DOES PAIR-MATCHING ON ORDERED BASELINE MEASURES INCREASE POWER: A SIMULATION STUDY

Yan Jin
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/399

This Thesis 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.
DOES PAIR-MATCHING ON ORDERED BASELINE MEASURES INCREASE POWER: A SIMULATION STUDY

A Thesis submitted in partial fulfillment of the requirements for the degree of
Master of Science at Virginia Commonwealth University.

by

YAN JIN
M.S. Suzhou Medical College, Suzhou, Jiangsu, China, 1992
M.S. University of Medicine and Dentistry of New Jersey, 2002

Major Director: Dr. Robert E. Johnson
Affiliate Associate Professor
Department of Biostatistics

Advisor: Dr. Roy T. Sabo
Assistant Professor
Department of Biostatistics

Virginia Commonwealth University
Richmond, Virginia
August, 2012
Acknowledgement

I would like to express my deepest gratitude to my advisor, Dr. Robert E Johnson, for his excellent guidance, caring, patience, and invaluable assistance. It would not have been possible for me to write this thesis without his kind help and encouragement. It is an honor for me to become his last student as he starts a new chapter of his life journey. I am most grateful to the members of my committee, Dr. Roy Sabo and Dr. D'Arcy P. Mays, for their valuable suggestions and comments. Their dedication and profound knowledge of statistics helped improve the quality of my thesis. Special thanks to Dr. Donna K. McClish and Russell Boyle for their constant guidance and encouragement. I am very thankful for Yvonne’s kind help during the last six years. I also would like to thank Lisa Shickle for her support and encouragement. I would never have been able to finish my thesis without the guidance of my advisor, committee members, help from fellow students, and support from my family.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgements</td>
<td>ii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>vi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>vii</td>
</tr>
<tr>
<td>Abstract</td>
<td>x</td>
</tr>
<tr>
<td>Chapter I Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Chapter II Background</td>
<td>3</td>
</tr>
<tr>
<td>2.1 Literature Review</td>
<td>4</td>
</tr>
<tr>
<td>2.2 Summary</td>
<td>6</td>
</tr>
<tr>
<td>Chapter III Methods</td>
<td></td>
</tr>
<tr>
<td>3.1 The Models</td>
<td>8</td>
</tr>
<tr>
<td>3.2 Simulation</td>
<td>10</td>
</tr>
<tr>
<td>3.2.1 Software to Use</td>
<td>10</td>
</tr>
<tr>
<td>3.2.2 Methods of Simulating the Datasets</td>
<td>11</td>
</tr>
<tr>
<td>3.2.3 Random Number Generator to Use</td>
<td>12</td>
</tr>
<tr>
<td>3.2.4 Treatment Effects Calculation</td>
<td>14</td>
</tr>
<tr>
<td>3.2.5 Power Analysis</td>
<td>15</td>
</tr>
<tr>
<td>3.3 Evaluation</td>
<td>15</td>
</tr>
</tbody>
</table>
Chapter IV  Results

4.1  Evaluation of Simulated Data

4.2  Evaluation of Pair-Matching on Ordered Baseline Measures

4.3  Evaluation of Power Results

4.4  The Six Test Cases Results

4.4.1  Case 3 vs. case 6 (N=60, n=30 for each group)

4.4.2  Case 3 vs. case 6 (N=40, 20, n=20, 10 for each group respectively)

4.4.3  Case 2 vs. case 5 (N=60, n=30 in each group)

4.4.4  Case 2 vs. case 5 (N=40, 20, n=20, 10 for each group respectively)

4.4.5  Case 1 vs. case 4 (N=60, n=30 for each group)

4.4.6  Case 1 vs. case 4 (N=40, 20, n=20, 10 for each group)

Chapter V  Conclusion

Chapter VI  Summary and Future Studies

References

Appendices

A  Figures of case 3 vs. case 6 (N=60, RHO=0.8)

B  Figures of case 3 vs. case 6 (N=60, RHO=0.2)

C  Figures of case 2 vs. case 5 (N=60, RHO=0.8)

D  Figures of case 2 vs. case 5 (N=60, RHO=0.2)
E  Figures of case 1 vs. case 4 (N=60, RHO=0.8).................................................. 87
F  Figures of case 1 vs. case 4 (N=60, RHO=0.2).................................................. 98
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Six Cases: Design and Analysis</td>
<td>6</td>
</tr>
<tr>
<td>Table 2a</td>
<td>Six Scenarios</td>
<td>13</td>
</tr>
<tr>
<td>Table 2b</td>
<td>The Parameters of the First Scenario in Each Unit Dataset</td>
<td>14</td>
</tr>
<tr>
<td>Table 3</td>
<td>The Data Structure of Baseline Pair-Matching</td>
<td>16</td>
</tr>
<tr>
<td>Table 4</td>
<td>Correlation Coefficients between Baseline and Post-baseline Measures</td>
<td>20</td>
</tr>
<tr>
<td>Table 5</td>
<td>Baseline Measures Pair-matching Evaluation</td>
<td>21</td>
</tr>
</tbody>
</table>
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Histograms of simulated baseline and post-baseline distribution (Group=1)</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>Normal probability plots of simulated baseline and post-baseline measures</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>Boxplots of simulated baseline and post-baseline measures across eight group levels</td>
<td>19</td>
</tr>
<tr>
<td>4</td>
<td>Histograms of simulated baseline and post-baseline distribution (Group=20)</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>Histograms of simulated baseline and post-baseline distribution (Group=22)</td>
<td>20</td>
</tr>
<tr>
<td>6a</td>
<td>Power curves of case 3 vs. case 6 (Tau=1, Delta=0, P=0.5)</td>
<td>22</td>
</tr>
<tr>
<td>6b</td>
<td>Power of case 3 vs. case 6 across Delta values (effect size=0)</td>
<td>22</td>
</tr>
<tr>
<td>7a</td>
<td>Curves of case 3 vs. case 6 (Tau=1, Delta=3, P=0.5)</td>
<td>23</td>
</tr>
<tr>
<td>7b</td>
<td>Power of case 3 vs. case 6 across Delta values (effect size=0.3)</td>
<td>23</td>
</tr>
<tr>
<td>8a</td>
<td>Power curves of case 3 vs. case 6 (Tau=1, Delta=6, P=0.5)</td>
<td>23</td>
</tr>
<tr>
<td>8b</td>
<td>Power of case 3 vs. case 6 across Delta values (effect size=0.5)</td>
<td>23</td>
</tr>
<tr>
<td>9a</td>
<td>Power curves of case 3 vs. case 6 (Tau=3, Delta=0, P=0.1)</td>
<td>24</td>
</tr>
<tr>
<td>9b</td>
<td>Power of case 3 vs. case 6 across Delta values (effect size=0)</td>
<td>24</td>
</tr>
<tr>
<td>10a</td>
<td>Power curves of case 3 vs. case 6 (Tau=3, Delta=3, P=0.1)</td>
<td>24</td>
</tr>
<tr>
<td>10b</td>
<td>Power of case 3 vs. case 6 across Delta values (effect size=0.3)</td>
<td>24</td>
</tr>
<tr>
<td>11a</td>
<td>Power curves of case 3 vs. case 6 (Tau=3, Delta=6, P=0.1)</td>
<td>25</td>
</tr>
<tr>
<td>11b</td>
<td>Power of case 3 vs. case 6 across Delta values (effect size=0.5)</td>
<td>25</td>
</tr>
<tr>
<td>12a</td>
<td>Power curves of case 3 vs. case 6 (Tau=1, Delta=0, P=0.5)</td>
<td>26</td>
</tr>
</tbody>
</table>
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>12b</td>
<td>Power of case 3 vs. case 6 across Delta values (effect size=0)</td>
<td>26</td>
</tr>
<tr>
<td>13a</td>
<td>Power curves of case 3 vs. case 6 (Tau=1, Delta=3, P=0.5)</td>
<td>26</td>
</tr>
<tr>
<td>13b</td>
<td>Power of case 3 vs. case 6 across Delta values (effect size=0.3)</td>
<td>26</td>
</tr>
<tr>
<td>14a</td>
<td>Power curves of case 3 vs. case 6 (Tau=1, Delta=6, P=0.5)</td>
<td>27</td>
</tr>
<tr>
<td>14b</td>
<td>Power of case 3 vs. case 6 across Delta values (effect size=0.5)</td>
<td>27</td>
</tr>
<tr>
<td>15a</td>
<td>Power curves of case 3 vs. case 6 (Tau=3, Delta=0, P=0.7)</td>
<td>27</td>
</tr>
<tr>
<td>15b</td>
<td>Power of case 3 vs. case 6 across Delta values (effect size=0)</td>
<td>27</td>
</tr>
<tr>
<td>16a</td>
<td>Power curves of case 3 vs. case 6 (Tau=3, Delta=3, P=0.7)</td>
<td>28</td>
</tr>
<tr>
<td>16b</td>
<td>Power of case 3 vs. case 6 across Delta values (effect size=0.3)</td>
<td>28</td>
</tr>
<tr>
<td>17a</td>
<td>Power curves of case 3 vs. case 6 (Tau=3, Delta=6, P=0.7)</td>
<td>28</td>
</tr>
<tr>
<td>17b</td>
<td>Power of case 3 vs. case 6 across Delta values (effect size=0.5)</td>
<td>28</td>
</tr>
<tr>
<td>18a</td>
<td>Power curves of case 3 vs. case 6 (Tau=1, Delta=0, P=0.5)</td>
<td>29</td>
</tr>
<tr>
<td>18b</td>
<td>Power curves of case 3 vs. case 6 (Tau=3, Delta=3, P=0.1)</td>
<td>29</td>
</tr>
<tr>
<td>19a</td>
<td>Power curves of case 3 vs. case 6 (Tau=2, Delta=6, P=0.1)</td>
<td>29</td>
</tr>
<tr>
<td>19b</td>
<td>Power curves of case 3 vs. case 6 (Tau=3, Delta=0, P=0.5)</td>
<td>29</td>
</tr>
<tr>
<td>20a</td>
<td>Power curves of case 2 vs. case 5 (Tau=1, Delta=0, P=0.5)</td>
<td>30</td>
</tr>
<tr>
<td>20b</td>
<td>Power curves of case 2 vs. case 5 (Tau=2, Delta=3, P=0.9)</td>
<td>30</td>
</tr>
<tr>
<td>21a</td>
<td>Power curves of case 2 vs. case 5 (Tau=2, Delta=3, P=0.1)</td>
<td>31</td>
</tr>
<tr>
<td>21b</td>
<td>Power curves of case 2 vs. case 5 (Tau=3, Delta=6, P=0.1)</td>
<td>31</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>Figure 22a</td>
<td>Power curves of case 2 vs. case 5 (Tau=1, Delta=0, P=0.5)</td>
<td>32</td>
</tr>
<tr>
<td>Figure 22b</td>
<td>Power curves of case 2 vs. case 5 (Tau=1, Delta=3, P=0.5)</td>
<td>32</td>
</tr>
<tr>
<td>Figure 23a</td>
<td>Power curves of case 2 vs. case 5 (Tau=1, Delta=6, P=0.5)</td>
<td>32</td>
</tr>
<tr>
<td>Figure 23b</td>
<td>Power of case 2 vs. case 5 across Delta values (effect size=0.5)</td>
<td>32</td>
</tr>
<tr>
<td>Figure 24a</td>
<td>Power curves of case 2 vs. case 5 (Tau=3, Delta=6, P=0.1)</td>
<td>33</td>
</tr>
<tr>
<td>Figure 24b</td>
<td>Power of case 2 vs. case 5 across Delta values (effect size=0.5)</td>
<td>33</td>
</tr>
<tr>
<td>Figure 25a</td>
<td>Power curves of case 2 vs. case 5 (Tau=2, Delta=6, P=0.9)</td>
<td>33</td>
</tr>
<tr>
<td>Figure 25b</td>
<td>Power of case 2 vs. case 5 across Delta values (effect size=0.5)</td>
<td>33</td>
</tr>
<tr>
<td>Figure 26a</td>
<td>Power curves of case 2 vs. case 5 (Tau=3, Delta=6, P=0.7)</td>
<td>34</td>
</tr>
<tr>
<td>Figure 26b</td>
<td>Power of case 2 vs. case 5 across Delta values (effect size=0.5)</td>
<td>34</td>
</tr>
<tr>
<td>Figure 27a</td>
<td>Power curves of case 1 vs. case 4 (Tau=1, Delta=0, P=0.5)</td>
<td>35</td>
</tr>
<tr>
<td>Figure 27b</td>
<td>Power curves of case 1 vs. case 4 (Tau=1, Delta=3, P=0.1)</td>
<td>35</td>
</tr>
<tr>
<td>Figure 28a</td>
<td>Power curves of case 1 vs. case 4 (Tau=3, Delta=6, P=0.7)</td>
<td>35</td>
</tr>
<tr>
<td>Figure 28b</td>
<td>Power of case 1 vs. case 4 across Delta values (effect size=0.5)</td>
<td>35</td>
</tr>
<tr>
<td>Figure 29a</td>
<td>Power curves of case 1 vs. case 4 (Tau=1, Delta=0, P=0.5)</td>
<td>36</td>
</tr>
<tr>
<td>Figure 29b</td>
<td>Power curves of case 1 vs. case 4 (Tau=1, Delta=3, P=0.5)</td>
<td>36</td>
</tr>
<tr>
<td>Figure 30a</td>
<td>Power curves of case 1 vs. case 4 (Tau=2, Delta=0, P=0.7)</td>
<td>36</td>
</tr>
<tr>
<td>Figure 30b</td>
<td>Power curves of case 1 vs. case 4 (Tau=2, Delta=3, P=0.7)</td>
<td>36</td>
</tr>
<tr>
<td>Figure 31a</td>
<td>Power curves of case 1 vs. case 4 (Tau=2, Delta=6, P=0.7)</td>
<td>37</td>
</tr>
<tr>
<td>Figure 31b</td>
<td>Power of case 1 vs. case 4 across Delta values (effect size=0.5)</td>
<td>37</td>
</tr>
</tbody>
</table>
Abstract

DOES PAIR-MATCHING ON ORDERED BASELINE MEASURES INCREASE POWER: A SIMULATION STUDY

By Yan Jin M.S.

A Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

Virginia Commonwealth University, 2012

Major Director: Dr. Robert E. Johnson
Affiliate Associate Professor
Department of Biostatistics

Advisor: Dr. Roy T. Sabo
Assistant Professor
Department of Biostatistics

It has been shown that pair-matching on an ordered baseline with normally distributed measures reduces the variance of the estimated treatment effect (Park and Johnson, 2006). The main objective of this study is to examine if pair-matching improves the power when the distribution is a mixture of two normal distributions. Multiple scenarios with a combination of different sample sizes and parameters are simulated. The power curves are provided for three cases, with and without matching, as follows: analysis of post-intervention data only, adding baseline as a covariate, and classic pre-post comparison. The study shows that the additional variance reduction provided by pair-
matching in the pre-post design is limited for high correlation. When correlation is low, there is a significant power increase. It is shown that the baseline pair-matching improves the power when the two means of a mixture normal distribution are widely spread. The pattern becomes clear for low correlation.
CHAPTER 1

Introduction

A randomized controlled trial (RCT) is considered as the gold standard for both clinical and biomedical research (Schulz and Grimes, 2002). There are three main experimental designs associated with the RCT: the completely randomized design, randomized block design, and matched-pair design. The completely randomized design is mostly used when the experimental units are comparatively homogenous. The block design is a useful technique to divide the heterogeneous experimental units into homogeneous blocks, and randomization is then conducted separately with each block, resulting in high precision for the treatment effects inference due to small experimental error within blocks. A matched-pair design is a special case of a block randomization and can be used whenever the study has only two treatment groups: subjects can be arranged into pairs based on their pre-treatment important confounding variables. Within each matched-pair, subjects are randomly allocated to one of the two treatment groups (Kutner et al, 2004; Schulz and Grimes, 2002).

In general, the matched-pair design is better than a completely randomized or a block randomized design. It is believed that the matched method will “improve the precision of the comparison and simplify the analysis” (Billewicz, 1965). However, Billewicz’s simulation study showed the matching efficiency depends on the data type of response variable.
It has been shown that pair-matching on ordered baseline with normally distributed measures reduces the variance of the estimated treatment effect (Park and Johnson, 2006). What is unknown is if Park’s statements hold true for a mixture distribution. The main objective of this study is to examine if pair-matching improves the power when the underlying distribution is a mixture of two normal distributions.
CHAPTER 2

Background

A randomization schema is a very important tactic for the randomized controlled trial. The three major advantages achieved from a proper randomization are listed as follows (Shen and Lu, 2006; Schulz and Grimes, 2002):

- Eliminates both selection and allocation bias
- Tends to make the treatment and control group comparable by balancing important confounding variables
- Provides a footing for valid statistical tests, which is an assumption satisfied for statistical test of the equality of treatments.

Sometimes, the simple randomization does not balance confounding variables well between treatment and control groups. In order to better balance the confounders, either a blocking or matching method is used and embedded in the randomization procedure, or the confounding variable is included in the statistical analysis as a covariable—such as Analysis of Covariance (ANCOVA) test.

But does a matching method improve the test’s precision and power? If it does, which matching method performs better in which scenario? Let us see how the literature records the knowledge evolution about the matching methodology.
2.1 Literature Review

Cochran studied the efficiency of pair-matching versus ANCOVA model in the linear regression setting and pointed out pair-matching has the advantage of simple computations, especially if the pair-matching is carried out on multiple independent variables (Cochran, 1953). Pair-matching will produce higher precision than an ANCOVA test especially when the relationship is nonlinear. Cochran concluded that “pairing is more efficient than covariance.”

However, the simulation study results from Billewicz are in conflict with Cochran’s statement (Billewicz, 1965). Billewicz studied the efficiency of matched methods by simulation study and concluded that the “matching method become less efficient in the quantitative response compared to ANCOVA test on simple random sampling no matter matching variables are either quantitative or qualitative. But, matching is a valuable tool for making groups comparable and improving comparison precision with all-or-none response (the factor of interested being two-level factor).”

Youkeles noted that the sign test has low power for small sample sizes below 20, but that the drawback of matched-pair disappears when the sample size reaches 30 or above. It seems the sample size matters to the usefulness of a matching technique (Youkeles, 1963).

Imai studied the relative efficiency of the matched-pair design against the completely randomized design using a nonparametric approach without the normal distribution assumption (Imai, 2008). He stated that “the variance estimator is unbiased if and only if the variance of the within-pair sample average treatment effect is zero. The relative
efficiency of the matched-pair design depends on whether matching induces positive or negative correlations regarding potential outcomes within each pair.” The positive correlation bears more efficient estimates comparing the matched-pair design to the completely randomized design.

Park and Johnson (2006) studied how pair-wise matching on baseline measures affects the treatment effect variance in six cases of design and analysis (listed in the Table 1) for both a RCT and a cluster randomized trial (CRT). They showed that the case 2 has smallest variance among the three traditional without matching and the case 3 would have a smaller variance only when the correlation coefficient (r) between baseline and post-baseline measures is larger than 0.5. They further focused on variance comparison, examining matching versus without matching and concluded that the pair-matching reduces variance comparing case 4 versus case 1 and case 6 versus case 3, but it is not true when comparing case 5 versus case 2. Case 2 has a smaller variance than case 5. In addition, the paper showed that case 2 has the smallest variance among all six cases.

Johnson and Park (2004) also investigated the centroid method with a stopping rule, deciding on how many clusters, for matching clustered baseline outcome measures. They found that this method is able to decrease variation for the outcome variables between clusters, which resulted in greater accuracy of estimates, more narrow confidence intervals, and increased statistical power to detect differences between study groups (Park and Johnson, 2004). Park and Johnson (2005) focused on the variance effect of stratification on baseline measurements. They concluded blocking clusters in pair-wise matching based on the ordered baseline outcome measures decreases the variance of treatment effects. In
Park’s dissertation (2006), she derived the variance formula of a treatment effect comparison for each of the six cases design (see Table 1) for both RCT and CRT studies, when the data follow a normal distribution. Park (2006) has also performed simulated power analyses for six cases in CRT design and found the ANCOVA test is the best method to control variance and holds the best power to detect the treatment effect. The pair-wise matching on baseline measures did not increase power on the ANCOVA test.

Table 1 Six Cases -- Design and Analysis

<table>
<thead>
<tr>
<th>Without matching</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Post-Baseline only</td>
</tr>
<tr>
<td>2. Post-Baseline analysis adjusting for baseline as a covariate</td>
</tr>
<tr>
<td>3. Baseline and Post-Baseline differences</td>
</tr>
<tr>
<td>With matching</td>
</tr>
<tr>
<td>4. Post-Baseline only</td>
</tr>
<tr>
<td>5. Post-Baseline analysis adjusting for baseline as covariate</td>
</tr>
<tr>
<td>6. Baseline and Post-Baseline differences</td>
</tr>
</tbody>
</table>

2.2 Summary

Most simulation studies have been performed on a single normal distribution (Billewicz, 1965; Park, 2006). What is unknown is if Park’s dissertation statements hold true for a mixture distribution which tends to have two distinct clusters as the distributions’ means diverge. It is hard to derive the analytical equations to answer our questions. Instead, a simulation study was designed to explore the six case scenarios and provide empirical information of the parameters we are interested in. A simulation study of power
analysis is conducted for the two mixed normal distribution under the matched-pair design where the baseline measures are paired based on their ordered values and the randomization is carried out within each matched pair.

The objective of our study is to investigate if pair-wise matching on the ordered baseline measurements will increase statistical power compared to a randomized design without matching. Do different test procedures provide different power in pair-wise matching design for the two mixed normally distributed data? Which test is the best in different scenarios?

We will consider the different scenarios in our study in combination of the sample size, correlation between baseline and post-baseline measures, proportion of each normal distribution in two mixed normal distributions, statistical test methods, equal variance versus unequal variance in two distributions, and the effect size.
CHAPTER 3

Methods

The methods are described to simulate datasets from a mixture of two normal distributions. The models and the statistical methods to be applied for statistical tests and evaluation are illustrated. It will include how to set the parameters such as the proportion of each normal distribution, the ratio of variance and the mean difference between two normal distributions. The randomization schema for both matching and without-matching scenarios will be illustrated also. Finally, we will justify how we evaluate the methods in the simulation.

3.1 The Model

Often in the RCT trials, each subject has at least two measurements – baseline before intervention and post-baseline after intervention. The data of each subject are simulated with baseline and post-baseline measurements from the mixture of normal distributions, then those two measurements for the same subject are correlated and treated as repeated measures.

The two randomization methods are applied to our same simulated datasets—the common simple random sampling (SRS) without matching and the SRS within each matched pair. The matching schema is to sort the subjects by their ascending baseline
measures from the mixture distribution, pair them based on sorting order prior to randomization, and then within each matched-pair, randomly assigned subjects to either treatment or control group. The corresponding post-baseline values become induced ordered statistics, which means the order of the post-baseline data is related to the baseline order statistics, but the order is not necessarily a true sorted order unless the pre-post data are perfectly and positively correlated.

The main statistical test of interest is to test if there is a statistically significant difference for the average means between treatment and control group after intervention. So, the null hypothesis is that the means of treatment and control are equal; the alternative hypothesis is that the means of treatment and control group are significantly different. The ANOVA and ANCOVA methods will be applied to test our hypotheses to the randomized data sample with and without matching. The test sequence is described below.

Without Matching (SRS)

Case 1 -- Analysis of variance (ANOVA) on the post-baseline measures only. It is a typical t-test for two independent samples. The baseline measure values are overlooked here. One major drawback is that the variance between subjects is not minimized.

Case 2 -- Analysis of variance (ANCOVA) on the post-baseline measures including

baseline measures as a covariate. The test improves the power due to a reduction in error variance.

Case 3 -- ANOVA on the difference measure values between post-baseline and baseline.
The test is also referred to within-subjects t-test and it reduces the random error variance within subjects. Cochran (1957) showed that the reduction in error variance using ANCOVA compared to ANOVA is determined by the quantity of the correlation between baseline and post-baseline measurements.

With Matching (SRS under the matched-pair design)

Case 4 -- Same as the Case 1 plus matched-pair randomization and adjusting pairing block in ANOVA analysis.

Case 5 -- Same as the Case 2 plus matched-pair randomization and adjusting pairing block in ANCOVA analysis.

Case 6 -- Same as the Case 3 plus the matched-pair randomization and adjusting pairing block in ANOVA analysis.

The first three test cases are the traditional designs and tests, which do not involve pair-matching before randomization. The last three cases are same as the first three cases plus the pair-matching during randomization and adjusting pairing block in ANOVA and ANCOVA analyses.

3.2 Simulation

3.2.1 Software to Use

The SAS Software is chosen to perform simulation due to the following reasons (Fan et al, 2003)
• Ease and flexible in data generation and transformation
• Portability of SAS program
• A reliable and a variety of random number generator functions
• Built-in and calibrated statistical procedures
• Report writing and graphs capabilities
• Efficiency enhanced by Macro facility

3.2.2 Methods of Simulating the Datasets

The 76 unit datasets were simulated completely independent for the different scenarios. The following six major parameters were taken into account whenever each unit set of data was simulated.

• The proportion is defined as the first normal distributed data weighted in the two mixed normal distributions, the range of proportion was between 0.1 and 0.9 by increasing 0.2 and was determined randomly from a binomial distribution. The two mixed normal distributions become one normal distribution when proportion is equal to 0 or equal to 1. The proportion was denoted as P in the simulated data set.

• The variance ratio between the first normal and the second normal distribution was denoted as TAU in the dataset. In addition, we always set the variance for the second normal distribution to 1. The range of TAU values were 1, 2, or 3.

• The means difference between the first and the second normal distribution was denoted as DELTA in the dataset. Also, we set the overall mean from the two
mixed normal distribution to zero \((p^*\mu_1 + (1 - p)^*\mu_2 = 0)\). The range of DELTA values were 0, 1, 3, or 6. Then, \(\mu_1 - \mu_2 = \text{DELTA} \); \(\mu_1 = (1 - p)^*\text{DELTA} \); and \(\mu_2 = -(p^*\text{DELTA})\).

- The correlation between baseline and post-baseline measures was denoted as RHO, the baseline and post-baseline measure values were generated using bivariate normal formula. The RHO values were set to low (0.2) and high (0.8) respectively.

- The sample size was evenly allocated to treatment and control groups. The total sample size was 60, 40, or 20, and then treatment and control groups will have sample size 30, 20, and 10 respectively.

- For the number of iterations for each scenario, the estimating event probability formula was used to calculate a sufficient number of iterates. Chose the number of iterations to be \(p \times (1 - p) \times \left(\frac{Z}{E}\right)^2\), \(Z\) was a standard normal variable and \(E\) was margin of error; chose 0.5 as feasible value \(p\) and \(E\) as 0.01. Then 9604 was the value returned, we chose 10,000 iterations as our choice for each scenario.

Note: when we set DELTA=0 and TAU=1, the simulated dataset becomes one standard normal distribution with mean 0 and variance 1 rather than the two mixed normal distributions no matter the proportion value is.

### 3.2.3 Random Number Generator to Use

The SAS RANNOR and RANBIN random-number generator functions were used to generate normal distribution and binomial distribution, respectively. The seed values
were set to be positive, allowing us to replicate the stream of random numbers by using the same data step. Therefore, the RANNOR and RANBIN functions were able to reproduce the identical data set when keeping the same positive seed starting value, which made the simulation study to be reproducible (SAS Online Dictionary, 2011). The unit set of simulated dataset was defined by combination set of proportion, TAU, and DELTA values. Each unit set of dataset was generated by new data step with different starting seed values, which ensured the independence of unit dataset (Burton et al, 2006). Each unit set data was used to compare analysis power for our six test cases.

While each unit dataset is determined by combination of proportion, TAU, and DELTA values, then each scenario is the different combination of total sample size and correlation coefficient between baseline and post-baseline measures. We have six scenarios shown in Table 2a and the first scenario is displayed in Table 2b. Basically, the first three scenarios are the high correlation (RHO=0.8) with three different sample sizes (60, 40, or 20) and the last three scenarios are the low correlation (RHO=0.2) with three different sample sizes. The small or large sample size and low or high correlation combination sets will be evaluated separately. In the process, we have documented the sample size, correlation value, proportion, TAU, and DELTA values for each unit set in detail.

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Sample Size</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>60</td>
<td>0.8</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>0.8</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>0.8</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>0.2</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>0.2</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>0.2</td>
</tr>
</tbody>
</table>
Table 2b The Parameters of the First Scenario in Each Unit Dataset.

<table>
<thead>
<tr>
<th>Set</th>
<th>Group</th>
<th>P</th>
<th>TAU</th>
<th>DELTA</th>
<th>Total N</th>
<th>RHÖ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>60</td>
<td>0.8</td>
</tr>
<tr>
<td>2</td>
<td>10-14</td>
<td>0.1-0.9 by 0.2</td>
<td>1</td>
<td>1</td>
<td>60</td>
<td>0.8</td>
</tr>
<tr>
<td>3</td>
<td>15-19</td>
<td>0.1-0.9 by 0.2</td>
<td>1</td>
<td>3</td>
<td>60</td>
<td>0.8</td>
</tr>
<tr>
<td>4</td>
<td>20-24</td>
<td>0.1-0.9 by 0.2</td>
<td>1</td>
<td>6</td>
<td>60</td>
<td>0.8</td>
</tr>
<tr>
<td>5</td>
<td>25-29</td>
<td>0.1-0.9 by 0.2</td>
<td>2</td>
<td>0</td>
<td>60</td>
<td>0.8</td>
</tr>
<tr>
<td>6</td>
<td>30-34</td>
<td>0.1-0.9 by 0.2</td>
<td>2</td>
<td>1</td>
<td>60</td>
<td>0.8</td>
</tr>
<tr>
<td>7</td>
<td>35-39</td>
<td>0.1-0.9 by 0.2</td>
<td>2</td>
<td>3</td>
<td>60</td>
<td>0.8</td>
</tr>
<tr>
<td>8</td>
<td>40-44</td>
<td>0.1-0.9 by 0.2</td>
<td>2</td>
<td>6</td>
<td>60</td>
<td>0.8</td>
</tr>
<tr>
<td>9</td>
<td>45-49</td>
<td>0.1-0.9 by 0.2</td>
<td>3</td>
<td>0</td>
<td>60</td>
<td>0.8</td>
</tr>
<tr>
<td>10</td>
<td>50-54</td>
<td>0.1-0.9 by 0.2</td>
<td>3</td>
<td>1</td>
<td>60</td>
<td>0.8</td>
</tr>
<tr>
<td>11</td>
<td>55-59</td>
<td>0.1-0.9 by 0.2</td>
<td>3</td>
<td>3</td>
<td>60</td>
<td>0.8</td>
</tr>
<tr>
<td>12</td>
<td>60-64</td>
<td>0.1-0.9 by 0.2</td>
<td>3</td>
<td>6</td>
<td>60</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Note: the dataset 1 is a standard normal distribution that is a special case here. Total N=60, then n=30 for treatment and control respectively.

3.2.4 Treatment Effects Calculation

The effect size list was created from 0.1 to 0.5 by increasing 0.1, and then the treatment effect list was calculated by adding effect size for the treatment group only. For the difference effect list between post-baseline and baseline, we used the newly calculated treatment effect list and subtracted baseline value for the treatment group only. This calculation step was applied to both SRS without matching and SRS with pair-matching data sample.
3.2.5 Power Analysis

The estimates and the relevant p-values obtained from each replication were saved in the permanent datasets using SAS/Output Delivery System (ODS), which permits estimation of power and comparisons.

The power of the test is defined as the probability of rejecting the null hypothesis when it is false. The power is estimated by the proportion of times the calculated p-value falls below an alpha level of 0.05; that is, the power is estimated as the proportion of the p-values that are significant.

The estimated powers from each of six test cases under each set of parameters were saved in a permanent SAS data set. Using this information, power curves were plotted to compare power among the test cases for different effect sizes or different proportions.

This simulation study has been accomplished efficiently using SAS/MACRO processing. The set of SAS MACRO programs were developed and connected through simulating the datasets, performing statistical tests, estimating the power from each test, and to generating power plots and other graphs.

3.3 Evaluation

For the simulated mixture of normal distributions, the Procedure CORR and the Procedure UNIVARIATE were applied to evaluate the parameters such as mean and standard deviation of each normal distribution, overall mean of the mixed distribution, and the correlation coefficient between baseline and post-baseline measures. The procedure BOXPLOT was used to generate a set of box-plots for comparing distributions of both
baseline and post-baseline measures across levels of a parameter grouping. This process makes sure data distribution resembles what is supposed to be simulated (Burton et al, 2006).

When performing randomization of pair-wise matching on ordered baseline measures, a new variable ‘rank’ was created to record the ascending order sequence of baseline measures after sorting them from the smallest to the largest. In addition, two other new variables were created to trace the pairing sequence numbers (pair_num, maximum value =N/2=n in each group) and the order within each pair (pair_seq=1, 2). An example is shown in Table 3.

Then, the PROC TRANSPOSE and data steps were used to check if the baseline measures with pair_seq value as 1 are always smaller than the baseline measures with pair_seq value as 2 within each pair. If it is true, it implies that the pair-wise matching on ordered baseline measures was done correctly.

Table 3 The Data Structure of Baseline Pair-Matching

<table>
<thead>
<tr>
<th>trt</th>
<th>p</th>
<th>Tau</th>
<th>Delta</th>
<th>iterate</th>
<th>id</th>
<th>baseline</th>
<th>rank</th>
<th>pair_num</th>
<th>pair_seq</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>22</td>
<td>-2.23</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>57</td>
<td>-2.108</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>17</td>
<td>-1.734</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>-1.69</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>-1.69</td>
<td>5</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>15</td>
<td>-1.463</td>
<td>6</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>-1.396</td>
<td>7</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>50</td>
<td>-1.175</td>
<td>8</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>-0.954</td>
<td>9</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>48</td>
<td>-0.857</td>
<td>10</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>
An R program was developed by Dr. Johnson to evaluate the power analysis by performing the same simulation as SAS programs do.
CHAPTER 4

Results

4.1 Evaluation of the Simulated Data

The results of evaluating the normality of the simulated datasets (for special case when group=1, TAU=1 and DELTA=0) are shown in Figures 1 through 3. From these histograms, normal probability plots, and box-plots, it can be seen that the simulated data follow the normal distribution. Based on the box-plots, it has been observed that the distributions across the group levels are comparable.

Figure 1  Histograms of simulated baseline and post-baseline distribution (group=1)

Figure 2  Normal probability plots of simulated baseline and post-baseline measures
The results of evaluating the simulated datasets with the two mixed normally distribution (TAU=1 and DELTA=6) are exhibited in Figure 4 (group=20, p=0.1) and Figure 5 (group=22, p=0.5). The Figure 4 shows a bimodal normal distribution with equal variance and unequal proportion, and the Figure 5 displays a bimodal normal distribution with equal variance and equal proportion. The results of evaluating correlation coefficient between baseline and post-baseline for group=1, 4, and 20 (see Table 2b) are shown in Table 4. The correlation coefficients of the three groups are simulated to the targeted values.
Figure 4  Histograms of simulated baseline and post-baseline distribution (group=20)

Figure 5  Histograms of simulated baseline and post-baseline distribution (group=22)

Table 4  Correlation Coefficients between Baseline and Post-Baseline Measures

<table>
<thead>
<tr>
<th></th>
<th>Group=1</th>
<th>Group=4</th>
<th>Group=20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline</td>
<td>Post</td>
<td>Baseline</td>
</tr>
<tr>
<td>Baseline</td>
<td>1</td>
<td>0.80</td>
<td>1</td>
</tr>
<tr>
<td>Post</td>
<td>0.80</td>
<td>1</td>
<td>0.20</td>
</tr>
</tbody>
</table>
4.2 Evaluation of Pair-Matching on Ordered Baseline Measures

Data steps and PROC TRANSPOSE were used to check if the second baseline measure is always larger than the first baseline measure within each pair, and to check if the data values of the pair_rank1 and pair_rank2 were sorted in ascending order. The results are shown in Table 5 and their difference (pair_rank2 – pair_rank1) is always greater than zero, which imply the baseline pair-matching on ordered baseline measures were done correctly.

Table 5  Baseline Measures Pair-matching Evaluation

<table>
<thead>
<tr>
<th>p</th>
<th>Tau</th>
<th>Delta</th>
<th>Iterate</th>
<th>pair_num</th>
<th>pair_rank1</th>
<th>pair_rank2</th>
<th>diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>-2.3019</td>
<td>-2.10813</td>
<td>0.12206</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>-1.73432</td>
<td>-1.69024</td>
<td>0.04408</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>-1.68983</td>
<td>-1.46304</td>
<td>0.22679</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>-1.39639</td>
<td>-1.17483</td>
<td>0.22156</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>-0.95449</td>
<td>-0.85681</td>
<td>0.09767</td>
</tr>
</tbody>
</table>

4.3 Evaluation of Power Results

An R program was developed by Johnson to perform the simulation with the same combination of parameters as used in the SAS programs. The 5% acceptable marginal error ranges were computed by the R program. The power results generated by SAS and R are comparable and the differences are within acceptable marginal error ranges.

4.4 The Six Test Cases Results

4.4.1 Case 3 vs. case 6 (N=60, n=30 for each group)
The power curves of case 3 vs. case 6 from a combination of different TAU, DELTA, and Proportion values with RHO=0.8 are displayed in Figures 6 to 11. It can be seen that the power curves of case 6 are just above the ones of case 3 when DELTA values are equal to 0 or 6 regardless of TAU and Proportion values, and the power curves of the two cases are overlapped when DELTA is equal to 3 and Proportion is equal to 0.5 regardless of TAU values. The power pattern is the same for both one normal distribution and the mixture distribution.

The variance of a post-baseline and baseline difference is already controlled well when the correlation between baseline and post-baseline measures is greater than 0.5. So, the additional variance reduction from pair-matching is limited when RHO is equal to 0.8.

Figure 6a  Power curves of case 3 vs. case 6 (Tau=1, Delta=0, P=0.5)
Figure 6b  Power of case 3 vs. case 6 across Delta values (effect size=0)
Figure 7a  Power curves of case 3 vs. case 6 (Tau=1, Delta=3, P=0.5)
Figure 7b  Power of case 3 vs. case 6 across Delta values (effect size=0.3)

Figure 8a  Power curves of case 3 vs. case 6 (Tau=1, Delta=6, P=0.5)
Figure 8b  Power of case 3 vs. case 6 across Delta values (effect size=0.5)
Figure 9a  Power curves of case 3 vs. case 6 (Tau=3, Delta=0, P=0.1)
Figure 9b  Power of case 3 vs. case 6 across Delta values (effect size=0)

Figure 10a  Power curves of case 3 vs. case 6 (Tau=3, Delta=3, P=0.1)
Figure 10b  Power of case 3 vs. case 6 across Delta values (effect size=0.3)
The power curves of case 3 vs. case 6 from a combination of different TAU, DELTA, and Proportion values with RHO=0.2 are shown in Figures 12 to 17. The power curves of case 6 are evidently improved compared to the power curves of case 3 for all scenarios, and the improvement is bigger than the one observed when RHO=0.8. The pattern is the same between one normal and the mixture distribution. The magnitude of the improvement decreases with an increase of TAU values, and overall power declines with increase of TAU values. Another interesting finding is that the power improvement of case 6 vs. case 3 is lowest when DELTA value is equal to 3, comparing to the DELTA values 0 or 6.
Figure 12a  Power curves of case 3 vs. case 6 (Tau=1, Delta=0, P=0.5)
Figure 12b  Power of case 3 vs. case 6 across Delta values (effect size=0)

Figure 13a  Power curves of case 3 vs. case 6 (Tau=1, Delta=3, P=0.5)
Figure 13b  Power of case 3 vs. case 6 across Delta values (effect size=0.3)
Figure 14a  Power curves of case 3 vs. case 6 (Tau=1, Delta=6, P=0.5)
Figure 14b  Power of case 3 vs. case 6 across Delta values (effect size=0.5)

Figure 15a  Power curves of case 3 vs. case 6 (Tau=3, Delta=0, P=0.7)
Figure 15b  Power of case 3 vs. case 6 across Delta values (effect size=0)
Figure 16a  Power curves of case 3 vs. case 6 (Tau=3, Delta=3, P=0.7)
Figure 16b  Power of case 3 vs. case 6 across Delta values (effect size=0.3)

Figure 17a  Power curves of case 3 vs. case 6 (Tau=3, Delta=6, P=0.7)
Figure 17b  Power of case 3 vs. case 6 across Delta values (effect size=0.5)

4.4.2  Case 3 vs. case 6 (N=40, 20, n=20, 10 for each group respectively)

When RHO = 0.8, the power curves of case 3 and case 6 are almost identical in the smallest sample group (N=20, n=10 for each group), which is not related to different TAU, Delta, and Proportion values. Their power curves are shown in Figure 18.
When RHO = 0.2, the power curves of these two cases are very close to each other in the smallest sample group (N=20, n=10 for each group). This is different from the pattern observed in the largest sample group (N=60, n=30 for each group) and may result from the overall low power. Their power curves are shown in Figure 19.

The power curve pattern between case 3 and case 6 in middle sample size group (N=40, n=20 for each group) is same as the pattern in largest sample group. The
improvement scale of case 6 over case 3 decreases with overall power fall when the sample size becomes small.

4.4.3 Case 2 vs. case 5 (N=60, n=30 in each group)

The power curves of case 2 vs. case 5 from a combination of different TAU, DELTA, and Proportion values with RHO=0.8 are displayed in Figure 20 and Figure 21. The power curves of case 2 are slightly higher than the ones of case 5 when Proportion is equal to 0.5 and Delta values are equal to 0 or 3, which happens for all groups with different TAU values. The two power curves are overlapped when Delta value is equal to 6 or Delta values are equal to 0 or 3 and Proportion is not equal to 0.5.

Figure 20a  Power curves of case 2 vs. case 5 (Tau=1, Delta=0, P=0.5)  
Figure 20b  Power curves of case 2 vs. case 5 (Tau=2, Delta=3, P=0.9)
Figure 21a  Power curves of case 2 vs. case 5 (Tau=2, Delta=3, P=0.1)  
Figure 21b  Power curves of case 2 vs. case 5 (Tau=3, Delta=6, P=0.1)  

The power curves of case 2 vs. case 5 from a combination of different TAU, DELTA, and Proportion values with RHO=0.2 are shown in Figures 22 to 24. The two power curves of case 2 and case 5 are overlapped or very close to each other when Delta values are equal to 0 or 3 regardless of TAU values. The power curves of case 5 are clearly better than the ones of case 2 when Delta is equal to 6 for all TAU values. The degree of improvement decreases when the TAU value becomes bigger, which is related to overall power reduces with increase of TAU value. Since the overall power also declines with increase of Proportion value at the same TAU group, the two power curves of case 2 and case 5 are very close to each other when TAU=3 and Proportion=0.9. The overall power of this scenario is approximately 15 percent.
Figure 22a  Power curves of case 2 vs. case 5 (Tau=1, Delta=0, P=0.5)
Figure 22b  Power curves of case 2 vs. case 5 (Tau=1, Delta=3, P=0.5)

Figure 23a  Power curves of case 2 vs. case 5 (Tau=1, Delta=6, P=0.5)
Figure 23b  Power of case 2 vs. case 5 across Delta values (effect size=0.5)
Figure 24a  Power curves of case 2 vs. case 5 (Tau=3, Delta=6, P=0.1)
Figure 24b  Power of case 2 vs. case 5 across Delta values (effect size=0.5)

4.4.4 Case 2 vs. case 5 (N=40, 20, n=20, 10 for each group respectively)

When RHO=0.8, the power curves of case 2 perform slightly better than the ones of case 5 for most scenarios in the smallest sample group (N=20, n=10 for each group). The two power curves are very close to each other when TAU value is equal to 3 and Proportion is larger than 0.5. The example figures are shown in Figure 25 and Figure 26.

Figure 25a  Power curves of case 2 vs. case 5 (Tau=2, Delta=6, P=0.9)
Figure 25b  Power of case 2 vs. case 5 across Delta values (effect size=0.5)
When RHO=0.2, the power curves of case 2 and case 5 are always identical in the smallest sample group (N=20, n=10 for each group). The pattern observed in the largest sample group disappears here, which could be caused by overall low powers when sample size decreases.

The pattern of the power curves comparison between case 2 and case 5 in the middle sample size group (N=40, n=20 for each group) is same as the pattern of the largest sample size group for both scenarios of high (RHO=0.8) and low (RHO=0.2) correlation.

4.4.5 Case 1 vs. case 4 (N=60, n=30 for each group)

The power curves of case 1 vs. case 4 from a combination of different TAU, DELTA, and Proportion values with RHO=0.8 are displayed in Figure 27 and Figure 28. The powers of case 4 are drastically higher compared to the powers of case 1 for all the scenarios. The pattern stays the same for the different TAU and Delta values. There is no difference between one normal and the two mixed normal distribution. The power curves
of case 1 go down when Delta value increases and power curves of case 4 stay the same. Since both case 1 and case 4 use only post-baseline measures, the pair-matching on ordered baseline measures definitely help improve the statistical powers.

Figure 27a  Power curves of case 1 vs. case 4 (Tau=1, Delta=0, P=0.5)
Figure 27b  Power curves of case 1 vs. case 4 (Tau=1, Delta=3, P=0.1)

Figure 28a  Power curves of case 1 vs. case 4 (Tau=3, Delta=6, P=0.7)
Figure 28b  Power of case 1 vs. case 4 across Delta values (effect size=0.5)

The power curves of case 1 vs. case 4 from a combination of different TAU, DELTA, and Proportion values with RHO=0.2 are shown in Figure 29 to Figure 31. The power curves of case 1 and case 4 are exactly same when Delta value is equal to zero.
regardless of TAU and Proportion values. The case 4 has better power when Delta value becomes larger than 1, the larger gap the bigger Delta value and the pattern is same for different TAU and Proportion values. The magnitude of the improvement decreases with an increase of TAU values, and overall power declines with increase of TAU values.

Figure 29a  Power curves of case 1 vs. case 4 (Tau=1, Delta=0, P=0.5)
Figure 29b  Power curves of case 1 vs. case 4 (Tau=1, Delta=3, P=0.5)

Figure 30a  Power curves of case 1 vs. case 4 (Tau=2, Delta=0, P=0.7)
Figure 30b  Power curves of case 1 vs. case 4 (Tau=2, Delta=3, P=0.7)
Figure 31a  Power curves of case 1 vs. case 4 (Tau=2, Delta=6, P=0.7)
Figure 31b  Power of case 1 vs. case 4 across Delta values (effect size=0.5)

4.4.6  Case 1 vs. case 4 (N=40, 20, n=20, 10 for each group)

The power curves of case 1 vs. case 4 in middle sample size group (N=40, n=20 for each group) hold the same pattern compared to the power curves in the largest sample size group (N=60, n=30 for each group) for both high (RHO=0.8) and low (RHO=0.2) correlation scenarios.

The power curves comparison between case 1 and case 4 in the smallest sample size group is similar to the ones of the largest sample size group when RHO is equal to 0.8. When RHO is equal to 0.2, the power line of case 1 and case 4 is very close to each other, it may be caused by overall very low power.
CHAPTER 5

Conclusion

When correlation is high (RHO=0.8), the simulation study has found that the additional variance reduction from pair-matching in pre-post design (case 6 vs. case 3) is limited. The power curves of case 5 have no improvement over case 2 for the ANCOVA test. The powers of case 4 are drastically higher compared to the powers of case 1 for all the scenarios. The pair-matching on ordered baseline measures definitely help improve the statistical powers for the test using only post-baseline data.

When correlation is low (RHO=0.2), there is significant power increase from pair-matching in pre-post design. The power curves of case 5 are clearly better than the ones of case 2 when Delta is equal to 6 regardless of TAU values. The case 4 has better power when Delta value becomes larger than 1.

It is shown that the baseline pair-matching randomization improves the power when the two means of a mixture normal distribution are widely spread and the sample size is greater than 20 in each group. The pattern becomes clear when the correlation between baseline and post-baseline is low. The pattern becomes vague when sample size drop to 10 in each group due to overall very low powers (less than 20 percent).

In conclusion, the ANCOVA remains the best test for the mixture distributions since the additional variance reduction from pair-matching in pre-post design is limited
when correlation is high (RHO=0.8). When correlation is low (RHO=0.2), there is significant power increase in pre-post design by pair-matching.
A combination of different parameters used to define the two mixed normal distributions in the simulation study cause a fluctuating overall variance. In the future, we could design and run a simulation study to confirm the findings by setting overall variance constant of the two mixed normal distributions. The mixture distribution can be extended to more complex normal distribution rather than the two mixed normal distribution, or other mixed distribution instead of the normal distribution. In addition, the study focused on baseline measures pair-matching without considering other baseline important confounding variables such as gender and age groups.

I have learned a lot through the study. It was a really different and invaluable experience to do literature review, design and run a simulation study in comparison to taking a class. I have learned the basic rules and principles of the simulation studies in design and execution of the study as well as evaluation of simulation results. In addition, my SAS program skills were also further enhanced through the study.

Through the M.S. Biostatistics program training, I have understood the basic statistical theories and methodologies including data management, study design and statistical inference. More importantly, I have learned from Dr. Johnson to “smile every day”, which I will cherish in the rest of my life.
List of References


List of References


APPENDIX A

The Figures from Case 3 vs. Case 6 (N=60, RHO=0.8)

When Tau=1, P=0.5 and delta=0   

When Tau=1, P=0.5 and delta=3   

effect size=0  

effect size=0.3
When Tau=1, P=0.5 and delta=6  

When Tau=1, P=0.1 and delta=3 

effect size=0.3  

When Tau=1, P=0.1 and delta=6  

effect size=0.5
When $\tau=1$, $P=0.7$ and $\delta=3$

effect size=0.3

When $\tau=1$, $P=0.7$ and $\delta=6$

effect size=0.5

When $\tau=2$, $P=0.5$ and $\delta=0$

effect size=0
When $\tau = 2$, $P = 0.5$ and $\delta = 3$

Effect size $= 0.3$

When $\tau = 2$, $P = 0.5$ and $\delta = 6$

Effect size $= 0.5$

When $\tau = 2$, $P = 0.1$ and $\delta = 0$

Effect size $= 0$
When Tau=2, P=0.1 and delta=3 

effect size=0.3

When Tau=2, P=0.1 and delta=6 

effect size=0.5

When Tau=2, P=0.7 and delta=0 

effect size=0
When $\tau=2$, $P=0.7$ and $\delta=3$  

effect size $= 0.3$

When $\tau=2$, $P=0.7$ and $\delta=6$  

effect size $= 0.5$

When $\tau=2$, $P=0.9$ and $\delta=0$  

effect size $= 0$
When \(\text{Tau}=2\), \(P=0.9\) and \(\text{delta}=3\)  
effect size=0.3

When \(\text{Tau}=2\), \(P=0.9\) and \(\text{delta}=6\)  
effect size=0.5

When \(\text{Tau}=3\), \(P=0.1\) and \(\text{delta}=0\)  
effect size=0
When $\tau=3$, $P=0.1$ and $\delta=3$, effect size $= 0.3$

When $\tau=3$, $P=0.1$ and $\delta=6$, effect size $= 0.5$

When $\tau=3$, $P=0.5$ and $\delta=0$, effect size $= 0$
When Tau=3, P=0.5 and delta=3  
effect size=0.3

When Tau=3, P=0.5 and delta=6  
effect size=0.5

When Tau=3, P=0.7 and delta=0  
effect size=0
When $\tau=3$, $P=0.7$ and $\delta=3$, effect size $=0.3$

When $\tau=3$, $P=0.7$ and $\delta=6$, effect size $=0.5$

When $\tau=3$, $P=0.9$ and $\delta=0$, effect size $=0$
When $\text{Tau}=3$, $P=0.9$ and $\text{delta}=3$  

Effect size $= 0.3$

When $\text{Tau}=3$, $P=0.9$ and $\text{delta}=6$  

Effect size $= 0.5$
APPENDIX B

The Figures from Case 3 vs. Case 6 (N=60, RHO=0.2)

When $\tau=1$, $P=0.5$ and $\delta=0$

effect size $= 0$

When $\tau=1$, $P=0.5$ and $\delta=3$

effect size $= 0.3$
When \( \tau = 1 \), \( \rho = 0.5 \) and \( \delta = 6 \) 

Effect size = 0.5

When \( \tau = 1 \), \( \rho = 0.1 \) and \( \delta = 3 \) 

Effect size = 0.3

When \( \tau = 1 \), \( \rho = 0.1 \) and \( \delta = 6 \) 

Effect size = 0.5
When \( \text{Tau}=1, \ P=0.7 \) and \( \delta=3 \)  

\( \text{effect size}=0.3 \)

---

When \( \text{Tau}=1, \ P=0.7 \) and \( \delta=6 \)  

\( \text{effect size}=0.5 \)

---

When \( \text{Tau}=2, \ P=0.1 \) and \( \delta=0 \)  

\( \text{effect size}=0 \)
When \( \tau = 2 \), \( P = 0.1 \) and \( \delta = 3 \)  
effect size = 0.3

When \( \tau = 2 \), \( P = 0.1 \) and \( \delta = 6 \)  
effect size = 0.5

When \( \tau = 2 \), \( P = 0.5 \) and \( \delta = 0 \)  
effect size = 0
When $\tau=2$, $P=0.5$ and $\delta=3$  

effect size = 0.3

When $\tau=2$, $P=0.5$ and $\delta=6$  

effect size = 0.5

When $\tau=2$, $P=0.7$ and $\delta=0$  

effect size = 0
When $\tau=2$, $P=0.7$ and $\delta=3$  
\[ \text{effect size}=0.3 \]

When $\tau=2$, $P=0.7$ and $\delta=6$  
\[ \text{effect size}=0.5 \]

When $\tau=2$, $P=0.9$ and $\delta=0$  
\[ \text{effect size}=0 \]
When Tau=2, P=0.9 and delta=3  
effect size=0.3

When Tau=2, P=0.9 and delta=6  
effect size=0.5

When Tau=3, P=0.1 and delta=0  
effect size=0
When $\tau=3$, $p=0.1$ and $\delta=3$  
effect size=0.3

When $\tau=3$, $p=0.1$ and $\delta=6$  
effect size=0.5

When $\tau=3$, $p=0.5$ and $\delta=0$  
effect size=0
When $\tau=3$, $P=0.5$ and $\delta=3$  
\[\text{effect size}=0.3\]

When $\tau=3$, $P=0.5$ and $\delta=6$  
\[\text{effect size}=0.5\]

When $\tau=3$, $P=0.7$ and $\delta=0$  
\[\text{effect size}=0\]
When $\tau=3$, $P=0.7$ and $\delta=3$

effect size=0.3

When $\tau=3$, $P=0.7$ and $\delta=6$

effect size=0.5

When $\tau=3$, $P=0.9$ and $\delta=0$

effect size=0
When Tau=3, P=0.9 and delta=3  

effect size=0.3

When Tau=3, P=0.9 and delta=6  

effect size=0.5
APPENDIX C

The Figures from Case2 vs. Case 5 (N=60, RHO=0.8)

When $\tau=1$, $P=0.5$ and $\delta=0$

$\text{effect size}=0$

When $\tau=1$, $P=0.5$ and $\delta=3$

$\text{effect size}=0.3$
When \( \text{Tau}=1 \), \( P=0.5 \) and \( \Delta=6 \)  

\text{effect size}=0.5

When \( \text{Tau}=1 \), \( P=0.1 \) and \( \Delta=3 \)  

\text{effect size}=0.3

When \( \text{Tau}=1 \), \( P=0.1 \) and \( \Delta=6 \)  

\text{effect size}=0.5
When Tau=1, P=0.7 and delta=3  
\[ \text{effect size} = 0.3 \]

When Tau=1, P=0.7 and delta=6  
\[ \text{effect size} = 0.5 \]

When Tau=2, P=0.1 and delta=0  
\[ \text{effect size} = 0 \]
When $\tau=2$, $P=0.1$ and $\delta=3$  

effect size=0.3

When $\tau=2$, $P=0.1$ and $\delta=6$  

effect size=0.5

When $\tau=2$, $P=0.5$ and $\delta=0$  

effect size=0
When Tau=2, P=0.5 and delta=3  

effect size=0.3

When Tau=2, P=0.5 and delta=6  

effect size=0.5

When Tau=2, P=0.7 and delta=0  

effect size=0
When $\tau=2$, $P=0.7$ and $\delta=3$  
\[\text{effect size}=0.3\]

When $\tau=2$, $P=0.7$ and $\delta=6$  
\[\text{effect size}=0.5\]

When $\tau=2$, $P=0.9$ and $\delta=0$  
\[\text{effect size}=0\]
When $\text{Tau}=2$, $P=0.9$ and $\text{delta}=3$  

effect size = 0.3

When $\text{Tau}=2$, $P=0.9$ and $\text{delta}=6$  

effect size = 0.5

When $\text{Tau}=3$, $P=0.1$ and $\text{delta}=0$  

effect size = 0
When \( \text{Tau}=3, \ P=0.1 \) and \( \delta=3 \)  
effect size=0.3

When \( \text{Tau}=3, \ P=0.1 \) and \( \delta=6 \)  
effect size=0.5

When \( \text{Tau}=3, \ P=0.5 \) and \( \delta=0 \)  
effect size=0
When $\tau=3$, $P=0.5$ and $\delta=3$ 

effect size $= 0.3$

When $\tau=3$, $P=0.5$ and $\delta=6$ 

effect size $= 0.5$

When $\tau=3$, $P=0.7$ and $\delta=0$ 

effect size $= 0$
When \( \tau = 3 \), \( P = 0.7 \) and \( \delta = 3 \)  
\[ \text{effect size} = 0.3 \]

When \( \tau = 3 \), \( P = 0.7 \) and \( \delta = 6 \)  
\[ \text{effect size} = 0.5 \]

When \( \tau = 3 \), \( P = 0.9 \) and \( \delta = 0 \)  
\[ \text{effect size} = 0 \]
When \( \tau=3 \), \( P=0.9 \) and \( \delta=3 \)

\[ \text{effect size}=0.3 \]

When \( \tau=3 \), \( P=0.9 \) and \( \delta=6 \)

\[ \text{effect size}=0.5 \]
APPENDIX D

The Figures from Case2 vs. Case 5 (N=60, RHO=0.2)

When Tau=1, P=0.5 and delta=0  

effect size=0

When Tau=1, P=0.5 and delta=3  

effect size=0.3
When \( \tau = 1 \), \( P = 0.5 \) and \( \delta = 6 \)  
\[ \text{effect size} = 0.5 \]

When \( \tau = 1 \), \( P = 0.1 \) and \( \delta = 3 \)  
\[ \text{effect size} = 0.3 \]

When \( \tau = 1 \), \( P = 0.1 \) and \( \delta = 6 \)  
\[ \text{effect size} = 0.5 \]
When Tau=1, P=0.7 and delta=3 
effect size=0.3

When Tau=1, P=0.7 and delta=6 
effect size=0.5

When Tau=2, P=0.1 and delta=0 
effect size=0
When \( \tau=2, P=0.1 \) and \( \delta=3 \)  
effect size=0.3

When \( \tau=2, P=0.1 \) and \( \delta=6 \)  
effect size=0.5

When \( \tau=2, P=0.5 \) and \( \delta=0 \)  
effect size=0
When $\tau=2$, $P=0.5$ and $\delta=3$, 
\[
effect\ size = 0.3
\]

When $\tau=2$, $P=0.5$ and $\delta=6$, 
\[
effect\ size = 0.5
\]

When $\tau=2$, $P=0.7$ and $\delta=0$, 
\[
effect\ size = 0
\]
When $\text{Tau}=2$, $P=0.7$ and $\Delta=3$  
\text{effect size}=0.3

When $\text{Tau}=2$, $P=0.7$ and $\Delta=6$  
\text{effect size}=0.5

When $\text{Tau}=2$, $P=0.9$ and $\Delta=0$  
\text{effect size}=0
When \( \text{Tau}=2 \), \( P=0.9 \) and \( \delta=3 \)  

\( \text{effect size}=0.3 \)

When \( \text{Tau}=2 \), \( P=0.9 \) and \( \delta=6 \)  

\( \text{effect size}=0.5 \)

When \( \text{Tau}=3 \), \( P=0.1 \) and \( \delta=0 \)  

\( \text{effect size}=0 \)
When $\tau=3$, $P=0.1$ and $\delta=3$  

effect size = 0.3

When $\tau=3$, $P=0.1$ and $\delta=6$  

effect size = 0.5

When $\tau=3$, $P=0.5$ and $\delta=0$  

effect size = 0
When $\text{Tau}=3$, $P=0.5$ and $\text{delta}=3$  
\text{effect size}=0.3

When $\text{Tau}=3$, $P=0.5$ and $\text{delta}=6$  
\text{effect size}=0.5

When $\text{Tau}=3$, $P=0.7$ and $\text{delta}=0$  
\text{effect size}=0
When $\text{Tau}=3$, $P=0.7$ and $\text{delta}=3$  
\text{effect size}=0.3

When $\text{Tau}=3$, $P=0.7$ and $\text{delta}=6$  
\text{effect size}=0.5

When $\text{Tau}=3$, $P=0.9$ and $\text{delta}=0$  
\text{effect size}=0
When \( \tau = 3 \), \( P = 0.9 \) and \( \delta = 3 \)  
effect size = 0.3

When \( \tau = 3 \), \( P = 0.9 \) and \( \delta = 6 \)  
effect size = 0.5
APPENDIX E

The Figures from Case 1 vs. Case 4 (N=60, RHO=0.8)

When Tau=1, P=0.5 and delta=0  
\text{effect size}=0

When Tau=1, P=0.5 and delta=3  
\text{effect size}=0.3
When $\tau = 1$, $P=0.5$ and $\delta=6$

effect size=$0.5$

When $\tau = 1$, $P=0.1$ and $\delta=3$

effect size=$0.3$

When $\tau = 1$, $P=0.1$ and $\delta=6$

effect size=$0.5$
When Tau=1, P=0.7 and delta=3  
\[ \text{effect size}=0.3 \]

When Tau=1, P=0.7 and delta=6  
\[ \text{effect size}=0.5 \]

When Tau=2, P=0.1 and delta=0  
\[ \text{effect size}=0 \]
When \( \tau = 2, \ P = 0.1 \) and \( \delta = 3 \)  

effect size = 0.3

When \( \tau = 2, \ P = 0.1 \) and \( \delta = 6 \)  

effect size = 0.5

When \( \tau = 2, \ P = 0.5 \) and \( \delta = 0 \)  

effect size = 0
When $\tau=2$, $P=0.5$ and $\delta=3$, effect size=0.3

When $\tau=2$, $P=0.5$ and $\delta=6$, effect size=0.5

When $\tau=2$, $P=0.7$ and $\delta=0$, effect size=0
When Tau=2, P=0.7 and delta=3
effect size=0.3

When Tau=2, P=0.7 and delta=6
effect size=0.5

When Tau=2, P=0.9 and delta=0
effect size=0
When $\tau=2$, $P=0.9$ and $\delta=3$  

effect size $=0.3$

When $\tau=2$, $P=0.9$ and $\delta=6$  

effect size $=0.5$

When $\tau=3$, $P=0.1$ and $\delta=0$  

effect size $=0$
When \( \tau = 3 \), \( p = 0.1 \) and \( \delta = 3 \)  
\text{effect size}=0.3

When \( \tau = 3 \), \( p = 0.1 \) and \( \delta = 6 \)  
\text{effect size}=0.5

When \( \tau = 3 \), \( p = 0.5 \) and \( \delta = 0 \)  
\text{effect size}=0
When \( \text{Tau}=3 \), \( P=0.5 \) and \( \text{delta}=3 \)  

\[
\text{effect size} = 0.3
\]

When \( \text{Tau}=3 \), \( P=0.5 \) and \( \text{delta}=6 \)  

\[
\text{effect size} = 0.5
\]

When \( \text{Tau}=3 \), \( P=0.7 \) and \( \text{delta}=0 \)  

\[
\text{effect size} = 0
\]
When $\tau=3$, $P=0.7$ and $\delta=3$  
  
  effect size=$0.3$

When $\tau=3$, $P=0.7$ and $\delta=6$  
  
  effect size=$0.5$

When $\tau=3$, $P=0.9$ and $\delta=0$  
  
  effect size=$0$
When $\tau=3$, $P=0.9$ and $\delta=3$  
\text{effect size}=0.3

When $\tau=3$, $P=0.9$ and $\delta=6$  
\text{effect size}=0.5
APPENDIX F

The Figures from Case 1 vs. Case 4 (N=60, RHO=0.2)

When Tau=1, P=0.5 and delta=0  
effect size=0

When Tau=1, P=0.5 and delta=3  
effect size=0.3
When $\tau=1$, $P=0.5$ and $\Delta=6$ 

effect size=$0.5$

When $\tau=1$, $P=0.1$ and $\Delta=3$ 

effect size=$0.3$

When $\tau=1$, $P=0.1$ and $\Delta=6$ 

effect size=$0.5$
When $\ Tau=1$, $P=0.7$ and $\Delta=3$  
\text{effect size}=0.3

When $\ Tau=1$, $P=0.7$ and $\Delta=6$  
\text{effect size}=0.5

When $\ Tau=2$, $P=0.1$ and $\Delta=0$  
\text{effect size}=0
When $\tau=2$, $P=0.1$ and $\Delta=3$  
\[\text{effect size}=0.3\]

When $\tau=2$, $P=0.1$ and $\Delta=6$  
\[\text{effect size}=0.5\]

When $\tau=2$, $P=0.5$ and $\Delta=0$  
\[\text{effect size}=0\]
When Tau=2, P=0.5 and delta=3  
effect size=0.3

When Tau=2, P=0.5 and delta=6  
effect size=0.5

When Tau=2, P=0.7 and delta=0  
effect size=0
When $\tau = 2$, $P = 0.7$ and $\delta = 3$

Effect size = 0.3

When $\tau = 2$, $P = 0.7$ and $\delta = 6$

Effect size = 0.5

When $\tau = 2$, $P = 0.9$ and $\delta = 0$

Effect size = 0
When $\text{Tau}=2$, $P=0.9$ and $\text{delta}=3$  
\text{effect size}=0.3

When $\text{Tau}=2$, $P=0.9$ and $\text{delta}=6$  
\text{effect size}=0.5

When $\text{Tau}=3$, $P=0.1$ and $\text{delta}=0$  
\text{effect size}=0
When \( \tau=3 \), \( P=0.1 \) and \( \delta=3 \), 

\text{effect size}=0.3

When \( \tau=3 \), \( P=0.1 \) and \( \delta=6 \), 

\text{effect size}=0.5

When \( \tau=3 \), \( P=0.5 \) and \( \delta=0 \), 

\text{effect size}=0
When $\tau=3$, $P=0.5$ and $\delta=3$  \hspace{1cm} \text{effect size}=0.3

When $\tau=3$, $P=0.5$ and $\delta=6$ \hspace{1cm} \text{effect size}=0.5

When $\tau=3$, $P=0.7$ and $\delta=0$ \hspace{1cm} \text{effect size}=0
When \( \text{Tau}=3, P=0.7 \) and \( \delta=3 \)  
\text{effect size}=0.3

When \( \text{Tau}=3, P=0.7 \) and \( \delta=6 \)  
\text{effect size}=0.5

When \( \text{Tau}=3, P=0.9 \) and \( \delta=0 \)  
\text{effect size}=0
When \( \tau = 3 \), \( P = 0.9 \) and \( \delta = 3 \)  

\text{effect size} = 0.3

When \( \tau = 3 \), \( P = 0.9 \) and \( \delta = 6 \)  

\text{effect size} = 0.5
VITA

Yan Jin was born on May 18, 1963, Suzhou, China. She first graduated from Zhengjiang Medical College, China with a Professional Diplomat in Medicine, and then she received a M.S. in Medicine from Suzhou Medical College.

At the United States, she received a M.S. in Biomedical Informatics from the University of Medicine and Dentistry of New Jersey. She has been working a Data Analyst in clinical trials and medical research fields for 12 years. She became a part-time student in Biostatistics Department at Virginia Commonwealth University in 2008 to receive her M.S. in Biostatistics.