



Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations

Graduate School

2006

Design and Analysis Methods for Cluster Randomized Trials with Pair-Matching on Baseline Outcome: Reduction of Treatment Effect Variance

Misook Park
Virginia Commonwealth University

Follow this and additional works at: <https://scholarscompass.vcu.edu/etd>



Part of the [Biostatistics Commons](#)

© The Author

Downloaded from

<https://scholarscompass.vcu.edu/etd/1535>

This Dissertation is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

© Misook Park 2006

All Rights Reserved

DESIGN AND ANALYSIS METHODS FOR CLUSTER RANDOMIZED TRIALS
WITH PAIR-MATCHING ON BASELINE OUTCOME: REDUCTION OF
TREATMENT EFFECT VARIANCE

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor
of Philosophy at Virginia Commonwealth University.

by

Misook Park

B.S. YonSei University, Seoul, Korea, 1979
M.S. in University of North Carolina ,1989
Chapel Hill, NC

Director:

Robert E. Johnson

Associate Professor, Department of Biostatistics

Virginia Commonwealth University
Richmond, VA
May, 2006

Acknowledgement

This research would not be possible without the painstaking guidance, patience, understanding and continual encouragement of my advisor, Dr. Robert E Johnson. He believed in me enough to place me as one of his students and taught me to love to learn and helped me grow into a fine statistician. I will always remember his dedicated mentoring as well as his friendship. It has been a pleasure to work with him.

I also want to thank the members of my committee for their thoughtful insights and encouragement. Drs. Viswanathan Ramakrishnan, Donna K. McClish for helpful suggestions and critiques to improve the quality of this research. They have been a positive source of knowledge and support. Dr. D'Arcy P. Mays for his taking time to sending me encouraging messages from time to time and thorough reading of the manuscripts for the valuable comments. Dr. Stephen F. Rothemich MD for his insightful input as a medical researcher and providing me a motivation for this work. Dr. Gennings, for her candid advice and being a good example as a successful woman statistician in the competitive professional world. Russ Boyle for giving me a chance to be a student in this program and thoughtful support throughout this whole process.

Special thanks go to my friends in the Academic Technology, Joanne Einsmann, Margie Rigsbee, Rick Raughton, Paula Freeman and Barbara Robison. Without your love and support, I would not be where I am today.

Without the love, support and constant encouragement of faculty, Yvonne and Denise, my fellow students as well as friends currently and previously in BIOS at VCU, my road to travel would have been much more difficult. I will miss them greatly after I leave school.

And the last and always first, I would like to acknowledge and thank the very special person in my life who made all my success possible, my husband, Il Hyun Park. His unselfishness and sacrifices he has made over the past 26 years in support of the fulfillment of my dream is appreciated beyond words. It also goes to my wonderful two children, Christopher (Jun Young) and Sun Young for their love and support through the fun and the not-so fun times. They gave me the motivation to succeed and the desire to set positive examples for them to see. Finally I must acknowledge the mercies provided by the loving God who has guided me and has opened up the numerous opportunities to meet wonderful people throughout my life.

Table of Contents

	Page
List of Tables.....	x
List of Figures.....	xi
Abstract.....	xii
 Chapter I	
Introduction.....	1
 Chapter II Background and Significance of Cluster Randomized Trials	8
2.1 Introduction and Examples.....	8
2.2 Background.....	13
2.3 Significance.....	20
 Chapter III Review of Relevant Statistical Topics	22
3.1 Introduction.....	22
3.2 Variance-Covariance Structure of Linear Models.....	23
3.2.1 Mean and Variance-Covariance of Random Vectors.....	23

3.2.2	Conditional Mean and Variance of Multivariate Normal Vector.....	24
3.3	Marginal and Conditional Moments of a Random Variable.....	25
3.3.1	Marginal Variance.....	25
3.3.2	Conditional Mean and Variance.....	25
3.3.3	Marginal Mean and Variance from Conditional Mean and Variance.....	25
3.4	Quadratic Forms and Relevant Distributions.....	26
3.4.1	Distributions of Quadratic Forms.....	26
3.4.2	Independence of Quadratic Forms.....	27
3.4.3	Noncentral χ^2 Distribution.....	28
3.4.4	Noncentral \mathcal{F} Distribution.....	28
3.5	Order Statistics.....	29
3.5.1	Distribution of Order Statistics.....	30
3.5.2	Asymptotic Distribution.....	30
3.5.3	Induced (Concomitant) Order Statistics.....	33
3.5.4	Standard Normal Order Statistics.....	34
3.5.5	Mean and Covariance of Ordered and Induced Ordered Normal Statistics.....	36
3.6	The Contrast of Treatment Effect Within a Block.....	37
3.7	Methods for Simulation.....	38

3.7.1	Simulation of a Two-Stage Sample from a Multivariate Normal Distribution.....	38
3.7.2	Verification of Variance Calculations.....	40
3.7.3	Power Calculations and Verification.....	40
3.8	Computing the Power.....	41
Chapter IV	Randomized Controlled Trials: 6-Case Scenarios.....	42
4.1	Introduction.....	42
4.2	Methods.....	45
4.2.1	Treatment Effect.....	45
4.2.2	Variance of the Treatment Effect.....	47
4.2.3	Methods for Significance Testing.....	49
4.2.4	Approximations to the F Distribution.....	50
4.3	Case Scenarios.....	53
4.3.1	Case 1: Posttest Data With No Matching.....	53
4.3.1.1	Variance of Treatment Effect.....	54
4.3.1.2	Inference on Treatment Effect.....	55
4.3.2	Case 2: Posttest Data, Baseline as Covariable, No Matching.....	56
4.3.2.1	Variance of Treatment Effect.....	58

4.3.2.2	Inference on Treatment Effect.....	59
4.3.3	Case 3: Pre-Post Difference, No Matching.....	60
4.3.3.1	Variance of Treatment Effect.....	61
4.3.3.2	Inference on Treatment Effect.....	61
4.3.4	Case 4: Posttest Data With Matching.....	62
4.3.4.1	Variance of Treatment Effect.....	64
4.3.4.2	Inference on Treatment Effect.....	66
4.3.5	Case 5: Posttest Data, Baseline as Covariable, With Matching	68
4.3.5.1	Variance of Treatment Effect.....	68
4.3.5.2	Inference on Treatment Effect.....	70
4.3.6	Case 6: Pre-Post Differences With Matching.....	71
4.3.6.1	Variance of Treatment Effect.....	72
4.3.6.2	Inference on Treatment Effect.....	74
Chapter V	Cluster Randomized Trial: 6-Case Scenarios.....	76
5.1	Introduction.....	76
5.2	Cluster-Randomized Pre-Post Design.....	76
5.3	Variance of Treatment Effect for the Six Cases.....	81
5.3.1	Case 1: Posttest Data With No Matching.....	81

5.3.2	Case 2: Posttest Data, Baseline as Covariable, No Matching.....	81
5.3.3	Case 3: Pre-Post Difference, No Matching.....	81
5.3.4	Case 4: Posttest Data With Matching.....	82
5.3.5	Case 5: Posttest Data, Baseline as Covariable, With Matching.....	82
5.3.6	Case 6: Pre-Post Differences With Matching.....	83
Chapter VI	Results.....	85
6.1	Introduction.....	85
6.2	Comparison of Cases 1, 2, and 3.....	86
6.2.1	Comparison of Variances between Cases 1, 2 and 3.....	88
6.2.1.1	Case 1 versus Case 2.....	89
6.2.1.2	Case 1 versus Case 3.....	89
6.2.1.3	Case 2 versus Case 3.....	90
6.2.2	Discussion.....	90
6.2.3	Preferred Analysis from the Design Stage.....	91
6.2.3.1	Case 1 versus Case 2.....	93
6.2.3.2	Case 1 versus Case 3.....	94
6.2.4	Summary.....	95

6.3	Comparison of Variances between Cases With and Without Matching	95
6.3.1	Post data only: Case1 and Case 4.....	97
6.3.2	Post Data with Baseline as Covariable: Case 2 and Case 5.....	98
6.3.3	Pre-post differences: Case 3 and Case 6.....	99
6.3.4	Discussion of Variance Comparisons.....	100
6.4	Power of test.....	101
Chapter VII	Conclusions and Discussion	109
7.1	Summary of Work.....	109
7.2	Contributions.....	111
7.3	Recommendations.....	112
7.3.1	Recommendations at Design Phase.....	112
7.3.2	Recommendations Post Data Collection.....	113
7.3.2.1	Data Collection Without Pair-Matching.....	114
7.3.2.2	Data Collection With Pair-Matching.....	114
7.4	Limitations.....	115

Chapter VIII	Future work	116
8.1	Introduction.....	116
8.2	General Grouping Methods to Produce Homogeneous Groups.....	116
8.3	Generalized Error Structure.....	117
8.4	Hierarchical or Deeply Nested Designs.....	118
8.5	Matching on Multiple Baseline Measures.....	119
8.6	Repeated Measures on Clusters.....	119
8.7	Optimal Allocation of Sample Sizes Across Time Points.....	120
References	121
Appendices.....		125
A	SAS Programs.....	125

List of Tables

	Page
Table 6.1: Differences in Variances between Cases 1, 2, and 3.....	89
Table 6.2: Differences in Variances: Comparing Case 1 with Cases 2 and 3.....	93
Table 6.3: The Variances for Cases 4-6.....	96
Table 6.4: Differences in Variances Between Cases 1-4, 2-5, and 3-6	96
Table 6.5: 4 Schemes Considered for Comparison.....	102

List of Figures

	Page
Figure	
4.1: Data and Parameters from Two Time Points	48
4.2: Paired-Matching on Baseline Data: Order and Induced Order Statistics.....	63
5.1: Cluster Parameters, Variation, and Time 1/2 Correlation.....	76
6.1: Functions of ω_p by p.....	97
6.2: Power Curves for 7 Case Scenarios Assuming p=10 N=50	
$\rho = 0.01 \psi = 0.2$	103
6.3: Power Curves for 7 Case Scenarios Assuming p=50 N=10	
$\rho = 0.01 \psi = 0.2$	104
6.4: Power Curves for 7 Case Scenarios Assuming p=10 N=50	
$\rho = 0.1 \psi = 0.8$	106
6.5: Power Curves for 7 Case Scenarios Assuming p=50 N=10	
$\rho = 0.1 \psi = 0.8$	107

Abstract

DESIGN AND ANALYSIS METHODS FOR CLUSTER RANDOMIZED TRIALS WITH PAIR-MATCHING ON BASELINE OUTCOME: REDUCTION OF TREATMENT EFFECT VARIANCE

By Misook Park

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2006

Research Director:

Dr. Robert E Johnson
Associate Professor
Department of Biostatistics

Cluster randomized trials (CRT) are comparative studies designed to evaluate interventions where the unit of analysis and randomization is the cluster but the unit of observation is individuals within clusters. Typically such designs involve a limited number of clusters and thus the variation between clusters is left uncontrolled. Experimental designs and analysis strategies that minimize this variance are required. In this work we focus on the CRT with pre-post intervention measures. By incorporating the baseline measure into the analysis, we can effectively reduce the variance of the treatment effect. Well known

methods such as adjustment for baseline as a covariate and analysis of differences of pre and post measures are two ways to accomplish this. An alternate way of incorporating baseline measures in the data analysis is to order the clusters on baseline means and pair-match the two clusters with the smallest means, pair-match the next two, and so on. Our results show that matching on baseline helps to control the between cluster variation when there is a high correlation between the pre-post measures. Six cases of designs and analysis are evaluated by comparing the variance of the treatment effect and the power of related hypothesis tests. We observed that—given our assumptions—the adjusted analysis for baseline as a covariate without pair-matching is the best choice in terms of variance. Future work may reveal that other matching schemes that reflect the natural clustering of experimental units could reduce the variance and increase the power over the standard methods

Chapter I

Introduction

Cluster randomized trials are common in primary care and educational research. Individuals in these populations of interest are naturally clustered, such as patients in a primary care clinic or students in an elementary school. When study treatments are applied to the entire cluster, then all subjects within a cluster receive the same treatment. In such a case, the clusters are the unit of assignment and analysis and the units of observation are typically individuals within those clusters. We are interested in analyzing differences in treatment effect versus control using clustered data such as these.

Our interest in this problem was motivated by a randomized cluster trial conducted by Dr. Stephen F. Rothemich, MD, MS, in collaboration with Dr. Robert E. Johnson and the research team of the Virginia Ambulatory Care Outcomes Research Network (ACORN). Rothemich asked if using smoking status as a vital sign—along with the usual vital signs such as blood pressure and weight—would increase counseling for

Smoking cessation in a busy primary care practice. The limited number of sampled primary care practices—clusters—were the units of analysis and randomization. A review of previous studies revealed a significant amount of variation among practices. In order to better control this variation, in the absence of selecting additional clusters, consideration was given to using baseline counseling rates to adjust for the post-intervention rates. The researchers designed a trial where the practices were first matched into pairs with similar baseline means and then randomized to study arms. This raises the question about the efficiency of such a design. Does matching provide greater control of the between cluster variance than does pre-post differences or using the baseline mean as a covariable? Will such matching lead to better power? The model for this motivating example has errors that are binomially distributed. However for this work, focus will remain on the classical Gaussian error.

Various methods to control for the between-cluster variation have been suggested in the literature. These include analyses on the following four cases:

Case 1. post intervention measures (this does not use the baseline to control the variance, but we include this case for comparison),

Case 2. post measures adjusted for covariate baseline measures,

Case 3. pre-post difference measures, and

Case 4. matching on baseline measures.

Other methods such as the analysis of percent change scores will not be directly considered in this work (though percent change scores could be converted to difference scores by log transformation); however, we will consider two additional cases:

Case 5. matching on baseline measures together with post measures adjusted for covariate baseline measures, and

Case 6. matching on baseline measures together with pre-post difference measures.

Each case, except Case 1, makes use of the baseline outcome measures to control variance. Typically one can analyze pre-post differences (Case 3) using a paired t-test or similar analysis. This is probably the simplest approach and it is widely used. Alternatively one may model the relationship between the baseline and post-intervention measures by using the baseline measure as a covariable (Case 2). The post-intervention variance matrix conditioned on the baseline generally results in a reduction of variance that exceeds the reduction achieved by the simpler analysis of pre-post differences.

Matching on baseline prior to randomization should reduce the variance in the usual way by *removing* the variance between matched blocks from the analysis and

focusing on the more homogeneous within block measures. It is less clear if adding matching to either the analysis of covariance or the analysis of differences leads to a reduction in variation or an improvement in power.

Although other matching or blocking schemes may be used, this work will focus only on creating matched blocks by pairing clusters with adjacent baseline means. This method has been suggested by some (Cox, 1957; Bonate, 2000; Murray, 1998; Donner 2000). Pairing may not be as efficient as creating blocks of varying size where blocks are determined by maximizing the variation between blocks while requiring the minimal block size to equal the number of study arms. This method is discussed in small detail for future research in Chapter 8.

For the purpose of this work it is assumed that the experiment consists of a one or two-staged sampling design. At the first stage clusters (subjects, primary care practices, schools) are sampled from a population of clusters. It is assumed that either the population is large relative to the sample size or the sampling is performed with replacement. Thus finite sampling corrections to the variance are not needed. Baseline outcome measures are drawn from each cluster as well as post intervention measures. For

each cluster it is assumed that the pair of measures follows a common bivariate normal distribution with a positive correlation between measures.

Two designs commonly employed in cluster randomized trials are cross-sectional and cohort designs. The essential feature of the nested cross-sectional design is that independent samples of members are measured at each time interval included in the study. On the other hand in the nested cohort design, the same group of members—one sample—is measured at all time intervals included in the study. Our study assumes a nested cross-sectional design. Each cluster may be measured more than once, but each member within cluster may be measured at a given time period only if selected for that time period's sample. The results of this work may not extend to the cohort design.

Two approaches take into consideration how the cluster measures are made. The first approach assumes that measures are made without further sampling. For example, a measure is computed for each primary care practice based on the totality of its patient encounters. This is termed the *subjects* case and the corresponding experimental design is a *randomized controlled trial* (RCT). The second approach assumes that measures taken from a random sample of patients within a cluster are averaged to represent the cluster value. This constitutes the second stage of sampling. For this work it is assumed that,

given the cluster, the second stage sample is a simple (independent) random sample from a common normal distribution and that finite sampling corrections are not required. This is termed the *clusters* case and the corresponding experimental design is a *cluster randomized trial* (CRT). With either approach, clusters (subjects) are randomized to intervention or control after the baseline measures are taken.

Because of the two stage design, an intraclass correlation (ICC) between measures within clusters is induced. The variance of a single individual sampled from the total population of subjects (all subjects within all clusters) is the sum of the variance between the population cluster means and the variance among subjects within the clusters. The ICC is the ratio of the variance between the cluster means to this total variance. Subjects may have an inherent correlation as well. Subjects within a cluster or other associated group may respond to the experiment under the influence of a common factor thus causing their responses to be correlated. The measures within a cluster may not be a simple random sample of subjects but rather repeated measures on one or more subjects. These repeated measures have an inherent correlation. In this study we assume there is no inherent correlation. However, due to the design, subjects within clusters will be correlated according to the ICC.

Relevant background is presented in Chapter 2. In Chapter 3 statistical and mathematical tools related to this work are given. The six cases mentioned above are presented in detail in Chapter 4 where the subjects-level randomization and analysis for simplicity and clarity is considered. In Chapter 5 an extension is given to the cluster-level analysis with both between and within cluster variation. The results of variance and power comparisons between the six cases are presented in Chapter 6. Included is a consideration of optimal allocation of sample size between the baseline and post-intervention periods. Lastly in Chapter 7 a discussion of conclusions is presented and Chapter 8 contains directions for future work.

Chapter II

Background and Significance of Cluster Randomized Trials

2.1 Introduction and Examples

Randomized designs utilizing clusters of subjects are common in research studies of primary care, health service, public health and education. Often termed *group randomized trial*, the cluster randomized trial (CRT) is the best experimental design available whenever the investigator wants to evaluate an intervention that operates at a group level, manipulates the social or physical environment, or cannot be delivered to individuals (Donner and Klar 2000; Murray 1998). Some notable examples are as follows:

- 1) School-based smoking prevention programs (Peterson et al. 2000).

This HSPP trial (The Hutchinson Smoking Prevention Project) was a very rigorous cluster randomized trial designed to evaluate the long-term impact of a theory-based, school-based social-influences, grade 3-12 intervention on smoking prevalence among youth. Forty Washington school districts were randomly assigned to the intervention or to

the control condition. Study participants were children enrolled in two consecutive 3rd grades in the 40 districts (n=8388).

2) Dietary change programs (Lytle et al. 2004).

This is a 2-year intervention study (TEENS- Teens Eating for Energy and Nutrition at School) conducted in 16 middle schools with the goal of increasing students' intakes of fruits, vegetables, and lower fat foods. The TEENS was a cluster randomized trial occurring in the Twin Cities, Minnesota, metropolitan area from 1997 to 2000. The primary outcome measures for evaluating the effectiveness of TEENS were student-level intake of fruits, vegetables, and energy from fat based on 24-hour dietary recalls. Schools were randomly assigned from within matched pairs to either control or intervention. Schools were matched on both the proportion of seventh graders expected to receive the TEENS curriculum and on the proportion receiving free or reduced-price school lunch. Randomization was constrained so that the four smallest schools were distributed with two in each of the two conditions. The eight intervention schools received the TEENS intervention and related training for 2 consecutive years beginning when the grade cohort was in the seventh grade (1998 to 1999) and continuing through the eighth-grade year

(1999 to 2000). The eight control schools received intervention materials and training after the follow-up survey (fall 2000).

3) Community-based initiatives to reduce mortality from cardiovascular disease (Luepker et al. 2000).

Named the Rapid Early Action for Coronary Treatment Trial (REACT), this trial was a cluster randomized study of community intervention to improve acute myocardial infarction (AMI) treatment by reducing patient delay to access the health care system. The trial was conducted from 1995 to 1997 in 20 US cities (10 pairs of communities matched on their demographic attributes; population range, 55,777-238,912) in 5 geographic areas (10 states) of the United States. A 4-month baseline data collection phase (December 1995 to March 1996) was followed by 18 months of intervention and concurrent evaluation (April 1996 to August 1997). After the initiation of baseline measures, 1 city in each pair was randomized to the community intervention and the other to reference status.

4) Worksite interventions designed to improve health and safety (Sorensen et al. 2002).

This study was a cluster randomized controlled trial with 15 manufacturing worksites assigned to a health promotion (HP) or a health promotion plus occupational health and safety intervention (HP/OHS), and compared from baseline (1997) to final (1999). This study assessed whether an intervention integrating health promotion with occupational health and safety results in significant and meaningful increases in smoking cessation and consumption of fruits and vegetables, compared to a standard health promotion intervention, for workers overall and for blue-collar workers in particular.

The completely randomized cluster design is most commonly used cluster randomized trials and it is suited when large numbers of clusters are randomized. An interesting example of a completely randomized design is the Indonesian study by Abdeljaber et al. (1991). This trial randomized a fairly large number of clusters, where 229 villages were assigned to the experimental group and 221 to the control group. This study was conducted to evaluate the effectiveness of vitamin A supplementation on symptoms on respiratory infection among Indonesian children aged one to five years. For smaller studies Donner (2000) recommends some matching or stratification.

The matched-pair design has been used widely in practice. The ability to create intervention groups that are perceived to be comparable at baseline requires a consensus of what matching variables are most important. A goal of this study is to have a matched pair of clusters appear to be similar by way of matching on the baseline measure of outcome.

Martin and Diher (1993) pointed out that the number of experimental units (communities) in health promotion programs is usually very small due to the costly activities. Investigators often match communities on demographic variables in order to improve the power of their studies. Matching is known to improve power in certain circumstances. However, if the number of communities is small, the matched design will probably have less power than the unmatched design. This is due primarily to the loss of degrees of freedom in the matched design which may outweigh the benefits of matching on any but the strongest correlation of changes in behavior. In the community intervention situation, even small differences in sample size between the matched and unmatched analyses can have expensive consequences.

2.2 Background

It is known that adopting a cluster randomization is less efficient and has weaker statistical power relative to a design individually randomizing the same number of subjects (Donner, 2000). The loss of efficiency arises because the responses of individuals in the same cluster tend to be more similar than the responses of individuals in different clusters. The degree of similarity among responses within a cluster is typically measured by a parameter known as the intraclass (intraclass) correlation coefficient. It is similarly stated by Donner (2000) and Murray (1998) that the intraclass correlation may be interpreted as the fraction of total variation in the data that can be accounted for by the among cluster variation. Donner also pointed out that if we assume the intraclass correlation cannot be negative (if no inherent correlation then this assumption holds), we may write $\rho = \sigma_A^2 / (\sigma_A^2 + \sigma_W^2) = \sigma_A^2 / \sigma_T^2$ (the induced ICC), where σ_A^2 is the among-cluster variation, σ_W^2 is the within cluster variation, and $\sigma_T^2 = \sigma_A^2 + \sigma_W^2$ is the total variation. With this interpretation, we may write $\sigma_W^2 = \sigma_T^2(1 - \rho)$, indicating that the effect of this positive intraclass correlation is to reduce the total variation among the subjects of the same cluster. In those regards, the variance of the study condition mean in

a cluster randomized trial is different from the variance that would be expected when the observations are independent.

The formulation of the study condition mean's variance in a cluster randomized trial has some important implications. It has an additional factor that Kish (1965), among others, has called the *design effect* or *variance inflation factor*. We should notice that the effect of clustering depends on the joint influence of both the number of subjects in each cluster and the intraclass correlation. Phrased another way, small values of intraclass correlation, which we often encounter in practice, combined with large cluster sizes can yield design effects of substantial magnitude. As Murray (1998) observes, the variance of the intervention (treatment) effect is almost always larger in a cluster-randomized trial than in a study based on random assignment of the same number of individuals to the study conditions. Thus standard analysis methods that assume independent errors are not appropriate for the cluster-randomized trials. Statistical testing procedures that do not account for the variation due to the clustering effect will result in an inflated Type I error rate. Statistical testing procedures that properly account for this inflated variation will have reduced power compared to a study where the observations are independent assuming all other factors are constant and the intraclass correlation is positive. When

any treatment effect is defined as a contrast among study condition means, the test of the treatment effect is associated with the error degrees of freedom. The variance of the study condition mean in cluster-randomized trials will involve a component of variance associated with the clustering and the error degrees of freedom will depend on the structure of the variance of the study condition mean, for instance, the degrees of freedom will be directly related to the number of clusters allocated to each study condition. As a result the error degrees of freedom for the test of the treatment effect will be limited. This will further reduce the power of a test compared to the case where we sample only subjects and use them as the units of analysis and randomization. Again, when the proper degrees of freedom are limited, application of methods that provide too many degrees of freedom will yield a Type I error rate that is greater than the nominal level (Murray et al., 1996). In summary, the variation in the data is usually larger and the degrees of freedom for the treatment effects are usually limited in cluster-randomized trials. Because of this we must consider adopting methods such as matching, stratification, pre-post measures, and regression adjustment for pretest measures aimed at improving the efficiency of design—reducing the variance of treatment comparisons.

One approach involves collecting baseline (pretest) outcome data. These are outcome measures made before the treatment that will be made again post treatment. This will allow us to use various methods to improve the efficiency of design by incorporating these pretest data into the statistical analysis. It is known that analysis of posttest data alone may not provide the most powerful method to analyze the data (Bonate, 2000). The underlying assumption of the cluster randomized trial using solely posttest data is that each study group has comparable baseline features as a result of randomly assigning clusters to each treatment. For example, common methods such as Student's t-test or analysis of variance on posttest data assume study groups are comparable at baseline prior to imposition of a treatment intervention. But randomization alone does not insure baseline comparability between study groups, especially when we have limited number of clusters. Randomization does yield unbiased comparisons, but in the presence of a small sample size, chance variation can obscure true intervention effects. In such a case ignoring pretest data may lead to less power to detect the true difference.

There are four elements which are used to define the statistical power of a test (Lipsey, 1990): the alpha level of significance, sample size, effect size, and the statistical test itself. If we assume the test statistic is fixed, one solution to increasing statistical

power is to increase the number of clusters. However this option may not be feasible in some studies. Another alternative is to increase the effect size between groups. Cohen (1988) defined the effect size as the difference between treatments in units of standard deviation: *Effect Size* = $(\mu_t - \mu_c) / \sigma$, where μ_t and μ_c are the means for treatment and control group respectively and σ is the common standard deviation. The sample mean and pooled standard deviation can be used to compute an estimate of the effect size, *effect size* = $(\bar{X}_t - \bar{X}_c) / S$, where $S = \sqrt{\frac{S_t + S_c}{d_t + d_c}}$. The quantities S_t and S_c are the sums of squares for the treatment and control group, respectively, and d_t and d_c are the degrees of freedom for the treatment and control group, respectively. If all other elements are equal, the larger effect size results in greater power. Increasing effect size is often very difficult to achieve in practice (Bonate, 2000). If we could control for differences among clusters in the course of designing the study, then we may be able to reduce the variation between clusters and increase the effect size. One method to control for differences among clusters is to have baseline measurements of the dependent variables of interest and then incorporate them into a data analysis.

It is expected that posttest data will be correlated with the pretest (baseline) data. A number of standard methods are available for using this correlation in order to improve the power to detect the treatment effect. Suppose that the final observation of interest is y and the preliminary observation is x . Cox (1957) suggested five standard methods as follows. For the first two methods the experimental units (clusters) are completely randomized to treatment groups.

Method 1. Use an index of response. A plausible form for the relation between y and

x may be considered for the index; for example, y/x , $y - x$, etc.

Method 2. Make an adjustment for regression on x by including x as a covariable to the treatment indicator.

Method 3. Rank the experimental units in ascending order on x and then group the data into homogeneous blocks. Within each block, units are randomly assigned to treatment groups.

Method 4. Combine the second and third methods to form the randomized block design plus adjustment for x as a covariate.

Method 5. Use additional variables with the preliminary x to define blocks. In this, grouping based on a criterion separate from x should be considered and then a covariance adjustment made for x .

Another approach suggested by Dalton and Overall (1977) is to use systematic nonrandom assignment of clusters to treatment groups. Their assignment algorithm, called the alternate ranks design (ARD), assigns subjects to treatment groups, thereby removing any possibility for a correlation between treatment and pretest. First, all items are sorted in descending order on their pretest data. Suppose we have two study arms labeled A and B. The highest score is assigned to A, the second and the third scores are assigned to B, the fourth and fifth scores are assigned to A, the next pair of scores is assigned to B, and this procedure continues until all items have been assigned to a treatment group. Each block has one item assigned to treatment and one item assigned to control. The study arm associated with the highest block score alternates from block to block. Dalton and Overall (1977) pointed out that “systematic nonrandom assignment based on the ranked values of the observed pretest scores has the effect of equating not only group means but the entire pretest distributions within the treatment groups.” The authors evaluated this ARD method by conducting a Monte Carlo simulation study and

they concluded that this method did not lead to biased estimates of treatment effect but the precision of these estimates was slightly smaller than those obtained from complete randomization. They also clearly stated that their purpose of this study was not to show that ARD is superior to randomization. However Bonate (2000) supported this method by saying that this method offers some advantages over random assignment. Because their algorithm is fixed, assignment to treatment groups removes any prejudices or biases the experimenter may have in assigning subjects to treatment groups. Of course randomization will achieve the same result.

2.3 Significance

The main interest in this work concerns cluster randomized pre-post designs that employ matching prior to randomization in order to increase the power to detect differences between subgroups by reducing the variance of the differences. The primary aim is to investigate how matching controls the variance when the matching variable is a function of the baseline measure. A related goal is to properly estimate the reduced variance and account for it in the data analysis. In pursuit of these goals, the common methods that

incorporate baseline measures as a way of controlling the variation between clusters without facilitating matching are uncovered. Comparisons are made between two approaches, matching and no matching, to understand how to best design the experimental study. The ultimate goal is to test for treatment effects with a design that maximizes the power.

Chapter III

Review of Relevant Statistical Topics

3.1 Introduction

This chapter provides a brief review of some statistical topics and tools that will be used in later chapters. In Section 2 the variance–covariance structure of random vectors is reviewed together with the conditional mean and variance of the multivariate normal distribution. Because the Gaussian error structure is considered in this work, the theory for the general linear model plays an important role. The relevant references are provided in Section 2. In Section 3, methods for determining marginal and conditional moments of random variables are presented. Section 4 contains a review of quadratic forms and related distributions. The matched-pair method used in this work leads to the utilization of order statistics. A brief review of relevant order statistics topics is presented in Section 5. The development of the treatment effect contrast in a block design is covered in Section 6. The method used for simulating multivariate normal data for verification of variance formulas and for assessing power is covered in Section 7. Lastly, Section 8

presents a method for determining the power to detect treatment differences for each of the studied design/analysis cases.

The theory of linear statistical models underlies several important and widely used methods such as univariate and multivariate regression analysis, analysis of variance, analysis of covariance, etc. Because thorough explanations of the theory of linear statistical models are available elsewhere (Graybill, 1976; Searle, 1971; Myers, 1998) the reader is referred to these sources and further details are provided only where needed.

3.2 Variance-Covariance Structure of Linear Models

3.2.1 Mean and Variance-Covariance of Random Vectors

Let $Y = (Y_1 \ Y_2 \ \cdots \ Y_k)'$ be a vector of random variables with

$$\begin{aligned} E[Y_i] &= \mu_i, \\ \text{Var}[Y_i] &= \sigma_{ii} = \sigma_i^2 = E(Y_i - \mu_i)^2, \ i = 1, 2, \dots, k; \text{ and} \\ \text{Cov}(Y_i, Y_j) &= \sigma_{ij} = E[(Y_i - \mu_i)(Y_j - \mu_j)], \ i \neq j. \end{aligned}$$

The mean vector of Y is given by $E[Y] = \mu = (\mu_1 \ \mu_2 \ \dots \ \mu_k)'$. The variance-covariance matrix of Y is the $k \times k$ matrix given by $\text{Var}[Y] = \Sigma = E[(Y - \mu)(Y - \mu)']$. The i^{th} diagonal element of Σ is the variance of Y_i and its ij^{th} off-diagonal element (for $i \neq j$) is the covariance of Y_i and Y_j . Let A be a $k \times k$ matrix. The variance matrix of a linear

transformation $Z = AY$ of Y_i 's is $Var[Z] = AVar(Y)A' = A\Sigma A'$ (Searle, 1971).

3.2.2 Conditional Mean and Variance of Multivariate Normal Vector

Suppose the random vector Y has a multivariate normal distribution with mean vector μ and positive definite covariance matrix Σ . That is $Y \sim MVN(\mu, \Sigma)$. [Here the assumption of a normal distribution (Gaussian error) is not strictly needed. However, since the application in this work is based on Gaussian error, normality is assumed.]

Partition Y , μ , and Σ conformably as $Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$, $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ with

$\Sigma_{12} = (\Sigma_{21})'$. Define

$$\Sigma_{2\cdot 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \quad (3.1)$$

where Σ_{11} is positive definite. Then it can be shown that the marginal distribution of

$Y_2 \sim MVN(\mu_2, \Sigma_{22})$ and the conditional distribution of Y_2 given Y_1 is

$$Y_{2\cdot 1} \sim MVN[\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(Y_1 - \mu_1), \Sigma_{2\cdot 1}] \quad (3.2)$$

(Rencher, 2000). The conditional mean of Y_2 given Y_1 is thus the unconditional mean of Y_2 corrected for its regression on Y_1 . The conditional covariance is a matrix of partial variances and covariances, partialing on Y_1 .

3.3 Marginal and Conditional Moments of a Random Variable

3.3.1 Marginal Variance

The definition of the (marginal) variance of a random variable Y is

$$Var(Y) = E\left[(Y - E(Y))^2\right] = E(Y^2) - [E(Y)]^2 \text{ (Bickel and Doksum, 1977).}$$

3.3.2 Conditional Mean and Variance

The conditional mean of Y given X is denoted by $E(Y|X)$. The conditional variance of Y

given X is the variance of the conditional distribution of $Y|X$ denoted and defined by

$$Var(Y|X) = E\left[(Y - E(Y|X))^2 | X\right]. \text{ Note that both } E(Y|X) \text{ and } Var(Y|X) \text{ are random}$$

variables in X (Bickel and Doksum, 1977).

3.3.3 Marginal Mean and Variance from Conditional Mean and Variance

The marginal mean and variance may be derived from the conditional mean and variance

by way of the following relationships (Mood et al. 1974).

$$\begin{aligned} E(Y) &= E[E(Y|X)] \\ Var(Y) &= E[Var(Y|X)] + Var[E(Y|X)] \end{aligned} \tag{3.3}$$

These relationships imply the follow interpretations.

1. The marginal mean is the average conditional mean.
2. The marginal variance is the sum of two components of variance: the expected value of the conditional variance and the variance of the conditional means.
3. Since variances are always non-negative, both $Var[E(Y|X)] \geq 0$ and $E[Var(Y|X)] \geq 0$.
4. Note that $Var(Y) = E[Var(Y|X)]$ if and only if $Var[E(Y|X)] = 0$.
5. $Var[E(Y|X)] = 0$ if and only if $E(Y|X) = E(Y)$ for all X .

3.4 Quadratic Forms and Relevant Distributions

3.4.1 Distributions of Quadratic Forms

Sums of squares in an analysis of variance may be expressed as quadratic forms of a vector of observations. A quadratic form in \mathbf{y} is denoted as $\mathbf{y}'A\mathbf{y}$ for some symmetric matrix A . This quadratic form $\mathbf{y}'A\mathbf{y}$ is a random variable and can be expressed as a weighted sum of squares and cross products of the \mathbf{y}' 's. Hypothesis testing and variance component estimation often utilize expected values of $\mathbf{y}'A\mathbf{y}$.

The expected value and variance of the quadratic form is given by (Myers, 1998)

as

$$\begin{aligned}
E[\mathbf{y}'A\mathbf{y}] &= \text{tr}(A\Sigma) + \mu' A \mu \\
\text{Var}(\mathbf{y}'A\mathbf{y}) &= 2\text{tr}(A\Sigma)^2 + 4\mu' A \Sigma A \mu
\end{aligned} \tag{3.4}$$

where $\text{Var}(\mathbf{y}) = \Sigma$. The variance relation requires that \mathbf{y} is normally distributed but the expected value does not.

When \mathbf{y} is normally distributed and $A\Sigma$ is idempotent, the quadratic form $\mathbf{y}'A\mathbf{y}$ has a non-central χ^2 distribution with $\text{rank}(A\Sigma)$ degrees of freedom and non-centrality parameter $\phi = \mu' A \mu / 2$. A symmetric matrix A is called idempotent when the matrix has the characteristic that its square equals itself: $AA = A$.

3.4.2 Independence of Quadratic Forms

The following theorem (Searle, 1971) gives necessary and sufficient conditions for quadratic forms of the type encountered in Gaussian linear models to be independent.

Theorem: Let \mathbf{y} be a multivariate normal random vector with mean μ and variance Σ . Let A and B be symmetric, $n \times n$ matrices of ranks d_1 and d_2 , respectively. The quadratic forms $\mathbf{y}'A\mathbf{y}$ and $\mathbf{y}'B\mathbf{y}$ are independent, if and only if $A\Sigma B = 0$.

In practice it is often assumed that y_1, y_2, \dots, y_n are independent, normally distributed random variables with common variance σ^2 . In this case, $\mathbf{y}'A\mathbf{y}$ and $\mathbf{y}'B\mathbf{y}$ are independent if and only if A and B are orthogonal to each other ($AB = 0$).

3.4.3 Noncentral χ^2 Distribution

Let the $n \times 1$ random vector \mathbf{y} be distributed as $N(\mu, 1)$. Then it can be shown that $U = \mathbf{y}'\mathbf{y}$ is distributed as a non-central chi-squared with n degrees of freedom and non-centrality parameter $\phi = \mu'\mu/2$ [Myers, 1998]. When the mean of \mathbf{y} is $\mu = 0$, then random variable U has a central chi-squared distribution with n degrees of freedom.

3.4.4 Noncentral \mathcal{F} Distribution

The ratio of a chi-squared random variable to an independent central chi-squared random variable, both divided by their degrees of freedom, is distributed as a central \mathcal{F} distribution. The \mathcal{F} distribution is called central or non-central according to the same label attached to the numerator chi-squared variable.

Consider the two quadratic forms $\mathbf{y}'A\mathbf{y}$ and $\mathbf{y}'B\mathbf{y}$ where the ranks of A and B are d_1 and d_2 , respectively. Further assume that $A\Sigma B = 0$ and that both $A\Sigma$ and $B\Sigma$ are idempotent. Then the statistic

$$F = \frac{\mathbf{y}'A\mathbf{y} / d_1}{\mathbf{y}'B\mathbf{y} / d_2} \sim \mathcal{F}(d_1, d_2, \mu'\Sigma\mu/2) \quad (3.5)$$

has a non-central \mathcal{F} distribution with d_1 and d_2 degrees of freedom and non-centrality parameter $\mu'\Sigma\mu/2$.

The usage of the noncentral \mathcal{F} distribution in this work is to provide an approximate (exact in some cases) distribution of test statistics and to evaluate the power of the tests.

3.5 Order Statistics

Suppose that X_1, X_2, \dots, X_n are independent, identically distributed random variables with continuous distribution function (*cdf*) F and probability density function (*pdf*) f .

The values of X_1, X_2, \dots, X_n may be rearranged in nondecreasing order of magnitude to form order statistics $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ such that $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$.

3.5.1 Distribution of Order Statistics

The r^{th} order statistic, $X_{(r)}$, has *cdf* (Armitage and Colton, 2005)

$$\Pr(X_{(r)} \leq x) = \sum_{i=r}^n \binom{n}{i} F^i(x) [1 - F(x)]^{n-i} \quad (3.6)$$

and *pdf*

$$\frac{n!}{(n-r)!(r-1)!} F^{r-1}(x) [1 - F(x)]^{n-r} f(x) \quad (3.7)$$

3.5.2 Asymptotic Distribution

It is not a simple task to work with the exact formulas (3.6) and (3.7) for the *pdf* and *cdf* of the r^{th} order statistic. However, a simple approximation of a central order statistic in (3.8) can be used to approximate the distribution of the r^{th} order statistic. Suppose that r is the integer part of pn , where $0 < p < 1$, and let x_p denote the p^{th} quantile of F , where $f(x_p) > 0$. Then in large samples, the distribution of $X_{(r)}$ is approximately normal:

$$X_{(r)} \sim N \left[x_p, \frac{p(1-p)}{nf^2(x_p)} \right] \quad (3.8)$$

Under mild conditions on F , (3.8) extends to any fixed number of central order statistics.

Consider $X_{(r_1)} \leq \dots \leq X_{(r_k)}$, where r_j is the integer part of np_j , and $0 < p_1 < \dots < p_k < 1$,

and for $j = 1, \dots, k$ let $x_{pj} = F^{-1}(p_j)$ and suppose that $f(x_{pj})$ is positive and finite. Then

the joint limiting distribution of $X_{(r_1)}, \dots, X_{(r_k)}$ is

$$X_{(r_1)}, \dots, X_{(r_k)} \sim MN \left[x_{p1}, \dots, x_{pk}, \frac{p_i(1-p_j)}{nf(x_{pi})f(x_{pj})} \right] \quad (3.9)$$

where $j \geq i$, x_{pi} is the mean of $X_{(ri)}$, and $\frac{p_i(1-p_j)}{nf(x_{pi})f(x_{pj})}$ is the covariance (variance) of

the $(i,j)^{\text{th}}$ pair.

For this work pre-post multivariate normal data were simulated and then ordered in nondescending order. (See Section 7 for more details on simulations.) Estimates of the means and covariance structure of order statistics using formula (3.9) were generated in relatively small samples and compared to values provided for sample sizes up to 50 by Tietjen, Kahaner, and Beckman (1977). The results show that the limiting distribution of order statistics seems to work fairly well for the small samples.

In this work it is assumed the random variables X_1, X_2, \dots, X_{2n} are a random sample from a normal distribution $N(\mu, \sigma^2)$. These values may be viewed as cluster-level measures at the baseline period. The corresponding order statistics are $X_{(i)}$'s arranged in nondecreasing order. Most frequently adopted grouping method in cluster randomization trials is to pair clusters by taking these order statistics and pairing the first two, next two, and so on. The i^{th} block (pair) has mean value, $Y_i = \frac{(X_{(2i-1)} + X_{(2i)})}{2}$, $i = 1, \dots, n$. There are two sources of variation to be considered. The within-block component of variance is given by

$$\frac{(X_{(2i-1)} - X_{(2i)})^2}{2} \quad (3.10)$$

and the component of variance among blocks given by

$$\frac{\sum_{i=1}^n (Y_i - \bar{X})^2}{n-1} \quad (3.11)$$

where \bar{X} is the overall mean of the X 's. Suppose that x_i, x_{i+1} are contiguous order statistics and are expressed as cluster values in a matched pair. The joint density function of these two order statistics is (Arnold et al, 1992)

$$f_{i,i+1,n}(x_i, x_{i+1}) = \frac{n!}{(i-1)!(n-i-1)!} [F(x_i)]^{i-1} [1-F(x_{i+1})]^{n-i-1} f(x_i) f(x_{i+1}); -\infty < x_i < x_{i+1} < \infty$$

From this joint density the distribution of the within-block differences can be—
theoretically—obtained.

Note that the within-block component of variance depends on i . Small or large values of i correspond to the tails of $N(\mu, \sigma^2)$. Thus the within-block component of variance from the tail of the distribution is not the same as that from the middle of the distribution; i.e., the variance is larger for the small or large i than for middle values.

3.5.3 Induced (Concomitant) Order Statistics

When baseline measures are treated as order statistics the related post-study outcomes are considered *concomitants* of these order statistics. The concept of concomitants of order statistics was first used by David (1973) and independently, under the name of *induced order statistics*, by Bhattacharya (1974). Many different aspects of the asymptotic theory of concomitants of order statistics have been developed (for example Sen (1976), Yang (1977)). Interesting applications of concomitants of order statistics, either as an underlying model or as a tool in statistical inferences, can be found Barnett et al. (1976), David et al. (1977), and Yang (1977).

The following discussion is paraphrased from Armitage and Colton (2005). Let (X_i, Y_i) , $i = 1, \dots, n$, be a random sample from an absolutely continuous bivariate distribution with *cdf* $F(x, y)$. Let $X_{r:n}$ denote the r^{th} order statistic of the X sample values.

The concomitants are obtained by arranging the data in the second component according to the ordering in the first component: the pairs are ordered by the X_i , and the Y -value associated with $X_{r:n}$ is denoted by $Y_{[r:n]}$. We call $Y_{[r:n]}$ the concomitant of the r^{th} order statistic.

3.5.4 Standard Normal Order Statistics

Order statistics form the basis for many inferential techniques and knowledge of associated moments provides information about performance characteristics (see David, 1981). Applications are found in methods associated with trimmed means and more generally L-statistics. Testing for departure from normality as presented by Shapiro and Wilk (1965), for example, relies on a table of coefficients that are defined in terms of the means, variances, and covariances of normal order statistics.

In this section some important properties of order statistics from the normal population will be presented. First consider the order statistics $X_{i:n}$ ($1 \leq i \leq n$) from the standard normal distribution. Parrish (1992) pointed out that expected values have been reported by several authors to varying degrees of accuracy and precision. Tietjen et al. (1977) computed the means, variances, and covariances of all order statistics for sample

sizes up to 50. However, their values contrast with those given by Harter and Yamauti (1972)—who reported expected values to 10 decimal places—in the fourth decimal place for a sample size 50, for example. Parrish (1991) used a numerical integration technique to provide high precision tables of expected values and standard deviations of order statistics to 25 decimal places, taken from final values believed to be accurate to at least 27 decimal places. The numerical evaluation technique is based on a simple Gauss-Legendre quadrature method for which computational routines are available in the literature. For examples, see Stroud and Secrest (1969), David and Rabinowitz (1984), and Lether (1978). Tables are made available for sample sizes of 2(1)50, 50(10)200, and 200(25)500. Parrish stated that a related method can be applied to obtain the covariances, although the computation of covariances is more complex and computationally intensive. The precision with which covariances can be practically computed is more limited, especially for larger sample sizes. He also reported that with respect to all other known tables, his results extend the accuracy and precision of variance and covariances of normal order statistics for sample sizes up to 50.

3.5.5 Mean and Covariance of Ordered and Induced Ordered Normal Statistics

Suppose \mathbf{y}_1 and \mathbf{y}_2 are the vector of p standard normal statistics such that let \mathbf{z}_1 be the vector of ordered \mathbf{y}_1 and \mathbf{z}_2 be the vector of induced ordered \mathbf{y}_2 , which is induced by \mathbf{z}_1 ,

$E \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $Cov(\mathbf{y}_1, \mathbf{y}_2) = \begin{pmatrix} 1 & \psi \\ \psi & 1 \end{pmatrix} \otimes I_p$. Then the means for standard form for \mathbf{z}_1

and \mathbf{z}_2 are given by (Lee and Viana, 1999)

$$E \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{c} \\ \psi \mathbf{c} \end{bmatrix} \quad (3.12)$$

where \mathbf{c} is the mean vector of p ordered independent standard normal variates.

Alternatively, \mathbf{y}_1 and \mathbf{y}_2 be such that $E \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}$ and $Cov(\mathbf{y}_1, \mathbf{y}_2) = \tau^2 \begin{pmatrix} 1 & \psi \\ \psi & 1 \end{pmatrix} \otimes I_p$.

Then the means of \mathbf{z}_1 and \mathbf{z}_2 in non-standard form are given by

$$E \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 + \tau \mathbf{c} \\ \boldsymbol{\mu}_2 + \tau \psi \mathbf{c} \end{bmatrix} \quad (3.13)$$

The variance-covariance matrix of \mathbf{z}_1 and \mathbf{z}_2 is

$$Cov(\mathbf{z}_1, \mathbf{z}_2) = \boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{\Gamma}_{11} & \boldsymbol{\Gamma}_{12} \\ \boldsymbol{\Gamma}_{21} & \boldsymbol{\Gamma}_{22} \end{bmatrix} = \begin{bmatrix} \tau^2 \mathbf{C} & \tau^2 \psi \mathbf{C} \\ \tau^2 \psi \mathbf{C} & \tau^2 \mathbf{I}_p + \tau^2 \psi^2 (\mathbf{C} - \mathbf{I}_p) \end{bmatrix} \quad (3.14)$$

where \mathbf{C} is the covariance matrix of independent standard normal variates (Parrish, 1992).

3.6 The Contrast of Treatment Effect Within a Block

Suppose the two block members are randomized to study arm, either treatment or control.

The within-block treatment contrast is the difference between the values, treatment minus

control. Let $\delta_i = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ observation is randomized to treatment} \\ 0 & \text{if the } i^{\text{th}} \text{ observation is randomized to control} \end{cases}$. Then the treatment

contrast in the first block may be defined as $(2\delta_1 - 1)(x_1 - x_2)$, noting that $\delta_2 = 1 - \delta_1$.

$$E[(2\delta_{2i-1} - 1)(x_{2i-1} - x_{2i})] = E[2\delta_{2i-1} - 1]E[x_{2i-1} - x_{2i}] = 0$$

$$\text{Cov}[(2\delta_{2i-1} - 1)(x_{2i-1} - x_{2i}), (2\delta_{2j-1} - 1)(x_{2j-1} - x_{2j})]$$

$$= E[(2\delta_{2i-1} - 1)(2\delta_{2j-1} - 1), (x_{2i-1} - x_{2i})(x_{2j-1} - x_{2j})]$$

$$= E[2\delta_{2i-1} - 1]E[2\delta_{2j-1} - 1]E[(x_{2i-1} - x_{2i})(x_{2j-1} - x_{2j})]$$

Then overall contrast is thus

$$= 0$$

so variance

$$\begin{aligned} \text{Var}[(2\delta_{2i-1} - 1)(x_{2i-1} - x_{2i})] &= E[(2\delta_{2i-1} - 1)^2 (x_{2i-1} - x_{2i})^2] \\ &= E[(x_{2i-1} - x_{2i})^2] \end{aligned}$$

$$E[(2\delta_{2i-1} - 1)(x_{2i-1} - x_{2i})] = E[2\delta_{2i-1} - 1]E[x_{2i-1} - x_{2i}] = 0$$

$$\text{Cov}[(2\delta_{2i-1} - 1)(x_{2i-1} - x_{2i}), (2\delta_{2j-1} - 1)(x_{2j-1} - x_{2j})]$$

$$= E[(2\delta_{2i-1} - 1)(2\delta_{2j-1} - 1), (x_{2i-1} - x_{2i})(x_{2j-1} - x_{2j})]$$

$$= E[2\delta_{2i-1} - 1]E[2\delta_{2j-1} - 1]E[(x_{2i-1} - x_{2i})(x_{2j-1} - x_{2j})]$$

$$= 0$$

These differences—the contrast of treatment effect within a block—are independent from

block to block. This is readily seen since the allocation is independent from the block

values and $E(2\delta_1 - 1) = 0$.

It can now be shown that the variance of the overall contrast is given by

$$\frac{\sum_{i=1}^n E[(x_{2i-1} - x_{2i})^2]}{n^2}. \text{ If the } x\text{'s are completely randomized, then each expectation is the}$$

same. However if the x 's correspond to order or induced order statistics, then as noted above the expectations are largest for i close to 1 or n .

3.7 Methods for Simulation

In the chapters that follow certain variances are developed for the six study cases of design and analysis. These variances are derived analytically but were checked for accuracy and consistency by way of simulation. Additionally simulation was used to compute the power of the tests related to the six cases. Some detail is provided below.

The SAS programs used to perform the simulation are found in Appendix A.

3.7.1 Simulation of a Two-Stage Sample from a Multivariate Normal Distribution

To simulate data for the purposes of verification of derived variances, a two-stage sampling scheme was employed. First a covariance matrix was formed to represent the first stage of sampling from a population of clusters:

$$\mathbf{V} = \tau^2 \begin{pmatrix} 1 & \psi \\ \psi & 1 \end{pmatrix} \quad (3.15)$$

Additionally the mean vector was assumed to contain all zeros. (The mean vector may be easily altered post-simulation by adding the appropriate constants to the simulated data.)

The SAS/IML (SAS Interactive Matrix Language) function VNORMAL was used to generate p bivariate observations representing the pre and posttest cluster means.

At the second stage the variance matrix of the within cluster sample was formed as

$$\sigma^2 I_n \quad (3.16)$$

where n represents either the baseline sample size or the posttest sample size. Note that conditioned on the cluster, the within cluster observations are independent with common variance. The mean vector for each cluster is based on the results from the first stage. That is, the first stage value for a cluster is used as the mean of the normal distribution from which the second stage sample is drawn. Again the function VNORMAL is used to generate the data.

3.7.2 Verification of Variance Calculations

The simulation generates 2500 iterates for each combination of sample size and parameter values. For each iterate, the treatment contrast is computed and stored. Additionally, the computed variance is stored. The analytic variance is confirmed by computing an estimate of the standard error of the treatment contrast and forming a 95% confidence interval. Numerous runs were made on varying parameters and in each case the computed variance was within the confidence limits of the estimated variance.

3.7.3 Power Calculations and Verification

The method for analytically estimating power is given below in Section 8. In addition to the above, the value of the test statistic is generated for each simulated iteration. The test statistic is compared to the appropriate 0.05 critical value taken from an \mathcal{F} distribution and an indicator variable (1=critical value is exceeded, 0=otherwise) is stored. The average value of this indicator variable over all 2500 iterates constitutes an estimate of the power. Estimates of the power along with 95% confidence intervals were computed. In each case the analytically computed power fell within the confidence limits.

3.8 Computing the Power

The power of the tests associated with six case analyses is derived from the associated \mathcal{F} distribution, which for some cases is an approximation. The noncentrality parameter is computed for various treatment effects and power curves are generated using the SAS function CDF with the “F” distribution option.

For Cases 1-3, the exact \mathcal{F} distribution is used. For Cases 4-6 the approximate \mathcal{F} distributions, noted in the associated sections of Chapter IV and V, are used. Both Cases 2 and 5 are conditional analyses. For each realization of the baseline data (for each simulated iterate) the power was computed. This constitutes the power of the test conditioned on the baseline data and the results will vary from iteration to iteration. The average, or estimate of the expected power, is computed across all of the iterates and is reported. Thus the power was computed without simulation for all Cases except Cases 2 and 5.

Chapter IV

Randomized Controlled Trials: 6-Case Scenarios

4.1 Introduction

The randomized controlled trial (RCT) is a form of clinical trial, or scientific procedure where we randomly select p subjects from a study population and randomly assign them to study conditions (treatment or control). It is a quantitative, comparative, controlled method and widely considered to provide the most reliable form of scientific evidence. It is the best known design for eliminating the variety of biases that regularly compromise the validity of medical research. Thus the RCT is one of the simplest and most powerful tools in clinical research.

RCTs are often used to establish average efficacy of a treatment as well as learn about its most frequently occurring side-effects. This is meant to address the following concerns. First, effects of a treatment may be small and therefore undetectable except when studied systematically on a large population. Second, biological organisms (including humans) are complex, and do not react to the same

stimulus in the same way, which makes inference from single clinical reports very unreliable and generally unacceptable as scientific evidence. Third, some conditions will spontaneously go into remission, with many extant reports of miraculous cures for no discernible reason. Finally, it is well-known and has been proven that the simple process of administering the treatment may have direct psychological effects on the patient, sometimes very powerful, creating what is known as the placebo effect. For example, a “cure” might be a successful counseling session for smoking cessation. Increases in such “cures” may occur outside of the influence of a controlled intervention (thus the need for a control group).

In this work, in addition to RCT structure, each subject is measured on the outcome of interest at two time points, that is, before and after assignment of treatment. The first measure is called the pretest or baseline measure and the second measure is called the posttest measure. These two measures on the same subject are correlated and viewed as repeated measures. To compare the results of a treatment with the results of a control group, sample means are contrasted between treatment and control groups; that is, the sample mean of the control group is compared with that of other treatment groups to detect treatment effects.

There are several standard designs and analysis methods that can be applied (Cases 1-3). The methods covered here are divided into two groups: without matching and with matching. The matching technique is to collect the baseline data, sort the subjects by their baseline values, pair them, and then randomize to treatment or control within the matched pair.

I. Without Matching (Standard method)

Case 1. Analysis of variance (ANOVA) on the posttest scores. The baseline data are ignored.

Case 2. Analysis of covariance (ANCOVA) on the posttest scores with pretest score as a covariable.

Case 3. ANOVA on the pre-post differences.

II. With Matching (Standard method augmented with matching)

Case 4. Same as Case 1, accounting for matching.

Case 5. Same as Case 2, accounting for matching.

Case 6. Same as Case 3, accounting for matching.

Note that each method adjusts in some way for the baseline except for the first method. In Cases 5 and 6 the baseline is accounted for in two ways.

In what follows, the derivation of the variance of the overall treatment

contrast and a method of significance testing will be presented for each of the six cases.

4.2 Methods

The general methodology for determining the overall treatment contrast and deriving its variance is given here. The methodology for constructing a test statistic with an \mathcal{F} distribution (or approximate) is presented. These methods are used in Section 3 for each of the six cases.

4.2.1 Treatment Effect

For Cases 1-3 assume without loss of generality that the first half of the data is the control group and the remainder is the treatment group. Let ℓ_t , the linear contrast that compares treatment and control group averages, be written as

$$\ell_t = \frac{2}{p} \begin{pmatrix} -1 \\ 1 \end{pmatrix} \otimes j_{p/2}, \quad (4.1)$$

where j_k is the k -dimensional vector of 1's. The treatment effect is the difference between treatment and control group sample means and it can be expressed as $\ell'_t y$. Under the null hypothesis of no treatment effect, the expected value of $\ell'_t y$ is $E(\ell'_t y) = 0$ and its variance is

$$\text{Var}(\ell'_i y) = \ell'_i \text{Var}(y) \ell_i \quad (4.2)$$

For Cases 4-5 the data are arranged in non-decreasing order on the baseline data. Let z_1 be the vector of ordered baseline measures and z_2 be the vector of induced ordered posttest measures, which is induced by z_1 . Taking matching on baseline measures into consideration, the sample means are contrasted between treatment and control subjects within each matched group. Were the elements of z_2 identically and independently distributed, then we may freely permute them and assign the first half to control and the second half to intervention. However, since the measures have a specific (induced) order, the vector of contrasts that denotes the treatment assignment must also reflect the randomness of the assignment.

Let ϕ_b be a random variable with equal probability of values ± 1 . A value of '1' denotes treatment and a value of '-1' denotes control as the assignment for the $(2b-1)^{\text{st}}$ subject—corresponding to the first subject in the b^{th} pair. Note this also indicates that the assignment for the second subject in pair b will be the opposite of the first subject. The ϕ_b 's are independent each with mean zero and variance 1. The covariance of ϕ_b and $-\phi_b$ is easily shown to be -1. Treatment effect is the difference of means between treatment and control groups and it can be written as $L'_2 z$ where

$$L'_2 = [1/(p/2)] \left((\phi_1, -\phi_1), (\phi_2, -\phi_2), \dots, (\phi_{p/2}, -\phi_{p/2}) \right) \quad (4.3)$$

The expected value of L_2 is $E[L_2] = \mathbf{0}$ and the variance is

$$Var[L_2] = V = I_{p/2} \otimes \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} / (p/2)^2 \quad (4.4)$$

To find the variance of the contrast we must account for the fact that both L_2' and z_2 are random. Using equation (3.3) from tools section (variance derived from the conditional variance and conditional expectation) we see that

$$Var(L_2'z) = Var(E[L_2'z | L_2]) + E(Var[L_2'z | L_2]) \quad (4.5)$$

The resolution of (4.5) depends on the definition of z , which varies from case to case, and the covariance matrix of z .

4.2.2 Variance of the Treatment Effect

In order to find the variance of the treatment effect the variance of the data vector (unordered or ordered) is required. Let $y_j = (y_{1j}, \dots, y_{pj})'$ be the vector of observed values, $i=1, \dots, p$, at time j where $j=1, 2$. The y 's at each time point are assumed to be independently, identically distributed as $N(\mu_j, \tau^2)$. Each time 1-2 pair of values, (y_{i1}, y_{i2}) is assumed to be bivariate normal with common correlation ψ (Figure 4.1). Therefore $Var((y_{i1}, y_{i2})) = \begin{pmatrix} \tau^2 & \tau^2\psi \\ \tau^2\psi & \tau^2 \end{pmatrix}$. The covariance matrix of y_1 and y_2 can be written as

$$\Sigma = Cov \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} \tau^2 & \tau^2\psi \\ \tau^2\psi & \tau^2 \end{pmatrix} \otimes I_p \quad (4.6)$$

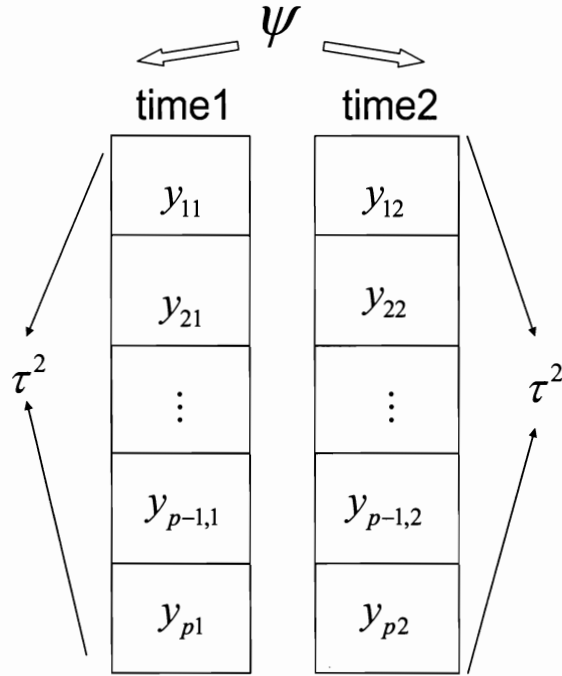


Figure 4.1: Data and Parameters from Two Time Points

Equation (4.6)—or a function of this equation—is used in (4.1) for Cases 1-3.

For Cases 4-5 the focus is on the ordered and induced ordered vectors z_1 and z_2 defined above. Drawing on equations (3.13) and (3.14) –from section 5.5 of the tools chapter, the means of z_1 and z_2 are given by

$$E \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} \mu_1 + \tau \mathbf{c} \\ \mu_2 + \tau \psi \mathbf{c} \end{bmatrix} \quad (4.7)$$

where \mathbf{c} is the mean vector of p ordered independent standard normal variates. The variance of z_1 and z_2 is given by

$$\Gamma = \text{Cov}(z_1, z_2) = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix} = \begin{bmatrix} \tau^2 C & \tau^2 \psi C \\ \tau^2 \psi C & \tau^2 I_p + \tau^2 \psi^2 (C - I_p) \end{bmatrix} \quad (4.8)$$

where \mathbf{C} is the covariance matrix of independent standard normal variates. Equation

(4.8)—or a function of this equation—is used in (4.5) for Cases 4-5.

4.2.3 Methods for Significance Testing

Define the elements of linear model as follows:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}, X = [X_1 : X_2], \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \text{ and } \varepsilon = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{bmatrix} \quad (4.9)$$

where β_1 is a $k \times 1$ vector of coefficients, the design matrix X has dimensions $p \times (k+1)$ and can be partitioned as $X = [X_1 : X_2]$ where X_1 consists of all but the last column—includes the vector of 1's—and X_2 consists of the last column representing the vector of indicator variables associated with treatment assignment. We want to test $H_0 : \beta_2 = 0$ versus $H_1 : \beta_2 \neq 0$ to assess the significance of the treatment effect.

The model under H_1 is called the full model. The model under H_0 is called the reduced model since X_2 is dropped. The two models are compared and the reduced model will be retained unless it is shown to be inadequate.

The method used to test the null hypothesis is analysis of variance (ANOVA).

When there is only control and one treatment group, the two-sample t-test is equivalent and may be more convenient to use. ANOVA is an analytic technique in which total sum of squares—which measures the total variation in the data—is divided into components that can be attributed to important sources. The regression sum of squares for the full model is $SSR_{full} = y'X(X'X)^{-1}X'y$. The regression sum of squares for the reduced model is $SSR_{reduced} = y'X_1(X_1'X_1)^{-1}X_1'y$. The difference between the two models is the amount of variation in the response that cannot be accounted for by the reduced model alone. The difference is expressed as $SSR_D = y'(X(X'X)^{-1}X' - X_1(X_1'X_1)^{-1}X_1')y$. When H_0 is true, most of variability in the response should be explained by the reduced model and the difference should be small. Thus this quadratic form will play a major role in the test statistic used to choose between H_0 and H_1 . The residual sum of squares can be expressed as $SSE = y'(I - X(X'X)^{-1}X')y$. Using these sums of squares, an F statistic may be formed as in equation (3.5) –from the section 4.4 in the tools chapter.

4.2.4 Approximations to the F Distribution

Commonly the application of the F distribution is only approximate. The underlying

data may be non-normal, the covariance structure has an unyielding eigen-structure, quadratic form matrices are not orthogonal, and other reasons. In this work paired matching on baseline measures leads to the underlying data being distributed as ordered-normal, rather than simply (unordered) normal. Such data will not lead to homogeneity of variance. The result is that the matrix of the quadratic form (the perpendicular projection, Q) times the covariance matrix of the data (ordered statistics) may not be idempotent. Accordingly the quadratic form may not have an exact chi-squared distribution. Finally the covariance structure is such that relevant matrices of the quadratic forms (perpendicular projections for regression and error) may not be orthogonal with respect to the covariance. Thus the two quadratic forms may be correlated.

The approach given here is not one founded in asymptotic theory, though appealing to such theory would be useful. Rather the approach is to note how the properties are approximately met to make the usual ratio statistics have approximate F distributions.

Consider the quadratic form $Y'QY$ where $Y \sim MVN(\mu, \Sigma)$ and Q is a perpendicular projection (and is thus idempotent). Note this means that r of the eigenvalues of Q are equal to one and the others are equal to zero, where r is the

dimension of Q . When $r=1$, $Q\Sigma$ will have a single non-zero eigenvalue, say κ , provided that $Q\Sigma \neq 0$. To show this, let the vector $u \in$ column span of Q such that its norm is one, $\|u\|=1$. Then

$$Q\Sigma u = (u'\Sigma u)u \quad (4.10)$$

since Q is a perpendicular projection. Hence u is an eigenvector of $Q\Sigma$ with corresponding eigenvalue

$$\kappa = u'\Sigma u \quad (4.11)$$

Thus $Q\Sigma/\kappa$ is idempotent and $Y'QY/\kappa$ will have a (possibly non-central) chi-squared distribution. However when $r > 1$ the positive eigenvalues of $Q\Sigma$ may vary and if so then no single scalar correction can be used to achieve the desired idempotency. However, if the eigenvalues have relatively small variation, then an average eigenvalue may be used for the divisor κ . The result will yield $Q\Sigma/\kappa$ to be approximately idempotent and it may be argued that $Y'QY/\kappa$ will be approximately chi-squared.

Let $Q_D = X(X'X)^{-1}X' - X_1(X_1'X_1)^{-1}X_1'$ be the perpendicular projection onto the orthogonal complement of the span of X_1 in the span of X . Then $SSR_D = y'Q_D y$. The rank of Q_D , $r(Q_D)$, is 1 and, assuming that $Var(y)$ is full rank, $r(Q_D Var(y))=1$. As noted above, $Q_D Var(y)$ has one positive eigenvalue (the

others are zero) denoted by κ . Thus $Q_D \text{Var}(y)/\kappa$ is idempotent and $SSR_D/\kappa = y'Q_D y/\kappa$ has a chi-squared distribution with noncentrality parameter $\lambda = \mu'Q_D \mu/2\kappa$.

Let $Q_E = I - X(X'X)^{-1}X'$ be the perpendicular projection onto the orthogonal complement of the span of X in R^n (n -dimension Euclidean space). Then $SSE = y'Q_E y$, $r(Q_E) = r(Q_E \text{Var}(y)) = p - k - 1$. This means $Q_E \text{Var}(y)$ has $p - k - 1$ non-zero eigenvalues. If all of these are the same, then we have two distinct eigenvalues, κ and 0, where the multiplicity of κ is $p - k - 1$. In this case $Q_E \text{Var}(y)/\kappa$ is idempotent and $SSE/\kappa = y'Q_E y/\kappa$ has a central chi-squared distribution.

4.3 Case Scenarios

4.3.1 Case 1: Posttest Data With No Matching

This is a completely randomized trial, involving no pre-stratification or matching of subjects according to baseline characteristics. Therefore with this approach the baseline measures are ignored. To find a treatment effect, using only posttest data, the sample mean of control group is compared with that of treatment group.

This type of analysis can be both weak and strong, depending on the

circumstances. The strength of this analysis is its simplicity in that it requires no baseline data collection. One major weakness is that the variance between subjects may be large relative to the variance within subjects. Any variation not attributable to study condition is left to residual error. It is assumed that randomization will cause the two groups to be similar at baseline. However when the sample size is small, the variation between subjects can overwhelm the difference caused by the treatment. Randomization can remove bias, but it does not reduce variation. This is a weak design unless a fairly large number of subjects are randomized to each study condition. So if baseline data is collected, this method would not be recommended. There may be situations where collecting baseline data is not feasible. Then—among the six cases—this is the only choice.

4.3.1.1 Variance of Treatment Effect

The treatment effect is the difference between treatment and control group sample means taken at time 2 and it can be written as $\ell'_i y_2$. Since $Var(y_2) = \Sigma_{22} = \tau^2 I_p$ then from (4.2) and (4.6) the variance is given by

$$Var(\ell'_i y_2) = \tau^2 \ell'_i \ell_i = \frac{4}{p} \tau^2 \quad (4.12)$$

4.3.1.2 Inference on Treatment Effect

The matrix X_1 is the vector of 1's since the model only contains the intercept term and a treatment indicator. The matrix Q_D is thus the perpendicular projection on the space spanned by ℓ_t . Since $Var(y_2) = \tau^2 I_p$, $\kappa = \tau^2$ is the positive eigenvalue of $Q_D \Sigma_{22}$. Therefore $SSR_D / \tau^2 \sim \chi^2_{(1, \lambda)}$ with non-centrality parameter $\lambda = \frac{1}{2\tau^2} \mu' Q_D \mu$.

The product $Q_E \Sigma_{22}$ has $p-2$ non-zero eigenvalues since $r(Q_E) = r(Q_E \Sigma_{22}) = p-2$. Since $Var(y_2) = \tau^2 I_p$, $\kappa = \tau^2$ is the one positive eigenvalue with multiplicity $p-2$. Thus $Q_E \Sigma_{22} / \tau^2$ is idempotent and SSE / τ^2 follows a central chi-squared distribution with $p-2$ degrees of freedom. The centrality is a result of the mean vector $\mu = X\beta$ being orthogonal to Q_E .

To summarize, SSR_D / τ^2 follows a noncentral chi-squared distribution with 1 degree of freedom and noncentrality parameter $\lambda = \frac{1}{2\tau^2} \mu' Q_D \mu$. Also SSE / τ^2 follows a central chi-squared distribution with $p-2$ degrees of freedom. The variables SSR_D and SSE are independent since $Q_E Q_D = \mathbf{0}$. Hence the ratio $F = \frac{SSR_D / \tau^2}{SSE / (p-2) \tau^2} = \frac{SSR_D}{SSE / (p-2)}$ follows a noncentral \mathcal{F} distribution with 1, $p-2$ degrees of freedom and noncentrality parameter λ .

4.3.2 Case 2: Posttest Data, Baseline as Covariable, No Matching

If we have baseline data collected, it is known that analysis of posttest data alone may not provide the most powerful method to analyze the data (Bonate, 2000). Because of this we must consider adapting methods such as matching, stratification, repeated measures and regression adjustment for covariates aimed at improving the efficiency of the test—reducing the variance of treatment comparisons.

In general it is expected that posttest data, in the absence of treatment effects, will be correlated with the baseline (pretest) data. A number of standard methods are available for using this correlation in order to improve the power to detect the treatment effect. One possible approach will be to have the baseline data as a covariate to improve precision in testing the effectiveness of the treatment. It allows us to control for differences among subjects. In the adjusted analysis, baseline data are used as a covariate that represents a source of variation which has not been controlled for in the experiment. This can reduce the confounding and improve the precision of the estimate of the treatment. Generally if baseline comparability among groups is uncertain or regression towards the mean greatly influences the posttest measures, ANCOVA is recommended as a method for data analysis.

The concept *regression towards the mean* refers to the following. The

posttest measure on the individual will tend to be greater than his corresponding baseline measure when his baseline measure is below the average, whereas posttest measure will tend to decrease when his baseline measure is above the average baseline measure, independent of any treatment effect. The farther a given baseline is from its mean, the greater the amount of regression towards the mean, which is independent of any treatment effect.

Analysis of covariance (ANCOVA) adjusts the dependent variable in order to remove the influence of the baseline on the posttest. Thus ANCOVA tests the hypothesis that the posttest mean in each treatment group is equal given that the mean baseline in each treatment group is equal. Compared to a completely randomized design, it can be seen that ANCOVA splits the error term into two variance components: one is due to regression of the baseline on the posttest data and another due to residual variation. Hence the mean square error term is smaller in ANCOVA than in a completely randomized design and ANCOVA provides a better estimate of the treatment effect. Cochran (1957) showed that the reduction in the mean square error using ANCOVA compared to a completely randomized design is determined by the amount of the correlation between baseline and posttest measures. Usually higher correlation between baseline and posttest measures results in a

smaller mean square error if we have a large degree of freedom so that the reduction in the error term mainly depends on the correlation between pre-post measurements.

4.3.2.1 Variance of Treatment Effect

Method of inferring about the difference between treatment and control means

Assuming y_1 represents the vector of baseline measurements and y_2 represents the vector of posttest measurements, the covariance matrix of y_1 and y_2 is given in (4.6). From equation (3.2) - section 2.2 in the tool chapter, we have that the conditional distribution of y_2 given y_1 is $y_{2.1} \sim MVN[\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(Y_1 - \mu_1), \Sigma_{2.1}]$.

From (4.6),

$$\begin{aligned}\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(y_1 - \mu_1) &= \mu_2 + \tau^2\psi I_p (\tau^{-2}I_p)(y_1 - \mu_1) \\ &= \mu_2 + \psi(y_1 - \mu_1)\end{aligned}\tag{4.13}$$

and

$$\begin{aligned}\Sigma_{2.1} &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \\ &= \tau^2 I_p - \tau^2\psi I_p (\tau^{-2}I_p) \tau^2\psi I_p \\ &= \tau^2(1 - \psi^2)I_p\end{aligned}\tag{4.14}$$

The conditional expectation and variance of the treatment contrast is thus

$$E[\ell'_i y_2 | y_1] = \ell'_i [\mu_2 + \psi(y_1 - \mu_1)]\tag{4.15}$$

and

$$\begin{aligned}
V(\ell'_i y_2 | y_1) &= \ell'_i \Sigma_{2,1} \ell_i \\
&= \ell'_i \tau^2 (1 - \psi^2) I_p \ell_i \\
&= \frac{4}{p} \tau^2 (1 - \psi^2)
\end{aligned} \tag{4.16}$$

4.3.2.2 Inference on Treatment Effect

The matrix X_1 contains the vector of 1's and the baseline data as well as the baseline vector y_1 . The matrix Q_D is thus the perpendicular projection on the space spanned by ℓ_i orthogonalized to y_1 . Since $\text{Var}(y_{2,1}) = \Sigma_{2,1} = \tau^2(1 - \psi^2)I_p$, $\kappa = \tau^2(1 - \psi^2)$ is the positive eigenvalue of $Q_D \Sigma_{2,1}$.

Therefore $SSR_D / (\tau^2(1 - \psi^2)) \sim \chi^2_{(1,\lambda)}$ with non-centrality parameter $\lambda = \frac{1}{2\tau^2(1 - \psi^2)} \mu' Q_D \mu$.

The product $Q_E \Sigma_{2,1}$ has $p-3$ non-zero eigenvalues since $r(Q_E) = r(Q_E \Sigma_{2,1}) = p-3$. Since $\text{Var}(y_{2,1}) = \tau^2(1 - \psi^2)I_p$, $\kappa = \tau^2(1 - \psi^2)$ is the one positive eigenvalue with multiplicity $p-3$. Thus $Q_E \Sigma_{2,1} / (\tau^2(1 - \psi^2))$ is idempotent and $SSE / (\tau^2(1 - \psi^2))$ follows a central chi-squared distribution with $p-3$ degrees of freedom. The centrality is a result of the mean vector $\mu = X\beta$ being orthogonal to Q_E .

To summarize, $SSR_D / (\tau^2(1 - \psi^2))$ follows a noncentral chi-squared distribution with 1 degree of freedom and noncentrality

parameter $\lambda = \frac{1}{2\tau^2(1-\psi^2)} \mu' Q_D \mu$. Also $SSE / (\tau^2(1-\psi^2))$ follows a central chi-squared distribution with $p-3$ degrees of freedom. The variables SSR_D and SSE are independent since $Q_E Q_D = \mathbf{0}$. Hence the ratio $F = \frac{SSR_D / (\tau^2(1-\psi^2))}{SSE / ((p-3)\tau^2(1-\psi^2))} = \frac{SSR_D}{SSE / (p-3)}$ follows a noncentral \mathcal{F} distribution with 1, $p-3$ degrees of freedom and noncentrality parameter λ .

4.3.3 Case 3: Pre-Post Difference, No Matching

As previously discussed, the inclusion of baseline data in a study design provides more flexibility in the selection of methods of data analysis and gives better power to detect the treatment differences. Another possible approach is when the treatment is assessed as a net difference of differences between baseline and posttest measures in the treatment group relative to the control group. This allows for an adjustment for baseline measures. The variance of differences is a function of both the variance of measures at each time point and the correlation between baseline and posttest measures. For this reason, precision is increased if we use differences instead of posttest measures for the data analysis, provided that the correlation is sufficiently large.

4.3.3.1 Variance of Treatment Effect

The treatment effect is the difference between treatment and control group sample pre-post mean differences and it can be written as $\ell'_i(y_2 - y_1)$. Define

$$\ell_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \otimes I_p \quad (4.17)$$

to be the contrast of posttest and pretest measures. Recall that $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$, the

baseline data over the posttest data. Then $\ell'_2 y$ is the vector of pre-post differences

which constitutes the data vector for this case. The treatment effect contrast is thus

$\ell'_i \ell'_2 y$. Since $\text{Var}(y) = \Sigma$ then from (4.6)

$$\begin{aligned} \text{Var}(\ell'_2 y) &= \ell'_2 \Sigma \ell_2 \\ &= \Sigma_{2-1} \\ &= \Sigma_{11} - 2\Sigma_{12} + \Sigma_{22} \\ &= 2\tau^2(1-\psi)I_p \end{aligned} \quad (4.18)$$

and from (4.2) and the treatment contrast variance is given by

$$\text{Var}(\ell'_i \ell'_2 y) = 2\tau^2(1-\psi)\ell'_i \ell_i = \frac{8}{p}\tau^2(1-\psi) \quad (4.19)$$

4.3.3.2 Inference on Treatment Effect

The matrix X_1 is the vector of 1's since the model only contains the intercept term

and a treatment indicator. The matrix Q_D is thus the perpendicular projection on the

space spanned by ℓ_i . The data for this case are pre-post differences in means, $\ell'_2 y$.

Since $\text{Var}(\ell'_2 y) = \Sigma_{2-1} = 2\tau^2(1-\psi)I_p$, $\kappa = 2\tau^2(1-\psi)$ is the positive eigenvalue of

$Q_D \Sigma_{2-1}$. Therefore $SSR_D / (2\tau^2(1-\psi)) \sim \chi^2_{(1,\lambda)}$ with non-centrality parameter

$$\lambda = \frac{1}{4\tau^2(1-\psi)} \mu' Q_D \mu.$$

The product $Q_E \Sigma_{2-1}$ has $p-2$ non-zero eigenvalues since $r(Q_E) = r(Q_E \Sigma_{2-1}) = p-2$. Since $Var(\ell'_2 y) = \Sigma_{2-1} = 2\tau^2(1-\psi)I_p$, $\kappa = 2\tau^2(1-\psi)$ is the one positive eigenvalue with multiplicity $p-2$. Thus $Q_E \Sigma_{2-1} / (2\tau^2(1-\psi))$ is idempotent and $SSE / (2\tau^2(1-\psi))$ follows a central chi-squared distribution with $p-2$ degrees of freedom. The centrality is a result of the mean vector $\mu = X\beta$ being orthogonal to Q_E .

To summarize, $SSR_D / (2\tau^2(1-\psi))$ follows a noncentral chi-squared distribution with 1 degree of freedom and noncentrality parameter $\lambda = \frac{1}{4\tau^2(1-\psi)} \mu' Q_D \mu$. Also $SSE / (2\tau^2(1-\psi))$ follows a central chi-squared distribution with $p-2$ degrees of freedom. The variables SSR_D and SSE are independent since $Q_E Q_D = \mathbf{0}$. Hence the ratio

$$F = \frac{SSR_D / (2\tau^2(1-\psi))}{SSE / [(p-2)(2\tau^2(1-\psi))]} = \frac{SSR_D}{SSE / (p-2)} \quad \text{follows a noncentral } \mathcal{F}$$

distribution with 1, $p-2$ degrees of freedom and noncentrality parameter λ .

4.3.4 Case 4: Posttest Data With Matching

It is known that adopting methods such as matching, stratification, repeated

measures and regression adjustment for covariates may improve the efficiency of the test—reducing the variance of treatment comparisons. We here consider incorporating matching on baseline measures in the data analysis.

Suppose that y_1, y_2, \dots, y_p are independently and identically distributed normal random variables that represent the baseline outcome measures. Let $y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(p)}$ be the order statistics obtained by arranging these random samples in nondecreasing order of magnitude. Clearly the order statistics are not independent. Prior to randomization, these values are paired by taking these order statistics and matching the first two, next two, and so on.

The corresponding values at time 2 represent induced ordered statistics by way of association with the time 1 ordered values. For example, the first value at time 1 is less than or equal to the second value; however the first value at time 2 need not be less than the second value, though this direction of order has increase likelihood due to its association with a correlated order in the time 1 data. After matching, one subject within each pair is randomized to the treatment arm and the other to the control arm (Figure 4.2).

Time1	Block	Time2
$Y_{(1)1}$	1	$Y_{[1]2}$
$Y_{(2)1}$		$Y_{[2]2}$
.	.	.
.		.
.		.
$Y_{(p-)1}$	p/2	$Y_{[p-]2}$
$Y_{(p)1}$		$Y_{[p]2}$

Treatment / Control

Figure 4.2. Paired-Matching on Baseline Data: Order and Induced Order Statistics

4.3.4.1 Variance of Treatment Effect

The treatment effect is the difference between treatment and control group sample means contrasted within each matched-pair at time 2 and it can be written as $L_2'z_2$ where L_2 is defined in (4.3). Recall that from (4.7) the mean of z_2 is given by

$$E(z_2) = \mu_2 + \tau\psi \mathbf{c} \quad (4.20)$$

and from (4.8) the variance of z_2 is given by

$$Var(z_2) = \Gamma_{22} = \tau^2 I_p + \tau^2 \psi^2 (C - I_p) \quad (4.21)$$

The conditional expectation of $L_2'z_2$ given the random allocation L_2

is

$$E[L_2'z_2 | L_2] = L_2'(\mu_2 + \tau\psi \mathbf{c}) \quad (4.22)$$

Each element of the time 2 mean vector μ_2 varies from the corresponding element of the time 1 mean vector μ_1 by a constant (temporal effect or placebo effect) and an amount due to the effect of the intervention, that is, if the given element was exposed

to the treatment. Assume that the added constant is zero (it will work its way out of the ensuing calculations anyway) and that the only difference is the possible treatment effect. The baseline mean μ_1 is assumed by the stage one model to be a multiple of the vector of 1's. Let

$$\delta = \frac{\mathbf{j}_p + (p/2)L_2}{2} \quad (4.23)$$

be the vector of indicators of treatment or control (1's and 0's). Let Δ be the treatment effect difference. Then

$$\mu_2 = \mu_1 + \Delta\delta \quad (4.24)$$

Thus (4.22) reduces to

$$\begin{aligned} E[L_2' z_2 | L_2] &= L_2' (\mu_1 + \Delta\delta + \tau\psi \mathbf{c}) \\ &= \Delta L_2' \delta + \tau\psi L_2' \mathbf{c} \\ &= \frac{\Delta}{2} (L_2' \mathbf{j}_p + (p/2) L_2' L_2) + \tau\psi L_2' \mathbf{c} \\ &= \Delta + \tau\psi L_2' \mathbf{c} \end{aligned} \quad (4.25)$$

since $L_2' \mu_1 = 0$, $L_2' \mathbf{j}_p = 0$, and $L_2' L_2 = 4/p$.

The conditional variance is

$$\begin{aligned} \text{Var}[L_2' z_2 | L_2] &= L_2' (\tau^2 I_p + \tau^2 \psi^2 (C - I_p)) L_2 \\ &= (\tau^2 - \tau^2 \psi^2) L_2' L_2 + \tau^2 \psi^2 L_2' C L_2 \\ &= \frac{4}{p} \tau^2 (1 - \psi^2) + \tau^2 \psi^2 L_2' C L_2 \end{aligned} \quad (4.26)$$

Recall that the expectation of L_2 is the zero vector and its variance is given by (4.4).

Using (4.5) we thus have that

$$\begin{aligned}
Var[L_2' z_2] &= Var(\Delta + \tau\psi L_2' \mathbf{c}) + E\left(\frac{4}{p}\tau^2(1-\psi^2) + \tau^2\psi^2 L_2' C L_2\right) \\
&= \tau^2\psi^2 \mathbf{c}' V \mathbf{c} + \frac{4}{p}\tau^2(1-\psi^2) + \tau^2\psi^2 trace(VC) \\
&= \frac{4}{p}\tau^2(1-\psi^2) + \tau^2\psi^2 trace\left(V(C + \mathbf{c}\mathbf{c}')\right)
\end{aligned} \tag{4.27}$$

The latter term (trace) is a function of the average squared difference between adjacent standard normal order statistics, which from (3.2) is related to the within-matched pair variation. Define

$$\omega_p = trace\left(V(C + \mathbf{c}\mathbf{c}')\right)/p \tag{4.28}$$

to be a function of p . Then (4.27) reduces to

$$Var[L_2' z_2] = \frac{4}{p}\tau^2(1-\psi^2) + p\tau^2\psi^2\omega_p \tag{4.29}$$

4.3.4.2 Inference on Treatment Effect

The matrix X_1 is the vector of 1's and $\frac{p}{2}-1$ indicator vectors that identify the matched-pair blocks. There are $\frac{p}{2}$ blocks but one indicator vector may be omitted so that X_1 remains full rank. The vector denoted by X_2 consists of the observed value of L_2 , the treatment contrast. The matrix Q_D , conditioned on the baseline data, is thus the perpendicular projection on the space spanned by L_2 orthogonalized to the span of X_1 . Since $Var(z_2) = \tau^2 I_p + \tau^2\psi^2(C - I_p)$, then by (4.11) $\kappa = \tau^2(1-\psi^2) + \tau^2\psi^2 \mathbf{u}' C \mathbf{u}$ where the span of the unit vector \mathbf{u} equals the span of

Q_D . Therefore $SSR_D / \kappa \sim \chi^2_{(1, \lambda)}$ with non-centrality parameter $\lambda = \frac{1}{2\kappa} \mu' Q_D \mu$.

The product $Q_E \Gamma_{22}$ has $\frac{P}{2} - 1$ non-zero eigenvalues since $r(Q_E) = r(Q_E \Sigma_{22}) = \frac{P}{2} - 1$. However it can be shown that these eigenvalues are distinct when $\tau^2 \psi^2 > 0$. Thus no scalar times $Q_E \Gamma_{22}$ is idempotent. However, computations show that the average eigenvalue is equal to (or nearly equal to) κ that was found in the previous paragraph. Thus $Q_E \Gamma_{22}$ is approximately idempotent and SSE / κ will approximately follow a central chi-squared distribution with $\frac{P}{2} - 1$ degrees of freedom. The centrality is a result of the mean vector $\mu = X\beta$ being orthogonal to Q_E .

To summarize, SSR_D / κ follows a noncentral chi-squared distribution with 1 degree of freedom and noncentrality parameter $\lambda = \frac{1}{2\kappa} \mu' Q_D \mu$. Also SSE / κ follows an approximate central chi-squared distribution with $\frac{P}{2} - 1$ degrees of freedom. The variables SSR_D and SSE are not independent since $Q_E \Gamma_{22} Q_D \neq 0$; however, computations show that all terms in this product are very near to zero (approximately $\leq 1E-05$) and thus they are essentially uncorrelated. Hence the ratio

$$F = \frac{SSR_D / \kappa}{SSE / \left(\frac{P}{2} - 1\right) \kappa} = \frac{SSR_D}{SSE / \left(\frac{P}{2} - 1\right)} \quad \text{follows an approximate noncentral } \mathcal{F}$$

distribution with 1, $\frac{P}{2} - 1$ degrees of freedom and noncentrality parameter λ .

4.3.5 Case 5: Posttest Data, Baseline as Covariable, With Matching

In the pre-post design, baseline measures are used as a covariate that represents a source of variation which may be controlled for in the experiment. The baseline measure is also used as a matching variable in an effort to control the variation between subjects. Here both methods are combined and incorporated into the data analysis to improve precision.

4.3.5.1 Variance of Treatment Effect

The treatment effect is the difference between treatment and control group sample means contrasted within each matched-pair at time 2 and conditioned on the baseline means. This may be written as $L'_2 z_{2,1}$ where L_2 is defined in (4.3) and $z_{2,1}$ represents the conditional value of z_2 given z_1 . The (conditional) mean of $z_{2,1}$ can be shown using (4.7) with (3.1) - (section 2.2 in the tools chapter) to be

$$\begin{aligned}
 E(z_2 | z_1) &= \mu_2 + \tau \psi \mathbf{c} + \Gamma_{21} \Gamma_{11}^{-1} (z_1 - \mu_1 - \tau \mathbf{c}) \\
 &= \mu_2 + \tau \psi \mathbf{c} + \tau^2 \psi \mathbf{C} (\tau^{-2} \mathbf{C}^{-1}) (z_1 - \mu_1 - \tau \mathbf{c}) \\
 &= \mu_2 + \tau \psi \mathbf{c} + \psi (z_1 - \mu_1 - \tau \mathbf{c}) \\
 &= \mu_2 + \psi (z_1 - \mu_1)
 \end{aligned} \tag{4.30}$$

and from (4.8) the (conditional) variance of $z_{2,1}$ is

$$\begin{aligned}
Var(z_2|z_1) &= \Gamma_{2,1} \\
&= \Gamma_{22} - \Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12} \\
&= \tau^2 I_p + \tau^2 \psi^2 (C - I_p) - \tau^2 \psi C (\tau^{-2} C^{-1}) \tau^2 \psi C \\
&= \tau^2 I_p + \tau^2 \psi^2 (C - I_p) - \tau^2 \psi^2 C \\
&= \tau^2 (1 - \psi^2) I_p
\end{aligned} \tag{4.31}$$

Note that (4.14) and (4.31) give the same result. The (conditional) expectation of

$L'_2 z_{2,1}$ given the random allocation L_2 , (4.24), and (4.30) is

$$\begin{aligned}
E[L'_2 z_{2,1} | L_2] &= L'_2 (\mu_2 + \psi(z_1 - \mu_1)) \\
&= \Delta + \psi L'_2 z_1
\end{aligned} \tag{4.32}$$

The (conditional) variance is

$$\begin{aligned}
Var[L'_2 z_{2,1} | L_2] &= L'_2 (\tau^2 (1 - \psi^2) I_p) L_2 \\
&= \tau^2 (1 - \psi^2) L'_2 L_2 \\
&= \frac{4}{p} \tau^2 (1 - \psi^2)
\end{aligned} \tag{4.33}$$

Using (4.5) we thus have that

$$\begin{aligned}
Var[L'_2 z_{2,1}] &= Var(\Delta + \psi L'_2 z_1) + E\left(\frac{4}{p} \tau^2 (1 - \psi^2)\right) \\
&= \psi^2 z_1' V z_1 + \frac{4}{p} \tau^2 (1 - \psi^2)
\end{aligned} \tag{4.34}$$

The conditional variance depends upon the baseline ordered values (compared to no dependence for Case 2). The expected conditional variance, averaging over all realizations of the baseline data, is given by

$$\begin{aligned}
E\left(Var\left[L_2'z_{2,1}\right]\right) &= \psi^2 E\left(z_1'Vz_1\right) + \frac{4}{p}\tau^2(1-\psi^2) \\
&= \psi^2\left(\tau^2 trace(VC) + (\mu_1 + \tau\mathbf{c})'V(\mu_1 + \tau\mathbf{c})\right) \\
&\quad + \frac{4}{p}\tau^2(1-\psi^2) \\
&= \frac{4}{p}\tau^2(1-\psi^2) + p\tau^2\psi^2\omega_p
\end{aligned} \tag{4.35}$$

where ω_p is given by (4.28). Note that the average variance in (4.35) is the same as the Case 4 variance given in (4.29).

4.3.5.2 Inference on Treatment Effect

The matrix X_1 is the vector of 1's, $\frac{p}{2}-1$ indicator vectors that identify the matched-pair blocks, and the ordered baseline data z_1 . The vector denoted by X_2 consist of the observed value of L_2 , the treatment contrast. The matrix Q_D , conditioned on the baseline data, is thus the perpendicular projection on the space spanned by L_2 orthogonalized to the span of X_1 . Since $Var(z_2|z_1) = \tau^2(1-\psi^2)I_p$, then $\kappa = \tau^2(1-\psi^2)$. Therefore $SSR_D / (\tau^2(1-\psi^2)) \sim \chi^2_{(1,\lambda)}$ with non-centrality parameter $\lambda = \frac{1}{2\tau^2(1-\psi^2)}\mu'Q_D\mu$.

The product $Q_E\Gamma_{2,1}$ has $\frac{p}{2}-2$ non-zero eigenvalues since $r(Q_E) = r(Q_E\Sigma_{2,1}) = \frac{p}{2}-2$. Since $Var(z_2|z_1) = \tau^2(1-\psi^2)I_p$, $\kappa = \tau^2(1-\psi^2)$ is the one positive eigenvalue with multiplicity $\frac{p}{2}-2$. Thus $Q_E\Gamma_{2,1}$ is idempotent

and SSE/κ has a central chi-squared distribution with $P/2 - 2$ degrees of freedom.

The centrality is a result of the mean vector $\mu = X\beta$ being orthogonal to Q_E .

To summarize, $SSR_D / (\tau^2(1-\psi^2))$ follows a noncentral chi-squared distribution with 1 degree of freedom and noncentrality parameter $\lambda = \frac{1}{2\tau^2(1-\psi^2)} \mu' Q_D \mu$. Also $SSE / (\tau^2(1-\psi^2))$ follows a central chi-squared distribution with $P/2 - 2$ degrees of freedom. The variables SSR_D and SSE are independent since $Q_E Q_D = \mathbf{0}$. Hence the ratio

$$F = \frac{SSR_D / (\tau^2(1-\psi^2))}{SSE / (P/2 - 2)(\tau^2(1-\psi^2))} = \frac{SSR_D}{SSE / (P/2 - 2)}$$

follows a noncentral \mathcal{F} distribution with 1, $P/2 - 2$ degrees of freedom and noncentrality parameter λ .

4.3.6 Case 6: Pre-Post Differences With Matching

In Case 3 the treatment is assessed as a net difference of differences between baseline and posttest measures in the treatment group relative to the control group.

This allows for one adjustment for baseline measures. Here matching is added as another method of adjustment. The variance of differences is a function of both the variance of measures at each time point and the correlation between baseline and posttest measures. On one hand it might be expected that matching would not have

any beneficial effect on the variation after taking differences from baseline into account. As will be discussed in Chapter VI, matching does decrease the variance over using differences from baseline alone.

4.3.6.1 Variance of Treatment Effect

The treatment effect is the difference between treatment and control group pre-post mean differences contrasted within each matched-pair at time 2 and it can be written

as $L'_2(z_2 - z_1)$, where L_2 is defined in (4.3). Recall that $z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$, the ordered

baseline data over the concomitant posttest data. Using (4.17) to define the pre-post

differences, it follows that $\ell'_2 z$ is the vector of pre-post differences which

constitutes the data vector for this case. The treatment effect contrast is thus $L'_2 \ell'_2 z$.

The expectation of $\ell'_2 z$ is given by

$$\begin{aligned} E(\ell'_2 z) &= \mu_2 + \tau\psi\mathbf{c} - \mu_1 - \tau\mathbf{c} \\ &= \mu_2 - \mu_1 - \tau(1 - \psi)\mathbf{c} \end{aligned} \tag{4.36}$$

Since $\text{Var}(z) = \Gamma$ then from (4.8)

$$\begin{aligned} \text{Var}(\ell'_2 z) &= \ell'_2 \Gamma \ell_2 \\ &= \Gamma_{2-1} \\ &= \Gamma_{11} - 2\Gamma_{12} + \Gamma_{22} \\ &= \tau^2 \mathbf{C} - 2\tau^2 \psi \mathbf{C} + \tau^2 (1 - \psi^2) I_p + \tau^2 \psi^2 \mathbf{C} \\ &= \tau^2 (1 - \psi^2) I_p + \tau^2 (1 - \psi)^2 \mathbf{C} \end{aligned} \tag{4.37}$$

The conditional expectation of $L_2' \ell_2' z$ given the random allocation L_2 and (4.24)

is

$$\begin{aligned} E[L_2' \ell_2' z | L_2] &= L_2' (\mu_2 - \mu_1 - \tau(1-\psi)\mathbf{c}) \\ &= L_2' (\Delta\delta - \tau(1-\psi)\mathbf{c}) \\ &= \Delta - \tau(1-\psi)L_2'\mathbf{c} \end{aligned} \quad (4.38)$$

The conditional variance is

$$\begin{aligned} Var[L_2' \ell_2' z | L_2] &= L_2' \left(\tau^2(1-\psi^2)I_p + \tau^2(1-\psi)^2\mathbf{C} \right) L_2 \\ &= \tau^2(1-\psi^2)L_2'L_2 + \tau^2(1-\psi)^2 L_2'CL_2 \\ &= \frac{4}{p}\tau^2(1-\psi^2) + \tau^2(1-\psi)^2 L_2'CL_2 \end{aligned} \quad (4.39)$$

Recall that the expectation of L_2 is the zero vector and its variance is given by (4.4).

Using (4.5), (4.28), and noting that

$$\begin{aligned} Var(L_2'\underline{\mathbf{c}}) &= \underline{\mathbf{c}}'V\underline{\mathbf{c}}, \\ Var(L_2'CL_2) &= trace(VC) + E[L_2']CE[L_2], \\ \text{and } E[L_2] &= 0 \end{aligned}$$

We thus have that

$$\begin{aligned} Var[L_2' z_2] &= Var(\Delta - \tau(1-\psi)L_2'\mathbf{c}) + E\left(\frac{4}{p}\tau^2(1-\psi^2) + \tau^2(1-\psi)^2 L_2'CL_2\right) \\ &= \tau^2(1-\psi)^2 \mathbf{c}'V\mathbf{c} + \frac{4}{p}\tau^2(1-\psi^2) + \tau^2(1-\psi)^2 trace(VC) \\ &= \frac{4}{p}\tau^2(1-\psi^2) + \tau^2(1-\psi)^2 trace\left(V(C + \mathbf{c}\mathbf{c}')\right) \\ &= \frac{4}{p}\tau^2(1-\psi^2) + p\tau^2(1-\psi)^2 \omega_p \end{aligned} \quad (4.40)$$

4.3.6.2 Inference on Treatment Effect

The matrix X_1 is the vector of 1's and $\frac{p}{2}-1$ indicator vectors that identify the matched-pair blocks. The vector denoted by X_2 consists of the observed value of L_2 , the treatment contrast. The matrix Q_D , conditioned on the baseline data, is thus the perpendicular projection on the space spanned by L_2 orthogonalized to the span of X_1 . The data for this case are pre-post differences in the ordered and induced ordered means, $\ell'_2 z$. Since $Var(\ell'_2 z) = \Gamma_{2-1} = \tau^2(1-\psi^2)I_p + \tau^2(1-\psi)^2 C$, then by (4.11) $\kappa = \tau^2(1-\psi^2) + \tau^2(1-\psi)^2 \mathbf{u}' C \mathbf{u}$ where the span of the unit vector \mathbf{u} equals the span of Q_D . Therefore $SSR_D / \kappa \sim \chi^2_{(1,\lambda)}$ with non-centrality

$$\text{parameter } \lambda = \frac{1}{2\kappa} \mu' Q_D \mu.$$

The product $Q_E \Gamma_{2-1}$ has $\frac{p}{2}-1$ non-zero eigenvalues since $r(Q_E) = r(Q_E \Gamma_{2-1}) = \frac{p}{2}-1$. However it can be shown that these eigenvalues are distinct when $\tau^2(1-\psi)^2 > 0$. Thus no scalar times $Q_E \Gamma_{2-1}$ is idempotent. However, computations show that the average eigenvalue is equal to (or nearly equal to) κ that was found in the previous paragraph. Thus $Q_E \Gamma_{2-1}$ is approximately idempotent and SSE / κ will approximately follow a central chi-squared distribution with $\frac{p}{2}-1$ degrees of freedom. The centrality is a result of the mean vector $\mu = X\beta$ being orthogonal to Q_E .

To summarize, SSR_D / κ follows a noncentral chi-squared distribution with 1 degree of freedom and noncentrality parameter $\lambda = \frac{1}{2\kappa} \mu' Q_D \mu$. Also SSE / κ follows a approximate central chi-squared distribution with $P/2 - 1$ degrees of freedom. The variables SSR_D and SSE are not independent since $Q_E \Gamma_{2-1} Q_D \neq \mathbf{0}$; however, computations show that all terms in this product are very near to zero (approximately $\leq 1E-05$) and thus they are essentially uncorrelated.

Hence the ratio $F = \frac{SSR_D / \kappa}{SSE / \left(\frac{P}{2} - 1\right) \kappa} = \frac{SSR_D}{SSE / \left(\frac{P}{2} - 1\right)}$ follows an approximate

noncentral \mathcal{F} distribution with 1, $P/2 - 1$ degrees of freedom and noncentrality parameter λ .

CHAPTER V

Cluster Randomized Trial: 6-Case Scenarios

5.1 Introduction

The methods applied to the randomized controlled trials can be extended naturally to the cluster randomized trial. Although random assignment ensures that treatment groups will be, on average, balanced across baseline values, imbalance may occur by chance. Such imbalance will be greater in cluster randomized trials because clusters are the units of randomization and they are of a relatively smaller number than the total number of sampled subjects.

Notation and the necessary assumptions for CRT pre-post designs are given in Section 2. This section also presents the variance-covariance matrix of this design. Section 3 explains how to estimate the treatment effect and find a closed form for the estimate of the treatment effect for Case 1 to Case 6.

5.2 Cluster-Randomized Pre-Post Design

Suppose the population of interest in a study consists of a large number of subjects who naturally cluster together. For example the population may be the adult primary care healthcare seekers who are affiliated with a large health system. The adults naturally cluster into patient populations of individual primary care practices. An intervention aimed at the patients by affecting a change in the practice requires that the practice

(cluster) be treated as the unit of treatment assignment and that all of the patients in each practice are exposed to the assigned treatment.

The sampling process is two-staged. First a random sample of p clusters is selected from the study population. In order to adjust for differences in baseline measure, a cross-sectional sample of patients selected from each sampled cluster is used to estimate the cluster baseline outcome. Next the clusters are randomly assigned to study condition. After the intervention another cross-sectional sample of patients is selected and the posttest cluster outcome value is estimated. This design is summarized in Figure 5.1

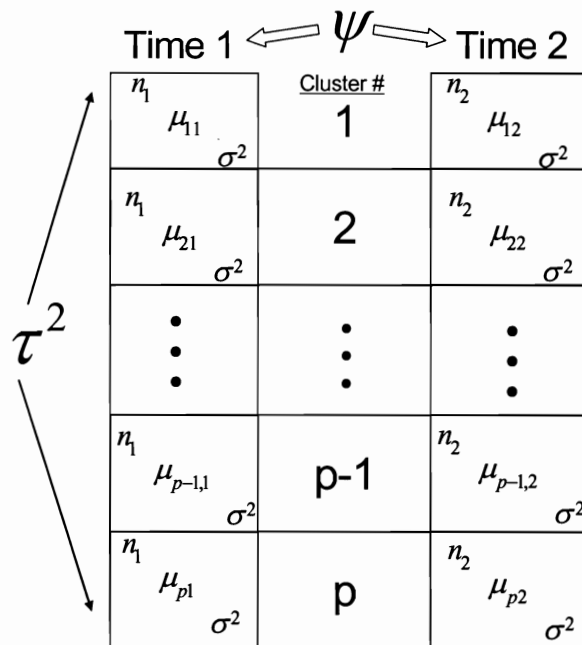


Figure 5.1: Cluster Parameters, Variation, and Time 1/2 Correlation

The variance between population cluster means is denoted by τ^2 . Given a cluster, the variance between subjects within a cluster is denoted by σ^2 . If a subject is taken at

random from the full population, then the (total) variance of the measure is $\tau^2 + \sigma^2$. This may be shown with an application of (3.3):

$$\begin{aligned} E(y|cluster) &= x \\ Var(y|cluster) &= \sigma^2 \\ Var(y) &= Var(E(y|cluster)) + E(Var(y|cluster)) = \tau^2 + \sigma^2 \end{aligned} \quad (5.1)$$

The variance of the within cluster sample mean can be found in similar fashion:

$$Var(\bar{y}) = \tau^2 + \sigma^2 / n_j \quad (5.2)$$

where n_j is the within cluster sample size at time j ($=1,2$). Note that the inclusion of τ^2 in (5.2) is predicated on the fact that clusters are chosen at random. If a set of clusters are purposely chosen (same set every time the experiment is—conceptually—repeated) or if all of the clusters are chosen, then τ^2 is dropped from (5.2).

A common measure of the relative degree of between cluster variation is the (induced) intraclass correlation (ICC). Defined as

$$\rho = \frac{\tau^2}{\tau^2 + \sigma^2} \quad (5.3)$$

the ICC is the proportion of total variance explained by the between cluster variation.

Note that

$$\begin{aligned} \tau^2 &= (\tau^2 + \sigma^2)\rho \\ \sigma^2 &= (\tau^2 + \sigma^2)(1 - \rho) \end{aligned} \quad (5.4)$$

Substituting (5.4) into (5.2) the variance of the sample mean is given by

$$Var(\bar{y}) = \frac{\tau^2 + \sigma^2}{n_j} (1 + (n_j - 1)\rho) \quad (5.5)$$

The term $(1 + (n_j - 1)\rho)$ is called the *design effect* or the *variance inflation factor*. If n_j subjects are sampled at random from the full population then the variance of the sample mean would be given by $\frac{\tau^2 + \sigma^2}{n_j}$. The design effect shows how the variance is inflated due to the fact that the sampling design samples clusters first and then patients within clusters.

For this work it is assumed that no inherent correlation exists between subjects within the same cluster. However, because the clusters are randomly sampled at the first stage, a correlation is induced on the subjects within a cluster. Given the clusters are randomly sampled, the covariance between two subjects in the same cluster is τ^2 . The (total) variance of the measure from a subject within the cluster is $\tau^2 + \sigma^2$ (5.1). The proportion of total variance explained by the variance between two subjects in the same cluster is thus $\frac{\tau^2}{\tau^2 + \sigma^2}$. Note that this is the (induced) intraclass correlation, ICC, as previously defined in (5.3).

The results in the paragraph above can be shown using a generalization of (3.3):

$$Cov(y_i, y_j) = E(Cov(y_i, y_j | cluster)) + Cov(E(y_i | cluster), E(y_j | cluster)) \quad (5.6)$$

Assume that (y_i, y_j) is a pair of observations drawn from the same cluster. Given the cluster, these observations are independent and both have (conditional) expectation equal to the cluster mean, say x . Thus

$$\begin{aligned} Cov(y_i, y_j) &= Cov(x, x) \\ &= Var(x) \\ &= \tau^2 \end{aligned}$$

In similar fashion the covariance between cluster means at time 1 and 2 can be expressed. Denote the sample mean for the i^{th} cluster at time j as \bar{y}_{ij} . Using (5.6) and noting that, given the cluster, the means at the two time points are independent,

$$\begin{aligned} Cov(\bar{y}_{i1}, \bar{y}_{i2}) &= Cov(x_1, x_2) \\ &= \tau^2 \psi \end{aligned} \tag{5.7}$$

In summary we assume the cluster sample means are normally distributed

$$\begin{aligned} \bar{y}_{ij} &\sim N\left(\mu_{ij}, \tau^2 + \frac{\sigma^2}{n_j}\right); i = 1, \dots, p; j = 1, 2 \\ Cov(\bar{y}_{ij}, \bar{y}_{i'j'}) &= \begin{cases} \tau^2 \psi & \text{if } i = i' \text{ and } j \neq j' \\ 0 & \text{if } i \neq i' \end{cases} \end{aligned}$$

Furthermore we define two terms.

1. The variance of the cluster sample mean at time j : $\sigma_j^2 = \tau^2 + \frac{\sigma^2}{n_j}$ and
2. The correlation between cluster sample means at time 1 and 2: $\lambda_{12} = \frac{\tau^2 \psi}{\sigma_1 \sigma_2}$.

5.3 Variance of Treatment Effect for the Six Cases

In this section the variances associated with the six cases developed in Chapter IV will be modified to reflect the parameters of the CRT. Basically certain substitutions are made into the Chapter IV formulas. For each time point the variance, τ^2 , used in the RCT is replaced with σ_j^2 , ψ is replaced with λ_{12} , and the sample size n is replaced with n_j .

5.3.1 Case 1: Posttest Data With No Matching

Only the time 2 variance is involved, thus simple substitution using (4.12) yields

$$Var = \frac{4}{p} \sigma_2^2 \quad (5.8)$$

5.3.2 Case 2: Posttest Data, Baseline as Covariable, No Matching

The conditional covariance matrix, with the substitutions, becomes

$$\begin{aligned} \Sigma_{2 \cdot 1} &= \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \\ &= \sigma_2^2 I_p - (\sigma_1 \sigma_2 \lambda_{12} I_p) (\sigma_1^{-2} I_p) (\sigma_1 \sigma_2 \lambda_{12} I_p) \\ &= \sigma_2^2 (1 - \lambda_{12}^2) I_p \end{aligned} \quad (5.9)$$

Thus the variance is given by

$$\begin{aligned} Var &= \ell_t' \Sigma_{2 \cdot 1} \ell_t \\ &= \frac{4}{p} \sigma_2^2 (1 - \lambda_{12}^2) \end{aligned} \quad (5.10)$$

5.3.3 Case 3: Pre-Post Difference, No Matching

The variance of the pre-post differences, with the substitutions, is

$$\begin{aligned}
Var(\ell'_2 y) &= \ell'_2 \Sigma \ell_2 \\
&= \Sigma_{2-1} \\
&= \Sigma_{11} - 2\Sigma_{12} + \Sigma_{22} \\
&= \sigma_1^2 I_p - 2\tau^2 \psi I_p + \sigma_2^2 I \\
&= 2\tau^2 (1 - \psi) + \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)
\end{aligned}$$

Thus the variance is given by

$$\begin{aligned}
Var &= Var(\ell'_i \ell'_2 y) \\
&= \frac{8}{p} \tau^2 (1 - \psi) + \frac{4}{p} \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)
\end{aligned} \tag{5.11}$$

5.3.4 Case 4: Posttest Data With Matching

In the conditional expectations of (4.25) and (4.26), the variance terms referred to the time 2 period. Therefore with simple substitutions we get the variance to be

$$Var = \frac{4}{p} \sigma_2^2 (1 - \lambda_{12}^2) + p \sigma_2^2 \lambda_{12}^2 \omega_p \tag{5.12}$$

5.3.5 Case 5: Posttest Data, Baseline as Covariable, With Matching

The conditional covariance matrix, with substitutions, is the same as the results in (5.9).

Thus with substitutions, the conditional variance is

$$Var_{conditional} = \lambda_{12}^2 z_1' V z_1 + \frac{4}{p} \sigma_2^2 (1 - \lambda_{12}^2) \tag{5.13}$$

Recall that the expected conditional variance for Case 5 is the same as the variance for Case 4. Therefore the expected conditional variance is given by

$$Var_{expected\ conditional} = \frac{4}{p} \sigma_2^2 (1 - \lambda_{12}^2) + p \sigma_2^2 \lambda_{12}^2 \omega_p \quad (5.14)$$

5.3.6 Case 6: Pre-Post Differences With Matching

The expectation of the pre-post differences, with substitutions is

$$\begin{aligned} E(\ell'_2 z) &= \mu_2 + \sigma_2 \lambda_{12} \mathbf{c} - \mu_1 - \sigma_1 \mathbf{c} \\ &= \mu_2 - \mu_1 - (\sigma_1 - \lambda_{12} \sigma_2) \mathbf{c} \end{aligned} \quad (5.15)$$

The variance of the pre-post differences, with substitutions, is

$$\begin{aligned} Var(\ell'_2 z) &= \ell'_2 \Gamma \ell_2 \\ &= \Gamma_{2-1} \\ &= \Gamma_{11} - 2\Gamma_{12} + \Gamma_{22} \\ &= \sigma_1^2 \mathbf{C} - 2\sigma_1 \sigma_2 \lambda_{12} \mathbf{C} + \sigma_2^2 (1 - \lambda_{12}^2) I_p + \sigma_2^2 \lambda_{12}^2 \mathbf{C} \\ &= \sigma_2^2 (1 - \lambda_{12}^2) I_p + (\sigma_1 - \lambda_{12} \sigma_2)^2 \mathbf{C} \end{aligned} \quad (5.16)$$

The conditional expectation of $L'_2 \ell'_2 z$ given the random allocation L_2 and (4.22) is

$$\begin{aligned} E[L'_2 \ell'_2 z | L_2] &= L'_2 (\mu_2 - \mu_1 - (\sigma_1 - \lambda_{12} \sigma_2) \mathbf{c}) \\ &= L'_2 (\Delta \delta - (\sigma_1 - \lambda_{12} \sigma_2) \mathbf{c}) \\ &= \Delta - (\sigma_1 - \lambda_{12} \sigma_2) \mathbf{c} L'_2 \mathbf{c} \end{aligned} \quad (5.17)$$

The conditional variance is

$$\begin{aligned} Var[L'_2 \ell'_2 z | L_2] &= L'_2 (\sigma_2^2 (1 - \lambda_{12}^2) I_p + (\sigma_1 - \lambda_{12} \sigma_2)^2 \mathbf{C}) L_2 \\ &= \sigma_2^2 (1 - \lambda_{12}^2) L'_2 L_2 + (\sigma_1 - \lambda_{12} \sigma_2)^2 L'_2 \mathbf{C} L_2 \\ &= \frac{4}{p} \sigma_2^2 (1 - \lambda_{12}^2) + (\sigma_1 - \lambda_{12} \sigma_2)^2 L'_2 \mathbf{C} L_2 \end{aligned} \quad (5.18)$$

Using (4.5) and (4.28) we thus have that

$$\begin{aligned}
Var &= Var\left(\Delta - (\sigma_1 - \lambda_{12}\sigma_2)\mathbf{c}L_2'\mathbf{c}\right) + E\left(\frac{4}{p}\sigma_2^2(1 - \lambda_{12}^2) + (\sigma_1 - \lambda_{12}\sigma_2)^2 L_2'CL_2\right) \\
&= (\sigma_1 - \lambda_{12}\sigma_2)^2 \mathbf{c}'V\mathbf{c} + \frac{4}{p}\sigma_2^2(1 - \lambda_{12}^2) + (\sigma_1 - \lambda_{12}\sigma_2)^2 trace(VC) \\
&= \frac{4}{p}\sigma_2^2(1 - \lambda_{12}^2) + (\sigma_1 - \lambda_{12}\sigma_2)^2 trace\left(V\left(C + \mathbf{c}\mathbf{c}'\right)\right) \\
&= \frac{4}{p}\sigma_2^2(1 - \lambda_{12}^2) + p(\sigma_1 - \lambda_{12}\sigma_2)^2 \omega_p
\end{aligned}$$

Thus the variance is given by using (4.38) and (5.14) in (4.40)

$$Var = \frac{4}{p}\left[\sigma_2^2(1 - \lambda_{12}^2) + (\sigma_1 - \lambda_{12}\sigma_2)^2 \omega_p\right] \quad (5.19)$$

CHAPTER VI

Results

6.1 Introduction

In this chapter results are presented of various comparisons aimed at uncovering which of the six cases yields the smallest treatment effect variance and the higher power. Focus begins on the variance of the estimated treatment effect and comparison of the classic cases, 1-3. Two approaches are taken, one from the point of view of an analyst who has been presented with collected data and the other from the point of view of a consultant who is giving advice on which design (case) to use. The key difference is that the consultant may decide about using a pre-post design or a post only design and allocate resources accordingly.

We next consider changes in the variance when adding paired blocking to the design. Here we compare Cases 1 to 4, 2 to 5, and 3 to 6 by computing the difference in variance and the conditions for which one case has the smaller variance. While the variance may be less for a given case, the difference may be small. The size of the

variance affects the magnitude of the power to detect treatment differences, but the degrees of freedom also play a role. We compare the power of selected cases under varying assumptions and sample sizes.

Lastly we explore for Cases 2-6 the optimal allocation of within cluster sample size across the two time points. The optimal allocation is different from case to case. Under optimal allocation we again compare the variance to discover if the relationships found earlier still hold.

6.2 Comparison of Cases 1, 2, and 3

Cluster randomized trials often employ only limited number of clusters as a unit of assignment. Thus the known methods such as matching, stratification, pre-post differences, and adjustment for covariates can be used to improve the efficiency of design. In this section we consider only completely randomized trials (not matched cases) and focus on most common methods found in literature. We first take a random sample of clusters and randomize these clusters to treatment and control. A random sample of subjects is selected from each cluster and baseline measures are collected. The treatment

is applied and we collect post-intervention measures. Thus we have the data from CRT with pre-post design to analyze.

Our question is what is the best way to analyze these data? We compare three methods on the variance of the estimated treatment effect. As a reminder, the three cases, which are the most common methods found in the literature, are:

Case 1: post data alone

Case 2: post data with adjustment for baseline as a covariate

Case 3: difference of pre-post measures

First, consider Case 1—post data only. The major weakness of this approach is that the baseline data are ignored. Any variation not attributable to study conditions is left to residual error. Case 1 will lead to a comparatively large variance of treatment-control contrasts and less power.

Second, Case 2 is an adjusted analysis where the collection of baseline cluster means is used as a covariate to improve the precision of the estimate of the treatment effect. This method differs from the Case 1 (unadjusted analysis) by addition of baseline measures as a covariate and by the loss of one degree of freedom. The post-intervention

cluster means are adjusted by regression on their baseline means. These adjusted observations will generally have a smaller variance than the unadjusted observations.

Third, Case 3 deals with differences of pre-post measures. These differences are contrasted between the treatment group relative to the control group—a net difference of differences—to assess the treatment effect. Murray (1998) asserts that this analysis often has less power than either Case 1 or Case 2 analysis. The degrees of freedom for condition is the same as Case 1 but the treatment effect in this analysis involves a net difference (difference of differences).

6.2.1 Comparison of Variances between Cases 1, 2 and 3

The analytic differences in variances between Cases 1, 2, and 3 are shown in Table 6.1.

We first assume that n_1 subjects have been sampled for each cluster and that baseline data have been collected. The sample size collected post-intervention is n_2 . Recall from

Chapter 5, equations (5.6)-(5.9) that the variances for Cases 1-3 are:

$$\text{Case 1: } \frac{4}{p} \sigma_2^2, \text{ where } \sigma_2^2 = \tau^2 + \frac{\sigma^2}{n_2}.$$

$$\text{Case 2: } \frac{4}{p} \sigma_2^2 (1 - \lambda_{12}^2), \text{ where } \lambda_{12} = \frac{\tau^2 \psi}{\sigma_1 \sigma_2}.$$

$$\text{Case 3: } \frac{4}{p} [\sigma_1^2 + \sigma_2^2 - 2\tau^2 \psi], \text{ where } \sigma_1^2 = \tau^2 + \frac{\sigma^2}{n_1}$$

Table 6.1: Differences in Variances between Cases 1, 2, and 3

Cases	Difference	Results
1 versus 2 (2-1)	$-\frac{4}{p}\sigma_2^2\lambda_{12}^2 \leq 0$	Case 2 variance is smaller when $\lambda_{12} \neq 0$
2 versus 3 (3-2)	$\frac{4}{p}[(\sigma_1 - \lambda_{12}\sigma_2)^2 + \sigma_2^2\lambda_{12}(1 - \lambda_{12})] > 0$	Case 2 variance is smaller
1 versus 3 (3-1)	$\sigma_1^2 - 2\tau^2\psi$	Case 3 variance is smaller when $\psi > \frac{1}{2} + \frac{\sigma^2}{2n_1\tau^2}$

Clearly Case 2 has the smallest variance. Details of the comparisons follow.

6.2.1.1 Case 1 versus Case 2

The variance for Case 2 is always less than for Case 1 since the difference $-\frac{4}{p}\sigma_2^2\lambda_{12}^2$ is negative. As expected, this difference diminishes when the correlation is small.

6.2.1.2 Case 1 versus Case 3

The variance for Case 3 is not always an improvement over Case 1. The variance of a difference involves the sum of the variances of the two components which must be overcome via the correlation. Here the correlation must be such that $\psi > \frac{1}{2} + \frac{\sigma^2}{2n_1\tau^2}$. This formula can be rewritten in terms of the ICC, ρ : $\psi > \frac{1}{2} + \frac{1-\rho}{2n_1\rho}$. With small ICC, the

correlation must be large to overcome the sum of the variances. Since the correlation is bounded above by 1, it can be seen that when ρ is less than $\frac{1}{n_1 + 1}$ the variance for Case 1 will be no larger than the variance for Case 3. When ρ is close to but above this value then ψ must be very large (close to 1) to make the variance for Case 3 less than for Case 1.

6.2.1.3 Case 2 versus Case 3

The variance of Case 2 is always less than for Case 3. The difference $\frac{4}{p}[(\sigma_1 - \lambda_{12}\sigma_2)^2 + \sigma_2^2\lambda_{12}(1 - \lambda_{12})]$ is positive, though will be small when the correlation is near 1.

6.2.2 Discussion

In summary we found, as expected, that Case 2 has a smaller variance than Case 1 and Case 3. Also for sufficiently large correlation, Case 3 will have a smaller variance than Case 1. Case 2 and Case 3 are similar for large correlation. For Case 2, the improvement in variance could be offset by an extra degree of freedom, but with the adequate sample

size, this should not make a difference. These results are consistent with other earlier studies (i.e. Murray, 1998 p 143).

Intuitively, we are saying that the time 1 and 2 measures have to be highly correlated in order for an analysis which uses the time 1 data to be an improvement over Case 1. Note that for both Cases 2 and 3 the correlation ψ comes into play through the product $\tau^2\psi$. Thus for the improvement to reduce the variance, there must exist a non-negligible variation between clusters relative to the variation within clusters. Note that the relative variation (between clusters versus within clusters) is represented by ρ (ρ is a function of this relative variance). The larger ρ , the greater the relative variance.

There may be situations where collecting baseline data is not feasible. Then—among three cases—Case 1 is the only choice.

6.2.3 Preferred Analysis from the Design Stage

In the preceding discussion it was given that both baseline data and posttest data were collected. Ignoring the baseline data by performing a Case 1 analysis was not recommended.

Suppose now that a consultant is to propose some guidelines for selecting a design that has obvious merit in being able to identify the optimal design among these three cases. Case 2 or Case 3 may not be always the best approach. If the baseline data is not required, then all resources could be put into collecting posttest data. So while Cases 2 and 3 must be compared with part of the data being collected at baseline, Case 1 may have all data collected post test. For further clarity, denote the full sample as $N (=n_1 + n_2)$.

Then the variances for Cases 1-3 are:

$$\text{Case 1: } \frac{4}{p} \left(\frac{\sigma^2}{N} + \tau^2 \right)$$

$$\text{Case 2: } \frac{4}{p} \sigma_2^2 (1 - \lambda_{12}^2), \text{ where } \lambda_{12} = \frac{\tau^2 \psi}{\sigma_1 \sigma_2} \text{ and } \sigma_i^2 = \tau^2 + \frac{\sigma^2}{n_i}.$$

$$\text{Case 3: } \frac{4}{p} [\sigma_1^2 + \sigma_2^2 - 2\tau^2 \psi]$$

Table 6.2: Differences in Variances: Comparing Case 1 with Cases 2 and 3

Cases	Difference	Results
1 versus 2 (1-2)	$\frac{4}{p} \left[\frac{n_2 - (1 - \lambda_{12}^2)N}{Nn_2} \sigma^2 + \tau^2 \lambda_{12}^2 \right]$	Case 2 variance is smaller, when $\psi^2 \geq \frac{\sigma_1^2 \sigma_2^2 n_1 \sigma^2}{\tau^4 N(\sigma^2 + n_2 \tau^2)}$ If $n_1 = n_2 = n$, then $\psi \geq \frac{1}{2} \frac{\sqrt{2(1+n\mathcal{R})}}{n\mathcal{R}},$ where $\mathcal{R} = \frac{\rho}{1-\rho}$
1 versus 3 (1-3)	$\frac{4}{p} \left[\left(\frac{1}{N} - \frac{1}{n_1} - \frac{1}{n_2} \right) \sigma^2 - \tau^2 (1 - 2\psi) \right]$	Case 3 variance is smaller when $\psi > \frac{1}{2} + \left(\frac{1}{n_1} + \frac{1}{n_2} - \frac{1}{N} \right) \frac{1}{2\mathcal{R}}$ If $n_1 = n_2 = n$, then $\psi \geq \frac{3}{4n\mathcal{R}} + \frac{1}{2}$

6.2.3.1 Case 1 versus Case 2

The difference between the variances for Cases 1 and 2 can be shown to be

$$\frac{4}{p} \left[\frac{n_2 - (1 - \lambda_{12}^2)N}{Nn_2} \sigma^2 + \tau^2 \lambda_{12}^2 \right] \quad (6.1)$$

For this to be a positive value, which implies Case 2 variance to be smaller than Case 1,

the following must hold

$$\psi^2 \geq \frac{\sigma_1^2 \sigma_2^2 n_1 \sigma^2}{\tau^4 N(\sigma^2 + n_2 \tau^2)} \quad (6.2)$$

To facilitate understanding, consider the special case when $n_1 = n_2 = n$. Thus

$\sigma_1^2 = \sigma_2^2 = \sigma_*^2 = \frac{\sigma^2}{n} + \tau^2$. Then (6.2) can be simplified as

$$\psi \geq \frac{1}{2} \frac{\sqrt{2(1+n\mathcal{R})}}{n\mathcal{R}} \quad (6.3)$$

where $\mathcal{R} = \frac{\rho}{1-\rho} = \frac{\tau^2}{\sigma^2}$ is the ratio of the between and within cluster variances.

Since the correlation is bounded above by 1, then, with some algebra, it can be shown

that Case 2 cannot have a smaller variance than Case 1 unless $\rho > \frac{1}{n+1}$ and (6.3) holds.

6.2.3.2 Case 1 versus Case 3

The difference between the variances for Cases 1 and 3 can be shown to be

$$\frac{4}{p} \left[\left(\frac{1}{N} - \frac{1}{n_1} - \frac{1}{n_2} \right) \sigma^2 - \tau^2 (1 - 2\psi) \right] \quad (6.4)$$

For this to be a positive value, which implies Case 3 variance to be smaller than Case 1,

the following must hold

$$\psi > \frac{1}{2\tau^2} \left[\left(\frac{1}{n_1} + \frac{1}{n_2} - \frac{1}{N} \right) \sigma^2 + \tau^2 \right] = \frac{1}{2} + \left(\frac{1}{n_1} + \frac{1}{n_2} - \frac{1}{N} \right) \frac{1}{2\mathcal{R}} \quad (6.5)$$

Consider the special case when $n_1 = n_2 = n$. Then (6.5) can be simplified as

$$\psi \geq \frac{3}{4n\mathcal{R}} + \frac{1}{2} \quad (6.6)$$

Since the correlation is bounded above by 1, then, with some algebra, it can be shown that Case 3 cannot have a smaller variance than Case 1 unless $\rho > \frac{3}{2n+3}$ and (6.6) holds.

6.2.4 Summary

In summary we considered two scenarios: (i) perform the Case 1 analysis given that we have collected the baseline data and the time two sample size is fixed as is or (ii) we are in the design phase and may decide to put all of our resources into the post-intervention data collection and use a Case 1 analysis or collect some baseline data and use one of the other analyses. In the first scenario Case 2 is the best choice, in terms of variance. However in the second scenario a smaller variance may be achieved with Case 1 unless the correlation is sufficiently large.

6.3 Comparison of Variances between Cases With and Without Matching

The purpose of pair-matching in the design of a CRT is to improve the power for detecting treatment effects. Matching is achieved by sorting the clusters on the baseline means, creating pairs of clusters having adjacent baseline means, and randomizing to treatment or control within the cluster pair. We then may use one of the three analyses

previously described. In the discussion that follows the comparison of cases with and without matching, specifically Cases 1-4, 2-5, 3-6, will be presented.

The variances for Cases 1-6 are shown below in Table 6.3:

Table 6.3: The Variances for Cases 4-6

<i>No matching</i>	Case 1	$\frac{4}{p}\sigma_2^2$
	Case 2	$\frac{4}{p}[\sigma_2^2(1-\lambda_{12}^2)]$
	Case 3	$\frac{4}{p}\left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\sigma^2 + 2\tau^2(1-\psi)\right] = \frac{4}{p}[\sigma_1^2 + \sigma_2^2 - 2\tau^2\psi]$
<i>Matching</i>	Case 4	$\frac{4}{p}\sigma_2^2(1-\lambda_{12}^2) + p\sigma_2^2\lambda_{12}^2\omega_p$, where $\omega_p = \frac{\text{trace}[V(C+cc')]}{p}$
	Case 5 [†]	$\frac{4}{p}\sigma_2^2(1-\lambda_{12}^2) + p\sigma_2^2\lambda_{12}^2\omega_p$
	Case 6	$\frac{4}{p}\sigma_2^2(1-\lambda_{12}^2) + p(\sigma_1 - \lambda_{12}\sigma_2)^2\omega_p$

[†] The expected conditional variance is shown for Case 5

Table 6.4: Differences in Variances between Cases 1-4, 2-5, and 3-6

Cases	Difference	Results
1 versus 4 (1-4)	$\frac{4}{p}\sigma_2^2\lambda_{12}^2\left[1 - \frac{p^2}{4}\omega_p\right] > 0$	Case 4 variance is smaller
2 versus 5 (2-5)	$-p\sigma_2^2\lambda_{12}^2\omega_p < 0$	Case 2 variance is smaller
3 versus 6 (3-6)	$\frac{4}{p}(\sigma_1 - \lambda_{12}\sigma_2)^2(1 - p\omega_p) > 0$	Case 6 variance is smaller

6.3.1 Post data only: Case 1 and Case 4

Case 1 is focused only on post data with no adjustment for baseline measures. Case 4 incorporates pairwise matching using baseline as the matching variable. Does matching help? The difference between Case 4 and Case 1 variances is given by

$$\frac{4}{p} \sigma_2^2 \lambda_{12}^2 \left[1 - \frac{p^2}{4} \omega_p \right] \quad (6.7)$$

The quantity ω_p is a function of the sum of second moments of the differences in adjacent standard normal order statistics. As such it is a small positive value that converges to zero as p increases. The factor $1 - \frac{p^2}{4} \omega_p$ in (6.7) converges to 1 from below as p increases. This is illustrated in Figure 6.1 for $p=2, \dots, 50$ by 2. Thus (6.7) is positive which implies that the variance of Case 4 is smaller than the variance of Case 1. Note that this comparison assumes equal sample size for the matched and unmatched. However ‘cost’ would not be the same though, because matching requires a baseline assessment.

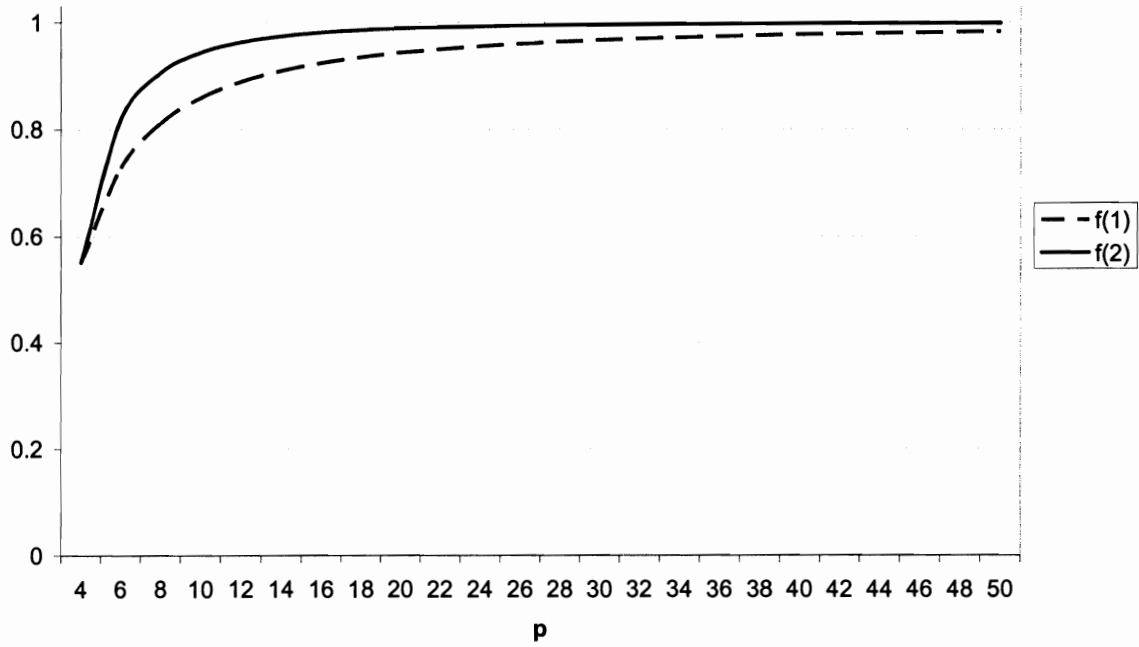


Figure 1: Functions of ω_p by p $f(1)=1-\frac{p^2}{4}\omega_p$ $f(2)=1-\omega_p$

6.3.2 Post Data with Baseline as Covariable: Case 2 and Case 5

Another way of adjusting for baseline to improve the power of the test is to use the baseline cluster means as a covariable. Case 2 adopts this method and Case 5 incorporates both matching and adjusting for baseline. Does matching in addition to adjusting decrease the variance? The difference between Case 5 and Case 2 variances is given by

$$-p\sigma_2^2\lambda_{12}^2\omega_p \quad (6.8)$$

This value is negative which implies that the variance of Case 5 is larger than the variance of Case 2.

6.3.3 Pre-post differences: Case 3 and Case 6

Both Cases 3 and 6 analyze the differences in cluster means between two time points. The difference between the two methods is that Case 6 incorporates additional pair-wise matching. The difference between Case 6 and Case 3 variances is given by

$$\frac{4}{p}(\sigma_1 - \lambda_{12}\sigma_2)^2(1 - p\omega_p) \quad (6.9)$$

The factor $1 - p\omega_p$ in (6.9) converges to 1 from below as p increases. This is illustrated in Figure 6.1 for $p=2, \dots, 50$ by 2. Thus (6.9) is positive which implies that the variance of Case 6 is smaller than the variance of Case 3.

For the comparison between Case 4 and case 5, the difference is

$\frac{\sigma_2^2 \lambda_{12}^2}{\sigma_1^2} [z_1' V z_1 - \sigma_1^2 \text{trace}[V(C + cc')]]$. This is a random variable since it depends on the baseline measures z_1 . Notice also that $E[z_1' V z_1]$ is $\sigma_1^2 \text{trace}[V(C + cc')]$. Thus we use the expected conditional variance for Case 5, then variances in Case 4 and 5 are the same.

6.3.4 Discussion of Variance Comparisons

Blocking helps to control the variation between clusters. If the variation between clusters is relatively small (ICC is small) then blocking may have little effect in reducing the variance. Even if the ICC is relatively large, we must have a ‘significantly’ large correlation between the time 1 and time 2 measures in order for any method that adjusts for baseline to effectively reduce the variance. This is seen more readily by looking at the variance between clusters relative to the variance of a cluster mean; that is, τ^2 relative to (divided by) $\tau^2 + \sigma^2/n$. When n is large (approaches infinity) then this variance converges to τ^2 and the ratio to 1. When n is small, the second component (σ^2/n) is a significant portion of $\tau^2 + \sigma^2/n$. The effect of the correlation can be expressed in some fashion for all of the cases in comparison to Case 1. For example, in comparing Case 2 to Case 1, look at the variance term for Case 2: $(\tau^2 + \sigma^2/n)(1 - \lambda_{12}^2)$. Small n makes the second term, $(1 - \lambda_{12}^2)$, larger because $\lambda_{12} = \frac{\tau^2}{\tau^2 + \sigma^2/n} \psi = \frac{n\rho}{1 + (n-1)\rho} \psi$ monotonically converges to ψ in n . This reduces the effect of the correlation ψ even when it is large.

We need large ψ , large ICC (ρ), and large n for Case 2 to have a lower variance in the cluster case such requirements will lead to an inflated design effect.. This seems counter to conventional recommendations. Donner (2000) and others purport that

in a CRT a small design effect is desired. Thus one should have small sample sizes within clusters, small ICC, and a large number of clusters. The latter is the key. With a large number of clusters (large p) the variance is greatly reduced for each case. When the ICC is small, the between cluster variation is relatively small and thus it is less important to control it. However when the ICC is small, the within cluster variation is relatively large (relative to the between variation) and an increased sample size within cluster is advantageous.

In subject case (RCT), the difference of variance between Case 1 and Case 2 is

$$\frac{4}{p}\tau^2 - \frac{4}{p}\tau^2(1-\psi^2) = \frac{4}{p}\tau^2\psi^2 > 0 .$$

Variance for Case 1 is always larger than variance for

Case 2. However in the cluster case (CRT) the magnitude of the variance depends on additional parameters and no case is globally the optimal one.

6.4 Power of test

Previously we explored our six cases by making comparisons between them in terms of variance. In addition the comparison between six cases can be re-examined from the point of the power of a test. For an illustration we set several different values for each parameter (N , p , ψ and ρ) involved in CRT and see how these parameter values are

related to each other and also effect the power of the test. We here present four schemes that have the most prominent features. Assuming total sample size is fixed at 1000, 500 per study period, we consider number of clusters (p) of 10 and 50 for the two different RCT settings. Then number of subjects in each cluster, where $p=10$ and 50, would be 50 and 10 respectively. We set the total variance to one, thus making $\tau^2 + \sigma^2 = 1$ and $\rho = \tau^2$. We consider correlation (ψ) between two time points of 0.2 (low) and 0.8 (high). Correlation within cluster (ρ) of 0.01 (low) and 0.1 (relatively high) is also considered. To view the effect of allocating all sample size resources to time 2, we divide Case 1 into two parts. Case 1A has the same time 2 sample size as the other cases. Case 1B pools all the sample size resources into time 2. To summarize the preceding information, Table 6.5 below assigns labels to the four notable schemes.

Table 6.5: 4 Schemes Considered for Comparison

P	N	n	ψ	ρ	
10	100	50	0.2	0.01	Figure 6.2
			0.8	0.1	Figure 6.4
50	20	10	0.2	0.01	Figure 6.3
			0.8	0.1	Figure 6.5

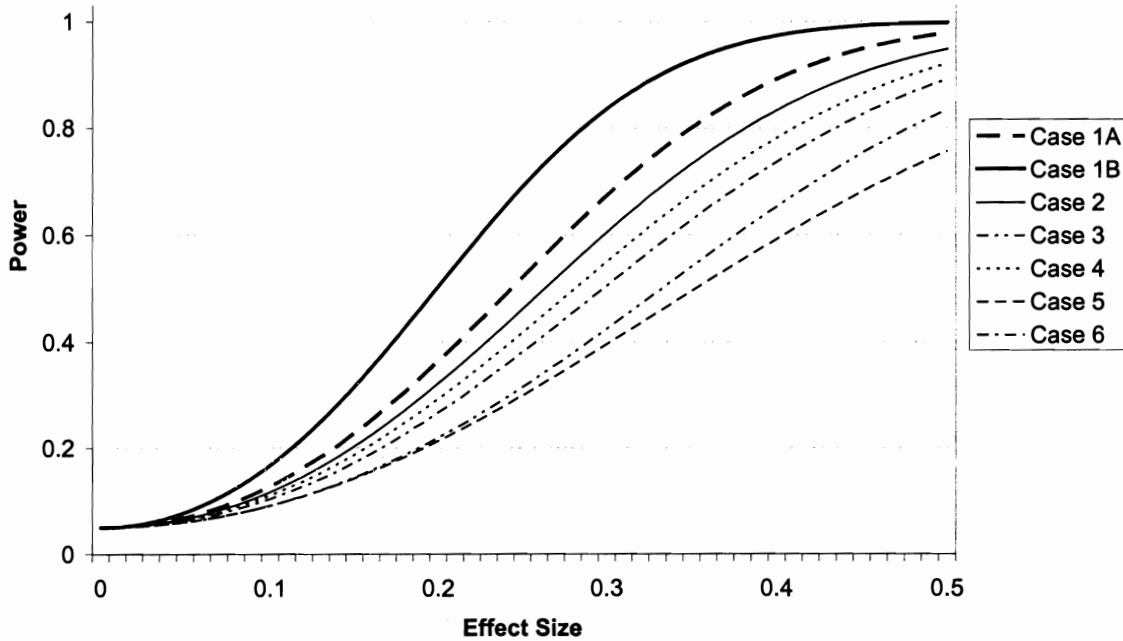


Figure 6.2 Power Curves for 7 case scenarios assuming $p=10$ $N=50$ $\rho = 0.01$ $\psi = 0.2$

The power curves in descending order correspond to Cases 1B, 1A, 2, 4, 6, 3, and 5. This has low correlation between measures in two time points ($\psi = 0.2$) and low correlation within cluster ($\rho = 0.01$), which means τ^2 (variation between clusters) is small. The quantity $1 - \lambda_{12}^2 = 0.995$ and thus there is essentially no correction to the Case 1 variance due to correlation between time 1 and 2. Since Case 2 has $p - 3 = 7$ degrees of freedom, one less than Case 1, then Case 2's power is diminished relative to Case 1. The variances for Cases 4-6 will be similar to Case 2 but these cases have lost even more degrees of

freedom (Case 4; df=4, Case 5; df=3, Case 6; df=4). Case 1B takes advantage of σ^2

being the dominating variance component and large $N=100$ that effectively controls σ^2 .

Note that the variances in 1A and 1B are $\frac{4}{p}(\tau^2 + \frac{\sigma^2}{n_2}) = \frac{4\tau^2}{p} + \frac{\sigma^2}{500}$ and

$\frac{4}{p}(\tau^2 + \frac{\sigma^2}{N}) = \frac{4\tau^2}{p} + \frac{\sigma^2}{1000}$ respectively. Thus the power is altered only through changes

in p and τ^2 .

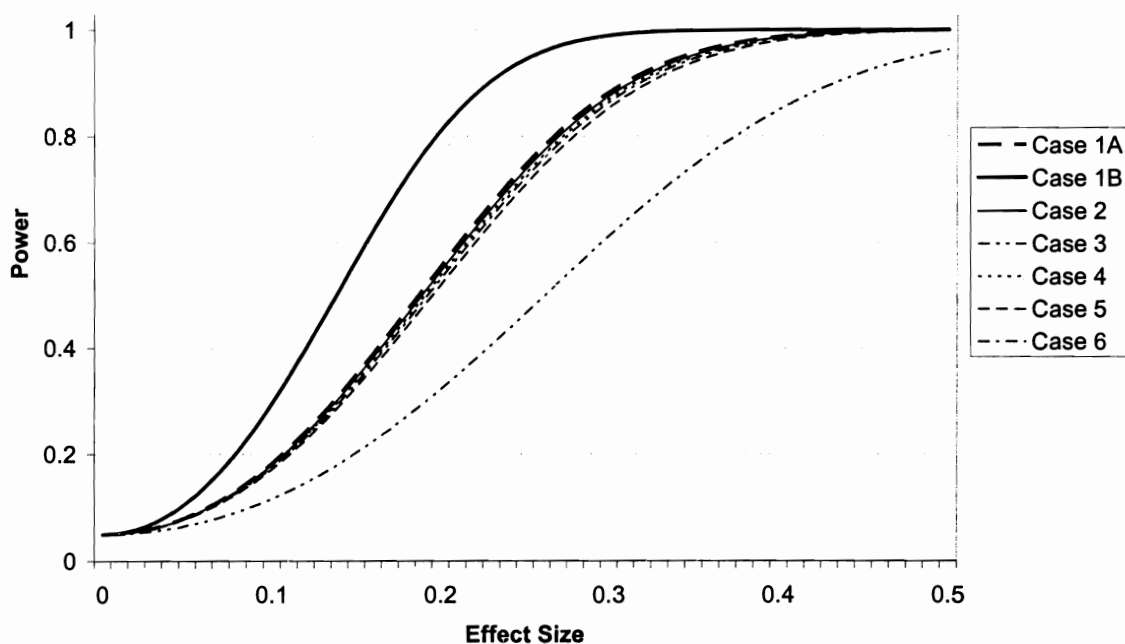


Figure 6.3 Power Curves for 7 case scenarios assuming $p=50$ $N=10$ $\rho = 0.01$ $\psi = 0.2$

The power curves in descending order correspond to Cases 1B, 1A, 2, 4, 6, 5, and 3,

though only 1B and 3 stand out. This has low correlation between measures in two time

points ($\psi = 0.2$) and low correlation within cluster ($\rho = 0.01$), which means τ^2 (variation between clusters) is small. The quantity $1 - \lambda_{12}^2 = 0.9997$ and thus there is essentially no correction to the Case 1 variance. Compared with panel A, each variance will decrease due to the increase in p . This in turn increases power over all. With relatively large p we know that the variances of Cases 2, 4, 5, and 6 are essentially the same. Also since p is large the difference in degrees of freedom has little effect on the power. Thus the curves for these cases coincide. Recall that Case 3 contains the sum of the variance from the two time points. Since the correlation is small, there is little to offset this sum. Thus the variance for Case 3 is larger than the other cases and the power curve is diminished.

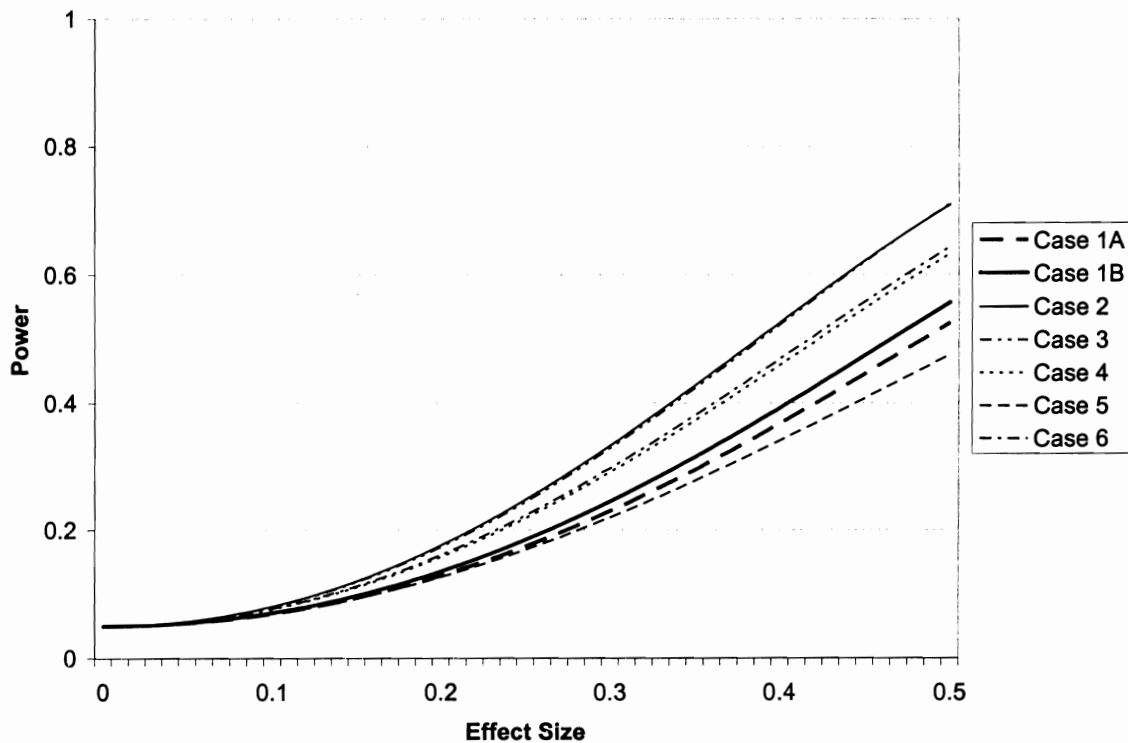


Figure 6.4 Power Curves for 7 case scenarios assuming $p=10$ $N=50$ $\rho = 0.1$ $\psi = 0.8$

The power curves in descending order correspond to Cases 2, 3, 6, 4, 1B, 1A, and 5, though Cases 2 and 3 are very close. This has high correlation between measures in two time points (0.8) and relatively high correlation within cluster ($\rho = 0.10$), which means τ^2 (variation between clusters) is relatively large. The quantity $1 - \lambda_{12}^2 = 0.54$ corresponds to an almost 50% reduction in the Case 1 variance. Compared with panel A, each variance will increase some due to the increase in the relative value of τ^2 . This in turn decreases the power over all. Since the correlation is large, Case 3's variance is

greatly reduced, though it is still somewhat higher than Case 2's variance. However, Case 2 has 7 degrees of freedom compared to 8 degrees of freedom for Case 3. Thus the power curves for these two cases are very close. It appears that Case 5 falls the others due to only having small degrees of freedom ($df=3$).

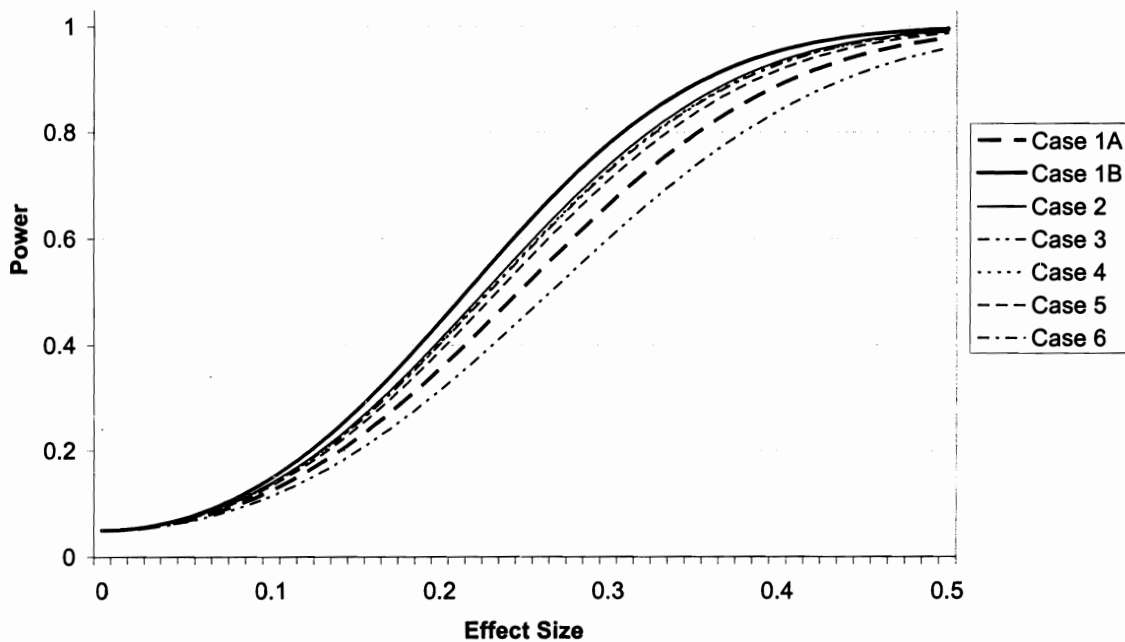


Figure 6.5 Power Curves for 7 case scenarios assuming $p=50$ $N=10$ $\rho = 0.1$ $\psi = 0.8$

The power curves in descending order correspond to Cases 1B, 2, 4, 5, (1A=6), and 3.

This has high correlation between measures in two time points (0.8) and relatively high correlation within cluster ($\rho = 0.10$), which means τ^2 (variation between clusters) is relatively large. The quantity $1 - \lambda_{12}^2 = 0.82$ corresponds to an almost 20% reduction in

the Case 1 variance, higher than in Panel C but much less than in Panels A and B.

Compared with panel C, each variance will decrease some due to the increase in the p .

This in turn increases the power over all.

CHAPTER VII

Conclusion and Discussion

7.1 Summary of Work

Over the last two decades increased attention has been given to identifying appropriate methods for the design and analysis of cluster randomized trials. There are certain characteristics that distinguish the CRT from the more familiar RCT:

- i) Clusters are formed not at random but rather through some type of connection among their members.
- ii) Unit of analysis is a cluster and the unit of observation is a cluster member.
- iii) Unit of assignment is a cluster. Different clusters are assigned to each study condition.
- iv) Commonly only a limited number of clusters are assigned to each study condition.

These characteristics create complexity for this design and the methods required for CRTs are not as simple as those required for RCTs.

A number of papers have appeared presenting general comparisons of analytic methods and methodological issues involved in CRTs. Recent books on the design and

analysis of CRTs have offered a comprehensive overview of relevant methods. After conducting a thorough literature search and recognizing much of the recent development of methodology, some doubts about these methods still remain. General explanations about possible solutions can be found in various articles and books but they fail to provide details about assertive evidence or analytically proven results.

In this work our focus was on a pre-post CRT design. Our primary aim was to improve the efficiency of the test when there is significant variation between clusters relative to within cluster variance. During the design of the VCU vital sign study (see Chapter I), the motivation for this work, a literature review of previous studies suggested wide variation between clusters. The researchers speculated that pair-matching on the baseline outcome measures would help to control for unmeasured covariates and reduce the variance of the treatment/control contrast. We investigated the performance of the most commonly used RCT methods adapted for a CRT pre-post design both with and without pair-matching. In order to have a clear understanding of fundamentals, we formulated the explicit terms of the variance estimate of treatment effect for each of six scenarios (Case 1—Case 6) and confirmed our results via simulation. We also computed and compared the power under varying conditions.

Details of our results are summarized in section 7.3 where we discuss recommendations for experimental design and analysis. In general we found that using the baseline means as a covariable provided the best control of the variance and the best power to detect differences in treatment and control effects. Pair-matching did not improve upon this method.

7.2 Contributions

We see three major contributions derived from this work.

- We derived the true variance structure of the treatment contrast when pair-matching is added to the classical design and analysis methods.
- The six cases were compared based on true variances and (approximate or true) power.
- Given the assumptions and limitations (see section 7.4), theory-based recommendations (see section 7.3) are given concerning the design and analysis of CRTs.

7.3 Recommendations

Our attention was given to incorporating baseline data in the analysis and the design of a CRT in order to reduce the variance of the treatment contrast. We approached this issue from the data analysis point of view assuming data was already collected. In addition, we approached this issue from the view of the design stage (prior to data collection). We considered well known methods, such as adjustment for baseline as a covariate (Case 2) and analysis of differences of pre and post measures (Case 3). For the comparison, the unadjusted data analysis (post data only: Case 1) was regarded as well. Explicit forms of the estimate of variance for each case were formulated, compared and tested via analytical approach. Power curves were computed and compared.

Based on the results of this study, we make recommendations given we are consulted at the design phase and recommendations given we are consulted after the data are collected.

7.3.1 Recommendations at the Design Phase

Suppose that we are approached by a research who wished to design a CRT. Questions relevant to our work include the choice of pre-post data collection versus post only. If

baseline data are collected, then should pair-matching be used? Should one also view the baseline mean as a covariate, analyze pre-post differences, or do neither?

Our results show that if there is low correlation between the two cluster measures or if the between cluster variance is small then we achieve the highest power by allocating all sample resources to the post intervention period. That is, the most powerful analysis is to collect only post intervention data and conduct a completely randomized trial (Case 1).

If there is moderate to high correlation or if there is moderate to high between cluster variance, then it is worthwhile to collect baseline data. The most powerful analysis is to conduct a completely randomized trial and use the baseline means as a covariable (Case 2).

7.3.2 Recommendations Post Data Collection

Now suppose that a researcher comes to use with data in hand from a CRT. The researcher collected baseline data as well as post intervention data. In the two scenarios below, the research either did or did not use pair-matching on baseline prior to randomization.

7.3.2.1 Data Collection Without Pair-Matching

The research used a completely randomized design with pre and post intervention data.

We are asked how the baseline data should be used. Our answer is based on the comparison of Cases 1-3. Case 2 has the best power and the smallest treatment contrast variance. Thus we recommend that the baseline means be used as a covariable in the analysis of the treatment contrast.

7.3.2.2 Data Collection With Pair-Matching

The researcher collected the baseline data, sorted the clustered on their baseline means, paired the clusters into matched groups, and then within each group randomized clusters to treatment or control. Our recommendation on the choice of analysis is based on the comparison of Cases 4-6. Case 6 has a lower variance and a higher power than Case 4 only when the correlation between time points is sufficiently greater than 0.5. Case 5's expected variance is the same as Case 4, but for a given realization of the baseline data, the variance may be smaller or larger. Given that Case 5 has one less degree of freedom for error and that for some realizations of the baseline data the variance is larger than the

Case 4 variance, we recommend that Case 4 be used when the correlation is low and Case 5 be used when the correlation is high.

7.4 Limitations

Several limitations of our study and results exist. We will summarize them here.

The matched-pair design has been widely used in community intervention trials as Murray (1998) and Donner (2000) have noted. However in spite of its abilities for creating comparable pairs of clusters, some analytic limitations of this design arise from the practical difficulty of obtaining close matches on important baseline factors and from the difficulty in estimating the intraclass correlation (ICC) from the matched-pair data. Recent studies report the ICC and related parameter values for the wide variety of groups, members, and endpoints so that researchers now have a better opportunity to find ICC estimates that are well matched to the particular circumstances of the trial they are planning.

Many researchers argue that the effectiveness of matching depends on the ability of the researcher to create pairs (groups) that vary as much as possible with respect to baseline matching factor(s). Otherwise matching may actually lead to a loss in statistical

efficiency. Another challenge is that variables highly related to outcome at the cluster level are frequently unknown at the outset of trial or they are not easy to match on if they known. In our study we have known baseline of the primary endpoint and we use it as a matching variable.

We assumed that the cluster means follow a normal distribution. As such, these means will tend to cluster around the single mode—mean—of the normal distribution, but will not tend to cluster at other locations. This may not reflect reality as cluster means may naturally group around multiple points. Matching clusters prior to randomization may prove to be of greater advantage when the clusters form natural groups.

Many applications require that the data be binomially distributed or distributed according to a member of the exponential family of distributions. The results of this research may to extend to non-normal error distributions.

CHAPTER VIII

Future Work

8.1 Introduction

In this chapter we discuss future work that builds on this dissertation. The list is not exhaustive, but we focus on a few ideas that should add to the knowledge of experimental designs for cluster randomized trials as well as general randomized controlled trials.

8.2 General Grouping Methods to Produce Homogeneous Groups

When we created groups that are relatively homogeneous on baseline measures prior to study arm randomization, we considered a simple method, grouping clusters into pairs.

While this pairing method is used in practice, it may not be the best approach to achieve the goal of improving efficiency and power. A beneficial alternative is to block on matched groups of clusters by matching clusters based on how their baseline values *cluster*. In our study we assumed a unimodal normal distribution for the cluster means. In essence, these means do not naturally cluster, unless you consider the middle of the

distribution. The reality may be that the cluster means do cluster and matching schemes will be beneficial in reducing the variation between clusters by matching like clusters. If the shape of underlying distribution of baseline data does not exhibit clustering then pairing could do little to reduce the variance. Arbitrary pairing simply takes away extra degrees of freedom without providing any benefits and it can be crucial to CRT setting since often a limited number of clusters is assigned to each study condition. A challenge of this method is to find groupings of homogeneous clusters where the number and size of the groups are only constrained by the requirement that each group has at least 2 clusters (or more generally, at least as many clusters as arms of the study). Determining the analytic variance or proper test may also prove an insurmountable challenge. For this study, we would assume a mixture of normal distributions, chosen to exhibit a degree of clustering, and depend on simulation to estimate variances and compare the difference case analyses.

8.3 Generalized Error Structure

In our work we assumed a Gaussian error structure. An extension to non-normally distributed data, e.g. binomial or other exponential-family distribution, would be useful in

practice. Methods for analyzing non-normal distributed data in cluster randomized trials are not as well established as methods for analyzing normally distributed and some other continuous outcome data. As an example, for binary outcome data, Donner (2000) mentioned that the analytic issues involved are complicated by the absence of a unique multivariate extension of binomial distribution analogous to the multivariate normal distribution and by the fact that there is no single approach that has uniformly superior properties. Further exploration on this issue would be valuable in practice.

8.4 Hierarchical or Deeply Nested Designs

Clustering is expected to arise more than at one level in practice. Our approach can be extended to more than two levels. For example, in the VCU vital sign study, two sources of variation beyond the variation between patients can be defined reflecting 1) the variation among clinics (σ_c^2), and 2) the variation among physician teams within clinics (σ_p^2). Then total variation (σ_T^2) is composed of those two variations and the within-physician component variance (σ_w^2): $\sigma_T^2 = \sigma_c^2 + \sigma_p^2 + \sigma_w^2$. Then ICC for patients within physicians is defined as

$$\rho_p = \frac{\sigma_p^2 + \sigma_c^2}{\sigma_T^2} \text{ while ICC for patients within clinic is defined as } \rho_c = \frac{\sigma_p^2}{\sigma_T^2} .$$

These definitions imply the intraclass correlation for patients within physicians is larger than that within clinics. This reflects the greater similarity to be assumed among patients under the same physician than among patients in simply under the same clinic. This one step further complexity can be incorporated into our frame of work.

8.5 Matching on Multiple Baseline Measures

We matched clusters at baseline on the pre-test value of the primary endpoint. It can be extended to matching on more than one factor, given that other additional matching variables should be highly related to outcome variable.

8.6 Repeated Measures on Clusters

We mainly focused on two time point measures and ANOVA/ANOCOVA was considered. This can be extended to more than two time points whereas random coefficient approach can be considered appropriate. Once the data are available, data may

be analyzed to determine whether there are differences among the conditions as well as whether those differences are patterned in time as the researcher would expect given the time course of their study.

8.7 Optimal Allocation of Sample Sizes Across Time Points

What is the optimal allocation of sample size between baseline and intervention periods?

We did some preliminary work on this but did not report the results in this dissertation.

Suffice it to say that the relationships observed between the various cases did not change when we used the optimal (minimal variance) allocation for each case. However, the need remains to recommend to researcher who wishes to employ a pre-post CRT the best allocation of resources between baseline and post-intervention samples.

List of References

- Abdeljaber, M.H., Monto, A.S., Tilden, R.L., Schork, M.A. and Tarwotjo, I. (1991). The impact of Vitamin A supplementation on morbidity: a randomized community intervention trial. *American Journal of Public Health* **81**, 1654-1656.
- Armitage, Peter and Colton, Theodore (2005), *Encyclopedia of Biostatistics*, Vol 6, John Wiley & Sons Inc.
- Arnold, Barry C., Balakrishnan, N., and Nagaraja, H.N. (1992). *First Course in Order Statistics*, Wiley Interscience.
- Barnett, V., Green, P.J., and Robinson, A. (1976). *Biometrika*, **63**, 323-328.
- Bhattacharya, P. K. (1974). *Annual Statistics*, **2**, 1034-1039.
- Bickel, Peter J. and Doksum, Kjell A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*, Duxbury.
- Bonate, Peter L (2000) *Analysis of Pretest-Posttest Designs*, Chapman&Hall/CRC, FL.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, Lawrence Erlbaum Associates, Hillsdale, NJ.
- Cox D. R. (1957). The Use of a Concomitant Variable in Selecting an Experimental Design *Biometrika*, Vol. **44**, No. 1/2. (Jun., 1957): 150-158.
- Dalton, S. and Overall, J.E. (1977). Nonrandom assignment in ANCOVA: the alternate ranks design, *J. Exp. Ed.*, **46**, 58 David, H. A. (1973). *Bull. Int. Statist.*, **45**(1), 295-300.
- David, H. A. (1981). *Order Statistics*. 2nd edition. Wiley, New York.
- David, H. A., O'Connell, M. J., and Yang, S. S. (1977). *Ann. Statist.*, **5**, 216-223.
- Davis, O. J. and P. Rabinowitz. (1984). *Methods of Numerical Integration*, Academic Press, New York.
- Donner, A., and N. Klar. 2000. *Design and analysis of cluster randomization trials in health research*. London, England: Arnold.

- Donner, A., Birkett, N., and Buck, C. 1981: Randomization by cluster: sample size requirements and analysis. *American Journal of Epidemiology* **114**, 906-914.
- Graybill, Franklin A.(1976). *Theory and Application of the Linear Model*, Duxbury.
- Kish, L. (1965). *Survey sampling*. New York,NY: John Wiley and Sons Inc.
- Klar, N., and A. Donner.(1997). The merits of matching in community intervention trials: A cautionary tale. *Statistics in Medicine* **16**:1753-64.
- Koepsell, T. D., E. H. Wagner, A. C. Cheadle, D. L. Patrick, D. C. Martin, P. H. Diehr, E. B.Perrin, A. R. Kristal, C. H. Allan-Andrilla, and L. J. Dey.(1992). Selected methodological issues in evaluating community-based health promotion and disease prevention programs. *Annual Review of Public Health* **13**:31-57.
- Lee Hak-Myung and Viana M. (1999) "The Joint Covariate structure of ordered symmetrically dependent observations and their concomitant of order statistics." *Statistics and Probability letters*, **43**:411-414.
- Lether, F. G. (1978). On the construction of Gauss-Legendre Quadrature rules. *JCompApMa* **4**, 47-52.
- Lipsey, M.W. (1990). *Design Sensitivity:Statistical Power for Experimental Research*, Sage Publications, Newbury Park, CA
- Luepker RV, Raczynski JM, Osganian S, Goldberg RJ, Finnegan JR Jr, Hedges JR, Goff DC Jr, Eisenberg MS, Zapka JG, Feldman HA, Labarthe DR, McGovern PG, Cornell CE, Proschan MA, Simons-Morton DG. 2000. Effect of a community intervention on patient delay and emergency medical service use in acute coronary heart disease: The Rapid Early Action for Coronary Treatment (REACT) Trial. *Journal of the American Medical Association*. Jul **5**; 284(1):60-7.
- Lytle, L., D. M. Murray, C. Perry, M. Story, A. Birnbaum, M. Kubik, and S. Varnell.(2004). School-based approaches to affecting adolescents' diets: Results from the TEENS study. *Health Education and Behavior* **31** (2): 270-87.
- Martin DC, Diehr P, Perrin EB, Koepsell TD.(1993). The effect of matching on the power of randomized community intervention studies: *Stat Med*. Feb;**12**(3-4):329-38.
- Mood , Alexander M., Graybill Franklin A., and Duane C. Boes. (1974). *Introduction to the Theory of Statistics*, third edition, McGraw-Hill, NY.

- Murray, DM. (1998). *Design and analysis of group-randomized trials*. New York: Oxford University Press.
- Murray, D. M., S. P. Varnell, and J. L. Blitstein.(2004). Design and analysis of group-randomized trials: A review of recent methodological developments. *American Journal of Public Health* **94** (3): 423-32.
- Myers, Raymond H., Milton, Janet S. (1998). *A First Course in the Theory of Linear Statistical Models*, McGraw-Hill Companies, Inc.
- Park, Misook and Johnson, Robert E. (2004). Methods for matching clusters on baseline outcome measures prior to randomization, *ASA Proceedings of the Joint Statistical Meetings*, 423-426 *American Statistical Association (Alexandria, VA.)*
- Parrish, Rudolph S. (1992) Communications in Statistics: Simulation and Computation, **21**:57-70.
- Parrish, Rudolph S. (1992) Communications in Statistics: Simulation and Computation, Peterson, A. V., K. A. Kealey, S. L. Mann, P. M. Marek, and I. G. Sarason. (2000). Hutchinson Smoking Prevention Project: Long-term randomized trial in school-based tobacco use prevention—Results on smoking. *Journal of the National Cancer Institute* **92** (24): 1979-91.
- Rencher Alvin C. (2000). *Linear Model in Statistics*, New York: Willy International Science **21**:71-101.
- Sen, P. K. (1976). *Ann. Prob.*, **4**, 474-479.
- Shapiro, S.S. and M. B. Wilk. (1965). An analysis of variance test for normality (complete samples). *Biometrics*, **52**, 591-611.
- Sorensen, G., A. Stoddard, A. D. LaMontagne, K. Emmons, M. K. Hunt, R. Youngstrom, D. McLellan, and D. C. Christiani. 2002. A comprehensive worksite cancer prevention intervention: behavior change results from a randomized controlled trial (United States). *Cancer Causes Control*. 2002 Aug;**13**(6):493-502.
- Stroud, A. H. and D. Secrest (1969). *Gaussian Integration Formulas*. Prentice-Hall, Englewood Cliffs, NJ.
- Searle, Shayle R. (1982). *Matrix Algebra Useful for Statistics*, John Willey & Sons.
- Searle, Shayle R. (1971). *Linear Models*, John Willey & Sons Inc.

SAS Institute Inc. (2004). SAS OnlineDoc®9.1.3. Cary, NC: SAS Institute Inc.

Tietjen, G. L., Kahaner, D. K., and Beckman, R. J. (1977). Variances and covariances of the normal order statistics of sample sizes 2 to 50. *Selected Tables in Mathematical Statistics* **5**, 1-73.

Varnell, S., D. M. Murray, J. B. Janega, and J. L. Blitstein.(2004). Design and analysis of group-randomized trials: A review of recent practices. *American Journal of Public Health* **94**(3): 393-99.

Yamauti, Ziro (1972). *Statistical Tables and Formulas with Computer Applications*. Japanese Standards Association, Tokyo. Table C1, 33-37.

Yang,S.S.(1977). *Ann. Statist.*, **5**, 996-1002.

```

*****
**
* This code computes the values of omega, which is defined as "trace(V*(bigC+lowc*lowc`))/p" ,given
*
* different p values. This omega is part of the variance terms of the treatment effect for the matched
*
* cases(case 4-6). This program also creates the plots of omega by p.
*
* *****
**;
```

libname order "C:\Misook Park - Dissertation\Normal Order Statistics Data";

```

proc iml;
create omega var {p omega};
do p=2 to 50 by 2;          ** Sample size;
    reset fuzz;
    **** Prepare the mean and covariance of the order statistics, given the number of clusters ****;
    hn=int(p/2); *Half p (hn);
    use order.means var {"n" "i" "x"} where(n=p);
    read all var {"x"} into mean;
    rx=mean; * reverse x;
    do i=1 to hn; rx[i]=mean[hn-i+1]; end;
    if p/2=hn then mean=-mean//rx;
    else mean=-mean//{0}//rx;          * Mean vector of std norm order statistics (c);

    cov=J(p,p,0);
    use order.cov var {"n" "i" "j" "v"} where (n=p);
    do k=1 to int(((p+1)**2)/4);
        read next var {"i" "j" "v"} into temp;
        cov[temp[1],temp[2]]=temp[3];
    end;
    do i=2 to p;
        do j=p to max(p-i+2,i) by -1;
            cov[i,j]=cov[p-j+1,p-i+1];
        end;
    end;
    cov=cov+cov`-diag(cov);          * Cov matrix of std norm order statistics (C);
    lowc=mean;
    bigC=cov;

    V=I(p/2)@{1 -1, -1 1}/(p/2)**2;          * Variance of the T-C randomized contrast;

```

```

        Compare=p/4 * trace(V*(bigC+lowc*lowc`));

        omega=trace(V*(bigC+lowc*lowc`))/p;

        without_p=trace(V*(bigC+lowc*lowc`));

*      Print Compare omega without_p;

        append;
end;
quit;

data order.omega;
    set omega;
    x=1-p**2/4*omega;
    y=1-omega;
    z=1-p*omega;
run;

proc gplot data=order.omega;
    plot x*p y*p/overlay;
    plot z*p;
run;

```

```

*****
* CASE ONE: Post Data only *
*This SAS code computes the variance of treatment effect for the Case one scenario *
* -post data only. It also calculates the power of the hypothesis test.We must run *
* the full data simulation program before implementing this program. *
*;
*****;

%let mu2=0.15;    * Treatment effect;

proc iml;
  create results var{iterate Beta3 seBeta3 t_stat p_value};
  critval=quantile('T',.975,&p-2);

  use themeans;

  do iterate=1 to &iterate;
    %let howmany=%eval(2*&p);
    read next &howmany var{time treatment meany} into temp;
    time=temp[,1]; treat=temp[,2]; y=temp[,3]; free temp;
    t2=treat[loc(time=2)];
    y1=y[loc(time=1)]; y2=y[loc(time=2)];
    y2[loc(t2=1)]=y2[loc(t2=1)]+&mu2;

    **      Case 1 **;

    X=J(&p,1)||t2;

    * Squared Projections;
    PR=y2`*y2;
    PJ=y2`*J(&p,1)*inv(J(&p,1)`*J(&p,1))*J(&p,1)`*y2;
    PX=y2`*X*inv(X`*X)*X`*y2;
    PT_X=y2`*X[,1]*inv(X[,1)`*X[,1])*X[,1)`*y2;

    * Estimates;
    Beta=inv(X`*X)*X`*y2;
    Beta3=beta[2];

    * Sums of Squares;
    SSTotal=PR-PJ;

```

```

SST=PX-PT_X;
SSE=PR-PX;

seBeta3=sqrt((SSE/(&p-2))@ (inv(X`X)[2,2]));

* Test Statistic - F-statistic;
F=SST/(SSE/(&p-2));
t_stat=sqrt(F)*sign(beta[2]);
p_value=(abs(t_stat) >= critval);
* Save the results;
append;
end;

* Non-centrality and power;
mean=&mu2@t2;
NCP_PX=mean`X*inv(X`X)*X`*mean;
NCP_PT_X=mean`X[,1]*inv(X[,1]`X[,1])*X[,1]`*mean;
NCP=(NCP_PX-NCP_PT_X)/(&tau2+&sig2/&n2);
Power=1-CDF('F',critval**2,1,&p-2,NCP);
print Power;

quit;

proc means data=results n mean var std stderr clm;
var Beta3 seBeta3 t_stat p_value;
run;

```



```

*****
*CASE THREE: ANOVA on the pre-post difference
*
*This SAS code performs the study for the Case three scenario. It computes the variance
*of the treatment effect as well as the associated power of the hypothesis test.We must
*run the full data simulation program to before this program.
*****;

```

```

%let mu2=0.15;      * Treatment effect;

```

```

proc iml;
  create results var{iterate Beta3 seBeta3 t_stat p_value};
  critval=quantile('T',.975,&p-2);

  use themeans;

  do iterate=1 to &iterate;
    %let howmany=%eval(2*&p);
    read next &howmany var{time treatment meany} into temp;
    time=temp[,1]; treat=temp[,2]; y=temp[,3]; free temp;

    t2=treat[loc(time=2)];
    y2=y[loc(time=2)]-y[loc(time=1)];

    y2[loc(t2=1)]=y2[loc(t2=1)]+&mu2;

    **      Case 3 **;

    X=J(&p,1)||t2;

    * Squared Projections;
    PR=y2`*y2;
    PJ=y2`*J(&p,1)*inv(J(&p,1)`*J(&p,1))*J(&p,1)`*y2;
    PX=y2`*X*inv(X`*X)*X`*y2;
    PT_X=y2`*X[,1]*inv(X[,1]`*X[,1])*X[,1]`*y2;

    * Estimates;
    Beta=inv(X`*X)*X`*y2;
    Beta3=beta[2];

    * Sums of Squares;
    SSTotal=PR-PJ;

```

```

SST=PX-PT_X;
SSE=PR-PX;

seBeta3=sqrt((SSE/(&p-2))@ (inv(X`*X)[2,2]));

* Test Statistic - F-stat;
F=SST/(SSE/(&p-2));
t_stat=sqrt(F)*sign(beta[2]);
p_value=(abs(t_stat) >= critval);
* Save the results;
append;
end;

* Non-centrality and power;
mean=&mu2@t2;
NCP_PX=mean`*X*inv(X`*X)*X`*mean;
NCP_PT_X=mean`*X[,1]*inv(X[,1]`*X[,1])*X[,1]`*mean;
NCP=(NCP_PX-NCP_PT_X)/((1/&n1+1/&n2)*&sig2+2*&tau2*(1-&psi));
Power=1-CDF('F',critval**2,1,&p-3,NCP);
print Power;

quit;

proc means data=results n mean var std stderr clm;
var Beta3 seBeta3 t_stat p_value;
run;

```

```

*****
*                                     CASE FIVE:                                     *
*ANOCOVA on the posttest scores with pretest score as a covariable, accounting for matching;*
*
*This SAS code performs the study for the Case five scenario. It computes the variance of *
*the treatment effect as well as the power of the hypothesis test.We must run the full data *
*simulation program before this program.                                           *
*****;
libname order "C:\Misook Park - Dissertation\Normal Order Statistics Data";

%let mu2=0.15;      * Treatment effect;

proc iml;

**** Prepare the mean and covariance of the order statistics, given the number of clusters ****;
    hn=int(&p/2); *Half p (hn);
    use order.means var {"n" "i" "x"} where(n=&p);
    read all var {"x"} into mean;
    rx=mean; * reverse x;
    do i=1 to hn; rx[i]=mean[hn-i+1]; end;
    if &p/2=hn then mean=-mean//rx;
    else mean=-mean//{0}//rx;                                                    * Mean vector of std norm order statistics (c);

    cov=J(&p,&p,0);
    use order.cov var {"n" "i" "j" "v"} where (n=&p);
    do k=1 to int(((&p+1)**2)/4);
        read next var {"i" "j" "v"} into temp;
        cov[temp[1],temp[2]]=temp[3];
    end;
    do i=2 to &p;
        do j=&p to max(&p-i+2,i) by -1;
            cov[i,j]=cov[&p-j+1,&p-i+1];
        end;
    end;
    cov=cov+cov`-diag(cov);                                                    * Cov matrix of std norm order statistics (C);
    print cov;
    lowc=mean;
    bigC=cov;

```

```

s1=&tau2+&sig2/&n1*(1+(&n1-1)*&rho);      * Sigma-1-squared: var of sample cluster mean, time 1;
s2=&tau2+&sig2/&n2*(1+(&n2-1)*&rho);      * Sigma-2-squared: var of sample cluster mean, time 2;
L=&tau2*&psi+&sig2*&rho;                  * Lambda: Cov of time 1 & 2 sample cluster means;
L12=L/sqrt(s1*s2);                      * Lambda12: Corr of time 1 & 2 sample cluster
means;
V=I(&p/2)@{1 -1, -1 1}/(&p/2)**2;      * Variance of the T-C randomized contrast;
omega=trace((cov+mean*mean`)*V)/&p;

var=(4/&p)*s2*(1-L12**2);

*****;

create results var{iterate Beta3 seBeta3 t_stat p_value Power SSE E_SSE};
critval=quantile('T',.975,&p/2-2);

use themeans;

do iterate=1 to &iterate;
  %let howmany=%eval(2*&p);
  read next &howmany var{time ptreatment meany group} into temp;
  print temp;
  time=temp[,1]; treat=temp[,2]; y=temp[,3]; group=temp[,4]; free temp;

  t2=treat[loc(time=2)];
  y1=y[loc(time=1)]; y2=y[loc(time=2)];
  group=design(group[loc(time=2)]);

  y2[loc(t2=1)]=y2[loc(t2=1)]+&mu2;

**      Case 5 **;

X=J(&p,1)||group[,2:%eval(&p/2)]||y1||t2;

X1=X[,1:%eval(&p/2+1)];

* Squared Projections;
  PR=y2`*y2;
  PJ=y2`*J(&p,1)*inv(J(&p,1)`*J(&p,1))*J(&p,1)`*y2;
  PX=y2`*X*inv(X`*X)*X`*y2;
  PT_X=y2`*X1*inv(X1`*X1)*X1`*y2;

```

```

* Estimates;
  Beta=inv(X`*X)*X`*y2;
  Beta3=beta[%eval(&p/2+2)];

* Sums of Squares;
  SSTotal=PR-PJ;
  SST=PX-PT_X;
  SSE=PR-PX;

  seBeta3=sqrt((SSE/(&p/2-2))@ (inv(X`*X)[%eval(&p/2+2),%eval(&p/2+2)]));
  E_SSE=(&p/2-2)*s2*(1-L12**2)+s2*L12**2*trace((I(&p)-X*inv(X`*X)*X`)*(bigC+lowc*lowc`));

* Test Statistic - F-stat;
  F=SST/(SSE/(&p/2-2));
  t_stat=sqrt(F)*sign(beta[%eval(&p/2+2)]);
  p_value=(abs(t_stat) >= critval);

* Non-centrality and power;
  mean=&mu2@t2 + sqrt(s2/s1)*L12*y1;
  NCP_PX=mean`*X*inv(X`*X)*X`*mean;
  NCP_PT_X=mean`*X1*inv(X1`*X1)*X1`*mean;
  NCP=(NCP_PX-NCP_PT_X)/var*(4/&p);
  Power=1-CDF('F',critval**2,1,&p/2-2,NCP);

* Save the results;
  append;
end;

quit;

proc means data=results n mean median var std stderr min max clm;
  var Beta3 seBeta3 t_stat p_value Power SSE E_SSE;
run;

```

```

*****
This SAS program simulates the data for cases 1-6 based on specific parameters.      *
The mean vector was assumed to contain all zeros. (The mean vector may be easily    *
altered post-simulation by adding the appropriate constants to the simulated data.)*
The SAS/IML (SAS Interactive Matrix Language) function VNORMAL was used to generate*
p bivariate observations representing the pre and posttest cluster means. This code *
could be easily modified for other parameters by simply setting different parameter *
values.
*****;
%let iterate=1000;
%let p=10;                                * Number of clusters;
%let n1=10;
%let n2=10;
%let sig2=.99;
%let rho=0;
%let eta=0;
%let tau2=.01;
%let psi=0.2;

proc iml;

    create thedata var {"iterate","cluster","treatment","time","y"};

    varmat1=tau2 @ (psi @ (J(2,2) - I(2)) + I(2));
    varmat211=sig2 @ (rho @ (J(n1,n1) - I(n1)) + I(n1));
    varmat222=sig2 @ (rho @ (J(n2,n2) - I(n2)) + I(n2));
    varmat212=eta @ sig2 * J(n1,n2);
    varmat221=varmat212`;
    varmat2=(varmat211 || varmat212) // (varmat221 || varmat222);

print varmat1;

test=(j(n1,1,0)/j(n2,1,1))/n2;
testvar=test`*varmat2*test;
print testvar;

    do k=1 to &iterate;
        call vnormal(x,j(2,1,0),varmat1,&p,0);
        do i=1 to &p;
            treat=(i > &p/2);                                * Treatment=1, Control=0, no blocks, CR
design;

            mean=(x[i,1]@j(n1,1))/(x[i,2]@j(n2,1));
            call vnormal(y0,mean,varmat2,1,0);

```

```

                y=y/(j(&n1+&n2,1,k)||j(&n1+&n2,1,i)||j(&n1+&n2,1,treat)||j(&n1,1,1)//j(&n2,1,2))
||y0`);
        end;
        append from y;
        free y;
    end;

quit;

proc summary data=thedata;
    class iterate cluster time;
    id treatment;
    var y;
    ways 3;
    output out=themeans mean=mean;
run;

proc sort data=themeans (where=(time=1)) out=paired;
    by iterate mean;
run;

data paired;
    retain ptreatment;
    set paired;
    by iterate;
    if first.iterate then
        do;
            group=0;
            ordered=0;
        end;
    ordered+1;
    if mod(_n_,2)=1 then
        do;
            ptreatment=(ranuni(0)<0.5);
            group+1;
        end;
    else ptreatment=1-ptreatment;
    keep iterate cluster group ptreatment ordered;
run;

proc sort data=paired;
    by iterate cluster;
run;

```

* Matched pair label;

```
data themeans;  
    merge themeans paired;  
    by iterate cluster;  
run;
```

```
data thedata;  
    merge thedata paired;  
    by iterate cluster;  
run;
```

```
proc sort data=themeans;  
    by iterate ordered;  
run;
```

```
proc sort data=thedata;  
    by iterate ordered;  
run;
```



```

*****
*CASE TWO: Post Data with Baseline Adjustment
*
*This SAS code performs the study for the Case two scenario. It computes the variance of *
*the treatment effect as well as associated power of the hypothesis test.We must run the *
*full data simulation program before this program.
*****;

```

```

%let mu2=0.15;      * Treatment effect;

```

```

proc iml;
  create results var{iterate Beta3 seBeta3 t_stat p_value power};
  critval=quantile('T',.975,&p-3);

  use themeans;

  L12=&tau2*&psi/sqrt((&tau2*&sig2/&n1)*(&tau2*&sig2/&n2));
  var=(4/&p)*(&tau2*&sig2/&n2)*(1-L12**2);

  do iterate=1 to &iterate;
    %let howmany=%eval(2*&p);
    time=temp[,1]; treat=temp[,2]; y=temp[,3]; free temp;

    t2=treat[loc(time=2)];
    y1=y[loc(time=1)]; y2=y[loc(time=2)];

    y2[loc(t2=1)]=y2[loc(t2=1)]+&mu2;

    **      Case 2**;

    X=J(&p,1)||y1||t2;

    * Squared Projections;
    PR=y2`*y2;
    PJ=y2`*J(&p,1)*inv(J(&p,1)`*J(&p,1))*J(&p,1)`*y2;
    PX=y2`*X*inv(X`*X)*X`*y2;
    PT_X=y2`*X[,1:2]*inv(X[,1:2]*X[,1:2])*X[,1:2]`*y2;

    * Estimates;
  end;
  results=results||var;
run;

```

```

        Beta=inv(X`*X)*X`*y2;
        Beta3=beta[3];

    * Sums of Squares;
        SSTotal=PR-PJ;
        SST=PX-PT_X;
        SSE=PR-PX;

        seBeta3=sqrt((SSE/(&p-3))@ (inv(X`*X)[3,3]));

    * Test Statistic - F-stat;
        F=SST/(SSE/(&p-3));
        t_stat=sqrt(F)*sign(beta[3]);
        p_value=(abs(t_stat) >= critval);

    * Non-centrality and power;
        mean=&mu2@t2;
        NCP_PX=mean`*X*inv(X`*X)*X`*mean;
        NCP_PT_X=mean`*X[,1:2]*inv(X[,1:2]`*X[,1:2])*X[,1:2]`*mean;
        NCP=(NCP_PX-NCP_PT_X)/(var*(&p/4));
        Power=1-CDF('F',critval**2,1,&p-3,NCP);

    * Save the results;
        append;
end;

quit;

proc means data=results n mean var std stderr min max clm;
    var Beta3 seBeta3 t_stat p_value power;
run;

```

```

*****
*CASE FOUR: ANOVA on the pre-post difference
*This SAS code performs the study for the Case four scenario. It computes the variance
*of the treatment effect as well as the power of the hypothesis test. We must run the full
*data simulation program before this program.
*****;
libname order "C:\Misook Park - Dissertation\Normal Order Statistics Data";

%let mu2=0.50;      * Treatment effect;

proc iml;

**** Prepare the mean and covariance of the order statistics, given the number of clusters ****;
    hn=int(&p/2); *Half p (hn);
    use order.means var {"n" "i" "x"} where(n=&p);
    read all var {"x"} into mean;
    rx=mean; * reverse x;
    do i=1 to hn; rx[i]=mean[hn-i+1]; end;
    if &p/2=hn then mean=-mean//rx;
    else mean=-mean//{0}//rx;          * Mean vector of standard normal order
statistics (c);

    cov=J(&p,&p,0);
    use order.cov var {"n" "i" "j" "v"} where (n=&p);
    do k=1 to int(((&p+1)**2)/4);
        read next var {"i" "j" "v"} into temp;
        cov[temp[1],temp[2]]=temp[3];
    end;
    do i=2 to &p;
        do j=&p to max(&p-i+2,i) by -1;
            cov[i,j]=cov[&p-j+1,&p-i+1];
        end;
    end;
    cov=cov+cov`-diag(cov);          * Cov matrix of standard normal order statistics
(C);
    lowc=mean;
    bigC=cov;

    s1=&tau2+&sig2/&n1*(1+(&n1-1)*&rho);      * Sigma-1-squared: var of sample cluster mean, time 1;
    s2=&tau2+&sig2/&n2*(1+(&n2-1)*&rho);      * Sigma-2-squared: var of sample cluster mean, time 2;
    L=&tau2*&psi+&sig2*&rho;                * Lambda: Cov of time 1 & 2 sample cluster
means;

```

```

L12=L/sqrt(s1*s2);
cluster means;
V=I(&p/2)@{1 -1, -1 1}/(&p/2)**2;
* Lambdal2: Corr of time 1 & 2 sample
* Variance of the T-C randomized contrast;

*****;

create results var{iterate Beta3 seBeta3 t_stat p_value SSE E_SSE};
critval=quantile('T',.975,&p/2-1);

use themeans;

do iterate=1 to &iterate;
  %let howmany=%eval(2*&p);
  read next &howmany var{time ptreatment meany group} into temp;

  time=temp[,1]; treat=temp[,2]; y=temp[,3]; group=temp[,4]; free temp;

  t2=treat[loc(time=2)];
  y1=y[loc(time=1)]; y2=y[loc(time=2)];
  group=design(group[loc(time=2)]);

  y2[loc(t2=1)]=y2[loc(t2=1)]+&mu2;

**      Case 4 **;

X=J(&p,1)||group[,2:%eval(&p/2)]||t2;

X1=X[,1:%eval(&p/2)];

* Squared Projections;
  PR=y2`*y2;
  PJ=y2`*J(&p,1)*inv(J(&p,1)`*J(&p,1))*J(&p,1)`*y2;
  PX=y2`*X*inv(X`*X)*X`*y2;
  PT_X=y2`*X1*inv(X1`*X1)*X1`*y2;

* Estimates;
  Beta=inv(X`*X)*X`*y2;
  Beta3=beta[%eval(&p/2+1)];

* Sums of Squares;
  SSTotal=PR-PJ;
  SST=PX-PT_X;
  SSE=PR-PX;

```

```

        seBeta3=sqrt((SSE/(&p/2-1))@ (inv(X`*X)[%eval(&p/2+1),%eval(&p/2+1)]));
        E_SSE=(&p/2-1)*s2*(1-L12**2)+s2*L12**2*trace((I(&p)-X*inv(X`*X)*X`)*
(bigC+lowc*lowc`));

        * Test Statistic - F-stat;
        F=SST/(SSE/(&p/2-1));
        t_stat=sqrt(F)*sign(beta[%eval(&p/2+1)]);
        p_value=(F >= critval**2);
        * Save the results;
        append;
end;

* Non-centrality and power;
mean=&mu2@t2;
var=(4/&p)*s2*(1-L12**2) + s2*L12**2*trace(V*(bigC+lowc*lowc`));
NCP_PX=mean`*X*inv(X`*X)*X`*mean;
NCP_PT_X=mean`*X1*inv(X1`*X1)*X1`*mean;
NCP=(NCP_PX-NCP_PT_X)/var*(4/&p);
Power=1-CDF('F',critval**2,1,&p/2-1,NCP);
print var Power NCP;
V_SSE=((&p/2-1)*s2*(1-L12**2)+s2*L12**2*trace((I(&p)-X1*inv(X1`*X1)*X1`-(&p/4)*V)*
(bigC+lowc*lowc`)))
/(&p/2-1)*4/&p;
IsV=(I(&p)-X1*inv(X1`*X1)*X1`-(&p/4)*V)/(&p/2-1)*4/&p;
Diff_IsV_V=fuzz(IsV-V);
Print V_SSE IsV,V Diff_IsV_V;

quit;

proc means data=results n mean median var std stderr clm;
    var Beta3 seBeta3 t_stat p_value SSE E_SSE;
run;

```

```

*****
*CASE SIX:
*ANOVA on the pre-post differences, accounting for matching;
*
*This SAS code performs the study for the Case six scenario. It computes the variance of
*the treatment effect as well as the power of the hypothesis test.We must run the full data
*simulation program before this program.
*****;

libname order "C:\Misook Park - Dissertation\Normal Order Statistics Data";

%let mu2=0.15;      * Treatment effect;

proc iml;

**** Prepare the mean and covariance of the order statistics, given the number of clusters ****;
    hn=int(&p/2); *Half p (hn);
    use order.means var {"n" "i" "x"} where(n=&p);
    read all var {"x"} into mean;
    rx=mean; * reverse x;
    do i=1 to hn; rx[i]=mean[hn-i+1]; end;
    if &p/2=hn then mean=-mean//rx;
    else mean=-mean//{0}//rx;
    print mean;

    cov=J(&p,&p,0);
    use order.cov var {"n" "i" "j" "v"} where (n=&p);
    do k=1 to int(((&p+1)**2)/4);
        read next var {"i" "j" "v"} into temp;
        cov[temp[1],temp[2]]=temp[3];
    end;
    do i=2 to &p;
        do j=&p to max(&p-i+2,i) by -1;
            cov[i,j]=cov[&p-j+1,&p-i+1];
        end;
    end;
    cov=cov+cov`-diag(cov);
    print cov;
    lowc=mean;
    bigC=cov;

    * Mean vector of std norm order statistics (c);

    * Cov matrix of std norm order statistics (C);

```

```

s1=&tau2+&sig2/&n1*(1+(&n1-1)*&rho);      * Sigma-1-squared: var of sample cluster mean, time 1;
s2=&tau2+&sig2/&n2*(1+(&n2-1)*&rho);      * Sigma-2-squared: var of sample cluster mean, time 2;
L=&tau2*&psi+&sig2*&rho;                  * Lambda: Cov of time 1 & 2 sample cluster means;
L12=L/sqrt(s1*s2);                      * Lambda12: Corr of time 1 & 2 sample cluster
means;
V=I(&p/2)@{1 -1, -1 1}/(&p/2)**2;      * Variance of the T-C randomized contrast;
omega=trace((cov+mean*mean`)*V)/&p;

*****;

create results var{iterate Beta3 seBeta3 t_stat p_value SSE E_SSE};
critval=quantile('T',.975,&p/2-1);

use themeans;

do iterate=1 to &iterate;
  %let howmany=%eval(2*&p);
  read next &howmany var{time ptreatment meany group} into temp;
  print temp;
  time=temp[,1]; treat=temp[,2]; y=temp[,3]; group=temp[,4]; free temp;

  t2=treat[loc(time=2)];
  y1=y[loc(time=1)]; y2=y[loc(time=2)];
  group=design(group[loc(time=2)]);
  y2=y[loc(time=2)]-y[loc(time=1)];

  y2[loc(t2=1)]=y2[loc(t2=1)]+&mu2;

  **      Case 6 **;

  X=J(&p,1)||group[,2:%eval(&p/2)]||t2;

  X1=X[,1:%eval(&p/2)];

  * Squared Projections;
  PR=y2`*y2;
  PJ=y2`*J(&p,1)*inv(J(&p,1)`*J(&p,1))*J(&p,1)`*y2;
  PX=y2`*X*inv(X`*X)*X`*y2;
  PT_X=y2`*X1*inv(X1`*X1)*X1`*y2;

```

```

* Estimates;
    Beta=inv(X`*X)*X`*y2;
    Beta3=beta[%eval(&p/2+1)];

* Sums of Squares;
    SSTotal=PR-PJ;
    SST=PX-PT_X;
    SSE=PR-PX;

    seBeta3=sqrt((SSE/(&p/2-1))@ (inv(X`*X)[%eval(&p/2+1),%eval(&p/2+1)]));
    E_SSE=(&p/2-1)*s2*(1-L12**2)+s2*L12**2*trace((I(&p)-X*inv(X`*X)*X`)*(bigC+lowc*lowc`));

* Test Statistic - F-stat;
    F=SST/(SSE/(&p/2-1));
    t_stat=sqrt(F)*sign(beta[%eval(&p/2+1)]);
    p_value=(abs(t_stat) >= critval);

* Save the results;
    append;

end;

* Non-centrality and power;
    mean=&mu2@t2 + sqrt(s2)*L12*lowc;
    var=(4/&p)*s2*(1-L12**2) + (sqrt(s2)*L12-sqrt(s1))**2*&p*omega;
    NCP_PX=mean`*X*inv(X`*X)*X`*mean;
    NCP_PT_X=mean`*X1*inv(X1`*X1)*X1`*mean;
    NCP=(NCP_PX-NCP_PT_X)/var*(4/&p);
    Power=1-CDF('F',critval**2,1,&p/2-1,NCP);
    print var Power;
    V_SSE=((&p/2-1)*s2*(1-L12**2)+s2*L12**2*trace((I(&p)-X1*inv(X1`*X1)*X1`-(&p/4)*V)*
(bigC+lowc*lowc`)))
        /(&p/2-1)*4/&p;
    IsV=(I(&p)-X1*inv(X1`*X1)*X1`-(&p/4)*V)/(&p/2-1)*4/&p;
    Diff_IsV_V=fuzz(IsV-V);
    Print V_SSE IsV,V Diff_IsV_V;

quit;

proc means data=results n mean median var std stderr clm;
    var Beta3 seBeta3 t_stat p_value SSE E_SSE;
run;

```



```

*****
***
This code computes the power of the test for case one. The power of the tests associated with six case
analyses
is derived from the associated F distribution, which for some cases is an approximation. The
noncentrality parameter
is computed for various treatment effects and power curves are generated using the SAS function CDF
with
the "F" distribution option. For Cases 1-3, the exact F distribution is used.
*****
****;

    %let ntime2 = n;  *** Uses 1/2 the total sample size at time two.;

proc iml;
  create power1A var{p n rho psi eff Power};

  do p=10 to 50 by 40;
  do n=10 to 50 by 40;
  do tau2_=0 to 10 by 5;

      tau2=tau2_/100;
      if tau2=0 then tau2=0.01;
      rho=tau2;
      sig2=1-tau2;

  do psi_=2 to 8 by 6;  psi=psi_/10;

  critval=quantile('T',.975,p-2);

  do i_mu2=0 to 50 by 1;
    mu2=i_mu2/100;      * Treatment effect;
    eff=mu2/(tau2+sig2);

    t2={0,1}@J(p/2,1);

    X=J(p,1)||t2;
    X1=X[,1];

    * Non-centrality and power;
    mean=mu2@t2;
    NCP_PX=mean`*X*inv(X`*X)*X`*mean;
    NCP_PT_X=mean`*X1*inv(X1`*X1)*X1`*mean;
  end;
  end;
  end;
end;

```

*** Use just n at time 1 ***;

```

        NCP=(NCP_PX-NCP_PT_X)/(tau2+sig2/(&ntime2));
        Power=1-CDF('F',critval**2,1,p-2,NCP);
    *      print Power;

        append;
    end;
end; end; end; end;
quit;
```

*** Uses the full sample size at time two.;

```

    %let ntime2 = 2*n;
proc iml;
    create power1B var{p n rho psi eff Power};
```

```

    do p=10 to 50 by 40;
    do n=10 to 50 by 40;
    do tau2_0=0 to 10 by 5;
    *      do tau2_100=100 to 100;
```

```

    tau2=tau2_/100; * Remember that rho=tau2 in this case (since
tau2+sig2=1);
```

```

        if tau2=0 then tau2=0.01;
        rho=tau2;
        sig2=1-tau2;
```

```

    do psi_2 to 8 by 6;    psi=psi_/10;
```

```

    critval=quantile('T',.975,p-2);
```

```

    do i_mu2=0 to 50 by 1;
        mu2=i_mu2/100;
        eff=mu2/(tau2+sig2);
```

```

        t2={0,1}@J(p/2,1);
```

```

        X=J(p,1)||t2;
        X1=X[,1];
```

```

    * Non-centrality and power;
        mean=mu2@t2;
        NCP_PX=mean`*X*inv(X`*X)*X`*mean;
        NCP_PT_X=mean`*X1*inv(X1`*X1)*X1`*mean;
```

```

2 ***;          NCP=(NCP_PX-NCP_PT_X)/(tau2+sig2/(&ntime2));    *** Use all of the data (2n) at time

          Power=1-CDF('F',critval**2,1,p-2,NCP);
          print Power;

          append;
          end;
          end; end; end; end;
quit;

** Crude power curve **;

proc gplot data=power1;
  by p n rho;
  where (psi=0.2);
  plot Power*eff;
  symbol v=none i=join;
run;
quit;

```

```

*****
This code computes the power of the test for Case two. The exact F distribution is used.
The noncentrality parameter is computed for treatment effects and power curves are
generated using the SAS function CDF with the "F" distribution option.
*****;

```

```
libname power "C:\Misook Park - Dissertation\Power Curves";
```

```
%let iterate=2500;
```

```
proc iml;
```

```
    create power2A var{p n rho psi eff Power};
```

```
    use power.themeans;
```

```
    do p=10 to 50 by 40;
```

```
        critval=quantile('T',.975,p-3);
```

```
    do n=10 to 50 by 40;
```

```
        do tau2_=0 to 10 by 5;
```

```
    *    do tau2_=100 to 100;
```

```
        tau2=tau2_/100; * Remember that rho=tau2 in this case (since tau2
```

```
+sig2=1);
```

```
            if tau2=0 then tau2=0.01;
```

```
            rho=tau2;
```

```
            sig2=1-tau2;
```

```
    do psi_=2 to 8 by 6;    psi=psi_/10;
```

```
        L12=tau2*psi/sqrt((tau2+sig2/n)*(tau2+sig2/n));
```

```
        var=(4/p)*(tau2+sig2/n)*(1-L12**2);
```

```
    do iterate=1 to &iterate;
```

```
        if p=10 then read next 20 var{time treatment meany} into temp;
```

```
        if p=50 then read next 100 var{time treatment meany} into temp;
```

```
        time=temp[,1]; treat=temp[,2]; y=temp[,3]; free temp;
```

```
        t2=treat[loc(time=2)];
```

```
        y1=y[loc(time=1)];
```

```
        X=J(p,1)||y1||t2;
```

```
    * Projections;
```

```

PX=X*inv(X`*X)*X`;
PT_X=X[,1:2]*inv(X[,1:2]*X[,1:2])*X[,1:2]`;

do i_mu2=0 to 50 by 1;
  mu2=i_mu2/100;      * Treatment effect;
  eff=mu2/(tau2+sig2);

  y2=y[loc(time=2)];
  y2[loc(t2=1)]=y2[loc(t2=1)]+mu2;

  * Estimates;
  Beta=inv(X`*X)*X`*y2;

  * Sums of Squares;
  SST=y2`*(PX-PT_X)*y2;
  SSE=y2`*(I(p)-PX)*y2;

  * Test Statistic - t-stat;
  t_stat=sqrt(SST/(SSE/(p-3)))*sign(beta[3]);
  p_value=(abs(t_stat) >= critval);

  * Non-centrality and power;
  mean=mu2@t2 + L12*y1;
  NCP=(mean`*(PX-PT_X)*mean)/(var*(p/4)); if NCP<0 then NCP=0;
  Power=1-CDF('F',critval**2,1,p-3,NCP);

  * Save the results;
  append;
end;
end; end; end; end;
quit;

proc summary data=power2A mean;
  class p n rho psi eff;
  var Power;
  output out=power2 mean=Power;
  ways 5;

```

run;

```

*****
This code computes the power of the test for Case 3. The exact F distribution is used.
The noncentrality parameter is computed for treatment effects and power curves are
generated using the SAS function CDF with the "F" distribution option. Note that rho=tau2 in this case
(since tau2+sig2=1);
*****;

```

```

*** Case 3 ***;

```

```

proc iml;
  create power3 var{p n rho psi eff Power};

  do p=10 to 50 by 40;
  do n=10 to 50 by 40;
  do tau2_ =0 to 10 by 5;
*  do tau2_ =100 to 100;

                                tau2=tau2_/100;
                                if tau2=0 then tau2=0.01;
                                rho=tau2;
                                sig2=1-tau2;

  do psi_ =2 to 8 by 6;    psi=psi_/10;

  critval=quantile('T',.975,p-2);

    do i_mu2=0 to 50 by 1;
      mu2=i_mu2/100;    * Treatment effect;
      eff=mu2/(tau2+sig2);

      t2={0,1}@J(p/2,1);

      X=J(p,1)||t2;
      X1=X[,1];

* Non-centrality and power;
      mean=mu2@t2;
      NCP_PX=mean`*X*inv(X`*X)*X`*mean;
      NCP_PT_X=mean`*X1*inv(X1`*X1)*X1`*mean;
      NCP=(NCP_PX-NCP_PT_X)/((1/n+1/n)*sig2+2*tau2*(1-psi));
      Power=1-CDF('F',critval**2,1,p-2,NCP);

    append;
  end;
end;

```

```
        end; end; end; end;  
quit;  
  
** Crude power curve **;  
  
proc gplot data=power3;  
    by p n rho psi;  
    plot Power*eff;  
    symbol v=none i=join;  
run;  
quit;
```



```

*****
This code computes the power of the test for Case 4. The approximate F distributions were used.
The power of the tests associated with six case analyses is derived from the associated F distribution,
which for
some cases is an approximation. The noncentrality parameter is computed for various treatment effects and
power
curves are generated using the SAS function CDF with the "F" distribution option.
*****;

```

```

libname order "C:\Misook Park - Dissertation\Normal Order Statistics Data";

```

```

*** Case 4 ***;

```

```

%let rho=0;

```

```

proc iml;

```

```

    create power4 var{p n rho psi eff Power};

```

```

    do p=10 to 50 by 40;

```

```

        **** Prepare the mean and covariance of the order statistics, given the number of clusters ****;

```

```

        hn=int(p/2); *Half p (hn);

```

```

        use order.means var {"n" "i" "x"} where(n=p);

```

```

        read all var {"x"} into mean;

```

```

        rx=mean; * reverse x;

```

```

        do i=1 to hn; rx[i]=mean[hn-i+1]; end;

```

```

        if p/2=hn then mean=-mean//rx;

```

```

        else mean=-mean//{0}//rx;

```

```

        * Mean vector of std norm order statistics (c);

```

```

        *
        print mean;

```

```

        cov=J(p,p,0);

```

```

        use order.cov var {"n" "i" "j" "v"} where (n=p);

```

```

        do k=1 to int(((p+1)**2)/4);

```

```

            read next var {"i" "j" "v"} into temp;

```

```

            cov[temp[1],temp[2]]=temp[3];

```

```

        end;

```

```

        do i=2 to p;

```

```

            do j=p to max(p-i+2,i) by -1;

```

```

        cov[i,j]=cov[p-j+1,p-i+1];
    end;
end;
cov=cov+cov`-diag(cov);
statistics (C);
* print cov;
  lowc=mean;
  bigC=cov;

critval=quantile('T',.975,p/2-1);

t2=(J(int(p/4+0.5),1)@{0,1,1,0})[1:p,]; * Invariant to the choice of t2? **;
X=J(p,1)||I(p/2)@{1,1})[,2:p/2]||t2;
X1=X[,1:p/2];

*** Projections ***;
PX=X*inv(X`*X)*X`;
PX1=X1*inv(X1`*X1)*X1`;
PT=PX-PX1;
u=PT*t2; u=u/sqrt(u`*u); *** Normalized vector contained in the 1-dimensional vector
space [X]||[X1] ***;

do n=10 to 50 by 40;
do tau2_=0 to 10 by 5;
* do tau2_=100 to 100;

        tau2=tau2_/100;
        if tau2=0 then tau2=0.01;
        rho=tau2;
        sig2=1-tau2;

do psi_=2 to 8 by 6;    psi=psi_/10;

        s1=tau2+sig2/n*(1+(n-1)*&rho); * Sigma-1-squared: var of sample cluster mean, time 1;
        s2=tau2+sig2/n*(1+(n-1)*&rho); * Sigma-2-squared: var of sample cluster mean, time 2;
        L=tau2*psi+sig2*&rho; * Lambda: Cov of time 1 & 2 sample cluster means;
        L12=L/sqrt(s1*s2); * Lambda12: Corr of time 1 & 2 sample cluster
means;

        V=I(p/2)@{1 -1, -1 1}/(p/2)**2; * Variance of the T-C randomized contrast;

do i_mu2=0 to 50 by 1;
    mu2=i_mu2/100; * Treatment effect;

```

```

        eff=mu2/(tau2+sig2);

    * Non-centrality and power;
        mean=mu2@t2;
        kappa=s2*(1-L12**2) + s2*L12**2*u`*bigC*u;
        NCP=mean`*PT*mean/kappa;
        Power=1-CDF('F',critval**2,1,p/2-1,NCP);
        append;
    end;

end; end; end; end;
quit;

** Crude power curve **;

proc gplot data=power4;
    by p n rho psi;
    plot Power*eff;
    symbol v=none i=join;
run;
quit;

```

```

*****
***
This code computes the power of the test for Case 5. The exact F distributions were used.
The noncentrality parameter is computed for treatment effects and power
curves are generated using the SAS function CDF with the "F" distribution option.
*****
****;

libname order "C:\Misook Park - Dissertation\Normal Order Statistics Data";
libname power "C:\Misook Park - Dissertation\Power Curves";

*** Case 5 ***;

%let iterate=2500;

options nonotes; *** Suppresses the notes on allocation of symbol space ***;

proc iml;
  create power5A var{p n rho psi eff Power};

  use power.themeans;

  do p=10 to 50 by 40;
    critval=quantile('T',.975,p/2-2);
    **** Prepare the mean and covariance of the order statistics, given the number of clusters
  ***;

    hn=int(p/2); *Half p (hn);
    use order.means var {"n" "i" "x"} where (n=p);
    read all var {"x"} into mean;
    close order.means;
    rx=mean; * reverse x;
    do i=1 to hn; rx[i]=mean[hn-i+1]; end;
    if p/2=hn then mean=-mean//rx;
    else mean=-mean//{0}//rx; * Mean vector of standard normal

  order statistics (c);

    cov=J(p,p,0);
    use order.cov var {"n" "i" "j" "v"} where (n=p);
    do k=1 to int(((p+1)**2)/4);
      read next var {"i" "j" "v"} into temp;
      cov[temp[1],temp[2]]=temp[3];
    end;
    do i=2 to p;

```

```

                do j=p to max(p-i+2,i) by -1;
                    cov[i,j]=cov[p-j+1,p-i+1];
                end;
            end;
            cov=cov+cov`-diag(cov);
normal order statistics (C);
            * print cov;
            lowc=mean;
            bigC=cov;
            V=I(p/2)@{1 -1, -1 1}/(p/2)**2;
contrast;
            *****;
            do n=10 to 50 by 40;
            do tau2_ =0 to 10 by 5;
            * do tau2_ =100 to 100;
                                tau2=tau2_/100;
                                * Note that rho=tau2 in
this case (since tau2+sig2=1);
                                if tau2=0 then tau2=0.01;
                                rho=tau2;
                                sig2=1-tau2;
            do psi_ =2 to 8 by 6;    psi=psi_/10;
                                s1=tau2+sig2/n*(1+(n-1)*&rho);
                                s2=tau2+sig2/n*(1+(n-1)*&rho);
                                L=tau2*psi+sig2*&rho;
                                L12=L/sqrt(s1*s2);
                                * Sigma-1-squared: var of sample cluster mean, time 1;
                                * Sigma-2-squared: var of sample cluster mean, time 2;
                                * Lambda: Cov of time 1 & 2 sample cluster means;
                                * Lambda12: Corr of time 1 & 2 sample cluster
means;
                                var=(4/p)*s2*(1-L12**2);

            setin power.themean;

            do iterate=1 to &iterate;
                if p=10 then read next 20 var{time ptreatment meany group} into temp;
                if p=50 then read next 100 var{time ptreatment meany group} into temp;
                time=temp[,1]; treat=temp[,2]; y=temp[,3]; group=temp[,4]; free temp;

                t2=treat[loc(time=2)];
                y1=y[loc(time=1)];
                group=design(group[loc(time=2)]);

                X=J(p,1)||group[,2:(p/2)]||y1||t2;
                X1=X[,1:(p/2+1)];

```

```

* Projections;
  PX=X*inv(X`*X)*X`;
  PT_X=X1*inv(X1`*X1)*X1`;

do i_mu2=0 to 50 by 1;
  mu2=i_mu2/100;      * Treatment effect;
  eff=mu2/(tau2+sig2);

  y2=y[loc(time=2)];
  y2[loc(t2=1)]=y2[loc(t2=1)]+mu2;

  * Non-centrality and power;
  mean=mu2@t2 + sqrt(s2/s1)*L12*y1;
  NCP=(mean`*(PX-PT_X)*mean)/(var*(p/4));
  if NCP<0 then NCP=0;
  Power=1-CDF('F',critval**2,1,p/2-2,NCP);

  * Save the results;
  append;
end;
end;
end; end; end;
quit;

options notes;

proc summary data=power5A mean;
  class p n rho psi eff;
  var Power;
  output out=power5 mean=Power;
  ways 5;
run;

```

```

*****
This code computes the power of the test for Case 6. The approximate F distributions were used.
The noncentrality parameter is computed for the treatment effects and power
curves are generated using the SAS function CDF with the "F" distribution option.
*****;

libname order "C:\Misook Park - Dissertation\Normal Order Statistics Data";

*** Case 6 ***;

%let rho=0;

proc iml;

    create power6 var{p n rho psi eff Power};

    do p=10 to 50 by 40;
        **** Prepare the mean and covariance of the order statistics, given the number of clusters ****;
        hn=int(p/2); *Half p (hn);
        use order.means var {"n" "i" "x"} where(n=p);
        read all var {"x"} into mean;
        rx=mean; * reverse x;
        do i=1 to hn; rx[i]=mean[hn-i+1]; end;
        if p/2=hn then mean=-mean//rx;
        else mean=-mean//{0}//rx;
        statistics (c);
        *
        print mean;

        cov=J(p,p,0);
        use order.cov var {"n" "i" "j" "v"} where (n=p);
        do k=1 to int((p+1)**2/4);
            read next var {"i" "j" "v"} into temp;
            cov[temp[1],temp[2]]=temp[3];
        end;
        do i=2 to p;
            do j=p to max(p-i+2,i) by -1;
                cov[i,j]=cov[p-j+1,p-i+1];
            end;
        end;
        cov=cov+cov`-diag(cov);
        * Mean vector of standard normal order
        * Covariance matrix of standard normal order
    end;
end;

```

```

statistics (C);
*
  print cov;
  lowc=mean;
  bigC=cov;

  critval=quantile('T',.975,p/2-1);

  V=I(p/2)@{1 -1, -1 1}/(p/2)**2;          * Variance of the T-C randomized contrast;
  omega=trace((bigC+lowc*lowc`)*V)/p;

  t2=(J(int(p/4+0.5),1)@{0,1,1,0})[1:p,];
  X=J(p,1)|| (I(p/2)@{1,1})[,2:p/2]||t2;
  X1=X[,1:p/2];

do n=10 to 50 by 40;
do tau2_=0 to 10 by 5;

      tau2=tau2_/100;
      if tau2=0 then tau2=0.01;
      rho=tau2;
      sig2=1-tau2;

do psi_=2 to 8 by 6;    psi=psi_/10;

  s1=tau2+sig2/n*(1+(n-1)*&rho);    * Sigma-1-squared: var of sample cluster mean, time 1;
  s2=tau2+sig2/n*(1+(n-1)*&rho);    * Sigma-2-squared: var of sample cluster mean, time 2;
  L=tau2*psi+sig2*&rho;             * Lambda: Cov of time 1 & 2 sample cluster means;
  L12=L/sqrt(s1*s2);               * Lambda12: Corr of time 1 & 2 sample cluster means;

do i_mu2=0 to 50 by 1;
  mu2=i_mu2/100;    * Treatment effect;
  eff=mu2/(tau2+sig2);

  * Non-centrality and power;
  mean=mu2@t2;
  var=(4/p)*s2*(1-L12**2) + (sqrt(s2)*L12-sqrt(s1))**2*p*omega;
  NCP_PX=mean`*X*inv(X`*X)*X`*mean;
  NCP_PT_X=mean`*X1*inv(X1`*X1)*X1`*mean;
  NCP=(NCP_PX-NCP_PT_X)/var*(4/p);
  Power=1-CDF('F',critval**2,1,p/2-1,NCP);

```



```
end;          append;  
end; end; end; end;  
quit;
```

```
*****
*This SAS code is to combine the Power Curves. First combine the data and then produce powerplots*
*****;
```

```
libname power "C:\Misook Park - Dissertation\Power Curves";
```

```
*** Build the data for these plots ***;
```

```
* 1; %include "C:\Misook Park - Dissertation\Power Curves\Power calculations for case 1.sas";
* 2; %include "C:\Misook Park - Dissertation\Power Curves\Power calculations for case 2.sas";
* 3; %include "C:\Misook Park - Dissertation\Power Curves\Power calculations for case 3.sas";
* 4; %include "C:\Misook Park - Dissertation\Power Curves\Power calculations for case 4.sas";
* 5; %include "C:\Misook Park - Dissertation\Power Curves\Power calculations for case 5.sas";
* 6; %include "C:\Misook Park - Dissertation\Power Curves\Power calculations for case 6.sas";
```

```
data powerback;
    set power.power;
run;
```

```
*** Use this code to combine the data ***;
```

```
data power.power;
    merge
        power1A
        power1B
        power2
        power3
        power4
        power5
        power6;
    drop _TYPE_ _FREQ_;
run;
```

```
*** Use this code to produce powerplots if you do not have the 7 individual
    data sets available... uses just power.power;
```

```
proc iml;
    use power.power;

    read next 1224 var{p n rho psi eff power};
    power1A=power; free power;
    read next 1224 var{power} into power1B;
    read next 1224 var{power} into power2;
    read next 1224 var{power} into power3;
    read next 1224 var{power} into power4;
    read next 1224 var{power} into power5;
```

```

read next 1224 var{power} into power6;

create power.powerplots var{p n rho psi eff power1A power1B power2 power3 power4 power5 power6};
append;
quit;

*** Use this code if you have all of the merged datasets available ***;
data power.powerplots;
    merge      power1A(rename=(power=power1A))
              power1B(rename=(power=power1B))
              power2 (rename=(power=power2))
              power3 (rename=(power=power3))
              power4 (rename=(power=power4))
              power5 (rename=(power=power5))
              power6 (rename=(power=power6));
    drop _TYPE_ _FREQ_;
run;
/**/

** Crude power curve **;
proc sort data=power.power;
    by p n rho psi;
run;

proc gplot data=power.power;
    by p n rho psi;
    plot Power*eff;
    symbol v=none i=join;
run;
quit;

proc freq data=power.power;
    tables case;
run;

```

```

*****
* This SAS code performs the simulation of full data for power curves.
*The value of the test statistic is generated for each simulated iteration. The test statistic is compared
*to the appropriate 0.05 critical value taken from an F distribution and an indicator variable
*(1=critical value is exceeded, 0=otherwise) is stored. The average value of this indicator variable over
*all 2500 iterates constitutes an estimate of the power. Estimates of the power along with 95% confidence
* intervals were computed. In each case the analytically computed power fell within the confidence limits.
*****;

libname power "C:\Misook Park - Dissertation\Power Curves";

%let iterate=2500;
%let rho=0;           * inherent correlation is zero for our applications;
%let eta=0;

proc iml;
    create thedata var {"p","n","rho","psi","iterate","cluster","treatment","time","y"};

do p=10 to 50 by 40;
do n=10 to 50 by 40;
    do tau2_=0 to 10 by 5;

        tau2=tau2_/100;
        if tau2=0 then tau2=0.01;
        rho=tau2;
        sig2=1-tau2;

do psi_=2 to 8 by 6;    psi=psi_/10;

        varmat1=tau2 @ (psi @ (J(2,2) - I(2)) + I(2));
        varmat211=sig2 @ (&rho @ (J(n,n) - I(n)) + I(n));
        varmat222=sig2 @ (&rho @ (J(n,n) - I(n)) + I(n));
        varmat212=&eta @ sig2 * J(n,n);
        varmat221=varmat212`;
        varmat2=(varmat211 || varmat212) // (varmat221 || varmat222);

print varmat1;

test=(j(n,1,0)//j(n,1,1))/n;
testvar=test`*varmat2*test;
print testvar;

```

```

do k=1 to &iterate;
    call vnormal(x,j(2,1,0),varmat1,p,0);

    do i=1 to p;
        treat=(i > p/2); * Treatment=1, Control=0, no blocks, CR design;
        mean=(x[i,1]@j(n,1))/(x[i,2]@j(n,1));
        call vnormal(y0,mean,varmat2,1,0);
        y=y/(j(2*n,1,k)||j(2*n,1,i)||j(2*n,1,treat)||j(n,1,1)/j(n,1,2)||y0`);
    *    print mean y0;
    end;
    y=( (p||n||rho||psi)@j(2*p*n,1) ) || y;
    append from y;
    free y;
end;

*    print y;
end;end;end;end;
quit;

proc summary data=thedata;
    class p n rho psi iterate cluster time;
    id treatment;
    var y;
    ways 7;
    output out=themean mean=mean;
run;

proc sort data=themean (where=(time=1)) out=paired;
    by p n rho psi iterate mean;
run;

data paired;
    retain ptreatment;
    set paired;
    by p n rho psi iterate;
    if first.iterate then
        do;
            group=0;
            ordered=0;

```

```

        end;
ordered+1;
if mod(_n_,2)=1 then
    do;
        ptreatment=(ranuni(0)<0.5);
        group+1;
        * Matched pair label;
    end;
else ptreatment=1-ptreatment;
keep p n rho psi iterate cluster group ptreatment ordered;
run;

proc sort data=paired;
    by p n rho psi iterate cluster;
run;

data themeans;
    merge themeans paired;
    by p n rho psi iterate cluster;
run;

data thedata;
    merge thedata paired;
    by p n rho psi iterate cluster;
run;

proc sort data=themeans out=power.themeans;
    by p n rho psi iterate ordered;
run;

proc sort data=thedata out=power.thedata;
    by p n rho psi iterate ordered;
run;

```

VITA

Misook Park was born on July 18, 1956, Seoul Korea. She graduated from YonSei University, Seoul, Korea with B.S. in Human Ecology. She received a M.S. in Biostatistics from the University of North Carolina, Chapel Hill in 1989. She taught College at Kumho National Institute of Technology, Kumi City in Korea before returning to school to earn her Ph.D. She began studies in Biostatistics at Virginia Commonwealth University in the fall of 2000.