetilde{\boldsymbol{F}}$ of $\widetilde{\boldsymbol{F}}$ with the matrix $\boldsymbol{C}$ as shown in (7.6) where $\boldsymbol{\delta}_{10,u}$ is a

$$\boldsymbol{C} = \begin{bmatrix} -\boldsymbol{\delta}'_{10,4} & \mathbf{0}'_{10} & \boldsymbol{\delta}'_{10,4} & \mathbf{0}'_{10} \\ -\boldsymbol{\delta}'_{10,5} & \mathbf{0}'_{10} & \boldsymbol{\delta}'_{10,5} & \mathbf{0}'_{10} \\ \mathbf{0}'_{10} & -\boldsymbol{\delta}'_{10,4} & \mathbf{0}'_{10} & \boldsymbol{\delta}'_{10,4} \\ \mathbf{0}'_{10} & -\boldsymbol{\delta}'_{10,5} & \mathbf{0}'_{10} & \boldsymbol{\delta}'_{10,5} \\ -\boldsymbol{\delta}'_{10,9} & \boldsymbol{\delta}'_{10,9} & \mathbf{0}'_{10} & \mathbf{0}'_{10} \\ -\boldsymbol{\delta}'_{10,10} & \boldsymbol{\delta}'_{10,10} & \mathbf{0}'_{10} & \mathbf{0}'_{10} \\ \mathbf{0}'_{10} & \mathbf{0}'_{10} & -\boldsymbol{\delta}'_{10,9} & \boldsymbol{\delta}'_{10,9} \\ \mathbf{0}'_{10} & \mathbf{0}'_{10} & -\boldsymbol{\delta}'_{10,10} & \boldsymbol{\delta}'_{10,10} \end{bmatrix} \qquad (7.6)$$

(10×1) vector with a 1 in the u-th position.

A consistent estimate for the covariance matrix of $\boldsymbol{c}$ is $\widehat{\boldsymbol{V}}_{\boldsymbol{c}} = \boldsymbol{C}\boldsymbol{V}_{\widetilde{\boldsymbol{F}}}\boldsymbol{C}'$. For $\boldsymbol{c} = (c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8)'$, the comparison c_1 and c_2 respectively pertain to T versus P on the left for responders and non-responders to P on the right; the comparisons c_5 and c_6 respectively pertain to T versus P on the right for responders and non-responders to P on the left; the comparisons c_3 and c_4 respectively pertain to T versus P on the left for responders and non-responders to T on the right; and the comparisons c_7 and c_8 respectively pertain to T versus P on the right for responders and non-responders to T on the left. Univariate test statistics for the overall comparison between T and P can be based on weighted linear combinations $c_w = \sum_{h=1}^{8} w_h c_h$ where $\boldsymbol{w} = (w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8)'$ is a vector of weights such at all $w_h \geq 0$ and $\sum_{h=1}^{8} w_h = 1$. With the weights $\boldsymbol{w}$, the test statistic for the overall null hypothesis $H_{0\boldsymbol{c}}$ for $\boldsymbol{E}_{\boldsymbol{A}}\{\boldsymbol{c}\} = 0$, where $\boldsymbol{E}_{\boldsymbol{A}}\{\ \}$ denotes asymptotic expected value, is $T_{w,\boldsymbol{c}} = \boldsymbol{w}'\boldsymbol{c}/\left(\boldsymbol{w}'\widehat{\boldsymbol{V}}_{\boldsymbol{c}}\boldsymbol{w}\right)^{0.5}$. Under $H_{0\boldsymbol{c}}$, $T_{w,\boldsymbol{c}}$ approximately has the normal distribution with mean 0 and variance 1. A $(1-\alpha)$ two-sided confidence interval based on c_w can be constructed as $\left[c_w - Z_{\alpha/2}\sqrt{\hat{v}_{c_w}}, c_w + Z_{\alpha/2}\sqrt{\hat{v}_{c_w}}\right]$ where $Z_{\alpha/2}$ is the $(1-\alpha/2)$ percentile of the standard normal distribution and $\hat{v}_{c_w} = \boldsymbol{w}'\widehat{\boldsymbol{V}}_{c_w}\boldsymbol{w}$. Since the comparisons of principal interest for the bilateral design are c_2, c_3, c_6, and c_7, a specification of equal weight for them and 0 weight for c_1, c_4, c_5, and c_8 is $\boldsymbol{w}_4 =$

$(0, 0.25, 0.25, 0, 0, 0.25, 0.25, 0)$. Alternatively, the use of other weights for other subsets for $\boldsymbol{c}$ is passible, with the scope including both equal weights and inverse covariance matrix weights.

7.2.2 Randomization-Based Covariance Adjusted Estimators

The constraints $\boldsymbol{c}_0 = \boldsymbol{C}_0\widetilde{\boldsymbol{F}}$ with $\boldsymbol{E}_{\boldsymbol{A}}\{\boldsymbol{c}_0\} = \boldsymbol{0}$ regardless of whether the previously noted overall null hypothesis H_0 applies are shown in (7.7) with the matrix $\boldsymbol{C}_0$ shown in (7.8), although

$$\boldsymbol{c}_0 = \begin{bmatrix} (\bar{Y}_{21*0} - \bar{Y}_{11*0}) \\ (\bar{Y}_{31*0} - \bar{Y}_{11*0}) \\ (\bar{Y}_{41*0} - \bar{Y}_{11*0}) \\ (\bar{Y}_{21} - \bar{Y}_{11}) \\ (\bar{Z}_{21} - \bar{Z}_{11}) \\ (\bar{Y}_{41} - \bar{Y}_{31}) \\ (\bar{Z}_{41} - \bar{Z}_{31}) \\ (\bar{Y}_{22*0} - \bar{Y}_{12*0}) \\ (\bar{Y}_{32*0} - \bar{Y}_{12*0}) \\ (\bar{Y}_{42*0} - \bar{Y}_{12*0}) \\ (\bar{Y}_{32} - \bar{Y}_{12}) \\ (\bar{Z}_{32} - \bar{Z}_{12}) \\ (\bar{Y}_{42} - \bar{Y}_{22}) \\ (\bar{Z}_{42} - \bar{Z}_{22}) \end{bmatrix} \tag{7.7}$$

$$\boldsymbol{C}_0 = \begin{bmatrix} -\boldsymbol{\delta}'_{10,1} & \boldsymbol{\delta}'_{10,1} & \boldsymbol{0}'_{10} & \boldsymbol{0}'_{10} \\ -\boldsymbol{\delta}'_{10,1} & \boldsymbol{0}'_{10} & \boldsymbol{\delta}'_{10,1} & \boldsymbol{0}'_{10} \\ -\boldsymbol{\delta}'_{10,1} & \boldsymbol{0}'_{10} & \boldsymbol{0}'_{10} & \boldsymbol{\delta}'_{10,1} \\ -\boldsymbol{\delta}'_{10,2} & \boldsymbol{\delta}'_{10,2} & \boldsymbol{0}'_{10} & \boldsymbol{0}'_{10} \\ -\boldsymbol{\delta}'_{10,3} & \boldsymbol{\delta}'_{10,3} & \boldsymbol{0}'_{10} & \boldsymbol{0}'_{10} \\ \boldsymbol{0}'_{10} & \boldsymbol{0}'_{10} & -\boldsymbol{\delta}'_{10,2} & \boldsymbol{\delta}'_{10,2} \\ \boldsymbol{0}'_{10} & \boldsymbol{0}'_{10} & -\boldsymbol{\delta}'_{10,3} & \boldsymbol{\delta}'_{10,3} \\ -\boldsymbol{\delta}'_{10,6} & \boldsymbol{\delta}'_{10,6} & \boldsymbol{0}'_{10} & \boldsymbol{0}'_{10} \\ -\boldsymbol{\delta}'_{10,6} & \boldsymbol{0}'_{10} & \boldsymbol{\delta}'_{10,6} & \boldsymbol{0}'_{10} \\ -\boldsymbol{\delta}'_{10,6} & \boldsymbol{0}'_{10} & \boldsymbol{0}'_{10} & \boldsymbol{\delta}'_{10,6} \\ -\boldsymbol{\delta}'_{10,7} & \boldsymbol{0}'_{10} & \boldsymbol{\delta}'_{10,7} & \boldsymbol{0}'_{10} \\ -\boldsymbol{\delta}'_{10,8} & \boldsymbol{0}'_{10} & \boldsymbol{\delta}'_{10,8} & \boldsymbol{0}'_{10} \\ \boldsymbol{0}'_{10} & -\boldsymbol{\delta}'_{10,7} & \boldsymbol{0}'_{10} & \boldsymbol{\delta}'_{10,7} \\ \boldsymbol{0}'_{10} & -\boldsymbol{\delta}'_{10,8} & \boldsymbol{0}'_{10} & \boldsymbol{\delta}'_{10,8} \end{bmatrix} \tag{7.8}$$

this specification does assume that the treatment on the left side does not affect the response on the right side, and vice versa. Let $\boldsymbol{c}_+ = [\boldsymbol{c}', \boldsymbol{c}_0'] = [\boldsymbol{C}', \boldsymbol{C}_0']'\widetilde{\boldsymbol{F}} = \boldsymbol{C}_+\widetilde{\boldsymbol{F}}$ denote the combined set of comparisons $\boldsymbol{c}$ pertaining to H_0 and constraints $\boldsymbol{c}_0$. The estimated covariance structure for $\boldsymbol{c}_+$ is $\widehat{\boldsymbol{V}}_{\boldsymbol{c}_+} = \boldsymbol{C}_+\widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}}\boldsymbol{C}_+$. Since $\boldsymbol{E}_{\boldsymbol{A}}\{\boldsymbol{c}_0\} = \boldsymbol{0}$, randomization-based covariance adjustment for $\boldsymbol{c}$ with respect to the constraints $\boldsymbol{c}_0$ is invoked by fitting the linear model in (7.9) by weighted least

$$\boldsymbol{E}_{\boldsymbol{A}}\{\boldsymbol{c}_+\} = \begin{bmatrix} \boldsymbol{I}_8 \\ \boldsymbol{0}_{14,8} \end{bmatrix}\boldsymbol{b} = \boldsymbol{A}\boldsymbol{b} \qquad (7.9)$$

squares with weights based on $\widehat{\boldsymbol{V}}_{\boldsymbol{c}_+}^{-1}$ and with "≙" meaning "is estimated by".

Accordingly, $\boldsymbol{b} = (b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8)'$ are covariance adjusted counterparts of $\boldsymbol{c}$. More specifically, $\boldsymbol{b} = \left(\boldsymbol{A}'\widehat{\boldsymbol{V}}_{\boldsymbol{c}_+}^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}'\widehat{\boldsymbol{V}}_{\boldsymbol{c}_+}^{-1}\boldsymbol{c}_+$. Also, a consistent estimator for the covariance matrix of $\boldsymbol{b}$ is $\widehat{\boldsymbol{V}}_{\boldsymbol{b}} = \left(\boldsymbol{A}'\widehat{\boldsymbol{V}}_{\boldsymbol{c}_+}^{-1}\boldsymbol{A}\right)^{-1}$. Covariance adjusted test statistics for H_0 can be based on weighted linear combinations $\boldsymbol{b}_w = \sum_{h=1}^{8} w_h b_h$ in ways similar to those discussed for $\boldsymbol{c}$ in Section 7.2. In this regard, test statistics based on $\boldsymbol{b}_w$ can have better power than those based on $\boldsymbol{c}_w$ because of their smaller variance via the structure shown in (7.10) for $\widehat{\boldsymbol{V}}_{\boldsymbol{b}}$ (Kawaguchi et al., 2009).

$$\widehat{\boldsymbol{V}}_{\boldsymbol{b}} = \widehat{\boldsymbol{V}}_{\boldsymbol{c}} - \left(\boldsymbol{C}\widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}}\boldsymbol{C}_0'\right)\left(\boldsymbol{C}_0\widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}}\boldsymbol{C}_0'\right)^{-1}\left(\boldsymbol{C}_0\widehat{\boldsymbol{V}}_{\widetilde{\boldsymbol{F}}}\boldsymbol{C}'\right) \qquad (7.10)$$

REFERENCES

Kawaguchi, A., Koch, G.G. and Ramaswamy, R. (2009) Applications of extensions of bivariate rank sum statistics to the crossover design to compare two treatments through four sequence groups. *Biometrics*, **65**, 979–988.

Koch, G.G., Landis, J.R., Freeman, J.L., Freeman, D.H., Jr., and Lehnen, R.G. (1977). A general methodology for the analysis of experiments with repeated measurements of categorical data. *Biometrics*, 33, 133-158.

Koch, G.G., Tangen, C.M., Jung, J.G. and Amara, I.A. (1998) Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. *Statistics in Medicine*, **17**, 1863–1892.