12-2021

A Practical Extension to the AB/BA Design

My T.A Nguyen
nguyenmt3@jacks.sfasu.edu

A Practical Extension to the AB/BA Design

Creative Commons License

GETTING THE MOST FROM YOUR CROSSOVER STUDIES, A PRACTICAL EXTENSION TO THE AB/BA DESIGN

by

My (Amy) Nguyen, B.S.

Presented to the Faculty of the Graduate School of

Stephen F. Austin State University

In Partial Fulfillment

of the Requirements

For the Degree of

Master of Science

STEPHEN F. AUSTIN STATE UNIVERSITY

December 2021

GETTING THE MOST FROM YOUR CROSSOVER STUDIES, A PRACTICAL EXTENSION TO THE AB/BA DESIGN

My (Amy) Nguyen, B.S.

APPROVED:

Jacob Turner, Ph.D., Thesis Director

Robert Henderson, Ph.D., Committee Member

Kent Riggs, Ph.D., Committee Member

Jeremy Becnel, Ph.D., Committee Member

Freddie Avant, Ph.D.
Interim Dean of Research and Graduate Studies

ABSTRACT

In this work, we take a close look at a general extension to the traditional AB/BA crossover design that is commonly used in clinical trials to determine the effectiveness of new candidate drugs. While the traditional crossover design requires each patient in the study to be measured on both treatment A and treatment B, we consider the possibility of additional measurements being available on each patient. This produces designs such as the AABB/BBAA design which has been used in previous studies. A general test statistic will be derived to test for treatment effects as well as its corresponding power function to aid in sample size determination to aid statistical planning. Lastly, we explore the theoretical power of our testing procedure and compare it to simulated power studies to verify how well sample size determinations will work in practice.

ACKNOWLEDGEMENTS

I owe special thanks to several people who made it possible for me to complete this thesis. My family: my parents and my brother were understanding and supportive throughout the years. Michael convinced me to go back to school and helped me identify what I wanted to do with my life. My best friend, Dung Nguyen, had faith in me. Konnor McLeod motivated me to be a successful, independent woman and to be a better person everyday. Especially, Dr. Jacob Turner enthusiastically guided me the thesis progress. And lastly, I wanted to say thank you to myself, for trying to be the best version of myself.

CONTENTS

LIST OF FIGURES

1 INTRODUCTION

Crossover designs have been widely used in clinical trials to determine the effectiveness of a drug. The most common crossover design is the AB/BA crossover. An AB/BA crossover design allows for each patient in the study to be given both treatment under consideration for comparison. The treatments are treatment A and treatment B. In many cases, treatment A is a proposed treatment of interest while treatment B is a placebo with the goal of determining if a difference between the two treatments exist.

Figure 1.1 illustrates the AB/BA crossover design. Patients are randomly assigned to two sequences: the AB sequence or the BA sequence. In the AB sequence, patients are administered treatment A first and a measurement is taken once the drug has had time to take effect. After some time to allow the drug to leave the person's system referred to as the "wash-out" period, the patient then "crosses-over" and is administered treatment B and a second measurement is recorded.

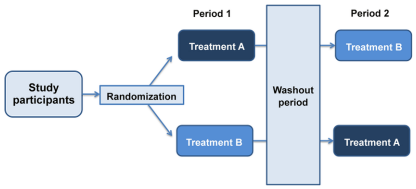

Figure 1.1: Crossover Design AB/BA [5]

A similar description holds for the patients in the BA sequence but the treatment order is reversed. In the BA sequence, patients are administered treatment B first and a measurement is taken once the drug has had time to take effect. After the "wash-out" period, the patient then "crosses over" and is administered treatment A

and a second measurement is recorded.

Taking measurements over time from the sequences are referred to as periods. When a patient is taking their first treatment, that is referred to as the first period. For example, if a patient is in the AB sequence, then they will receive treatment A during the first period and treatment B during the second period. In contrast, if a patient is in the BA sequence, then they will receive treatment B during the first period and treatment A during the second period.

The main benefit of the crossover design is that each patient serves as their own control effectively allowing for a paired analysis. If the outcomes for the pairs are positively correlated, then the tests will be more statistically powerful than if a simple randomized treatment/control study were used. However, the design also comes with its own considerations and pitfalls. For example, when patients crossover, a certain time must pass between the periods for the treatment to leave the patient's system before the next treatment is applied. If researchers rush into a second treatment without allocating enough time for the "wash-out", this is commonly referred to as the carryover effect and can result in a biased result of the treatment effect when performing a hypothesis test.

Cross-over designs have received numerous extensions in order to handle more complicated research models and designs. The extensions include various combinations of allowing for more than two treatment groups to be investigated, adding additional sequences to better handle potential carryover effects, and adding additional periods. Each extension has their own pros and cons. A less commonly discussed extension of AB/BA crossover design is AABB/BBAA crossover. Figure 1.2 demonstrates an example of an AABB/BBAA crossover study.

Following along with Figure 1.2, group 1 represents the AABB sequence where patients are given ReVeRe.D first (treatment A) followed by Examiner (treatment B). Group 2 represents the BBAA sequence where patients are given Examiner (treatment

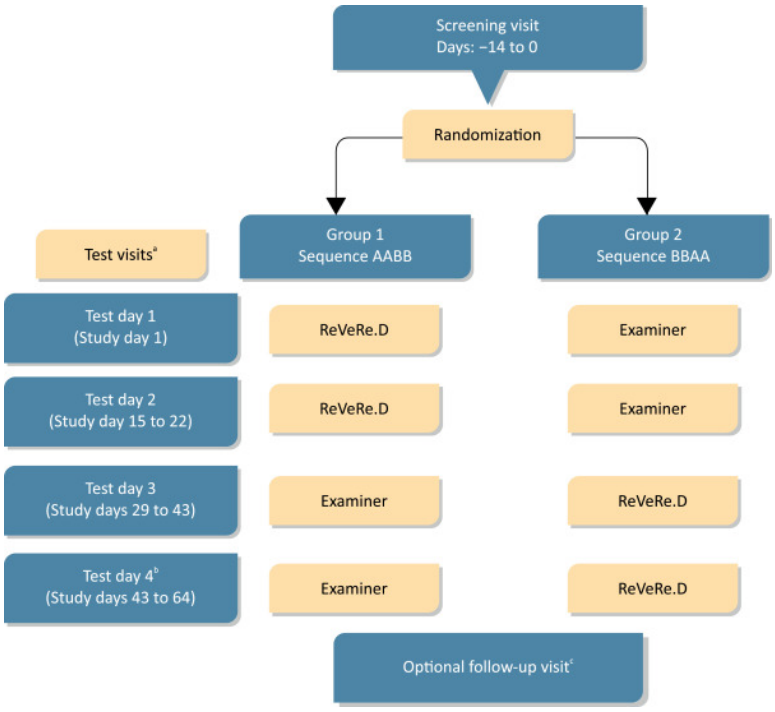

Figure 1.2: Crossover Design AABB/BBAA [2]

B) first then followed by ReVeRe.D first (treatment A). The test days represent the periods in which data is collected from the patients under their corresponding treatment. This highlights they key difference of the extended design from the classic AB/BA design. Under the AABB/BBAA design, multiple measurements are recorded from the patients under each treatment. While it is not noted in Figure 1.2, it is assumed that an appropriate wash out period is conducted between test day 2 and test day 3 (periods 2 and 3). [2]

1.1 Hypothesis Tests in AB/BA Designs

First, we summarize the hypothesis test of the classic traditional AB/BA design. The traditional cross over design has numerous ways to express the model and its

effects. Under the most common parameterization, the cross over design is simply a Latin square design. For the purpose of this thesis, we will introduce the model through a multivariate lens motivated by a discussion by Senn [8]. In summary, the AB/BA design results in collecting data under four conditions which correspond to what sequence, period, and treatment the sample were collected from. Figure 1.3 provides a cell means table, considering the sample averages of the four conditions along with their expected values as discussed by Senn [8].

Sequence	**Period 1**	**Period 2**
AB	$\mu + \tau_A + \pi_1$ $\overline{Y_{11}}$	$\mu + \tau_B + \pi_2 + \lambda_{AB}$ $\overline{Y_{12}}$
BA	$\mu + \tau_B + \pi_1$ $\overline{Y_{21}}$	$\mu + \tau_A + \pi_2 + \lambda_{BA}$ $\overline{Y_{22}}$

Figure 1.3: Cells Means and Expectations for an AB/BA Cross-over

The parameters defined for the expected values in Figure 1.3 are defined as follows. The treatment effect is $\tau_A - \tau_B$, the period effect is $\pi_1 - \pi_2$, and the carry-over effect is $\lambda_A - \lambda_B$. The sample mean of the first period of sequence AB is denoted $\bar{Y}_{11}$ whose expectation is $\mu + \tau_A + \pi_1$. The sample mean of the second period of sequence AB is denoted $\bar{Y}_{12}$ with expectation $\mu + \tau_B + \pi_2 + \lambda_{AB}$. Also, in the second sequence BA, the sample mean of the first period is denoted $\bar{Y}_{21}$ whose expectation is $\mu + \tau_B + \pi_1$. The final sample mean, $\bar{Y}_{22}$, is with respect to second period of sequence BA and has expectation $\mu + \tau_A + \pi_2 + \lambda_{BA}$.

Since the four averages are random variables and are correlated since multiple measurements are taken on each subject, we can use multivariate distribution theory to organize all the information about the averages. The four averages can be written as a random vector denoted below as:

$$\bar{\boldsymbol{Y}} = \begin{pmatrix} \bar{Y}_{11} \\ \bar{Y}_{12} \\ \bar{Y}_{21} \\ \bar{Y}_{22} \end{pmatrix} \tag{1.1}$$

with expected value

$$E(\bar{\boldsymbol{Y}}) = \begin{pmatrix} \mu + \tau_A + \pi_1 \\ \mu + \tau_B + \pi_2 + \lambda_{AB} \\ \mu + \tau_B + \pi_1 \\ \mu + \tau_A + \pi_2 + \lambda_{BA} \end{pmatrix}. \tag{1.2}$$

Following Senn's notation, the variance-covariance matrix for the random vector, $\bar{\boldsymbol{Y}}$ is expressed as

$$Var(\bar{\boldsymbol{Y}}) = \boldsymbol{\Sigma} = \begin{pmatrix} \frac{\sigma^2}{n_1} & \rho\frac{\sigma^2}{n_1} & 0 & 0 \\ \rho\frac{\sigma^2}{n_1} & \frac{\sigma^2}{n_1} & 0 & 0 \\ 0 & 0 & \frac{\sigma^2}{n_2} & \rho\frac{\sigma^2}{n_2} \\ 0 & 0 & \rho\frac{\sigma^2}{n_2} & \frac{\sigma^2}{n_2} \end{pmatrix} \tag{1.3}$$

where σ^2 is the variance, assumed constant for each observation across all four conditions. The parameter ρ is the correlation between the subjects and only exists between averages that are taken on the same subjects. The number of each subjects in the first (AB) and second (BA) sequence are n_1 and n_2, respectively.

To perform a test on the treatment effect, $\tau_A - \tau_B$, there are two possible tests to consider. These tests are constructed using specific contrasts, or linear combinations of the cell means $\bar{Y}$. The first contrast is referred to as "par" and only uses sample means that occur prior to the cross over. Let the contrast be a four by one row vector defined by $\boldsymbol{c}'_{par} = \frac{1}{2}(1, 0, -1, 0)$. The expected value of $\boldsymbol{c}'_{par}\bar{\boldsymbol{Y}}$ is $E(\boldsymbol{c}'_{par}\bar{\boldsymbol{Y}}) = \boldsymbol{c}'_{par}E(\bar{\boldsymbol{Y}}) = \tau_A - \tau_B$, and is thus and unbiased estimator for the treatment effect. The variance of $\boldsymbol{c}'_{par}\bar{\boldsymbol{Y}}$ can be obtained by computing $\boldsymbol{c}'_{par}\boldsymbol{\Sigma}\boldsymbol{c}_{par}$ which reduces to $(1/n_1 +$

$1/n_2)\sigma^2$. Assuming normality, the general test statistic can be constructed using the traditional z-score: $z = \frac{\boldsymbol{c}'_{par}\bar{\boldsymbol{Y}} - 0}{\sqrt{\boldsymbol{c}'_{par}\boldsymbol{\Sigma}\boldsymbol{c}_{par}}}$, under the null hypothesis of no treatment effect, the null distribution of Z is $N(0,1)$.

The second contrasts is referred to as "cros" and utilizes all four sample averages. Here the 4×1 contrast vector is defined as $\boldsymbol{c}'_{cros} = \frac{1}{2}(1,-1,-1,1)$. The expect value of $\boldsymbol{c}'_{cros}\bar{\boldsymbol{Y}}$ is $E(\boldsymbol{c}'_{cros}\bar{\boldsymbol{Y}}) = \boldsymbol{c}'_{cros}E(\bar{\boldsymbol{Y}}) = (\tau_A - \tau_B) - \frac{(\lambda_{AB} - \lambda_{BA})}{2}$. While on its face, the contrast for cros does not look particularly helpful since it is not an unbiased estimator for the treatment effect. However, if the researchers provide adequate amount of time for the treatment to leave the patients system before crossing over, then it is reasonable to assume that $\lambda_{AB} - \lambda_{BA} = 0$. If so, then the cros contrast is an unbiased estimator for the treatment effect. The variance of $\boldsymbol{c}'_{cros}\bar{\boldsymbol{Y}}$ is computed similarly to par and reduces to $\frac{\sigma^2(1-\rho)}{2}(1/n_1 + 1/n_2)$. Assuming normality again, the general test statistic can be constructed using the traditional z-score, $z = \frac{\boldsymbol{c}'_{cros}\bar{\boldsymbol{Y}} - 0}{\sqrt{\boldsymbol{c}'_{cros}\boldsymbol{\Sigma}\boldsymbol{c}_{cros}}}$.

The advantages of using the test involving cross is clear as long as the study ensures enough time for the drug to wash out. Comparing the variance for each of the tests, the variance of $\boldsymbol{c}_{cros}\bar{\boldsymbol{Y}}$ differs by a multiplication factor of $1 - \rho$. Since ρ is the correlation between repeated measurements on the same patient, it is expected that ρ will be positive, thus creating a smaller standard error which will produce more liberal statistical results in terms of smaller probability values and tighter confidence intervals for the treatment effect.

In practice, the parameters σ^2 and ρ will not be known in advance and must be estimated from data. Using consistent estimators for these parameters and plugging them into $\boldsymbol{\Sigma}$, denoted $\hat{\boldsymbol{\Sigma}}$, the test statistic is converted to a t-score: $t = \frac{\boldsymbol{c}'\bar{\boldsymbol{Y}} - 0}{\sqrt{\boldsymbol{c}'\hat{\boldsymbol{\Sigma}}\boldsymbol{c}}}$ where c is either the par or cros contrast. The estimates of σ^2 and ρ are typically obtained via restricted maximum likelihood (REML) obtained using a mixed model framework [4]. For the simple AB/BA design, the mixed model can be defined as

$$y_{ij} = \mu_j + \gamma_i + \epsilon_i$$

where i is the i^{th} patient in the study measured under the j^{th} condition $j = 1, 2, 3, 4$. The mean parameters μ_j are defined as they are in Figure 1.3. It should be noted that for each patient, only 2 of the 4 levels of μ_j will be estimated due to the cross over nature of the study. The γ_i's are subject specific random variables that are independent and assumed to be normally distributed with mean 0 and a subject specific variance, σ^2_{sub}. The errors, ϵ_i, are standard regression error terms that are independent and normally distributed with mean 0 with variance σ^2. Each ϵ_i are assumed to be independent from the γ_i's. Under the method of REML, estimates for mean parameters as well as the variance terms can be obtained . Under this particular model, the estimate of ρ is $\hat{\rho} = \frac{\hat{\sigma}^2_{sub}}{\hat{\sigma}^2_{sub}+\hat{\sigma}^2}$

1.2 Analysis example using a synthetic data set

In this section, an example of an analysis of an AB/BA cross-over design with a simulated data set is introduced. To simulate the data, we first specified the parameters. The common mean value $\mu = 100$. The treatment parameters were selected such that the difference between treatment A and treatment B, is $\tau_A - \tau_B = 10$. Similarly the period effect parameters were chosen such that $\pi_1 - \pi_2 = 3$. Assuming there is no carry-over effect, $\lambda_{AB} = \lambda_{BA} = 0$. The sample sizes for the two sequences AB and BA were both set to 10, $n_1 = n_2 = 10$. Lastly, we specified the variability due to measurement error ($\sigma_e = 7$) and due to subject error ($\sigma_{sub} = 7$) so that the correlation between subject is $\rho = \frac{7}{7+7} = 0.5$ and total variation is $\sigma = \sigma_e + \sigma_{sub} = 14$. With the parameters initialized, the data from two patients, 1 from each sequence, is simulated using a multivariate normal distribution with mean $E(\bar{\boldsymbol{Y}})$ as specified in Equation (1.2) and Variance-Covarinace Matrix defined as in Equation (1.3) but setting the sample sizes equal to 1 to reflect single observations rather than averages,

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & \rho\sigma^2 & 0 & 0 \\ \rho\sigma^2 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & \rho\sigma^2 \\ 0 & 0 & \rho\sigma^2 & \sigma^2 \end{pmatrix}.$$

Upon simulating the observations, we formatted the data so that it could be appropriately fitted by a linear model in the R statistical software. The first few rows are depicted in Figure 1.4 for reference. There are five columns presented indicating observation information such as the sequence (Seq: AB or BA), the period (Per: 1 or 2), subject id (1 to 20), treatment group (trt: treatment 1 or treat 2), and the response values.

	Seq	Per	Sub	Trt	Response
1	AB	1	1	2	111.55201
2	AB	1	2	2	110.44147
3	AB	1	3	2	107.70335
4	AB	1	4	2	106.63096
5	AB	1	5	2	108.19420
6	AB	1	6	2	118.63936
7	AB	1	7	2	107.95278
8	AB	1	8	2	110.47111
9	AB	1	9	2	110.22180
10	AB	1	10	2	107.09992
11	AB	2	1	1	102.84092
12	AB	2	2	1	99.59513
13	AB	2	3	1	103.01967
14	AB	2	4	1	100.84574
15	AB	2	5	1	99.46509

Figure 1.4: A Part of Simulated Data for Cross-over Design AB/BA

Figure 1.5 provides a plot of the full simulated data set. The x-axis presents the

treatments such as treat 1 or treatment 2. The y-axis presents the responses. The green lines illustrate all the patients that are on sequence AB. On the other hand, the red lines express all the patient on sequence BA. Each line represents each patient in the experimental test design. We can see that there is a mild correlation between all the pairs which perfectly fits since ρ is specified to be 0.5.

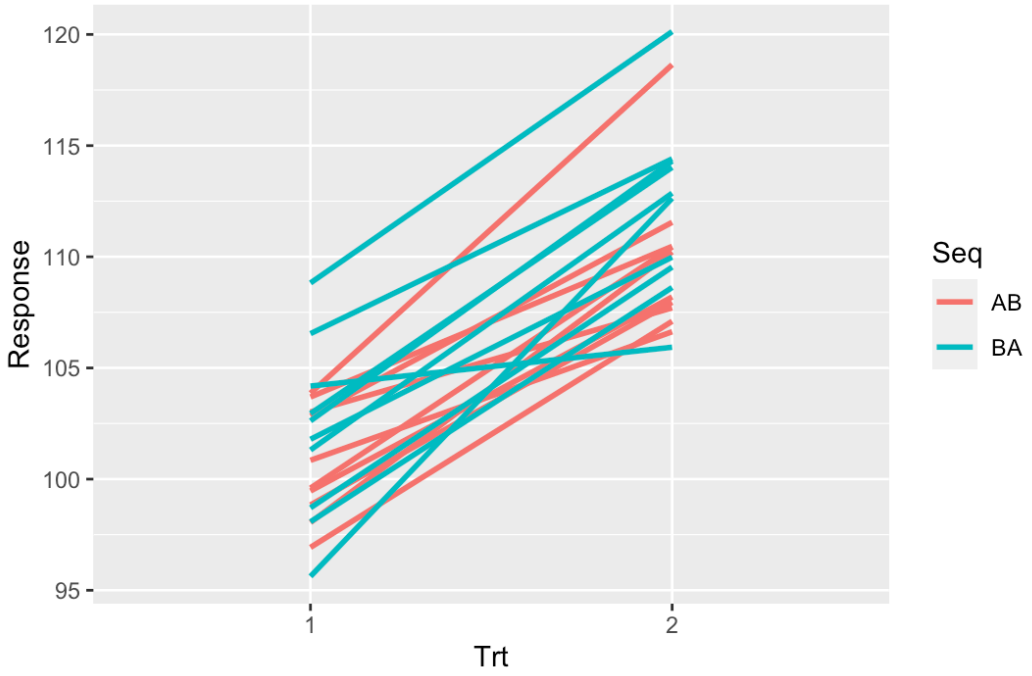

Figure 1.5: Graph of The AB/BA Cross-over Design

Upon fitting a linear mixed model to the data using the nlme package in R statistical software, the estimated cell means and their standard errors are presented in Table 1.1.

Table 1.1: Mean Result of the AB/BA Cross-over Design

Groups	Mean Value	Std. Error	df
Group Y_{11}	109.09266	1.041476	17
Group Y_{12}	101.35146	1.041476	17
Group Y_{21}	97.93285	1.041476	17
Group Y_{22}	113.55337	1.041476	17

The estimated covariance matrix of $\bar{\boldsymbol{Y}}$ of the AB/BA cross-over design from the simulated data is:

$$\hat{\Sigma} = \begin{pmatrix} 1.0847 & 0.702 & 0 & 0 \\ 0.702 & 1.0847 & 0 & 0 \\ 0 & 0 & 1.0847 & 0.702 \\ 0 & 0 & 0.702 & 1.087 \end{pmatrix}.$$

With the estimates $\bar{\boldsymbol{Y}}$ and $\hat{\boldsymbol{\Sigma}}$ obtained, we computed the t-statistic using the linear contrasts of both par and cros. The estimated treatment effect using the par contrast was 11.16 with a t-statistic equal to 7.577. The resulting two sided p-value is less than 0.0001. The estimated treatment effect using cros is 11.681. The t-statistic is equal to 18.876 and yielded a two sided p-value less then 0.0001. The t-statistics of both cros and par contrast indicates that there is significant evidence of a possible treatment effect. As we discussed in Section 1.1, t-ratios utilizing the cros contrast will typically be larger than that of par contrast because of the smaller variance of cros when positive correlation exists.

Table 1.2: Result of the Sample AB/BA Cross-over Design

Contrast	Estimate	SE	df	t-statistic	p-value
Par	11.16	1.472	17	7.577	< 0.0001
Cros	11.681	0.619	17	18.876	< 0.0001

1.3 Investigating Extended Cross-Over Designs.

There are two main goals of this thesis. The first goal is to derive a general test statistic for the extended cross-over design where the number of periods increases by an even number as described earlier in this chapter. Upon doing so, we will provide a closed form expression for the test statistic to gain insight to the relationship between whether or not it would be better to increase the total sample sizes of a study by either increasing the number of patients or the number of periods. It seems reasonable to assume that increasing the number of patients would be more valuable as each new subject recruited brings independent information to help estimate the treatment effect. However, for some patient populations, it can be quite costly to recruit patients yet quite easy to take additional measurements by increasing the number of periods. Understanding how these design parameters impact study design is of key importance. This knowledge will allow for better design choices under various real world constraints.

The final goal is to perform simulation studies to determine if our general statistic can be used to obtain reliable sample size determinations to achieve a specified statistical power. The simulations verifies that our theoretical results are indeed correct and also investigate the validity of computing sample size determinations assuming σ and ρ are known, when the actual study will be estimating the parameters from data. If any discrepancies are found, these scenarios will be noted and alternative recommendations will be made.

2 Methods

In this chapter we will generalize the test for treatment effect for a general cross-over design when additional periods are included in the study design. In the first section, we will introduce mathematical notation to appropriately describe the general framework for constructing the contrast. The next section will provide technical details for a general test of the treatment effect. Lastly, we will introduce the notion of statistical power, describe its usefulness in study design, and give insight how the generalized test can impact study design decisions.

2.1 Introductory Notation

To extend the AB/BA crossover design to allow for multiple repeated measures, we will use a different notation when describing the cell means of the cross over design introduced in Chapter 1. Recall the Cell Means table, originally discussed in Chapter 1 provided in Figure 1.3. The subscripts on $\bar{\boldsymbol{Y}}_{\boldsymbol{ij}}$ indicate the average of the data observed from the patients on the i^{th} sequence and j^{th} period.

Sequence	Period 1	Period 2
AB	$\mu + \tau_A + \pi_1$ $\overline{Y_{11}}$	$\mu + \tau_B + \pi_2 + \lambda_{AB}$ $\overline{Y_{12}}$
BA	$\mu + \tau_B + \pi_1$ $\overline{Y_{21}}$	$\mu + \tau_A + \pi_2 + \lambda_{BA}$ $\overline{Y_{22}}$

Figure 2.1: Cells Means and Expectations for an AB/BA Cross-over

Following the notation of Chapter 1, the random vector of averages provided in Equation (1.1) is organized such that the first sequence averages are listed first

followed by the averages from the second sequence. Rather than double scripting the averages, we will simply index the averages from one to four with the understanding that the first two averages correspond to the first sequence and in ascending order,

$$\bar{\boldsymbol{Y}} = \begin{pmatrix} \bar{Y}_{11} \\ \bar{Y}_{12} \\ \bar{Y}_{21} \\ \bar{Y}_{22} \end{pmatrix} = \begin{pmatrix} \bar{Y}_{1} \\ \bar{Y}_{2} \\ \bar{Y}_{3} \\ \bar{Y}_{4} \end{pmatrix}.$$

Under this notation, we can generalize the cell means vector to accommodate additional cases where the number of periods, l, increases by multiples of two. For example, when $l = 2$, we have the standard AB/BA design. When $l = 4$, we have the AABB/BBAA design and the vector of averages is written as

$$\bar{\boldsymbol{Y}} = \begin{pmatrix} \bar{Y}_{1} \\ \bar{Y}_{2} \\ \bar{Y}_{3} \\ \bar{Y}_{4} \\ \bar{Y}_{5} \\ \bar{Y}_{6} \\ \bar{Y}_{7} \\ \bar{Y}_{8} \end{pmatrix}.$$

We partitioned the vector into two smaller vectors to highlight the fact that the first 4 averages correspond to first sequence (AABB), while the remaining 4 correspond to the second sequence (BBAA). In general, for a design with l periods, the cell mean vector is written as

$$\bar{\boldsymbol{Y}} = \begin{pmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \vdots \\ \bar{Y}_l \\ \bar{Y}_{l+1} \\ \bar{Y}_{l+2} \\ \vdots \\ \bar{Y}_{2l} \end{pmatrix}.$$

The expected value and variance-covariance matrix for $\bar{\boldsymbol{Y}}$ can also be generalized. For the expected value, the cell means can be expressed by adding column vectors of parameters. Let $\boldsymbol{\mu_l}$ be a column vector of length l containing the scalar μ for each entry. Let $\boldsymbol{\tau_{ABl}}$ be a vector of treatment effects whose first $l/2$ entries are the τ_A and the remaining entries are τ_B. $\boldsymbol{\tau_{BAl}}$ is defined similarly but the first $l/2$ entries are τ_B and the remaining entries are τ_A. The period effects can be expressed as $\boldsymbol{\pi_l} = (\pi_1, \pi_2, ..., \pi_l)$. The carry over effects can be expressed $\boldsymbol{\lambda_{ABl}}$ and $\boldsymbol{\lambda_{BAl}}$, whose first $l/2$ entries are all 0 and remaining entries are λ_{AB} and λ_{BA} respectively. With this frame work, the expected value of $\bar{\boldsymbol{Y}}$ can be expressed as

$$\boldsymbol{E}(\bar{\boldsymbol{Y}}) = \begin{bmatrix} \boldsymbol{\mu_l} \\ \boldsymbol{\mu_l} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\tau_{ABl}} \\ \boldsymbol{\tau_{BAl}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\pi_l} \\ \boldsymbol{\pi_l} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\lambda_{ABl}} \\ \boldsymbol{\lambda_{BAl}} \end{bmatrix}. \tag{2.1}$$

When setting $l = 2$, it is easily verified that the resulting expectations for the four averages reduces to the AB/BA cell means provided by Senn [8] and reported in Figure 1.3 and Equation (1.2). We recognize the notation in the general case is somewhat cumbersome. The difficulty arises due to the cross over nature of the design and the various effects due to treatment and carry-over depend on the period and the sequence. Writing the effects using vectors of length l will prove useful in upcoming sections.

A generalization of the variance-covariance matrix can be expressed upon examination of the simple 2 period design. Recall that the variance-covariance matrix for the AB/BA design is expressed as:

$$Var(\bar{\boldsymbol{Y}}) = \Sigma = \begin{pmatrix} \frac{\sigma^2}{n_1} & \rho\frac{\sigma^2}{n_1} & 0 & 0 \\ \rho\frac{\sigma^2}{n_1} & \frac{\sigma^2}{n_1} & 0 & 0 \\ 0 & 0 & \frac{\sigma^2}{n_2} & \rho\frac{\sigma^2}{n_2} \\ 0 & 0 & \rho\frac{\sigma^2}{n_2} & \frac{\sigma^2}{n_2} \end{pmatrix}.$$

Upon examination, there are four 2×2 block matrices. The off diagonal blocks are simply a matrix of 0's and reflects the fact that the first 2 averages (Sequence AB) are computed from patients that are independent from the last 2 averages (Sequence BA). The diagonal blocks are essentially identical except for the fact that the sample sizes, n_1 and n_2, could be different. When extending to additional periods, l, the four blocks will expand with the same structure but will each have dimension $l \times l$.

Let $\boldsymbol{V_k}$ be a $k \times k$ matrix whose diagonal elements are 1 and off diagonal elements ρ, and denote $\mathbf{0_k}$ a $k \times k$ matrix of 0's. The variance-covariance matrix for $\bar{\boldsymbol{Y}}$ for a crossover design with l periods can be expressed as:

$$Var(\bar{\boldsymbol{Y}}) = \sigma^2 \left[\begin{array}{c|c} \frac{1}{n_1}\boldsymbol{V_l} & \mathbf{0_l} \\ \hline \mathbf{0_l} & \frac{1}{n_2}\boldsymbol{V_l} \end{array} \right]. \tag{2.2}$$

As with the expected value of $\bar{\boldsymbol{Y}}$, setting $l = 2$ and distributing σ^2, n_1, and n_2 back into $\boldsymbol{V_2}$ yields the variance-covariance matrix for the AB/BA design.

2.1.1 Additional Notation and Discussion

In the next section we will provide a generalized version of the "cros" contrasts discussed in Chapter 1, and derive the mean and variance of the estimator for constructing a test for the treatment effect. In doing so, we will utilize some additional vectors and matrices.

First for any even integer k, let $\mathbf{1_k}$ be a column vector of 1's with length k and let $\boldsymbol{d_k}$ be a column vector of length k whose first $k/2$ elements are 1 and the remaining $k/2$ elements are -1. Thus, $\boldsymbol{d_k}$ can be written using two blocks,

$$\boldsymbol{d_k'} = [\mathbf{1_{k/2}'} - \mathbf{1_{k/2}'}]. \tag{2.3}$$

Additionally let, $\boldsymbol{W_k}$ be a $k \times k$ matrix with all entries equal to ρ. Note that $\boldsymbol{V_k}$ can be partitioned into four $k/2 \times k/2$ blocks using itself and $\boldsymbol{W_k}$.

$$\boldsymbol{V_k} = \begin{bmatrix} \boldsymbol{V_{k/2}} & \boldsymbol{W_{k/2}} \\ \boldsymbol{W_{k/2}} & \boldsymbol{V_{k/2}} \end{bmatrix}. \tag{2.4}$$

2.2 A General Test for Treatment Effect

In this section we propose a contrast that generalizes the cros contrast of the AB/BA design. We will show that this new contrast is an unbiased estimator for a treatment effect if there is no drug carry over and will derive its variance. These results can then be combined to derive a general test statistic in which p-values and confidence intervals can be generated.

Recall that for AB/BA design the preferred contrasts for testing a treatment effect is the cros contrast, $\boldsymbol{c'_{cros}} = \frac{1}{2}(1, -1, -1, 1)$. Here the two averages obtained from observations receiving treatment A are contrasted against the two averages obtained from observations receiving treatment B. When one considers the four period design, $l = 4$ (AABB/BBAA), it seems reasonable to denote the cros contrast, as $\boldsymbol{c'_{cros}} = \frac{1}{4}(1, 1, -1, -1, -1, -1, 1, 1)$, again averaging across the estimates associated with treatment A and contrasting it against averaging across the estimates from treatment B. For the general case of any even period design, l, denote the general contrast as:

$$\boldsymbol{c'} = \frac{1}{l}\begin{bmatrix} \mathbf{1_{l/2}'} & -\mathbf{1_{l/2}'} & -\mathbf{1_{l/2}'} & \mathbf{1_{l/2}'} \end{bmatrix}. \tag{2.5}$$

Note here that we've dropped the subscript as we will not be considering any other contrasts throughout the rest of this manuscript. It turns out that this contrast provides an unbiased estimator for the treatment effect if no carry over effect is present. Because of this, a test statistic can be derived, similar to the AB/BA design, as long as the variance of the estimator can be estimated. These results will be proven in the following Theorem, but first we present a corollary.

Corollary 1. Let $\boldsymbol{d_k}$ be a k by one vector (k is even) whose elements are defined by Equation (2.3) and $\boldsymbol{V_k}$ is a k x k symmetric matrix as defined by Equation (2.4). Then

$$\boldsymbol{d_k'V_kd_k} = k(1-\rho).$$

Proof. Expanding $\boldsymbol{d_k'V_kd_k}$ into block matrix form using Equations (2.3) and (2.4), we have:

$$\boldsymbol{d_k'V_kd_k} = \begin{bmatrix} \mathbf{1'_{k/2}} & \mathbf{-1'_{k/2}} \end{bmatrix} \begin{bmatrix} \boldsymbol{V_{k/2}} & \boldsymbol{W_{k/2}} \\ \boldsymbol{W_{k/2}} & \boldsymbol{V_{k/2}} \end{bmatrix} \begin{bmatrix} \mathbf{1'_{k/2}} \\ \mathbf{-1'_{k/2}} \end{bmatrix}.$$

After performing block matrix multiplication, we have:

$$\begin{aligned} \boldsymbol{d_k'V_kd_k} &= \mathbf{1'_{k/2}V_{k/2}1_{k/2}} - \mathbf{1'_{k/2}W_{k/2}1_{k/2}} - \mathbf{1'_{k/2}W_{k/2}1_{k/2}} + \mathbf{1'_{k/2}V_{k/2}1_{k/2}} \\ &= 2(\mathbf{1'_{k/2}V_{k/2}1_{k/2}} - \mathbf{1'_{k/2}W_{k/2}1_{k/2}}) \\ &= 2\mathbf{1'_{k/2}}(\boldsymbol{V_{k/2}} - \boldsymbol{W_{k/2}})\mathbf{1_{k/2}}. \end{aligned}$$

The matrix $\boldsymbol{(V_{k/2} - W_{k/2})}$ is sandwiched between two vectors of one's, and therefore simplifies to just the sum of the elements of $\boldsymbol{(V_{k/2} - W_{k/2})}$. Note that $\boldsymbol{(V_{k/2} - W_{k/2})} = (1-\rho)\boldsymbol{I_{k/2}}$. The sum of the elements in the matrix is simply adding up $(1-\rho)$ $k/2$ times. After substitution of this simplification, we have the result:

$$\boldsymbol{d_k'V_kd_k} = 2(k/2)(1-\rho) = k(1-\rho).$$

□

Theorem 2.1. *Let $\bar{\boldsymbol{Y}}$ denote the estimated cell averages obtained from a general crossover design with l even periods and let $\boldsymbol{c'}$ be the linear contrast defined in Equation* (2.5) *. The mean and variance of $\boldsymbol{c'}\bar{\boldsymbol{Y}}$ is:*

$$E(\boldsymbol{c'}\bar{\boldsymbol{Y}}) = (\tau_A - \tau_B) - \frac{\lambda_{AB} - \lambda_{BA}}{2}.$$
$$Var(\boldsymbol{c'}\bar{\boldsymbol{Y}}) = \frac{\sigma^2(1-\rho)}{l}(1/n_1 + 1/n_2).$$

Proof. We can express $\boldsymbol{c'}$ as a block vector, $\boldsymbol{c'} = \frac{1}{l}(\boldsymbol{d'_l}, -\boldsymbol{d'_l})$. Since expectation is a linear operator, we have:

$$E(\boldsymbol{c'}\bar{\boldsymbol{Y}}) = \boldsymbol{c'}E(\bar{\boldsymbol{Y}}) = \frac{1}{l}[\boldsymbol{d'_l} - \boldsymbol{d'_l}]E(\bar{\boldsymbol{Y}}).$$

Recall earlier in the chapter, $E(\bar{\boldsymbol{Y}})$ is expressed as:

$$E(\bar{\boldsymbol{Y}}) = \begin{bmatrix} \boldsymbol{\mu_l} \\ \boldsymbol{\mu_l} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\tau_{ABl}} \\ \boldsymbol{\tau_{BAl}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\pi_l} \\ \boldsymbol{\pi_l} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\lambda_{ABl}} \\ \boldsymbol{\lambda_{BAl}} \end{bmatrix}.$$

Therefore, the expectation can be reduced using block matrix multiplication:

$$\begin{aligned} \boldsymbol{c'}E(\bar{\boldsymbol{Y}}) &= \frac{1}{l}[\boldsymbol{d'_l} - \boldsymbol{d'_l}]\left(\begin{bmatrix} \boldsymbol{\mu_l} \\ \boldsymbol{\mu_l} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\tau_{ABl}} \\ \boldsymbol{\tau_{BAl}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\pi_l} \\ \boldsymbol{\pi_l} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\lambda_{ABl}} \\ \boldsymbol{\lambda_{BAl}} \end{bmatrix} \right) \\ &= \frac{1}{l}\Big(\boldsymbol{d'_l\mu_l} - \boldsymbol{d'_l\mu_l} + \boldsymbol{d'_l\tau_{ABl}} - \boldsymbol{d'_l\tau_{BAl}} + \boldsymbol{d'_l\pi_l} - \boldsymbol{d'_l\pi_l} + \boldsymbol{d'_l\lambda_{ABl}} - \boldsymbol{d'_l\lambda_{BAl}} \Big) \\ &= \frac{1}{l}\Big(\boldsymbol{d'_l\tau_{ABl}} - \boldsymbol{d'_l\tau_{BAl}} + \boldsymbol{d'_l\lambda_{ABl}} - \boldsymbol{d'_l\lambda_{BAl}} \Big). \end{aligned}$$

Recall that $\boldsymbol{\tau_{ABl}}$ is a vector of treatment effects whose first $l/2$ entries are τ_A and the remaining entries are τ_B. $\boldsymbol{\tau_{BAl}}$ is defined similarly but the first $l/2$ entries are τ_B and the remaining entries are τ_A. The period effects can be expressed as $\boldsymbol{\pi'_l} = (\pi_1, \pi_2, ..., \pi_l)$. The carry over effects can be expressed $\boldsymbol{\lambda_{ABl}}$ and $\boldsymbol{\lambda_{BAl}}$, whose first $l/2$ entries are all 0 and remaining entries are λ_{AB} and λ_{BA} respectively. The multiplication of $\boldsymbol{d'_l}$ to each of the components simplifies to adding up the first $l/2$

elements and subtracting it from the sum of the last $l/2$ elements. Applying the multiplication to each component yields $\boldsymbol{d_l'\tau_{ABl}} = \frac{l}{2}(\tau_A - \tau_B)$, $\boldsymbol{-d_l'\tau_{BAl}} = \frac{l}{2}(\tau_A - \tau_B)$, $\boldsymbol{d_l'\lambda_{ABl}} = -\frac{l}{2}(\lambda_{AB})$, and $\boldsymbol{-d_l'\lambda_{BAl}} = \frac{l}{2}(\lambda_{AB})$.

Replacing the scalar results for each component, we have the final result:

$$\begin{aligned} E(\boldsymbol{c'\bar{Y}}) &= \frac{1}{l}\left[\frac{l}{2}(\tau_A - \tau_B) + \frac{l}{2}(\tau_A - \tau_B) - \frac{l}{2}(\lambda_{AB}) + \frac{l}{2}(\lambda_{BA})\right] \\ &= (\tau_A - \tau_B) - \frac{\lambda_{AB} - \lambda_{BA}}{2}. \end{aligned}$$

To prove the variance, we take a similar approach.

$$\begin{aligned} Var(\boldsymbol{c'\bar{Y}}) &= \boldsymbol{c'Var(\bar{Y})c} \\ &= \frac{\sigma^2}{l^2}\left[\begin{array}{cc} \boldsymbol{d_l'} & \boldsymbol{-d_l'} \end{array}\right]\left[\begin{array}{c|c} \frac{1}{n_1}\boldsymbol{V_l} & \boldsymbol{0_{lxl}} \\ \hline \boldsymbol{0_{lxl}} & \frac{1}{n_2}\boldsymbol{V_l} \end{array}\right]\left[\begin{array}{c} \boldsymbol{d_l} \\ \boldsymbol{-d_l} \end{array}\right] \\ &= \frac{\sigma^2}{l^2}\left[\boldsymbol{d_l'}(\frac{1}{n_1}\boldsymbol{V_l})\boldsymbol{(d_l)} + \boldsymbol{d_l'}(\frac{1}{n_2}\boldsymbol{V_l})\boldsymbol{d_l}\right] \\ &= \frac{\sigma^2}{l^2}\left[\frac{1}{n_1}\boldsymbol{d_l'V_l d_l} + \frac{1}{n_2}\boldsymbol{d_l'V_l d_l}\right]. \end{aligned}$$

Using Corollary 1, $\boldsymbol{d_l'V_l d_l} = l(1-\rho)$, and the result follows.

$$\begin{aligned} Var(\boldsymbol{c'\bar{Y}}) &= \frac{\sigma^2}{l^2}\left[\frac{1}{n_1}\boldsymbol{d_l'V_l d_l} + \frac{1}{n_2}\boldsymbol{d_l'V_l d_l}\right] \\ &= \frac{\sigma^2}{l^2}\left[\frac{1}{n_1}l(1-\rho) + \frac{1}{n_2}l(1-\rho)\right] \\ &= \frac{\sigma^2}{l}(1-\rho)\left[\frac{1}{n_1} + \frac{1}{n_2}\right]. \end{aligned}$$

□

From the result of Theorem 2.1, the expected value of $E(\boldsymbol{c'\bar{Y}})$ is $(\tau_A - \tau_B) - \frac{\lambda_{AB} - \lambda_{BA}}{2}$. And the variance of $Var(\boldsymbol{c'\bar{Y}})$ can be obtained by computing $\frac{\sigma^2}{l}\ (1-\rho)\left[\frac{1}{n_1} + \frac{1}{n_2}\right]$. Assuming normality and no carry-over effect, a test statistic can be constructed using the traditional z-score, $Z = \frac{\boldsymbol{c'\bar{Y}} - (\tau_A - \tau_B)}{\sqrt{\frac{\sigma^2(1-\rho)}{l}(1/n_1 + 1/n_2)}}$. Under the null

hypothesis of no treatment effect: $H_0 : \tau_A = \tau_B$, the null distribution of Z is $N(0,1)$. In practice, the parameters σ^2 and ρ are unknown and must be estimated from data. Therefore, the test statistic is converted to a t-score: $t = \frac{\boldsymbol{c'}\bar{\boldsymbol{Y}} - (\tau_A - \tau_B)}{\sqrt{\frac{\hat{\sigma}^2(1-\hat{\rho})}{l}(1/n_1 + 1/n_2)}}$.

2.3 Study Design Implications

The result of the variance term, although simple, is not intuitive for the following reason. Consider the following two designs with periods $l = 2$ and $l = 4$. When $l = 2$, $\boldsymbol{Var(c'\bar{Y})} = \frac{\sigma^2}{2}(1-\rho)\left[\frac{1}{n_1} + \frac{1}{n_2}\right]$.

For the 4 period ($l = 4$), the variance becomes $\frac{\sigma^2}{4}(1-\rho)\left[\frac{1}{n_1} + \frac{1}{n_2}\right]$. However, we can rewrite the 4 period variance in terms of the variance of a two period design as follows:

$$\frac{\sigma^2}{2}(1-\rho)\frac{1}{2}\left[\frac{1}{n_1} + \frac{1}{n_2}\right] = \frac{\sigma^2}{2}(1-\rho)\left[\frac{1}{2n_1} + \frac{1}{2n_2}\right]$$

The result above reflects that there is a relationship between the number of period and sample size. By doubling the number of periods in the design from an AB/BA design to an AABB/BBAA design, the efficiency of $\boldsymbol{c'\bar{Y}}$ for the AABB/BBAA design is equivalent to a standard AB/BA design with double the amount of patients, suggesting that there is no difference between increasing patients versus increasing periods. This goes against conventional thought as collecting measurements on the same subject typically does not dampen the variance as rapidly as independent observations do. This result allows for researchers to reconsider their crossover design based on the difficulty of collecting measurements on patients (periods) as well as the difficulty of recruiting patients.

For example, suppose a researcher has determined that a sample size of 120 patients in each sequence is required to detect a practically meaningful treatment effect using a standard AB/BA design. However, the funding is such that recruiting the sample size required is not feasible. Based on our results, conducting an

AABB/BBAA design using just 60 patients or an AAABBB/BBBAAA using just 40 patients are all statistically equivalent with respect to the variance of the estimate for treatment effect.

Determining sample sizes as discussed in the previous example typically is done by examining the power of a hypothesis testing procedure. Statistical power is the probability of rejecting the null hypothesis when the alternative hypothesis is in fact true. That is, in the case of our scenario, the probability of detecting a true difference in means between the two treatment groups.

For a one sided test assuming no carry-over, with alternative $H_a : \tau_A > \tau_B$ and significance level $\alpha = 0.05$, the power obtained by a shift of $\tau_A - \tau_B = \Delta$ can be computed as:

$$\begin{aligned}
\beta(\Delta) &= P(Z > z_{1-\alpha} | \tau_A - \tau_B = \Delta) \\
&= P\left(\frac{\boldsymbol{c'}\bar{\boldsymbol{Y}}}{\sqrt{\frac{\sigma^2(1-\rho)}{l}(1/n_1 + 1/n_2)}} > z_{1-\alpha} | \tau_A - \tau_B = \Delta\right) \\
&= P\left(\frac{\boldsymbol{c'}\bar{\boldsymbol{Y}} - \Delta}{\sqrt{\frac{\sigma^2(1-\rho)}{l}(1/n_1 + 1/n_2)}} + \frac{\Delta}{\sqrt{\frac{\sigma^2(1-\rho)}{l}(1/n_1 + 1/n_2)}} > z_{1-\alpha} | \tau_A - \tau_B = \Delta\right) \\
&= P\left(Z > z_{1-\alpha} - \frac{\Delta}{\sqrt{\frac{\sigma^2(1-\rho)}{l}(1/n_1 + 1/n_2)}} | \tau_A - \tau_B = \Delta\right),
\end{aligned}$$

where $z_{1-\alpha}$ is the normal quantile. The notation $\beta(\Delta)$ is referred to as the power function and is typically referred to as a function of the effect size, $\tau_A - \tau_B = \Delta$ with the additional information such as the number of patients, σ^2, and ρ known in advance. Using the power function as presented and assuming one has a good idea about what the number of patients, σ^2, and ρ actually are, investigators can compute the power of numerous scenarios, changing sample sizes and changing Δ to determine a good design with a high chance of detecting the difference sought.

While the theoretical result allows for simple power calculations for planning new experiments and determining appropriate sample sizes, it is important to study the performance of the theoretical results by way of simulation since the theoretical power is obtained assuming σ^2 and ρ are known. It is important to verify that when performing the t-test using estimates of σ^2 and ρ, the statistical power of the test is the same or very close to the theoretical power obtained using the power function which assumes they are known. The statistical power of the t-test we define to be as empirical power since we have to verify its performance on simulated data sets to allow for σ^2 and ρ to be estimated.

Additionally, while theoretically there is no difference in statistical power between a design with 20 patients in each sequence with 2 periods versus 10 patients using 4 periods, the empirical power of the t-test procedures might differ due to how well the estimators of σ^2 and ρ are behaving. We suspect that adding periods will help estimate ρ while adding sample sizes will help estimate σ^2 better and this could intern create a difference in the empirical power while theoretically there is no difference. The next chapter investigates these concerns by way of simulation.

3 Results and Final Remarks

As discussed at the end of Chapter 2, it is important to validate the performance of our generalized test for treatment effect. While theoretical computations of power are easily computed using the standard normal distribution, analysis in practice will use the t-test version of the test because ρ and σ^2 must be estimated from the data. The power of the actual test may not be consistent with the theoretical power.

An additional goal of this study is to investigate if equivalent crossover designs are consistent in terms of their power when applied in practice. We hypothesized that designs that have more periods might behave differently compared to an equivalent design that has more patients but less periods, since these parameters might have an impact on how ρ and σ^2 are estimated. To assess these two primary goals, we performed Monte Carlo simulations under various scenarios to estimate power of the t-test.

3.1 Simulation Design

For all of the simulations considered in this report, we varied the treatment effect $\Delta = \tau_A - \tau_B$ from 0 to 1.2 by increments of 0.2, all period effects were set to 3 ($\pi_1 = \pi_2 = ... = \pi_l = 3$), and no carry-over was assumed ($\lambda_{AB} = \lambda_{BA} = 0$). We also set $\sigma^2 = 14$. To address our first goal, we examined four scenarios:

1. $\rho = 0.5$, $n_1 = n_2 = 20$, $l = 2, 4, 6$
2. $\rho = 0.5$, $n_1 = n_2 = 5$, $l = 2, 4, 6$
3. $\rho = 0.8$, $n_1 = n_2 = 5$, $l = 2, 4, 6$
4. $\rho = 0.2$, $n_1 = n_2 = 5$, $l = 2, 4, 6$

Under these 4 scenarios, we can explore low, moderate, and high theoretical power and compare that to empirical estimates of power using the t-test derived from our contrast. To address our second goal, we considered 6 additional scenarios (2 sets of 3) whose contrasts to test for the treatment effect have the same variance and therefore should have the same power. Scenarios 5 through 7 are listed below and are equivalent in regards to the construction of their z-statistic.

5. $\rho = 0.5$, $n_1 = n_2 = 5$, $l = 8$

6. $\rho = 0.5$, $n_1 = n_2 = 10$, $l = 4$

7. $\rho = 0.5$, $n_1 = n_2 = 20$, $l = 2$

Similarly, scenarios 8 through 10 are equivalent to each other, are essentially the same as scenarios 5, 6, and 7 except for the fact that the strength of ρ is increased to 0.8.

8. $\rho = 0.8$, $n_1 = n_2 = 5$, $l = 8$

9. $\rho = 0.8$, $n_1 = n_2 = 10$, $l = 4$

10. $\rho = 0.8$, $n_1 = n_2 = 20$, $l = 2$

To simulate one data set from a given scenario, we followed the procedure described in Chapter 1 using the general mean and variance for $\bar{\boldsymbol{Y}}$. Briefly, the data from two patients, 1 from each sequence (AB and BA), is simulated using a multivariate normal distribution with mean $E(\bar{\boldsymbol{Y}})$ as specified in Equation (2.1) and Variance-Covarinace Matrix defined as in Equation (2.2) and setting the sample sizes equal to 1 to reflect single observations rather than averages. Additional random draws are taken to increase the number of patients in each group. The data was then formatted so that it could be fed into a general linear model framework inside of the program R.

To estimate the power of the contrast constructed through the t-test for a single set of parameters, we performed a Monte Carlo simulation using the following pseudo code:

1. For i in $1:5000$

    a. Create new data set based on the simulation parameters n_1, n_2, σ^2, ρ, l, and $\tau_A - \tau_B = \Delta$.

    b. Fit a linear mixed model to estimate the $2l$ cell means, σ^2, and ρ.

    c. Perform the t-test and compute the p-value using our proposed general contrast $\boldsymbol{c'}$.

    d. Record if the test is rejected and count it as 1 if rejected, 0 otherwise.

2. End Loop

3. Calculate the empirical power $\hat{p} = \frac{\#Rejections}{5000}$

4. Simulation margin of error is computed as $1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{5000}}$

3.2 Simulation Results

Figure 3.1 provides our first look at the performance of the t-test contrast compared to the theoretical power of the test assuming ρ and σ^2 are known. The x-axis presents the value of $\Delta = \tau_A - \tau_B$. The y-axis presents the power of the test, except for when $\Delta = 0$, where it can be interpreted as the type-I error rate. The empirical method is illustrated by the dotted lines while the theoretical method is displayed with solid lines. The different colors presents the different number of periods used in the design. The blue, green, and black colors correspond to $l = 2, 4, 6$, respectively. Upon examination of the figure, the power of the empirical method in all three designs

stays within simulation error of the true theoretical power as indicated by the simulation error bars plotted around the empirical estimates. As expected, while holding the patient sample sizes fixed, increasing the periods increases the statistical power.

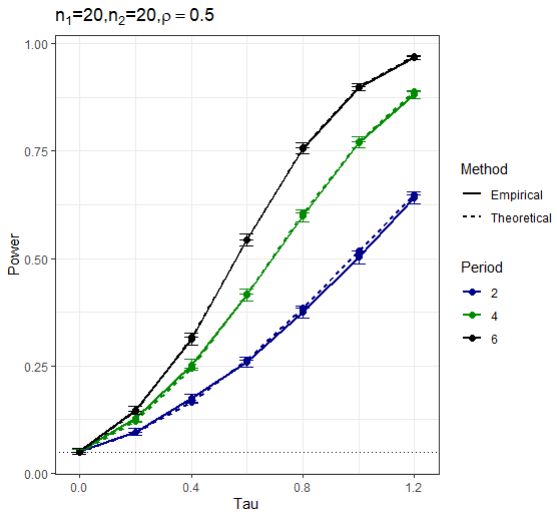

Figure 3.1: Theoretical and Empirical Power Curve for Scenario 1

Figure 3.2 provides some additional detail not recognized in the first scenario. In scenario 2, the patient sample size is decreased to $n_1 = 5$ and $n_2 = 5$. When examining the classic AB/BA design when the number of the period is 2, the empirical power at the higher values of Δ, are slightly lower than the theoretical ones and are not within simulation error. This suggests that if researchers are planning a traditional AB/BA design with smaller sample sizes, an empirical estimate of power via simulation would be more trustworthy than using the theoretical power. However, as the number of the periods increases to 4 or to 6, and the empirical power and the theoretical power become more consistent.

In scenario 3, figure 3.3 , we keep the sample sizes at $n_1 = 5$ and $n_2 = 5$, but we increased ρ to 0.8. Similarly to scenario 2, when the number of the period is 2 and for larger values of Δ, the empirical power is lower than that of the theoretical one. In this case, however the discrepancy in the power seems to have gotten worse than

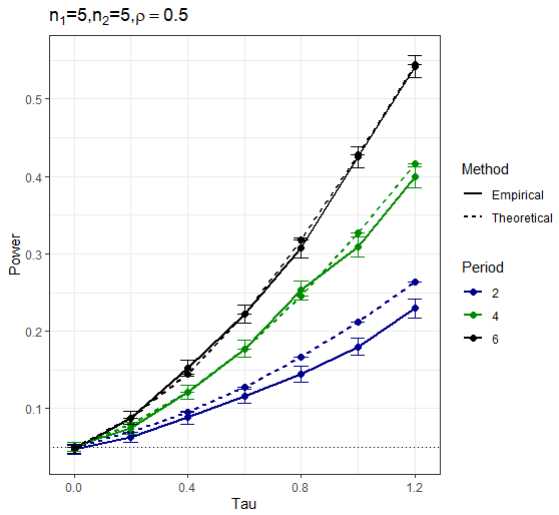

Figure 3.2: Theoretical and Empirical Power Curve for Scenario 2

in scenario 2. For example, when $\Delta = 1.2$, the difference between the theoretical and empirical power is only about 0.04 when $\rho = 0.5$, but the difference is closer to 0.9 when $\rho = 0.8$. Again, as the number of periods increases from 2, we see nice agreement between theoretical and empirical power. As expected the power is much higher in this scenario when comparing to scenario 2 as a larger ρ value will dampen the variance estimate used in the contrast test statistic.

In scenario 4, we again kept the sample sizes at $n_1 = 5$ and $n_2 = 5$ but we decreased ρ to 0.2. Figure 3.4 provides a similar behavior as the others. Again the main discrepancy in power is observed when there are just two periods. However, with a much smaller value for ρ, the discrepancy is less compared to when ρ was 0.5 and 0.8. As compared previously with scenario 2 and 3, when $\Delta = 1.2$, the observed difference in power is only 0.025 for the t-test vs z-test.

Table 3.1 provides the power estimates for scenarios 5, scenario 6, and scenario 7. There are three main columns which include the scenario, patient sample size (of each group), and true difference Δ. In the true difference Δ column, there are seven

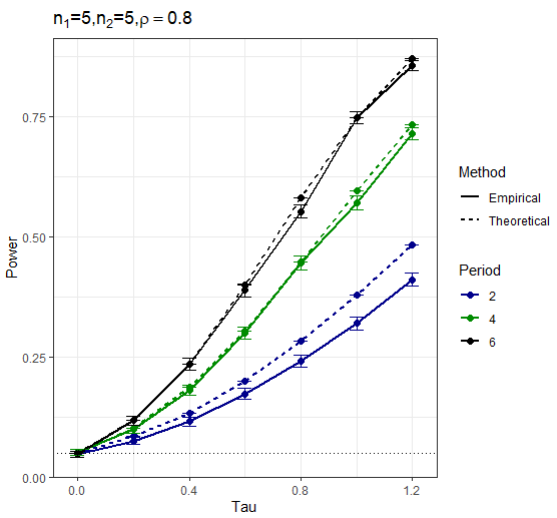

Figure 3.3: Theoretical and Empirical Power Curve for Scenario 3

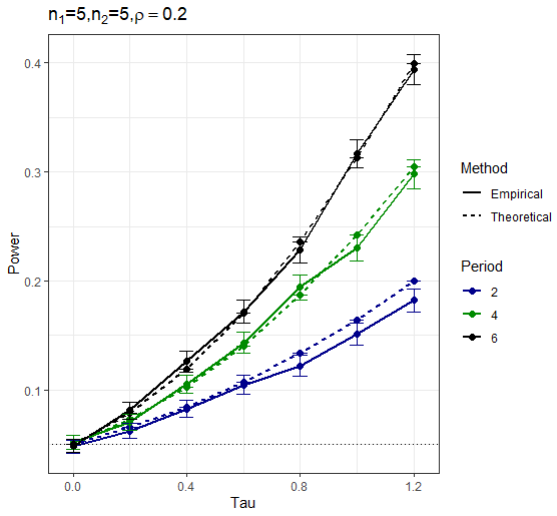

Figure 3.4: Theoretical and Empirical Power Curve for Scenario 4

smaller nested columns that present the different values of Δ, such as $\Delta = 0.00$, $\Delta = 0.20$, $\Delta = 0.40$, $\Delta = 0.60$, $\Delta = 0.80$, $\Delta = 1.00$, and $\Delta = 1.20$. The bold power estimates within the table indicate that the empirical estimates of power falls within

the simulation error of the theoretical power listed in the first row. While the results are not unanimous, we believe this is strong evidence that a researcher can truly consider alternate equivalent study designs without any fear of discrepancy when it comes to the power of the test in practice.

Table 3.1: Power estimates for Scenario 5, Scenario 6, and Scenario 7 ($\rho = 0.5$)

Scenario	Sample Size	True Difference Δ						
	per group	0.00	0.20	0.40	0.60	0.80	1.00	1.20
Theoretical	NA	**0.050**	**0.096**	**0.166**	**0.264**	**0.385**	**0.518**	**0.649**
5 (8 Periods)	5	**0.049**	**0.096**	**0.162**	**0.261**	**0.382**	**0.519**	**0.643**
6 (4 Periods)	10	**0.050**	**0.100**	**0.166**	**0.256**	0.370	0.500	**0.640**
7 (2 Periods)	20	**0.048**	0.106	**0.162**	**0.254**	**0.375**	**0.505**	0.630

Similarly to the set-up of Table 3.1, Table 3.2 provides the power estimate results for the equivalent scenarios 8, 9, and 10. These scenarios are identical to 5,6, and 7 with the exception that ρ is set to 0.8. With the exception of one, all empirical estimates fall within simulation error of the the theoretical result.

Table 3.2: Power estimates for Scenario 8, Scenario 9, and Scenario 10 ($\rho = 0.8$)

Scenario	Sample Size	True Difference Δ						
	per group	0.00	0.20	0.40	0.60	0.80	1.00	1.20
Theoretical	NA	**0.050**	**0.133**	**0.282**	**0.484**	**0.689**	**0.848**	**0.941**
8 (8 Periods)	5	**0.049**	**0.134**	**0.267**	**0.478**	**0.686**	**0.845**	**0.938**
9 (4 Periods)	10	**0.050**	**0.133**	**0.283**	**0.470**	0.678	**0.829**	**0.937**
10 (2 Periods)	20	**0.048**	**0.134**	**0.267**	**0.478**	**0.686**	**0.845**	**0.938**

3.3 Final Conclusions and Discussion

This thesis has provided a few key and interesting facts in regards to the AB/BA design and its extensions by way of simply increasing the number of periods for each sequence. In Chapter 2, we provided the necessary mathematical framework to generalize traditional crossover test for treatment effect, and provided a simple and compact result to construct a test by way of Theorem 2.1. This theorem highlights an interesting, and somewhat non-intuitive, property of the general crossover design with respect to the traditional AB/BA design. The result simply states that increasing the number of periods in the design by a factor of m while decreasing the number of patients they need to recruit by a factor of m indeed are equivalent in terms of the construction of our general test statistic. This provides a very quick rule of thumb for researchers who can determine sample sizes in the simple AB/BA design case and if the sample sizes are too difficult to implement due to rarity of the population or recruitment costs, the researcher can simply consider extending the number of periods with fewer subjects without loss of statistical power. These results were verified in our simulation studies in Table 3.1 and Table 3.2.

Based on our simulations, the theoretical power formula we provided in chapter two can be used by researchers to determine adequate sample sizes in the future for any design scenario considered in this report. The only situation that should give pause is if the the number of periods is 2 and the sample size is low. We recommend that when a theoretical result is suggesting less than 15 patients in each sequence, that simulations such as the ones we conducted would give researchers a better sense of what to expect and their design would be more accurately powered to meet their needs. It might be helpful for additional simulations to be conducted to ensure that no other issues with accuracy of power exist. We leave that as future work.

BIBLIOGRAPHY

[1] Simon H. Bishop and Byron Jones, *A review of higher order crossover designs*, Journal of Applied Statistics **11** (1984), no. 1, 29–50.

[2] J. Bogert, H. Rofael, K Mosca, R.O Ross, G Callaerts, D. Wang, J Jaeger, V. A Narayan, W. C Drevets, and R. L. Morrison, *A randomized, multicenter, crossover psychometric evaluation study of an ipad-administered cognitive test battery in participants with major depressive disorder who responded to treatment with oral antidepressants*, Journal of Affective Disorders **292** (2021), no. 1, 261–269.

[3] B. Jones and M.G Kenward, *Design and analysis of cross-over trials*, Chapman & Hall/CRC, 2015.

[4] N. M Laird and J.H Ware, *Random-effects models for longitudinal data*, Biometrics **38** (1982), no. 4, 963–974.

[5] T. Li, T. Yu, B. Hawkins, and K. Dickersin, *Design, analysis, and reporting of crossover trials for inclusion in a meta-analysis*, PloS one **10** (2015), no. 4, e0133023.

[6] James F. III Reed, *Extension of grizzle's classic crossover design*, Journal of Modern Applied Statistical Methods **10** (2011), no. 1, 322–328.

[7] R. Schiffmann, M. E. Wallace, D. Rinaldi, I. Ledoux, M. P. Luton, S. Coleman, H. O. Akman, K. Martin, J. Y. Hogrel, D. Blankenship, J. Turner, and F Mochel, *A double-blind, placebo-controlled trial of triheptanoin in adult polyglucosan body disease and open-label, long-term outcome*, Journal of inherited metabolic disease **45** (2017), no. 5, 877–883.

[8] S. Senn, *Cross-over trials in clinical research*, J. Wiley, 2002.

VITA

I received a Bachelor's degree in Biotechnology from University of Houston in December 2016. I started my first professional job in 2017 working for an accounting firm. In this job, I gained more understanding about the economics, and financial sides. At this job, I realized that understand the numbers was somewhat important in any kind of business. So I decided to go back to school for Statistics in beginning of 2019. In 2019, I start working for a CPA firm. At this job, I have had a change to analyzed a huge amount of different financial statements from different types of domestic and international businesses. The best part of this job is I get paid to learn and fix others' people mistakes.

Permanent Address: 22607 Hunters Tree Drive
Spring, TX 77389

The style manual used in this thesis is A Manual For Authors of Mathematical Papers published by the American Mathematical Society.

This thesis was prepared by My Nguyen using LaTeX.