

COMPARISON OF MIXED MODELS AND PAIRED T-TEST FOR ANALYZING CROSSOVER
CLINICAL TRIALS IN THE PRESENCE OF MISSING DATA

by

Anthony Vicenti

Copyright © Anthony Vicenti 2020

A Thesis Submitted to the Faculty of the

GRADUATE INTERDISCIPLINARY PROGRAM

IN STATISTICS AND DATA SCIENCE

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

2020

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Master's Committee, we certify that we have read the thesis prepared by: Anthony Vicenti
titled: COMPARISON OF MIXED MODELS AND PAIRED T-TEST FOR ANALYZING CROSSOVER CLINICAL TRIALS IN THE PRESENCE OF MISSING DATA

and recommend that it be accepted as fulfilling the thesis requirement for the Master's Degree.

Melanie L Bell

Melanie L Bell

Date: Jan 19, 2021

Joseph C. Watkins

JOE WATKINS

Date: Jan 19, 2021

Jin Zhou

Jin Zhou

Date: Jan 21, 2021

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to the Graduate College.

I hereby certify that I have read this thesis prepared under my direction and recommend that it be accepted as fulfilling the Master's requirement. 

Melanie L Bell

Melanie L Bell

Thesis Committee Chair
Epidemiology and Biostatistics

Date: Jan 19, 2021

ARIZONA

Acknowledgments

I want to extend a special thank you to my advisor *Professor Melanie Bell* for her guidance throughout this project. She has been a mentor and a friend, and I will forever be thankful for all she has done. I would also like to thank my other committee members *Professor Joe Watkins* and *Professor Jin Zhou* for all their time, effort, and expertise. Lastly, I would like to thank my family and friends for all their support and love, especially my wife, children, and parents.

Table of Contents

List of Tables	5
Abstract	6
Chapter 1 – Background	7
1.1 AB/BA Crossover Clinical Trial Overview.....	7
1.2 Missing Data Overview.....	7
1.3 Analysis and Previous Research.....	8
Chapter 2 – Methods	9
1.4 Data-generating Mechanisms.....	10
1.5 Missingness Mechanisms.....	10
1.6 Analysis.....	11
1.7 Performance Measures.....	11
Chapter 3 – Results	13
1.8 Complete Data.....	13
1.9 MCAR Data.....	13
1.10 MAR data.....	13
Chapter 4 – Discussion	17
1.11 Bias.....	17
1.12 Power.....	18
1.13 Other Performance Measures.....	18
1.14 Correlation.....	18
1.15 Strengths and Limitations.....	19
1.16 Conclusion.....	19
Appendix A – Code	21
References	51

List of Tables

3.1	Results of paired t-test and mixed model analysis under all missing structures using correlation $\rho = .3$, $n=100$ individuals, and $N=1000$ simulations.....	13
3.2	Results of paired t-test and mixed model analysis under all missing structures using correlation $\rho = .7$, $n=56$ individuals, and $N=1000$ simulations.....	14

Abstract

AB/BA crossover clinical trials are popular designs that can achieve high power with a lower number of subjects than other randomized control trial designs. They are often analyzed using paired t-test or mixed models, and like many clinical trials, are often impacted by missing data. Mixed models have been shown to produced more powerful and unbiased results in the presence of missing data than t-tests for other designs, but these two approaches have not been compared in crossover trials.

We conducted a simulation study to compare the bias and power of paired t-tests and mixed models when analyzing an AB/BA crossover clinical trial in the presence of missing data. Several different missing structures were simulated under two within-subject correlations, $\rho = 0.3$ and $\rho = 0.7$.

Both methods performed similarly when analyzing complete data, but the mixed model produced both equal or less bias estimates and higher power than the paired t-test under all simulation scenarios. In the worst-case scenario we considered, the t-tests resulted in percent bias up to -105% and power as low as 5% compared the mixed model's percent bias of 1% and 57% power. In less severe cases, both methods had 0% bias, but mixed models still achieved an absolute power gain of 2%-6%.

In the presence of missing data, the mixed model achieved higher power than the paired t-test under all simulated scenarios. The mixed model also achieved equal or less bias under all simulated scenarios. Therefore, mixed models should be used over paired t-test when analyzing AB/BA crossover clinical trial in the face of missing data.

Chapter 1

Background

1.1 AB/BA Crossover Clinical Trial Overview

In AB/BA crossover clinical trials, individuals are randomly assigned a starting treatment at period one. After a washout period, the individuals are given the other treatment in period two. The washout period is designed to be long enough to eliminate any carry-over effects the treatments from period one might have on period two. This design is popular for several reasons [1]. Since individuals take both treatments, they act as their own control. This negates between subject variance and requires fewer individuals for the same power as other designs [2]. Crossover clinical trials are ideal when the treatments have a short-term effect and the targeted response does not change much over time, such as chronic conditions like asthma [1]. Crossover clinical trials are not ideal when the treatment effects are long term, irreversible, or the targeted response does change significantly over time, such as experiments involving pregnancy [3]. This design usually assumes there is no carry-over effect; however, this may not be the case without an extensive washout period [2].

1.2 Missing Data Overview

Missing data is a problem in most clinical trials [4], [5] and can come in three different forms according to Rubin [6]. These include missing completely at random (MCAR), where the probability of a missing observation is not based on any current or former data, missing at random (MAR), where the probability of a missing observation is based on observed data, and missing not at random (MNAR), where the probability of a missing observation is based on unobserved missing data.

Other possible missing mechanisms include different missing rates between treatments and the direction of the missing mechanism (whether high values or low values are more likely to go missing). Direction may be a factor when treatments are so poor that individuals stop participating in the study or when treatments do so

well that individuals decide they no longer need the study. In extreme cases, two treatments may be missing in different directions. In real clinical trials, we expect the data to be some mixture of these missing mechanisms.

1.3 Analysis and Previous Research

AB/BA crossover clinical trials are often analyzed using mixed models [7] or paired t-tests [3]. In the presence of missing data, mixed models use maximum likelihood to implicitly impute the missing data [8]. According to Rosenkranz [7], this is likely to produce unbiased estimates in crossover clinical trials when data are MCAR and MAR. Paired t-test do not impute missing data and can only analyze subjects with complete data. Not incorporating individuals with missing data can result in less power and potentially induce bias. Bell et al. [9] showed t-tests produce biased results in longitudinal randomized control trials when data are MAR with different missing rates or different directions. Ashbeck et al. [10] showed that mixed models provide less bias and more power compared to t-tests when analyzing longitudinal randomized control trials with data that are MCAR and MAR. There has not been a direct comparison between paired t-test and mixed models in the crossover setting in the presence of these missing mechanisms.

We designed a simulation study to compare the power and bias for continuous outcomes between paired t-tests and mixed models when analyzing AB/BA crossover clinical trials in the presence of data that are MCAR and MAR. We varied within-subject correlation, and the missing mechanism.

Chapter 2

Methods

2.1 Data Generating Mechanisms

We performed a simulation study based on an AB/BA crossover design clinical where half of the individuals were randomly selected to receive treatment A at period one and the other half were given treatment B at period one. The individuals received the other treatment for period two. We assumed there was no carry-over effect and no sequence effect.

The following model was used to generate the simulated data, based on the model described in Rosenkranz [7].

$$Y_{ij} = u + \pi_j + t_{ij} + U_i + e_{ij} \quad (1.1)$$

Where Y_{ij} is the response of the i -th individual at the j -th period, u denotes overall mean, π_j is the effect of period j , t_{ij} is the treatment effect of the i -th individual at the j -th period, $U_i \sim N(0, \sigma_b^2)$ is the between-subject effect where σ_b^2 is the between-subject variance, and $e_{ij} \sim N(0, \sigma_e^2)$ is the within-subject effect and σ_e^2 is the within-subject variance for $i=1, 2, \dots, n$ and $j = 1, 2$.

Mean and variance used for the simulations were based on peak expiratory flow measurements from an asthma crossover study reported in Senn et al. [11], with u set to 100 and total variance set to 144. The treatment effect was chosen to have a standardized effect size (Cohen's d) of 0.3, a small effect size, corresponding to an unstandardized difference of 4 between the treatments [12]. The period effect was chosen to have standardized effect size (Cohen's d) of 0.03, corresponding to an unstandardized difference of 0.4 between the periods.

We calculated two sets of between-subject and within-subject variances based on the total variance and two chosen correlations of $\rho = 0.3$ and $\rho = 0.7$. The between-subject and within-subject variance sum to the total variance ($\sigma_b^2 + \sigma_e^2 = \sigma^2 = 144$) and correlation $\rho = \sigma_b^2 / (\sigma_b^2 + \sigma_e^2)$. Therefore, for $\rho = 0.3$, $\sigma_b = 6.57$ and $\sigma_e = 10.04$ and for $\rho = 0.7$, $\sigma_b = 10.04$ and $\sigma_e = 6.57$.

Two sample sizes were calculated for each of the different correlations to achieve a baseline power of 80% for the complete data using a paired t-test. For $\rho = 0.3$, with mean difference of 4 and standard deviation of 12, $n=100$ provides ~80% power. For $\rho = 0.7$, with mean difference of 4 and standard deviation of 12, $n=46$ provides ~80% power.

2.2 Missingness Mechanisms

Missingness structures were simulated under three overarching scenarios. Without loss of generality, we assumed only period two had missing data.

1. *Complete (No Missing)*. Under this scenario, nothing was done to the complete datasets.
2. *Missing Completely at Random (MCAR)*. Under this scenario, the probability of an observation missing is not based on any current or former values of Y. To simulate this, we took a simple random sample of the data in period two at 70% and 50% to create 30% and 50% missing data.
3. *Missing at Random (MAR)*: Under this scenario, the probability of an observation missing is based on observed data. To simulate this, we made the probability of data missing in period two a function of data in period one. $P(Y_{i,2} = \text{Missing}) = c(1 - \Phi(Y_{i,1}, \bar{Y}_1, \sigma))$ was used to make higher values go missing and $P(Y_{i,2} = \text{Missing}) = 1 - c(1 - \Phi(Y_{i,1}, \bar{Y}_1, \sigma))$ was used to make lower values go missing. Φ was the cumulative normal distribution, c was a tuning parameter used to generate the correct percentage of missing data, \bar{Y}_1 was the mean during period 1, and σ was the total standard deviation.

The final MAR structures varied the amount of dropout and the direction of the missingness. For equal dropout between treatments, there were four total structures consisting of 30% missing and in the same

direction (higher values went missing for both treatments) and 30% missing and in different directions (lower values from treatment A went missing and higher values from treatment B went missing). The same structures were applied to 50% missingness.

There were also four total structures for unequal dropout between treatments. These consisted of 20% missing from treatment A and 40% missing from treatment B in the same direction (higher values for both treatments went missing) and 20% missing from treatment A and 40% missing from treatment B in different directions (lower values from treatment A went missing and higher values from treatment B went missing). The same structures were applied to a 40%/60% split.

2.3 Analysis

Two different methods of analysis were undertaken for each of the missing structures across all ρ . The first was a simple paired t-test that compared the mean of treatment A and treatment B using only the complete cases. The second was a mixed model that incorporated a fixed treatment and period effect and a random subject effect. The mixed model used maximum likelihood to implicitly impute the missing data. Both analyses were run in SAS University Edition version 9.4. Code is available in the Appendix.

2.4 Performance Measures

This study targeted two performance measures, the percent bias of the estimated difference between treatments and the power of detecting treatment effects. For simulated difference $\theta = 4$ and mean of estimated difference across all simulations $\bar{\theta}$, the relative bias was calculated as $\frac{|\bar{\theta} - \theta|}{\theta} \times 100\%$. The power was calculated by counting the number of tests with p-value < 0.05 and dividing by the total number of simulations $N = 1000$ as follows: Power = $\frac{\text{Number of } p\text{-values} < 0.05}{N} \times 100\%$. Other measures included Monte Carlo standard error, coverage, empirical standard error, and model standard error as described in Morris [13].

The Monte Carlo standard error for the percent bias was calculated as $\sqrt{\frac{1}{N(N-1)} \sum_{i=1}^N (\hat{\theta}_i - \bar{\theta})^2} \times 100\%$,

where $N = 1000$ was the number of simulations, $\hat{\theta}_i$ was the estimate of the difference for each simulation, $\bar{\theta}$ was

the mean estimate of the difference across all simulations, and $\theta = 4$ was the simulated mean. The Monte Carlo standard error for power was calculated as $\sqrt{\frac{\widehat{power}(1-\widehat{power})}{N}}$, where \widehat{power} was the estimate of the power for each simulation and $N = 1000$ was the number of simulations.

Coverage was calculated by counting the number of simulations where the confidence interval for $\hat{\theta}$, the difference, included the true value $\theta = 4$ and dividing by the total number of simulations $N=1000$. Empirical standard error was calculated as the standard deviation of the difference estimates, $\hat{\theta}$, across all simulations. The model standard error was calculated as the average standard error across all simulations.

The best method will have low bias, high power, low Monte Carlo standard error, coverage at or above 95%, and empirical standard error that is equivalent to the model standard error.

Chapter 3

Results

Performance measures calculated from the $\rho = 0.3$ dataset are presented in Table 1 for all missing mechanisms. Table 2 contains the same measurements from the $\rho = 0.7$ datasets.

3.1 Complete Data

Analysis of the complete data yielded similar performance between the paired t-test and mixed model under all ρ . Both methods resulted in 0% bias and both achieved the expected power of 80%. Both methods also achieved or exceeded the expected coverage of 95% and both had empirical and model standard errors that roughly matched.

3.2 MCAR Data

Analysis of the MCAR data resulted in percent bias of 0% for both methods under all MCAR structures and under all ρ . Although both methods saw a drop in power compared to the complete data, the mixed model achieved slightly higher power under all MCAR scenarios. The absolute increase in power ranged from 2% to 6% for the mixed model compared to the paired t-test. Coverage achieved or exceeded the expected 95% and the empirical and model standard errors roughly matched under all MCAR scenarios.

3.3 MAR Data

For the MAR data, the mixed model resulted in lower or equal percent bias and higher power under all MAR structures and under all ρ .

The paired t-test resulted in the highest percent bias and lowest power when the data were unequal and in different directions. In the case of 40%/60% MAR in different directions when $\rho=.7$, the paired t-test resulted in -45% bias compared to the in the mixed model's 3% bias and achieved 19% power compared to the mixed

Table 1 Results of paired t-test and mixed model analysis under all missing structures using correlation $\rho = .3$, $n=100$ individuals, and $N=1000$ simulations.

	$\rho = 0.3$					
	Estimated Difference	Bias % (MC SE) ¹	Power % (MC SE) ²	Coverage %	Model SE	Empirical SE
Complete						
paired t-test	4.0	0 (1.1)	80 (1.3)	95	1.4	1.4
mixed model	4.0	0 (1.1)	80 (1.3)	95	1.4	1.4
MCAR 30%						
paired t-test	4.0	0 (1.3)	64 (1.5)	95	1.7	1.7
mixed model	4.0	0 (1.2)	70 (1.4)	96	1.6	1.6
MCAR 50%						
paired t-test	4.0	0 (1.6)	51 (1.6)	95	2.0	2.0
mixed model	4.0	0 (1.4)	57 (1.6)	96	1.8	1.8
MAR 30%						
Same Direction						
paired t-test	4.2	5 (1.4)	68 (1.5)	94	1.7	1.7
mixed model	4.0	0 (1.3)	70 (1.4)	95	1.6	1.6
Different Direction						
paired t-test	2.2	-45 (1.4)	25 (1.4)	80	1.7	1.7
mixed model	4.0	0 (1.3)	68 (1.5)	94	1.6	1.7
MAR 20/40%						
Same Direction						
paired t-test	3.5	-13 (1.4)	52 (1.6)	95	1.7	1.7
mixed model	4.0	0 (1.3)	69 (1.5)	95	1.6	1.6
Different Direction						
paired t-test	2.1	-48 (1.4)	24 (1.4)	80	1.7	1.7
mixed model	4.0	0 (1.3)	68 (1.5)	94	1.6	1.7
MAR 50%						
Same Direction						
paired t-test	4.4	10 (1.6)	55 (1.6)	95	2.0	2.0
mixed model	4.0	0 (1.5)	57 (1.6)	94	1.8	1.9
Different Direction						
paired t-test	-0.2	-105 (1.6)	5 (0.7)	41	1.9	2.0
mixed model	4.1	3 (1.6)	57 (1.6)	92	1.8	2.0
MAR 40/60%						
Same Direction						
paired t-test	3.5	-13 (1.6)	39 (1.5)	94	2.0	2.0
mixed model	4.0	0 (1.5)	55 (1.6)	93	1.8	1.9
Different Direction						
paired t-test	-0.2	-105 (1.6)	5 (0.7)	43	1.9	2.0
mixed model	4.0	1 (1.6)	57 (1.6)	92	1.8	2.1

¹Monte Carlo standard error for percent bias.

²Monte Carlo standard error for power.

Table 2 Results of paired t-test and mixed model analysis under all missing structures using correlation $\rho = .7$, $n=46$ individuals, and $N=1000$ simulations.

$\rho = 0.7$						
	Estimated Difference	Bias % (MC SE ¹)	Power % (MC SE ²)	Coverage %	Model SE	Empirical SE
Complete						
paired t-test	4.0	0 (1.1)	81 (1.2)	96	1.4	1.3
mixed model	4.0	0 (1.1)	81 (1.2)	96	1.4	1.3
MCAR 30%						
paired t-test	4.0	0 (1.3)	65 (1.5)	95	1.6	1.6
mixed model	4.0	0 (1.2)	67 (1.5)	95	1.6	1.6
MCAR 50%						
paired t-test	4.0	0 (1.5)	49 (1.6)	96	1.9	1.9
mixed model	4.0	0 (1.5)	53 (1.6)	96	1.9	1.9
MAR 30%						
Same Direction						
paired t-test	4.1	3 (1.3)	68 (1.5)	96	1.6	1.6
mixed model	4.0	0 (1.3)	69 (1.5)	96	1.6	1.6
Different Direction						
paired t-test	3.2	-20 (1.3)	48 (1.6)	92	1.6	1.7
mixed model	4.0	0 (1.3)	69 (1.5)	95	1.6	1.6
MAR 20/40%						
Same Direction						
paired t-test	3.8	-5 (1.3)	61 (1.5)	95	1.6	1.6
mixed model	4.0	0 (1.3)	67 (1.5)	96	1.6	1.6
Different Direction						
paired t-test	3.2	-20 (1.3)	47 (1.6)	91	1.6	1.7
mixed model	4.0	0 (1.3)	66 (1.5)	94	1.6	1.7
MAR 50%						
Same Direction						
paired t-test	4.1	3 (1.5)	50 (1.6)	96	1.9	1.9
mixed model	4.0	0 (1.5)	51 (1.6)	95	1.9	1.9
Different Direction						
paired t-test	2.2	-45 (1.5)	19 (1.2)	85	1.9	1.9
mixed model	4.1	3 (1.6)	53 (1.6)	93	1.9	2.1
MAR 40/60%						
Same Direction						
paired t-test	3.7	-8 (1.5)	46 (1.6)	96	1.9	1.9
mixed model	4.0	0 (1.5)	52 (1.6)	95	1.9	2.0
Different Direction						
paired t-test	2.2	-45 (1.5)	19 (1.2)	84	1.9	2.0
mixed model	4.1	3 (1.7)	51 (1.6)	93	1.9	2.1

¹Monte Carlo standard error for percent bias.

²Monte Carlo standard error for power.

model's 51%. This trend was similar under all unequal and in different direction MAR structures, with higher percent missingness leading to higher percent bias and lower power.

The paired t-test was relatively closer in percent bias and power to the mixed model when the missingness was equal and in the same direction. In the case of 30% MAR in the same direction when $\rho=.3$, the paired t-test resulted in 5% bias compared to the mixed model's 0% bias and achieved 68% power compared to the mixed model's 70%. The trend was similar under most equal and in the same direction MAR structures, with higher percent missingness leading to higher percent bias and lower power overall.

Although the trends were similar when $\rho=.3$ and $\rho=.7$, $\rho=.3$ resulted in higher bias for the paired t-test compared to $\rho=.7$ under all MAR structures. The empirical and model standard errors roughly matched under all MAR scenarios; however, coverage was below the expected 95% when bias was very high.

Chapter 3

Discussion

In the presence of missing data, mixed models outperformed paired t-tests in analyzing the data by having lower percent bias, higher power, and in many cases both. The paired t-test had the worst performance when missingness was unequal and in different directions but performed reasonably with complete data or data that were MCAR.

4.1 Bias

When data were MAR and in different directions, with high scores from treatment B and low scores from treatment A went missing. This missingness structure pushes the overall difference between treatments to be smaller. The estimates in this scenario ranged from -0.2 to 3.2 for the paired t-tests where mixed models had a range of 4.0-4.1 (the true value was 4.0). Paired t-test were not able to produce unbiased results, where mixed models were able to account for this structure through the maximum likelihood imputation of missing observations.

There was a similar trend when data were MAR with unequal dropout in the same direction. Under this structure, high values from both treatments went missing, but Treatment B had more missing data than Treatment A. This structure forced the mean of treatment B to decrease more than treatment A and pushed the difference in treatments to be smaller. The estimates of the difference where missingness was unequal and in the same direction ranged from 3.5-3.7 for the paired t-tests where mixed models all estimated 4.0. Again, the maximum likelihood imputation of missing observations allowed for unbiased estimation in the face of missingness where the t-test gave biased results.

For data that were MAR with equal dropout in the same direction, high values from both treatment A and treatment B went missing at the same proportion. Therefore, their difference should have stayed the same although their overall values decreased. These results still showed some bias from the t-test because the

cumulative normal distribution used for inducing missingness used the total mean and variance instead of treatment specific mean and variance. Since treatment A's values were inherently lower than B's, removing high values skewed towards the highest values. On the other hand, treatment B's values were higher, on average, than A's, meaning removing high values had a more even distribution. The end effect is that treatment A decreased more than treatment B, resulting in the difference in treatment to be larger. The t-test estimated differences of 4.1-4.4 under this structure while the mixed model was able to estimate 4.0 for all the cases.

4.2 Power

Under all missing structures, not including the complete structure where power was equal, mixed models achieved greater power than paired t-tests. The increase in power follows the same trends as the increase in bias for similar reasons. For the missing in different directions, the power ranged from 5% to 48% for the paired t-tests where the mixed models ranged from 51%-68%. For missing in unequal directions and in the same direction, the power ranged from 39% -61% for the paired t-tests where the mixed models ranged from 52%-69%. For MCAR and missingness that was equal and in the same direction, the power ranged from 49%-68% for the paired t-tests where the mixed models ranged from 51%-70%. There should be care when analyzing power in the face of bias. In this case, the estimates with high bias towards a smaller difference resulted in less power.

4.3 Other Performance Measures

Model standard error and empirical standard error seemed relatively equivalent across the board and coverage was at our above 95% most of the time. Coverage was only significantly lower when bias was high. This is expected since estimates that bias away from the true estimate have confidence intervals that also move away from the true estimate.

4.4 Correlation

In general, percent bias was higher when $\rho = .3$ compared to when $\rho = .7$ for the paired t-tests. This can be attributed to the crossover design where the between-subject effect cancels out when you take the difference

of treatment. This leaves only within-subject variance. Since $\rho = .3$ contains more within-subject variance and $\rho = .7$ contains more between-subject variance, $\rho = .3$ is more affected by missingness and therefore leads to more biased results.

These results are in line with longitudinal randomized control trial simulation studies. Bell et al's [9] study found similar trends in a t-test's ability to handle missingness that was in different directions and at different rates between treatments. Ashbeck et al's [10] study also found similar trends, except their bias increased as ρ increased. This illustrates how the canceling out of between-subject effect from a crossover design gives within-subject variance a greater role compared to a standard longitudinal randomized control trial.

4.5 Strengths and Limitations

While other studies, such as Rosenkranz [7], have explored the performance of using mixed models to analyze AB/BA crossover clinical trials in the face of missingness, they did not consider the direction of the missingness and missingness at different rates between treatments. There have also not been any direct comparisons on the performance between mixed models and paired t-test. We consider these strengths of this simulation study.

This study only looked at two sample sizes $n=46$, and $n=100$, which are fairly large for a crossover clinical trial. This study also did not look at how repeated measures within each treatment would affect performance. While there is reason to believe that MNAR data would yield poor performance for both models, this missing structure was not investigated.

Future studies in this area can explore how different sample sizes may affect percent bias and power. They can also explore the effect of repeated measures and include the MNAR structure.

4.6 Conclusion

While some of these missing structures are extreme, real world missingness probably lies somewhere between the best-case and worst-case scenarios. Therefore, in analyzing AB/BA crossover clinical trials with

missing data, mixed models will at minimum result in better power than paired t-tests and are likely to yield less biased results.

Appendix A - Code

```
/*Calculate Power p=.3*/
```

```
/*
```

```
proc power;
```

```
  pairedmeans test=diff
```

```
  meandiff= 4
```

```
  corr=.3
```

```
  stddev = 12
```

```
  npairs= .
```

```
  power = .8;
```

```
run;
```

```
*/
```

```
/*p=.3 parameters*/
```

```
/*
```

```
%LET Simulations = 1000;
```

```
%LET N=100;
```

```
%LET TMean=100;
```

```
%LET WithinVaraince=10.04;
```

```
%LET BetweenVaraince=6.57;
```

```
%LET NMiss30 = 70;
```

```
%LET NMiss50 = 50;
```

```
*/
```

```
/*Calculate Power p=.7*/
```

```
proc power;
```

```
  pairedmeans test=diff
```

```
  meandiff= 4
```

```
  corr=.7
```

```
  stddev = 12
```

```
  npairs= .
```

```
  power = .8;
```

```
run;
```

```
/*p=.7 parameters*/
```

```
%LET Simulations = 1000;
```

```
%LET N=46;
```

```
%LET TMean=100;
```

```
%LET WithinVaraince=6.57;
```

```
%LET BetweenVaraince=10.04;
```

```
%LET NMiss30 = 32;
```

```
%LET NMiss50 = 23;
```

```
/*Simulate Data*/
```

```

data Simulated(keep=y treat period person seq s B E);

call streaminit(54321);

do s = 1 to &Simulations;

    do i = 1 to &N/2;

        person =i;
        B=rand("Normal",0,&BetweenVaraince);
        seq = 0;
        do j = 0 to 1;
            treat=j;
            if j=0 then period = 0;
            if J=1 then period =1;
            E=rand("Normal",0,&WithinVaraince);

            y=&TMean+4*treat+.4*Period+B+E;
            output;

        end;
    end;

do i = &N/2+1 to &N;

    person =i;
    B=rand("Normal",0,&BetweenVaraince);
    seq =1;
    do j = 0 to 1;
        treat=j;
        if j=0 then period = 1;
        if J=1 then period =0;
        E=rand("Normal",0,&WithinVaraince);

        y=&TMean+4*treat+.4*Period+B+E;
        output;

    end;
end;

run;

/*Sort Data To Convert From Long To Wide Format*/
proc sort data=Simulated out=Simulated_Sort;
by s person;
run;

```

```

Data PairSet;
set Simulated_Sort;
by s person;
Ly=lag(y);

if last.person and treat = 0 then
do;
    Treat0 = y;
    Treat1 = Ly;
    output;
end;
else if last.person and treat = 1 then
do;
    Treat1 = y;
    Treat0 = Ly;
    Output;
end;
run;

```

```

/*Missing Data Generation*/

```

```

/*****

```

```

*****/

```

```

/*Missing Completely at Random 30%*/

```

```

proc surveyselect data=Simulated method=srs n=&NMiss30 out=MCAR_30_P1;
where period =1;
strata s;
run;

```

```

Data MCAR_30_P0;

```

```

set Simulated;
if period = 0;
run;

```

```

Data MCAR_30;

```

```

set MCAR_30_P0 MCAR_30_P1;
run;

```

```

proc sort data = MCAR_30 out = MCAR_30;
by s;
run;

```

```

/*Missing Completely at Random 50%*/

```

```

proc surveyselect data=Simulated method=srs n=&NMiss50 out=MCAR_50_P1;
where period =1;
strata s;
run;

```

```

Data MCAR_50_P0;
set Simulated;
if period = 0;
run;

Data MCAR_50;
set MCAR_50_P0 MCAR_50_P1;
run;

proc sort data = MCAR_50 out = MCAR_50;
by s;
run;

/*MAR*/
/*****
*****/
/*Missing at random, Equal same direction 30*/

proc sort data=Simulated out=Simulated_Sort;
by s person period;
run;

proc means data = Simulated_Sort p25 p75 mean std;
by s;
var y;
where period = 0;
ods output summary= SummarySimulated;
run;

Data Simulated_WithMedians;
merge Simulated_Sort SummarySimulated;
by s;
run;

data MAR_EqualSame30 (keep=s person treat period y p q B E);
set Simulated_WithMedians;
by s person period;
Ly=lag(y);

if first.person then output;

ELSE if last.person then
do;
    if(treat = 0) then
    do;
        q=.66*(1-cdf('NORMAL',Ly,y_Mean,y_StdDev));
        p=rand("Uniform");
        if p > q then output;
    end;
end;

```

```

else if(treat = 1) then
do;
    p=rand("Uniform");
    q=.545*(1-cdf('NORMAL',Ly,y_Mean,y_StdDev));

    if p > q then output;
end;
end;

run;

/*Missing at random, Equal different direction 30*/

proc sort data=Simulated out=Simulated_Sort;
by s person period;
run;

proc means data = Simulated_Sort p25 p75 mean std;
by s;
var y;
where period = 0;
ods output summary= SummarySimulated;
run;

Data Simulated_WithMedians;
merge Simulated_Sort SummarySimulated;
by s;
run;

data MAR_EqualDif30 (keep=s person treat period y p q B E);
set Simulated_WithMedians;
call streaminit(123456);
by s person period;
Ly=l原因(y);

if first.person then output;

ELSE if last.person then
do;
    if(treat = 0) then
do;
    q=.55*(cdf('NORMAL',Ly,y_Mean,y_StdDev));

    p=rand("Uniform");
    if p > q then output;
end;

    else if(treat = 1) then
do;

```

```

        p=rand("Uniform");
        q=.55*(1-cdf('NORMAL',Ly,y_Mean,y_StdDev));

        if p > q then output;
    end;
end;

run;

/*Missing at random, unequal 20% 40% same direction */

proc sort data=Simulated out=Simulated_Sort;
by s person period;
run;

proc means data = Simulated_Sort p25 p75 mean std;
by s;
var y;
where period = 0;
ods output summary= SummarySimulated;
run;

Data Simulated_WithMedians;
merge Simulated_Sort SummarySimulated;
by s;
run;

data MAR_UnequalSame2040 (keep=s person treat period y p q B E);
set Simulated_WithMedians;
call streaminit(123456);
by s person period;
Ly=lag(y);

if first.person then output;

ELSE if last.person then
do;
    if(treat = 0) then
    do;
        /*p=.7 and p=.3*/
        q=.44*(1-cdf('NORMAL',Ly,y_Mean,y_StdDev));

        p=rand("Uniform");
        if p > q then output;
    end;

    else if(treat = 1) then
    do;
        p=rand("Uniform");

```

```

                /*p=.7 and p=.3*/
q=.73*(1-cdf('NORMAL',Ly,y_Mean,y_StdDev));
    if p > q then output;
    end;
end;

run;

/*Missing at random, unequal 20% 40% different direction */

proc sort data=Simulated out=Simulated_Sort;
by s person period;
run;

proc means data = Simulated_Sort p25 p75 mean std;
by s;
var y;
where period = 0;
ods output summary= SummarySimulated;
run;

Data Simulated_WithMedians;
merge Simulated_Sort SummarySimulated;
by s;
run;

data MAR_UnequalDif2040 (keep=s person treat period y p q B E);
set Simulated_WithMedians;
call streaminit(123456);
by s person period;
Ly=lag(y);

if first.person then output;

ELSE if last.person then
do;
    if(treat = 0) then
    do;
        /*p=.7 and p=.3*/
q=.37*(cdf('NORMAL',Ly,y_Mean,y_StdDev));
p=rand("Uniform");
    if p > q then output;
    end;

    else if(treat = 1) then
    do;
        p=rand("Uniform");
        /*p=.7 and p=.3*/
q=.73*(1-cdf('NORMAL',Ly,y_Mean,y_StdDev));
    if p > q then output;

```

```

end;
end;

run;

/*MAR 50%*/
/*****
*****/
/*Missing at random, Equal same direction 50*/

proc sort data=Simulated out=Simulated_Sort;
by s person period;
run;

proc means data = Simulated_Sort p25 p75 mean std;
by s;
var y;
where period = 0;
ods output summary= SummarySimulated;
run;

Data Simulated_WithMedians;
merge Simulated_Sort SummarySimulated;
by s;
run;

data MAR_EqualSame50 (keep=s person treat period y p q B E);
set Simulated_WithMedians;
call streaminit(123456);
by s person period;
Ly=lag(y);

if first.person then output;

ELSE if last.person then
do;
    if(treat = 0) then
do;
/*p=.7 and p=.3*/
q=1.11*(1-cdf('NORMAL',Ly,y_Mean,y_StdDev));
p=rand("Uniform");
if p > q then output;
end;

    else if(treat = 1) then
do;
        p=rand("Uniform");
/*p=.7 and p=.3*/
q=.91*(1-cdf('NORMAL',Ly,y_Mean,y_StdDev));
if p > q then output;

```

```

        end;
end;

run;

/*Missing at random, Equal different direction 50*/

proc sort data=Simulated out=Simulated_Sort;
by s person period;
run;

proc means data = Simulated_Sort p25 p75 mean std;
by s;
var y;
where period = 0;
ods output summary= SummarySimulated;
run;

Data Simulated_WithMedians;
merge Simulated_Sort SummarySimulated;
by s;
run;

data MAR_EqualDif50 (keep=s person treat period y p q B E);
set Simulated_WithMedians;
call streaminit(123456);
by s person period;
Ly=lag(y);

if first.person then output;

ELSE if last.person then
do;
    if(treat = 0) then
    do;
        /*p=.7 and p=.3*/
        q=.91*(cdf('NORMAL',Ly,y_Mean,y_StdDev));
        p=rand("Uniform");
        if p > q then output;
        end;

    else if(treat = 1) then
    do;
        p=rand("Uniform");
        /*p=.7 and p=.3*/
        q=.91*(1-cdf('NORMAL',Ly,y_Mean,y_StdDev));
        if p > q then output;
        end;
end;
end;

```

```

run;

/*Missing at random, unequal 40% 60% same direction */

proc sort data=Simulated out=Simulated_Sort;
by s person period;
run;

proc means data = Simulated_Sort p25 p75 mean std;
by s;
var y;
where period = 0;
ods output summary= SummarySimulated;
run;

Data Simulated_WithMedians;
merge Simulated_Sort SummarySimulated;
by s;
run;

data MAR_UnequalSame4060 (keep=s person treat period y p q B E);
set Simulated_WithMedians;
call streaminit(123456);
by s person period;
Ly=lag(y);

if first.person then output;

ELSE if last.person then
do;
    if(treat = 0) then
    do;
        /*p=.7 and p=.3*/
        q=.88*(1-cdf('NORMAL',Ly,y_Mean,y_StdDev));
        p=rand("Uniform");
        if p > q then output;
        end;

    else if(treat = 1) then
    do;
        p=rand("Uniform");
        /*p=.7 and p=.3*/
        q=1.1*(1-cdf('NORMAL',Ly,y_Mean,y_StdDev));
        if p > q then output;
        end;
end;

run;

```

```
/*Missing at random, unequal 50% 60% different direction */
```

```
proc sort data=Simulated out=Simulated_Sort;  
by s person period;  
run;
```

```
proc means data = Simulated_Sort p25 p75 mean std;  
by s;  
var y;  
where period = 0;  
ods output summary= SummarySimulated;  
run;
```

```
Data Simulated_WithMedians;  
merge Simulated_Sort SummarySimulated;  
by s;  
run;
```

```
data MAR_UnequalDif4060 (keep=s person treat period y p q B E);  
set Simulated_WithMedians;  
call streaminit(123456);  
by s person period;  
Ly=lag(y);
```

```
if first.person then output;
```

```
ELSE if last.person then  
do;
```

```
    if(treat = 0) then  
    do;  
        q=.73*(cdf('NORMAL',Ly,y_Mean,y_StdDev));  
        p=rand("Uniform");  
        if p > q then output;  
    end;
```

```
    else if(treat = 1) then  
    do;  
        p=rand("Uniform");  
        q=1.1*(1-cdf('NORMAL',Ly,y_Mean,y_StdDev));  
        if p > q then output;  
    end;
```

```
end;
```

```
run;
```

```
/*Analysis*/
```

```
/******  
*****/
```

```
/*Complete*/
```

```

/*Paired T-test*/
proc ttest data=PairSet alpha=0.05 h0=0 plots=none;
    title "Paired sample t-test";
    by s;
    paired Treat0*Treat1;
    ods output Conflimits=OUTPUT.conflim1;
    ods output TTests=OUTPUT.pvalue1;
    ods output Statistics=OUTPUT.differences1;
run;

/*Proc mixed*/
Proc Mixed Data=Simulated_Sort;
by s;
Class treat period;
Model y = treat period treat*period;
random int/subject = person;
estimate 'Treat 1 vs Treat 2' treat 1 -1 / cl;
ods output Estimates=OUTPUT.Estimates1;
Run;

/*Analysis MCAR*/

/*Paired T-test 30*/
/*Sort Data To Convert From Long To Wide Format*/
proc sort data=MCAR_30 out=MCAR_30_Sort;
by s person;
run;

Data PairSet2;
set MCAR_30_Sort;
by s person;
Ly=lag(y);

if first.person and last.person then
do;
    Treat0 = y;
    Treat1 = .;
    output;
end;

else if last.person and treat = 0 then
do;
    Treat0 = y;
    Treat1 = Ly;
    output;
end;
else if last.person and treat = 1 then
do;
    Treat1 = y;

```

```

        Treat0 = Ly;
        Output;
end;
run;

proc ttest data=PairSet2 alpha=0.05 h0=0 plots=none;
    title "Paired sample t-test" ;
    by s;
    paired Treat0*Treat1;
    ods output ConfLimits=OUTPUT.conflim2;
    ods output TTests=OUTPUT.pvalue2;
    ods output Statistics=OUTPUT.differences2;
run;

/*Proc mixed*/
Proc Mixed Data=MCAR_30_Sort;
by s;
Class treat period;
Model y = treat period treat*period;
random int/subject = person;
estimate 'Treat 1 vs Treat 2' treat 1 -1 / cl;
ods output Estimates=OUTPUT.Estimates2;
Run;

/*Analysis MCAR*/

/*Paired T-test 50*/
/*Sort Data To Convert From Long To Wide Format*/
proc sort data=MCAR_50 out=MCAR_50_Sort;
by s person;
run;

Data PairSet3;
set MCAR_50_Sort;
by s person;
Ly=lag(y);

if first.person and last.person then
do;
    Treat0 = y;
    Treat1 = .;
    output;
end;

else if last.person and treat = 0 then
do;
    Treat0 = y;
    Treat1 = Ly;
    output;

```

```

end;
else if last.person and treat = 1 then
do;
    Treat1 = y;
    Treat0 = Ly;
    Output;
end;
run;

proc ttest data=PairSet3 alpha=0.05 h0=0 plots=none;
    title "Paired sample t-test" ;
    by s;
    paired Treat0*Treat1;
    ods output Conflimits=OUTPUT.conflim3;
    ods output TTests=OUTPUT.pvalue3;
    ods output Statistics=OUTPUT.differences3;
run;

/*Proc mixed*/
Proc Mixed Data=MCAR_50_Sort;
by s;
Class treat period;
Model y = treat period treat*period;
random int/subject = person;
estimate 'Treat 1 vs Treat 2' treat 1 -1 / cl;
ods output Estimates=OUTPUT.Estimates3;
Run;

/*Analysis MAR_EqualSame30*/

/*Paired T-test */
/*Sort Data To Convert From Long To Wide Format*/
proc sort data=MAR_EqualSame30 out=MAR_EqualSame30_Sort;
by s person;
run;

Data PairSet4;
set MAR_EqualSame30_Sort;
by s person;
Ly=lag(y);

if first.person and last.person then
do;
    Treat0 = y;
    Treat1 = .;
    output;
end;

else if last.person and treat = 0 then

```

```

do;
    Treat0 = y;
    Treat1 = Ly;
    output;
end;
else if last.person and treat = 1 then
do;
    Treat1 = y;
    Treat0 = Ly;
    Output;
end;
run;

proc ttest data=PairSet4 alpha=0.05 h0=0 plots=none;
    title "Paired sample t-test" ;
    by s;
    paired Treat0*Treat1;
    ods output Conflimits=OUTPUT.conflim4;
    ods output TTests=OUTPUT.pvalue4;
    ods output Statistics=OUTPUT.differences4;
run;

/*Proc mixed*/
Proc Mixed Data=MAR_EqualSame30_Sort;
by s;
Class treat period;
Model y = treat period treat*period;
random int/subject = person;
estimate 'Treat 1 vs Treat 2' treat 1 -1 / cl;
ods output Estimates=OUTPUT.Estimates4;
Run;

/*Analysis MAR_EqualDif30*/

/*Paired T-test */
/*Sort Data To Convert From Long To Wide Format*/
proc sort data=MAR_EqualDif30 out=MAR_EqualDif30_Sort;
by s person;
run;

Data PairSet5;
set MAR_EqualDif30_Sort;
by s person;
Ly=lag(y);

if first.person and last.person then
do;
    Treat0 = y;
    Treat1 = .;
    output;

```

```

end;

else if last.person and treat = 0 then
do;
    Treat0 = y;
    Treat1 = Ly;
    output;
end;
else if last.person and treat = 1 then
do;
    Treat1 = y;
    Treat0 = Ly;
    Output;
end;
run;

proc ttest data=PairSet5 alpha=0.05 h0=0 plots=none;
    title "Paired sample t-test" ;
    by s;
    paired Treat0*Treat1;
    ods output Conflimits=OUTPUT.conflim5;
    ods output TTests=OUTPUT.pvalue5;
    ods output Statistics=OUTPUT.differences5;
run;

/*Proc mixed*/
Proc Mixed Data=MAR_EqualDif30_Sort;
by s;
Class treat period;
Model y = treat period treat*period;
random int/subject = person;
estimate 'Treat 1 vs Treat 2' treat 1 -1 / cl;
ods output Estimates=OUTPUT.Estimates5;
Run;

/*Analysis UnequalSame2040*/

/*Paired T-test */
/*Sort Data To Convert From Long To Wide Format*/
proc sort data=MAR_UnequalSame2040 out=MAR_UnequalSame2040_Sort;
by s person;
run;

Data PairSet6;
set MAR_UnequalSame2040_Sort;
by s person;
Ly=lag(y);

if first.person and last.person then
do;

```

```

        Treat0 = y;
        Treat1 = .;
        output;
end;

else if last.person and treat = 0 then
do;
    Treat0 = y;
    Treat1 = Ly;
    output;
end;
else if last.person and treat = 1 then
do;
    Treat1 = y;
    Treat0 = Ly;
    Output;
end;
run;

proc ttest data=PairSet6 alpha=0.05 h0=0 plots=none;
    title "Paired sample t-test" ;
    by s;
    paired Treat0*Treat1;
    ods output Conflimits=OUTPUT.conflim6;
    ods output TTests=OUTPUT.pvalue6;
    ods output Statistics=OUTPUT.differences6;
run;

/*Proc mixed*/
Proc Mixed Data=MAR_UnequalSame2040_Sort;
by s;
Class treat period;
Model y = treat period treat*period;
random int/subject = person;
estimate 'Treat 1 vs Treat 2' treat 1 -1 / cl;
ods output Estimates=OUTPUT.Estimates6;
Run;

/*Analysis MAR_UnequalDif2040*/

/*Paired T-test */
/*Sort Data To Convert From Long To Wide Format*/
proc sort data=MAR_UnequalDif2040 out=MAR_UnequalDif2040_Sort;
by s person;
run;

Data PairSet7;
set MAR_UnequalDif2040_Sort;
by s person;
Ly=lag(y);

```

```

if first.person and last.person then
do;
    Treat0 = y;
    Treat1 = .;
    output;
end;

else if last.person and treat = 0 then
do;
    Treat0 = y;
    Treat1 = Ly;
    output;
end;
else if last.person and treat = 1 then
do;
    Treat1 = y;
    Treat0 = Ly;
    Output;
end;
run;

proc ttest data=PairSet7 alpha=0.05 h0=0 plots=none;
    title "Paired sample t-test" ;
    by s;
    paired Treat0*Treat1;
    ods output Conflimits=OUTPUT.conflim7;
    ods output TTests=OUTPUT.pvalue7;
    ods output Statistics=OUTPUT.differences7;
run;

/*Proc mixed*/
Proc Mixed Data=MAR_UnequalDif2040_Sort;
by s;
Class treat period;
Model y = treat period treat*period;
random int/subject = person;
estimate 'Treat 1 vs Treat 2' treat 1 -1 / cl;
ods output Estimates=OUTPUT.Estimates7;
Run;

/*Analysis MAR_EqualSame50*/

/*Paired T-test */
/*Sort Data To Convert From Long To Wide Format*/
proc sort data=MAR_EqualSame50 out=MAR_EqualSame50_Sort;
by s person;
run;

Data PairSet8;

```

```

set MAR_EqualSame50_Sort;
by s person;
Ly=lag(y);

if first.person and last.person then
do;
    Treat0 = y;
    Treat1 = .;
    output;
end;

else if last.person and treat = 0 then
do;
    Treat0 = y;
    Treat1 = Ly;
    output;
end;
else if last.person and treat = 1 then
do;
    Treat1 = y;
    Treat0 = Ly;
    Output;
end;
run;

proc ttest data=PairSet8 alpha=0.05 h0=0 plots=none;
    title "Paired sample t-test" ;
    by s;
    paired Treat0*Treat1;
    ods output Conflimits=OUTPUT.conflim8;
    ods output TTests=OUTPUT.pvalue8;
    ods output Statistics=OUTPUT.differences8;
run;

/*Proc mixed*/
Proc Mixed Data=MAR_EqualSame50_Sort;
by s;
Class treat period;
Model y = treat period treat*period;
random int/subject = person;
estimate 'Treat 1 vs Treat 2' treat 1 -1 / cl;
ods output Estimates=OUTPUT.Estimates8;
Run;

/*Analysis MAR_EqualDif50*/

/*Paired T-test */
/*Sort Data To Convert From Long To Wide Format*/
proc sort data=MAR_EqualDif50 out=MAR_EqualDif50_Sort;
by s person;

```

```

run;

Data PairSet9;
set MAR_EqualDif50_Sort;
by s person;
Ly=lag(y);

if first.person and last.person then
do;
    Treat0 = y;
    Treat1 = .;
    output;
end;

else if last.person and treat = 0 then
do;
    Treat0 = y;
    Treat1 = Ly;
    output;
end;
else if last.person and treat = 1 then
do;
    Treat1 = y;
    Treat0 = Ly;
    Output;
end;
run;

proc ttest data=PairSet9 alpha=0.05 h0=0 plots=none;
    title "Paired sample t-test" ;
    by s;
    paired Treat0*Treat1;
    ods output Conflimits=OUTPUT.conflim9;
    ods output TTests=OUTPUT.pvalue9;
    ods output Statistics=OUTPUT.differences9;
run;

/*Proc mixed*/
Proc Mixed Data=MAR_EqualDif50_Sort;
by s;
Class treat period;
Model y = treat period treat*period;
random int/subject = person;
estimate 'Treat 1 vs Treat 2' treat 1 -1 / cl;
ods output Estimates=OUTPUT.Estimates9;
Run;

/*Analysis MAR_UnequalSame4060*/

/*Paired T-test */

```

```

/*Sort Data To Convert From Long To Wide Format*/
proc sort data=MAR_UnequalSame4060 out=MAR_UnequalSame4060_Sort;
by s person;
run;

```

```

Data PairSet10;
set MAR_UnequalSame4060_Sort;
by s person;
Ly=lag(y);

```

```

if first.person and last.person then
do;
    Treat0 = y;
    Treat1 = .;
    output;
end;

```

```

else if last.person and treat = 0 then
do;
    Treat0 = y;
    Treat1 = Ly;
    output;
end;

```

```

else if last.person and treat = 1 then
do;
    Treat1 = y;
    Treat0 = Ly;
    Output;
end;
run;

```

```

proc ttest data=PairSet10 alpha=0.05 h0=0 plots=none;
    title "Paired sample t-test" ;
    by s;
    paired Treat0*Treat1;
    ods output Conflimits=OUTPUT.conflim10;
    ods output TTests=OUTPUT.pvalue10;
    ods output Statistics=OUTPUT.differences10;
run;

```

```

/*Proc mixed*/
Proc Mixed Data=MAR_UnequalSame4060_Sort;
by s;
Class treat period;
Model y = treat period treat*period;
random int/subject = person;
estimate 'Treat 1 vs Treat 2' treat 1 -1 / cl;
ods output Estimates=OUTPUT.Estimates10;
Run;

```

```

/*Analysis MAR_UnequalDif4060*/

/*Paired T-test */
/*Sort Data To Convert From Long To Wide Format*/
proc sort data=MAR_UnequalDif4060 out=MAR_UnequalDif4060_Sort;
by s person;
run;

Data PairSet11;
set MAR_UnequalDif4060_Sort;
by s person;
Ly=lag(y);

if first.person and last.person then
do;
    Treat0 = y;
    Treat1 = .;
    output;
end;

else if last.person and treat = 0 then
do;
    Treat0 = y;
    Treat1 = Ly;
    output;
end;
else if last.person and treat = 1 then
do;
    Treat1 = y;
    Treat0 = Ly;
    Output;
end;
run;

proc ttest data=PairSet11 alpha=0.05 h0=0 plots=none;
    title "Paired sample t-test" ;
    by s;
    paired Treat0*Treat1;
    ods output ConfLimits=OUTPUT.conflim11;
    ods output TTests=OUTPUT.pvalue11;
    ods output Statistics=OUTPUT.differences11;
run;

/*Proc mixed*/
Proc Mixed Data=MAR_UnequalDif4060_Sort;
by s;
Class treat period;
Model y = treat period treat*period;
random int/subject = person;
estimate 'Treat 1 vs Treat 2' treat 1 -1 / cl;

```

```
ods output Estimates=OUTPUT.Estimates11;
Run;
```

```
/******
******/
```

```
/*T-test results*/
```

```
proc means data=OUTPUT.differences1 mean n std nmiss;
title "T-Test Non-Missing Mean";
var Mean StdErr;
run;
```

```
proc means data=OUTPUT.pvalue1 n;
title "T-Test Non-Missing Number of p-values that are significant";
var Probt;
where Probt<.05;
run;
```

```
proc means data=OUTPUT.conflim1 n ;
title "T-Test NonMissing Coverage";
var LowerCLMean UpperCLMean;
where LowerCLMean <= -4 and UpperCLMean >= -4;
run;
```

```
proc means data=OUTPUT.differences2 mean n std var nmiss;
title "MCAR 30 Mean";
var Mean StdErr;
run;
```

```
proc means data=OUTPUT.pvalue2 n;
title "MCAR 30 Number of p-values that are significant";
var Probt;
where Probt<=.05;
run;
```

```
proc means data=OUTPUT.conflim2 n ;
title "MCAR 30 Coverage";
var LowerCLMean UpperCLMean;
where LowerCLMean <= -4 and UpperCLMean >= -4;
run;
```

```
proc means data=OUTPUT.differences3 mean n std var nmiss;
title "MCAR 50 Mean";
var Mean StdErr;
run;
```

```
proc means data=OUTPUT.pvalue3 n;
title "MCAR 50 Number of p-values that are significant";
var Probt;
where Probt<=.05;
```

```

run;

proc means data=OUTPUT.conflim3 n ;
title "MCAR 50 Coverage";
var LowerCLMean UpperCLMean;
where LowerCLMean <= -4 and UpperCLMean >= -4;
run;

proc means data=OUTPUT.differences4 mean n std var nmiss;
title "MAR Equal Same 30 Mean";
var Mean StdErr;
run;

proc means data=OUTPUT.pvalue4 n;
title "MAR Equal Same 30 Number of p-values that are significant";
var Probt;
where Probt<=.05;
run;

proc means data=OUTPUT.conflim4 n ;
title "MAR Equal Same 30 Coverage";
var LowerCLMean UpperCLMean;
where LowerCLMean <= -4 and UpperCLMean >= -4;
run;

proc means data=OUTPUT.differences5 mean n std var nmiss;
title "MAR Equal Dif 30 Mean";
var Mean StdErr;
run;

proc means data=OUTPUT.pvalue5 n;
title "MAR Equal Dif 30 Number of p-values that are significant";
var Probt;
where Probt<=.05;
run;

proc means data=OUTPUT.conflim5 n ;
title "MAR Equal Dif 30 Coverage";
var LowerCLMean UpperCLMean;
where LowerCLMean <= -4 and UpperCLMean >= -4;
run;

proc means data=OUTPUT.differences6 mean n std var nmiss;
title "MAR Unequal Same 2040 Mean";
var Mean StdErr;
run;

proc means data=OUTPUT.pvalue6 n;
title "MAR Unequal Same 2040 Number of p-values that are significant";
var Probt;

```

```
where Probt<=.05;
run;
```

```
proc means data=OUTPUT.conflim6 n ;
title "MAR Unequal Same 2040 Coverage";
var LowerCLMean UpperCLMean;
where LowerCLMean <= -4 and UpperCLMean >= -4;
run;
```

```
proc means data=OUTPUT.differences7 mean n std var nmiss;
title "MAR Unequal Dif 2040 Mean";
var Mean StdErr;
run;
```

```
proc means data=OUTPUT.pvalue7 n;
title "MAR Unequal Dif 2040 Number of p-values that are significant";
var Probt;
where Probt<=.05;
run;
```

```
proc means data=OUTPUT.conflim7 n ;
title "MAR Unequal Dif 2040 Coverage";
var LowerCLMean UpperCLMean;
where LowerCLMean <= -4 and UpperCLMean >= -4;
run;
```

```
proc means data=OUTPUT.differences8 mean n std var nmiss;
title "MAR Equal Same 50 Mean";
var Mean StdErr;
run;
```

```
proc means data=OUTPUT.pvalue8 n;
title "MAR Equal Same 50 Number of p-values that are significant";
var Probt;
where Probt<=.05;
run;
```

```
proc means data=OUTPUT.conflim8 n ;
title "MAR Equal Same 50 Coverage";
var LowerCLMean UpperCLMean;
where LowerCLMean <= -4 and UpperCLMean >= -4;
run;
```

```
proc means data=OUTPUT.differences9 mean n std var nmiss;
title "MAR Equal Dif 50 Mean";
var Mean StdErr;
run;
```

```
proc means data=OUTPUT.pvalue9 n;
title "MAR Equal Dif 50 Number of p-values that are significant";
```

```
var Probt;  
where Probt<=.05;  
run;
```

```
proc means data=OUTPUT.conflim9 n ;  
title "MAR Equal Dif 50 Coverage";  
var LowerCLMean UpperCLMean;  
where LowerCLMean <= -4 and UpperCLMean >= -4;  
run;
```

```
proc means data=OUTPUT.differences10 mean n std var nmiss;  
title "MAR Unequal Same 4060 Mean";  
var Mean StdErr;  
run;
```

```
proc means data=OUTPUT.pvalue10 n;  
title "MAR Unequal Same 4060 Number of p-values that are significant";  
var Probt;  
where Probt<=.05;  
run;
```

```
proc means data=OUTPUT.conflim10 n ;  
title "MAR Unequal Same 4060 Coverage";  
var LowerCLMean UpperCLMean;  
where LowerCLMean <= -4 and UpperCLMean >= -4;  
run;
```

```
proc means data=OUTPUT.differences11 mean n std var nmiss;  
title "MAR Unequal Dif 4060 Mean";  
var Mean StdErr;  
run;
```

```
proc means data=OUTPUT.pvalue11 n;  
title "MAR Unequal Dif 4060 Number of p-values that are significant";  
var Probt;  
where Probt<=.05;  
run;
```

```
proc means data=OUTPUT.conflim11 n ;  
title "MAR Unequal Dif 4060 Coverage";  
var LowerCLMean UpperCLMean;  
where LowerCLMean <= -4 and UpperCLMean >= -4;  
run;
```

```
/*Mixed Model results*/
```

```
proc means data=OUTPUT.Estimates1 mean n std var nmiss;  
title "Nonmissing Mixed Mean";  
var Estimate StdErr;  
run;
```

```
proc means data=OUTPUT.Estimates1 n;  
title "Nonmissing Mixed Number of p-values that are significant";  
var Probt;  
where Probt<.05;  
run;
```

```
proc means data=OUTPUT.Estimates1 n;  
title "Nonmissing Mixed Coverage";  
var Lower Upper;  
where Lower <= -4 and Upper >= -4;  
run;
```

```
proc means data=OUTPUT.Estimates2 mean n std var nmiss;  
title "MCAR 30 Mixed Mean";  
var Estimate StdErr;  
run;
```

```
proc means data=OUTPUT.Estimates2 n;  
title "MCAR 30 Mixed Number of p-values that are significant";  
var Probt;  
where Probt<.05;  
run;
```

```
proc means data=OUTPUT.Estimates2 n;  
title "MCAR 30 Mixed Coverage";  
var Lower Upper;  
where Lower <= -4 and Upper >= -4;  
run;
```

```
proc means data=OUTPUT.Estimates3 mean n std var nmiss;  
title "MCAR 50 Mixed Mean";  
var Estimate StdErr;  
run;
```

```
proc means data=OUTPUT.Estimates3 n;  
title "MCAR 50 Mixed Number of p-values that are significant";  
var Probt;  
where Probt<.05;  
run;
```

```
proc means data=OUTPUT.Estimates3 n;  
title "MCAR 50 Mixed Coverage";  
var Lower Upper;  
where Lower <= -4 and Upper >= -4;  
run;
```

```
proc means data=OUTPUT.Estimates4 mean n std var nmiss;  
title "MAR Equal Same 30 Mixed Mean";  
var Estimate StdErr;
```

```
run;
```

```
proc means data=OUTPUT.Estimates4 n;  
title "MAR Equal Same 30 Mixed Number of p-values that are significant";  
var Probt;  
where Probt<.05;  
run;
```

```
proc means data=OUTPUT.Estimates4 n;  
title "MAR Equal Same 30 Mixed Coverage";  
var Lower Upper;  
where Lower <= -4 and Upper >= -4;  
run;
```

```
proc means data=OUTPUT.Estimates5 mean n std var nmiss;  
title "MAR Equal Dif 30 Mixed Mean";  
var Estimate StdErr;  
run;
```

```
proc means data=OUTPUT.Estimates5 n;  
title "MAR Equal Dif 30 Mixed Number of p-values that are significant";  
var Probt;  
where Probt<.05;  
run;
```

```
proc means data=OUTPUT.Estimates5 n;  
title "MAR Equal Dif 30 Mixed Coverage";  
var Lower Upper;  
where Lower <= -4 and Upper >= -4;  
run;
```

```
proc means data=OUTPUT.Estimates6 mean n std var nmiss;  
title "MAR Unequal Same 2040 Mixed Mean";  
var Estimate StdErr;  
run;
```

```
proc means data=OUTPUT.Estimates6 n;  
title "MAR Unequal Same 2040 Mixed Number of p-values that are significant";  
var Probt;  
where Probt<.05;  
run;
```

```
proc means data=OUTPUT.Estimates6 n;  
title "MAR Unequal Same 2040 Mixed Coverage";  
var Lower Upper;  
where Lower <= -4 and Upper >= -4;  
run;
```

```
proc means data=OUTPUT.Estimates7 mean n std var nmiss;  
title "MAR Unequal Dif 2040 Mixed Mean";
```

```
var Estimate StdErr;  
run;
```

```
proc means data=OUTPUT.Estimates7 n;  
title "MAR Unequal Dif 2040 Mixed Number of p-values that are significant";  
var Probt;  
where Probt<.05;  
run;
```

```
proc means data=OUTPUT.Estimates7 n;  
title "MAR Unequal Dif 2040 Mixed Coverage";  
var Lower Upper;  
where Lower <= -4 and Upper >= -4;  
run;
```

```
proc means data=OUTPUT.Estimates8 mean n std var nmiss;  
title "MAR Equal Same 50 Mixed Mean";  
var Estimate StdErr;  
run;
```

```
proc means data=OUTPUT.Estimates8 n;  
title "MAR Equal Same 50 Mixed Number of p-values that are significant";  
var Probt;  
where Probt<.05;  
run;
```

```
proc means data=OUTPUT.Estimates8 n;  
title "MAR Equal Same 50 Mixed Coverage";  
var Lower Upper;  
where Lower <= -4 and Upper >= -4;  
run;
```

```
proc means data=OUTPUT.Estimates9 mean n std var nmiss;  
title "MAR Equal Dif 50 Mixed Mean";  
var Estimate StdErr;  
run;
```

```
proc means data=OUTPUT.Estimates9 n;  
title "MAR Equal Dif 50 Mixed Number of p-values that are significant";  
var Probt;  
where Probt<.05;  
run;
```

```
proc means data=OUTPUT.Estimates9 n;  
title "MAR Equal Dif 50 Mixed Coverage";  
var Lower Upper;  
where Lower <= -4 and Upper >= -4;  
run;
```

```
proc means data=OUTPUT.Estimates10 mean n std var nmiss;
```

```
title "MAR Unequal Same 4060 Mixed Mean";  
var Estimate StdErr;  
run;
```

```
proc means data=OUTPUT.Estimates10 n;  
title "MAR Unequal Same 4060 Mixed Number of p-values that are significant";  
var Probt;  
where Probt<.05;  
run;
```

```
proc means data=OUTPUT.Estimates10 n;  
title "MAR Unequal Same 4060 Mixed Coverage";  
var Lower Upper;  
where Lower <= -4 and Upper >= -4;  
run;
```

```
proc means data=OUTPUT.Estimates11 mean n std var nmiss;  
title "MAR Unequal Dif 4060 Mixed Mean";  
var Estimate StdErr;  
run;
```

```
proc means data=OUTPUT.Estimates11 n;  
title "MAR Unequal Dif 4060 Mixed Number of p-values that are significant";  
var Probt;  
where Probt<.05;  
run;
```

```
proc means data=OUTPUT.Estimates11 n;  
title "MAR Unequal Dif 4060 Mixed Coverage";  
var Lower Upper;  
where Lower <= -4 and Upper >= -4;  
run;
```

References

- [1] E. J. Mills, A.-W. Chan, P. Wu, A. Vail, G. H. Guyatt, and D. G. Altman, “Design, analysis, and presentation of crossover trials,” *Trials*, vol. 10, no. 1, p. 27, 2009, doi: 10.1186/1745-6215-10-27.
- [2] S. Wellek and M. Blettner, “On the proper use of the crossover design in clinical trials: part 18 of a series on evaluation of scientific publications,” *Dtsch Arztebl Int*, vol. 109, no. 15, pp. 276–281, 2012, doi: 10.3238/arztebl.2012.0276.
- [3] D. R. Elbourne, D. G. Altman, J. P. T. Higgins, F. Curtin, H. V Worthington, and A. Vail, “Meta-analyses involving cross-over trials: methodological issues,” *Int J Epidemiol*, vol. 31, no. 1, pp. 140–149, 2002, doi: 10.1093/ije/31.1.140.
- [4] M. L. Bell, M. Fiero, N. J. Horton, and C.-H. Hsu, “Handling missing data in RCTs,” *BMC Med Res Methodol*, vol. 14, no. 1, p. 118, 2014, doi: 10.1186/1471-2288-14-118.
- [5] B. A. Rabe, S. Day, M. H. Fiero, and M. L. Bell, “Missing data handling in non-inferiority and equivalence trials: A systematic review,” *Pharm Stat*, vol. 17, no. 5, pp. 477–488, 2018, doi: 10.1002/pst.1867.
- [6] D. B. RUBIN, “Inference and missing data,” *Biometrika*, vol. 63, no. 3, pp. 581–592, 1976, doi: 10.1093/biomet/63.3.581.
- [7] G. K. Rosenkranz, “Analysis of cross-over studies with missing data,” *Stat Methods Med Res*, vol. 24, no. 4, pp. 420–433, 2015, doi: 10.1177/0962280214521349.
- [8] P. Allison, “Handling Missing Data by Maximum Likelihood,” 2012.
- [9] M. L. Bell, M. G. Kenward, D. L. Fairclough, and N. J. Horton, “Differential dropout and bias in randomised controlled trials: when it matters and when it may not,” *BMJ*, vol. 346, no. jan21 1, pp. e8668–e8668, 2013, doi: 10.1136/bmj.e8668.
- [10] E. L. Ashbeck and M. L. Bell, “Single time point comparisons in longitudinal randomized controlled trials: power and bias in the presence of missing data,” *BMC Med Res Methodol*, vol. 16, no. 1, p. 43,

2016, doi: 10.1186/s12874-016-0144-0.

- [11] S. J. Senn, P. Auclair, and S. Johnson, “The graphical representation of clinical trials with particular reference to measurements over time,” *Stat Med*, vol. 9, no. 11, pp. 1287–1302, 1990, doi: 10.1002/sim.4780091108.
- [12] J. Cohen, *Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, N.J.: Hillsdale, N.J. : L. Erlbaum Associates, 1988.
- [13] T. P. Morris, I. R. White, and M. J. Crowther, “Using simulation studies to evaluate statistical methods,” *Stat Med*, vol. 38, no. 11, pp. 2074–2102, 2019, doi: 10.1002/sim.8086.