Behavioral Carry-Over Effects and Power Considerations in Crossover Trials

Danni Shi

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2023

Reading Committee:

Ting Ye, Chair

James Hughes

Program Authorized to Offer Degree:

Biostatistics

University of Washington

Abstract

Behavioral Carry-Over Effects and Power Considerations in Crossover Trials

Danni Shi

Chair of the Supervisory Committee:
Ting Ye
Department of Biostatistics

The carry-over effect remains an outstanding concern in crossover trials when washout periods cannot sufficiently diminish their impacts. This can occur in comparative effectiveness research where carry-over effects are behavioral rather than biological. We first investigate crossover trials with and without carry-over effects under the potential outcome framework. We find that when carry-over effects exist and satisfy some sign condition, the classic estimator underestimates the treatment effect, which does not inflate the type I error of one-sided tests but decreases the power and leads to a power trade-off between crossover and parallel trials. We derive the condition under which crossover trials do not have type I error inflation and are more powerful than parallel trials. We further develop covariate adjustment methods for crossover trials and illustrate their performance using data from a real analgesic study and a hypothetical HIV prevention trial. We also compare the total investigation times and the total trial times of crossover and parallel trials.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank many people who supported me during the work on this thesis and throughout my master's studies. I will always treasure their support during these very unusual two to three years.

First of all, I would like to thank Ting for her mentorship, support, and encouragement throughout the process of this project. I am very fortunate to have her as my mentor and her expertise and dedication will always be an inspiration to me. I have learned a lot from her, especially what it truly means to be a professional biostatistician and a professional scholar.

I would like to thank Jim Hughes for taking the time to be my second reader and bringing up the most valuable comments and suggestions.

I would like to thank the HIV Prevention Network (HPTN) for allowing me to use data from HPTN 084. I would like to thank Thomas Richardson and Carlos Cinelli from the Department of Statistics for their comments and suggestions for this project.

I would like to thank Minh Vo, Gitana Garofalo and Ken Rice for their administrative support, especially when I was on leave from graduate school.

I would like to thank the many other graduate students in the Department of Biostatistics and Statistics for their help and companionship.

Finally, I would like to thank my parents, for teaching me how to love and how to take care of who I love, and for being good examples.

Chapter 1. INTRODUCTION

A crossover trial is a longitudinal study in which patients are randomized to different sequences of treatments with the intention of comparing the effects of different treatments.[1,2,3] In the simplest two-treatment two-period crossover design, patients are randomly allocated into two groups: one group receives treatment 1 at period 1 and treatment 0 at period 2, and the other group receives the treatment sequence of the reverse order. The main advantage of a crossover design is that every patient can serve as their own control, which eliminates between-patient variability and enables the crossover design to be more efficient than the traditional parallel-group (parallel for short) design.

A key assumption for the crossover design is that there is no *carry-over effect*: that is, the treatment assigned during period 1 does not interfere with the outcome in period 2. Therefore, a crossover design is most common for treatments whose effect vanishes when the study is discontinued or has non-absorbing endpoints. Examples include early phase trials (e.g., pharmacokinetic (PK) studies and dose-finding studies) and phase III studies with chronic conditions (e.g., hypertension, pain, and asthma); see Jones and Lewis (1995)[4] for a review. Furthermore, a washout period is often designed between period 1 and period 2, which can effectively reduce the impact of the *biological carry-over effect* engendered by treatments taken at period 1. However, the carry-over effect remains an outstanding concern when a washout period is unethical or cannot sufficiently diminish the persisting impact from period 1. For instance, in an HIV prevention trial that is currently being planned, a dual prevention pill (DPP, a daily oral tenofovir disoproxil fumarate-emtricitabine (TDF-FTC) and a combined oral contraceptive) will be compared against a two-pill regime (2PR) using a two-period crossover design to test the

hypothesis that a DPP could increase adherence to TDF-FTC. If participants are more likely to develop habits of adherence to the assigned treatment in period 1, such habits may carry over into period 2 and affect the adherence to the assigned treatment in period 2. This type of carry-over effect is non-biological and is hard to be eliminated by a washout period or a washout period might be unethical: in this HIV prevention trial, a washout period means withholding HIV prevention medication. We refer to this type of carry-over effect as the *behavioral carry-over effect*, which is the impact that a treatment has on the outcomes at the subsequent period due to the altering of a patient's behavior. It can be an important consideration in other HIV prevention trials (e.g., MTN 034 study (NCT03593655) and TRIO study[5]) and more broadly in comparative effectiveness research that uses the crossover design.[6]

The presence of carry-over effects can bias the estimation of the treatment effect. This motivates Grizzle (1965)'s[7] two-stage procedure, which first tests whether the carry-over effect exists or not, and the testing result determines whether it is necessary to use the period 2 data for analysis. This two-stage procedure was then criticized by Freeman (1989)[8] because it has lower power and can inflate the type I error for the following analysis. Another approach is to model the carry-over effect,[9,10,11,12,13] but this approach is sensitive to model misspecification.

This thesis proceeds as follows: Section 2.1 focuses on crossover trials without carry-over effects, where we investigate the crossover trial using minimal statistical assumptions and a potential outcomes framework.[14,15] We then examine the classic estimator and derive its asymptotic properties and show that the crossover design is more efficient than the parallel one. Section 2.2 extends the results in 2.1 to the case with the carry-over. We observe that when the carry-over effect satisfies a sign condition, the classic estimator is biased and underestimates the treatment effect, which does not inflate the type I error of one-sided tests of superiority or non-

inferiority but negatively affects the statistical power. This leads to a power trade-off between the crossover design and the parallel design, and we derive the condition under which the crossover design does not lead to type I error inflation and is still more powerful than the parallel design. Section 2.3 proposes a covariate adjustment method to the crossover design and studies the asymptotic properties of the covariate-adjusted estimator. We show that this estimator is guaranteed to be more powerful than its unadjusted counterpart. In Section 3, we apply the above methods through data analysis: Section 3.1 examines the power trade-off of the crossover and the parallel design through simulation. Section 3.2 applies the proposed estimators to a real veterinary data example. Section 3.3 presents a simulation study based on resampling data from an HIV prevention trial. Section 4 first shows how carry-over effects and washout periods could impact the total investigation time and the total trial time, then compares the recruitment rates of crossover and parallel trials and ends with a conclusion.

Chapter 2. METHODS

2.1 Crossover Trials with No Carry-over Effect

2.1.1 *Setup and Assumptions*

Consider a two-treatment two-period crossover trial. A sample of n subjects are randomly allocated to two treatment sequences, where $A_i = 1$ denotes that subject i receives treatment 1 at period 1 and then treatment 0 at period 2, and $A_i = 0$ denotes the reverse order. Let $Y_{i1}^{(j)}$ be the potential treatment outcome at period 1 had the subject been exposed to treatment j at period 1,

for $j = 0, 1$. Let $Y_{i2}^{(jk)}$ be the potential outcome at period 2 had the subject been exposed to treatment k at period 2, for $j, k = 0, 1$. The observed outcome for subject i at period t is Y_{it}. Throughout this thesis, we make the *consistency assumption* that links the observed outcomes to the potential outcomes: when $A_i = 1$, $Y_{i1} = Y_{i1}^{(1)}, Y_{i2} = Y_{i2}^{(10)}$; when $A_i = 0$, $Y_{i1} = Y_{i1}^{(0)}, Y_{i2} = Y_{i2}^{(01)}$. Let $\boldsymbol{X}_i$ be a vector of observed baseline covariates for subject i. We assume that $\left(A_i, \boldsymbol{X}_i, Y_{i1}^{(j)}, Y_{i2}^{(jk)}, j, k = 0,1\right), i = 1, \dots, n$ are independent and identically distributed with finite second-order moments, and the covariance matrices $\mathrm{Var}(\boldsymbol{X}_i)$ and $\mathrm{Var}\left(Y_{i1}^{(j)}, Y_{i2}^{(jk)}\right)$ are both positive definite.

Simple randomization assigns subjects to the two treatment sequences completely at random. This is summarized in Assumption 1.

- *Assumption 1* (Simple randomization): $A_i \perp \left(\boldsymbol{X}_i, Y_{i1}^{(j)}, Y_{i2}^{(jk)}\right)$ for $j, k = 0,1$ and $P(A_i = 1) = \pi_1$ where $0 < \pi_1 < 1$ is known and $\pi_0 = 1 - \pi_1$.

Assumption 2 is the key assumption typically imposed in crossover trials. It says that the treatment at time 1 does not have a direct effect on the outcome at period 2; see Figure 1 for an illustration. Many crossover trials would plan a sufficiently long washout period between the two time periods to make this no-carry-over assumption more plausible.

- *Assumption 2* (No carry-over effect): For $k = 0,1, Y_{i2}^{(0k)} = Y_{i2}^{(1k)} \coloneqq Y_{i2}^{(k)}$ almost surely.

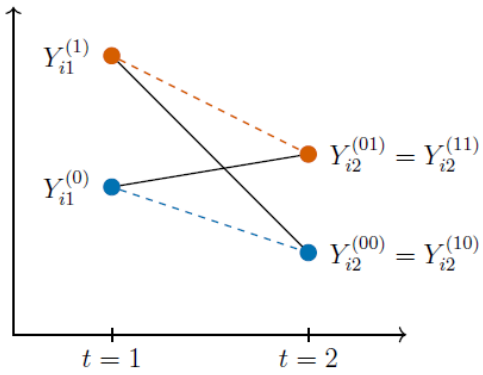

Figure 1: Six potential outcomes under Assumption 2

Under Assumption 2, we can simply let $Y_{i2}^{(k)}, k = 0,1$ denote the potential outcome at period 2. We are interested in the average treatment effects $2^{-1}(\theta_1 + \theta_2)$, where $\theta_1 = E\left(Y_{i1}^{(1)} - Y_{i1}^{(0)}\right)$ and $\theta_2 = E\left(Y_{i2}^{(1)} - Y_{i2}^{(0)}\right)$. Table 1 represents the parameterization of the expected values of the six potential outcomes $Y_{i1}^{(1)}, Y_{i1}^{(0)}, Y_{i2}^{(11)}, Y_{i2}^{(10)}, Y_{i2}^{(01)}, Y_{i2}^{(00)}$ without carry-over effects; the two shaded rows correspond to the two potential outcomes that are never observed. Note that it is common to assume that the treatment effect is time-invariant so that $\theta_1 = \theta_2$.[1] However, as shown in Theorem 1 in Section 2.1.2, the time-invariant treatment effect assumption is not necessary for studying treatment effect in crossover trials.

Table 1: Parameterization of the six potential outcome means with Assumption 2

Potential outcomes	Expected values
$E\left(Y_{i1}^{(0)}\right)$	μ
$E\left(Y_{i1}^{(1)}\right)$	$\mu + \theta_1$
$E\left(Y_{i2}^{(00)}\right)$	$\mu + \tau$
$E\left(Y_{i2}^{(10)}\right)$	$\mu + \tau$
$E\left(Y_{i2}^{(11)}\right)$	$\mu + \tau + \theta_2$
$E\left(Y_{i2}^{(01)}\right)$	$\mu + \tau + \theta_2$

2.1.2 *The Classic Estimator*

From the full two crossover periods, we can find the difference in outcomes for every subject under treatment arm 1 and arm 0. The *classic estimator* is commonly used, which first calculates the arm-specific average outcome difference, and then averages the two means from the two arms:[1,2,3]

$$\hat{\theta}_{\text{cr}} = \frac{1}{2}\left\{\frac{1}{n_1}\sum_{i=1}^{n} A_i(Y_{i1} - Y_{i2}) + \frac{1}{n_0}\sum_{i=1}^{n}(1 - A_i)(Y_{i2} - Y_{i1})\right\} \coloneqq \frac{\overline{\Delta}_1 - \overline{\Delta}_0}{2}, \qquad (1)$$

where $\overline{\Delta}_a = n_a^{-1}\sum_{A_i=a}\Delta_i$ is the average of Δ_i's for subjects with $A_i = a, \Delta_i = Y_{i1} - Y_{i2}$ and n_a is the number of subjects with $A_i = a$ for $a = 0,1$. Theorem 1 summarizes the statistical properties of the classic estimator $\hat{\theta}_{\text{cr}}$.

- *Theorem 1*: Under Assumptions 1-2,
  - *(a)* $E\left(\hat{\theta}_{cr}\right) = 2^{-1}(\theta_1 + \theta_2)$, where $\theta_t = E\left(Y_{it}^{(1)} - Y_{it}^{(0)}\right)$ for $t = 1,2$.
  - *(b)* $\sqrt{n}\left\{\hat{\theta}_{cr} - 2^{-1}(\theta_1 + \theta_2)\right\} \xrightarrow{d} N(0, \sigma_{cr}^2)$ where $\sigma_{cr}^2 = (4\pi_1)^{-1}\text{Var}\left(Y_{i1}^{(1)} - Y_{i2}^{(0)}\right) + (4\pi_0)^{-1}\text{Var}\left(Y_{i2}^{(1)} - Y_{i1}^{(0)}\right)$

Proof of Theorem 1 and all other proofs are in Appendix A1. Let $\tau = E\left(Y_{i2}^{(0)} - Y_{i1}^{(0)}\right)$ denote the expected temporal change in outcome in the absence of the treatment. The expected change in outcome for $A_i = 1$ is the average treatment effect at period 1 minus τ, i.e., $E(Y_{i1} - Y_{i2}|A_i = 1) = E\left(Y_{i1}^{(1)} - Y_{i2}^{(0)}\right) = E\left(Y_{i1}^{(1)} - Y_{i1}^{(0)} + Y_{i1}^{(0)} - Y_{i2}^{(0)}\right) = \theta_1 - \tau$; similarly, the expected change in outcome for $A_i = 0$ is the average treatment effect at period 2 plus τ, i.e., $E(Y_{i1} - Y_{i2}|A_i = 0) = E\left(Y_{i2}^{(1)} - Y_{i1}^{(0)}\right) = E\left(Y_{i2}^{(1)} - Y_{i2}^{(0)} + Y_{i2}^{(0)} - Y_{i1}^{(0)}\right) = \theta_2 + \tau$. At the first glance, it might appear that both group-specific average changes in outcomes are biased by τ. However, randomization balances out the effect of time trend that is not due to the treatment, and

thus the overall average change in outcome for either period remains unbiased for $2^{-1}(\theta_1 + \theta_2)$. Moreover, with a given sample size n, based on Theorem 1(b), we can obtain the optimal choice of π_1 that minimizes the asymptotic variance. For instance, if $\mathrm{Var}\left(Y_{i1}^{(1)} - Y_{i2}^{(0)}\right) = \mathrm{Var}\left(Y_{i1}^{(0)} - Y_{i2}^{(1)}\right)$, the optimal π_1 would follow equal allocation, i.e., $\pi_1 = \pi_0 = \frac{1}{2}$. A consistent estimator of σ_{cr}^2 is

$$\hat{\sigma}_{\mathrm{cr}}^2 = \frac{1}{4\pi_1} S_{\Delta_1}^2 + \frac{1}{4\pi_0} S_{\Delta_0}^2, \qquad (2)$$

where $S_{\Delta_1}^2, S_{\Delta_0}^2$ are the sample variance of Δ_i with $A_i = a$ for $a = 0,1$ respectively.

Note that there is another estimator that seems like the classic estimator defined in (1), but it has larger asymptotic variance and could be biased in some cases, which is discussed in Hills and Armitage (1979).[1] Consider $\hat{\theta}_{\mathrm{cr}}^{\mathrm{alt}} = \frac{1}{n}\sum_{i=1}^{n}\{A_i(Y_{i1} - Y_{i2}) + (1 - A_i)(Y_{i2} - Y_{i1})\}$. $\hat{\theta}_{\mathrm{cr}}$ and $\hat{\theta}_{\mathrm{cr}}^{\mathrm{alt}}$ are the same under the randomization scheme that enforces $n_1 = n_0$, but they are not the same under the simple randomization considered in this thesis. Under simple randomization with equal allocation $\pi_1 = \pi_0 = \frac{1}{2}$, $\hat{\theta}_{\mathrm{cr}}^{\mathrm{alt}}$ is unbiased for $2^{-1}(\theta_1 + \theta_2)$ as well; however, its asymptotic variance of $\sqrt{n}\{\hat{\theta}_{\mathrm{cr}}^{\mathrm{alt}} - 2^{-1}(\theta_1 + \theta_2)\} = 2^{-1}\mathrm{Var}\left(Y_{i1}^{(1)} - Y_{i2}^{(0)}\right) + 2^{-1}\mathrm{Var}\left(Y_{i1}^{(0)} - Y_{i2}^{(1)}\right) + 4^{-1}(\theta_1 - \theta_2 - 2\tau)^2$, which exceeds the asymptotic variance by an additional component $4^{-1}(\theta_1 - \theta_2 - 2\tau)^2$. This additional component is due to the random effect of the temporal trend τ and the heterogeneity of the treatment effect at two periods. Under simple randomization with unequal allocation, $\hat{\theta}_{\mathrm{cr}}^{\mathrm{alt}}$ is biased due to the temporal effect τ. Hence, we do not consider this alternative estimator $\hat{\theta}_{\mathrm{cr}}^{\mathrm{alt}}$ in the rest of this thesis.

2.1.3 *Efficiency Comparison between the Crossover and Parallel Design*

When Assumption 2 holds, i.e., when there is no carry-over effect, a crossover design is known for being typically more efficient than a parallel one. This is because each subject can serve as their own control, which allows a crossover trial to have twice as many measurements per patient; additionally, within-subject comparison can further remove the inter-subject variability.[16]

Suppose a randomized control trial is aimed to demonstrate the superiority or non-inferiority of an investigational treatment. First, for a crossover trial, consider the null hypothesis $H_0: 2^{-1}(\theta_1 + \theta_2) = \theta^*$ versus the alternative hypothesis $H_A: 2^{-1}(\theta_1 + \theta_2) > \theta^*$ for some pre-specified θ^*, where $\theta^* = 0$ for test of superiority and $\theta^* > 0$ for test of non-inferiority. The test statistic based on the classic estimator is $T_{\mathrm{cr}} = \frac{\sqrt{n}(\hat{\theta}_{\mathrm{cr}} - \theta^*)}{\hat{\sigma}_{\mathrm{cr}}}$. From Theorem 1, $T_{\mathrm{cr}} \xrightarrow{\mathrm{d}} N(0,1)$ under H_0 and thus, we reject H_0 if and only if $T_{\mathrm{cr}} > z_{1-\alpha}$, where α is the significance level and $z_{1-\alpha}$ is the $(1-\alpha)$ th quantile of the standard normal distribution. Under the local alternative, $\sqrt{n}\{2^{-1}(\theta_1 + \theta_2) - \theta^*\} \to \gamma_{\mathrm{cr}}$ for some positive constant γ_{cr}, and the asymptotic approximation of the corresponding power of T_{cr} is:

$$\mathrm{Power}_{\mathrm{cr}} \approx \Phi\left(-z_{1-\alpha} + \frac{\gamma_{\mathrm{cr}}}{\sigma_{\mathrm{cr}}}\right), \qquad (3)$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

Now for a parallel trial, i.e., when we only use data at period 1, the parallel counterpart of $\hat{\theta}_{\mathrm{cr}}$ is a simple mean difference: $\hat{\theta}_{\mathrm{pr}} = \frac{1}{n_1}\sum_{i=1}^{n} A_i Y_{i1} - \frac{1}{n_0}\sum_{i=1}^{n}(1 - A_i)Y_{i1}$. Under Assumption 1, we can derive its asymptotic properties: $E\left(\hat{\theta}_{\mathrm{pr}}\right) = \theta_1$ and $\sqrt{n}\left(\hat{\theta}_{\mathrm{pr}} - \theta_1\right) \xrightarrow{\mathrm{d}} N\left(0, \sigma_{\mathrm{pr}}^2\right)$, where $\sigma_{\mathrm{pr}}^2 = 2\mathrm{Var}\left(Y_{i1}^{(1)}\right) + 2\mathrm{Var}\left(Y_{i1}^{(0)}\right)$. When we have the one-sided test with the null hypothesis $H_0: \theta_1 = \theta^*$ versus $H_A: \theta_1 > \theta^*$, the test statistic is $T_{\mathrm{pr}} = \frac{\sqrt{n}(\hat{\theta}_{\mathrm{pr}} - \theta^*)}{\hat{\sigma}_{\mathrm{pr}}}$, where $\hat{\sigma}_{\mathrm{pr}}^2 = 2S_1^2 + 2S_0^2$

being a consistent estimator for σ_{pr}^2 with S_1^2, S_0^2 being the sample variance for subjects under $A_i = 1$ and $A_i = 0$ respectively. Under the local alternative, $\sqrt{n}(\theta_1 - \theta^*) \to \gamma_{\text{pr}}$ for some positive constant γ_{pr}, and the asymptotic approximation of the corresponding power of T_{pr} is:

$$\text{Power}_{\text{pr}} \approx \Phi\left(-z_{1-\alpha} + \frac{\gamma_{\text{pr}}}{\sigma_{\text{pr}}}\right), \qquad (4)$$

Now equate (3) and (4), i.e., let $\text{Power}_{\text{cr}} = \text{Power}_{\text{pr}} = 1 - \beta$. We obtain the required sample sizes to achieve $1 - \beta$ power using the two designs:

$$\frac{n_{\text{cr}}}{n_{\text{pr}}} = \frac{\sigma_{\text{cr}}^2}{\sigma_{\text{pr}}^2} \cdot \frac{(\theta_1 - \theta^*)^2}{\{2^{-1}(\theta_1 + \theta_2) - \theta^*\}^2},$$

When $\theta_1 = \theta_2$, the two aforementioned null hypotheses are the same. Let $\theta_1 = \theta_2 = \theta$, and the corresponding Pitman asymptotic relative efficiency, i.e., the ratio of the sample sizes to achieve the same power using the two tests, is simply $\frac{n_{\text{cr}}}{n_{\text{pr}}} = \frac{\sigma_{\text{cr}}^2}{\sigma_{\text{pr}}^2}$.

For illustration, consider a simple case where $\theta_1 = \theta_2 = \theta, \text{Var}\left(Y_{it}^j\right) = \sigma^2$ for $t = 1,2$ and $j = 0, 1,$ and $\text{Cov}\left(Y_{i1}^{(j)}, Y_{i2}^{(1-j)}\right) = \rho\sigma^2,$ where $\rho \in [0,1]$ is the intraclass correlation coefficient (ICC). Then we can have $\sigma_{\text{cr}}^2 = 2(1-\rho)\sigma^2, \sigma_{\text{pr}}^2 = 4\sigma^2,$ and $\frac{n_{\text{cr}}}{n_{\text{pr}}} = \frac{1-\rho}{2}$. This implies that the crossover design only requires $\frac{1-\rho}{2}$, i.e., at most half of the sample size required by the parallel design while obtaining twice as many data points of measurements.

When the sample size is small, we could also test $H_0: 2^{-1}(\theta_1 + \theta_2) = \theta^*$ vs. $H_A: 2^{-1}(\theta_1 + \theta_2) > \theta^*$ and obtain test statistics that follow some t-distributions; see Appendix A.2 for details.

2.2 Crossover Trials with Carry-over Effect

2.2.1 *Setup and Assumptions*

When there exist carry-over effects in a crossover trial, the treatment at period 1 may interfere with the outcome at period 2: see Figure 2 for an illustration of the six possible potential outcomes $Y_{i1}^{(1)}, Y_{i1}^{(0)}, Y_{i2}^{(11)}, Y_{i2}^{(10)}, Y_{i2}^{(01)}, Y_{i2}^{(00)}$ with carry-over effects. Here, Assumption 1 still holds by the act of randomization, but Assumption 2 is violated.

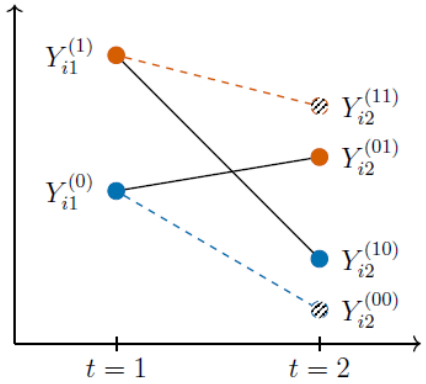

Figure 2: Six potential outcomes without Assumption 2

Table 2 presents the parameterization of the expected values of the six potential outcomes with carry-over effects and the two shaded rows correspond to the two potential outcomes that are never observed. $\tilde{\theta}_2 = E\left(Y_{i2}^{(11)} - Y_{i2}^{(00)}\right)$ denotes the average treatment effect at period 2, which compares the potential outcome had the subject stayed on treatment 1 to the one had the subject stayed on treatment 0, and the treatment effect is $2^{-1}\left(\theta_1 + \tilde{\theta}_2\right)$. Note that $\tilde{\theta}_2 = \theta_2$ when Assumption 2 holds. $\tilde{\tau} = E\left(Y_{i2}^{(0)} - Y_{i1}^{(0)}\right)$ denotes the expected temporal change in outcome had the subject stayed on treatment 0. Since the carry-over effect is the persisting impact of treatment a at period 1 on treatment $1 - a$ at period 2 for $a = 1,0$, we here define $\lambda_0 = E\left(Y_{i2}^{(10)} - Y_{i2}^{(00)}\right)$

and $\lambda_1 = E\left(Y_{i2}^{(11)} - Y_{i2}^{(01)}\right)$ to be the expected carry-over effects under the two treatment regimes $A_i = 0$ and $A_i = 1$ respectively.

Table 2: Parameterization of the six potential outcome means without Assumption 2

Potential outcomes	Expected values
$E\left(Y_{i1}^{(0)}\right)$	μ
$E\left(Y_{i1}^{(1)}\right)$	$\mu + \theta_1$
$E\left(Y_{i2}^{(00)}\right)$	$\mu + \tilde{\tau}$
$E\left(Y_{i2}^{(10)}\right)$	$\mu + \tilde{\tau} + \lambda_0$
$E\left(Y_{i2}^{(11)}\right)$	$\mu + \tilde{\tau} + \tilde{\theta}_2$
$E\left(Y_{i2}^{(01)}\right)$	$\mu + \tilde{\tau} + \tilde{\theta}_2 - \lambda_1$

2.2.2 *The Classic Estimator*

The classic estimator $\hat{\theta}_{\text{cr}}$ remains the same as defined in (1) and owing to carry-over effects, it is no longer unbiased for the true treatment effect. We derive its statistical properties without Assumption 2:

- *Theorem 2*: Under Assumption 1 and without Assumption 2,
  - *(a)* $E\left(\hat{\theta}_{\text{cr}}\right) = 2^{-1}\left(\theta_1 + \tilde{\theta}_2 - \lambda_0 - \lambda_1\right)$
  - *(b)* $\sqrt{n}\left\{\hat{\theta}_{\text{cr}} - 2^{-1}\left(\theta_1 + \tilde{\theta}_2 - \lambda_0 - \lambda_1\right)\right\} \xrightarrow{\text{d}} N(0, \tilde{\sigma}_{\text{cr}}^2)$, where $\tilde{\sigma}_{\text{cr}}^2 = (4\pi_1)^{-1}\text{Var}\left(Y_{i1}^{(1)} - Y_{i2}^{(10)}\right) + (4\pi_0)^{-1}\text{Var}\left(Y_{i2}^{(01)} - Y_{i1}^{(1)}\right)$

See Appendix A1.2 for the proof. Theorem 1 now is a special case of Theorem 2 under Assumption 2. Similar to the proof of Theorem 1(a) (see Appendix A1.2), the expected change in outcome for $A_i = 1$ is $E(Y_{i1} - Y_{i2}|A_i = 1) = E\left(Y_{i1}^{(1)} - Y_{i2}^{(10)}\right) = E\left(Y_{i1}^{(1)} - Y_{i1}^{(0)} + Y_{i1}^{(0)} -\right.$

$Y_{i2}^{(00)} + Y_{i2}^{(00)} - Y_{i2}^{(10)}\Big) = \theta_1 - \tilde{\tau} - \lambda_0$, and the one for $A_i = 0$ is $E(Y_{i2} - Y_{i1}|A_i = 0) = E\Big(Y_{i2}^{(01)} - Y_{i1}^{(0)}\Big) = E\Big(Y_{i2}^{(01)} - Y_{i2}^{(11)} + Y_{i2}^{(11)} - Y_{i2}^{(00)} + Y_{i2}^{(00)} - Y_{i1}^{(1)}\Big) = \tilde{\theta}_2 + \tilde{\tau} - \lambda_1$, and thus the overall average change in treatment outcome is $2^{-1}\big(\theta_1 + \tilde{\theta}_2 - \lambda_0 - \lambda_1\big)$. Note that from Theorem 2(a), under the assumption that $\lambda_0 + \lambda_1 = 0$, $\hat{\theta}_{\text{cr}}$ is still an unbiased estimator of the treatment effect $2^{-1}\big(\theta_1 + \tilde{\theta}_2\big)$, and this assumption is weaker than the individual-level no carry-over effect assumption stated in Assumption 2. Also, $\hat{\sigma}_{\text{cr}}^2$ defined in (2) is still a consistent estimator of $\tilde{\sigma}_{\text{cr}}^2$ when there are carry-over effects.

2.2.3 *Type I Error and Power Analysis*

Consider a test of superiority with the null hypothesis $H_0: 2^{-1}\big(\theta_1 + \tilde{\theta}_2\big) = \theta^*$ versus the alternative hypothesis $H_A: 2^{-1}\big(\theta_1 + \tilde{\theta}_2\big) > \theta^*$ for some pre-specified θ^* and the test statistic $T_{\text{cr}} = \sqrt{n}\big(\hat{\theta}_{\text{cr}} - \theta^*\big)/\hat{\sigma}_{\text{cr}}$. Before we move forward to the power analysis for the crossover and parallel designs, the first question we need to answer is whether T_{cr} would lead to type I error inflation. The corresponding type I error of T_{cr} is:

$$P_{H_0}(T_{\text{cr}} > z_{1-\alpha}) = P_{H_0}\left(\sqrt{n} \cdot \frac{\hat{\theta}_{\text{cr}} - E\big(\hat{\theta}_{\text{cr}}\big) + E\big(\hat{\theta}_{\text{cr}}\big) - \theta^*}{\hat{\sigma}_{\text{cr}}} > z_{1-\alpha}\right)$$

$$\approx \Phi\left(-z_{1-\alpha} - \sqrt{n} \cdot \frac{2^{-1}(\lambda_0 + \lambda_1)}{\tilde{\sigma}_{\text{cr}}}\right)$$

When $\lambda_0 + \lambda_1 = 0$, $\hat{\theta}_{\text{cr}}$ is an unbiased estimator for the treatment effect $2^{-1}\big(\theta_1 + \tilde{\theta}_2\big)$ and the type I error rate of T_{cr} is α. When $\lambda_0 + \lambda_1 > 0$, $\hat{\theta}_{\text{cr}}$ underestimates the treatment effect and the type I error rate of $T_{\text{cr}} < \alpha$, and then the test would be conservative. When $\lambda_0 + \lambda_1 < 0$, $\hat{\theta}_{\text{cr}}$ overestimates the treatment effect and the type I error rate of $T_{\text{cr}} > \alpha$, and then the test would be

invalid. Thus, the crossover design can still control the type I error rate of a one-sided test when $\lambda_0 + \lambda_1 \geq 0$.

However, the carry-over effect that does not inflate the type I error can decrease the power. Consider the case when $\lambda_0 + \lambda_1 \geq 0$, under the local alternative $\sqrt{n}\{2^{-1}(\theta_1 + \tilde{\theta}_2) - \theta^*\} \to \gamma_{\text{cr}}$ for a positive constant γ_{cr}, the power of T_{cr} is:

$$\text{Power}_{\text{cr}} \approx \Phi\left(-z_{1-\alpha} + \frac{\gamma_{\text{cr}} - \sqrt{n} \cdot 2^{-1}(\lambda_0 + \lambda_1)}{\tilde{\sigma}_{\text{cr}}}\right), \qquad (5)$$

So, the required sample size to achieve power of $1 - \beta$ is:

$$\tilde{n}_{\text{cr}} = \frac{\left(z_{1-\alpha} + z_{1-\beta}\right)^2 \tilde{\sigma}_{\text{cr}}^2}{\left\{2^{-1}\left(\theta_1 + \tilde{\theta}_2 - \lambda_0 - \lambda_1\right) - \theta^*\right\}^2}$$

For illustration, consider a test of superiority where $H_0: 2^{-1}\left(\theta_1 + \tilde{\theta}_2\right) = \theta_{\text{Alt}} = \theta^* = 0$ versus $H_A: 2^{-1}\left(\theta_1 + \tilde{\theta}_2\right) = \theta_{\text{Alt}} > \theta^* = 0$, and suppose $\theta_1 = \tilde{\theta}_2 = \theta_{\text{alt}} > 0$. The Pitman asymptotic relative efficiency between the crossover and the parallel design is:

$$\frac{\tilde{n}_{\text{cr}}}{n_{\text{pr}}} = \frac{\tilde{\sigma}_{\text{cr}}^2}{\sigma_{\text{pr}}^2} \cdot \left(1 - \frac{\lambda_0 + \lambda_1}{2\theta_{\text{Alt}}}\right)^{-2}. \qquad (6)$$

When $\theta_1 + \tilde{\theta}_2 > \lambda_0 + \lambda_1 \geq 0$, the carry-over effects are non-negative so that the type I error rate is controlled, and the carry-over effects are less than the treatment effects. Under this scenario, the difference between n_{pr} and $\tilde{n}_{\text{cr}}$ can be expressed as:

$$\tilde{n}_{\text{cr}} - n_{\text{pr}} = c\left\{\frac{2^{-1}(\lambda_0 + \lambda_1)}{\theta_{\text{Alt}}} - \left(1 - \frac{\tilde{\sigma}_{\text{cr}}}{\sigma_{\text{pr}}}\right)\right\},$$

where c is some positive constant. To guarantee that the crossover design is more efficient than the parallel one, we need to have the carry-over effect to be small and

$$2^{-1}(\lambda_0 + \lambda_1) < \theta_{\text{Alt}}\left(1 - \frac{\tilde{\sigma}_{\text{cr}}}{\sigma_{\text{pr}}}\right). \qquad (7)$$

For illustration, consider the simple example at the end of Section 2.1.3: when $\tilde{\sigma}_{\mathrm{cr}}/\sigma_{\mathrm{pr}} = \sqrt{(1-\rho)/2}$ and $2^{-1}(\lambda_0+\lambda_1) > 0$, we need $0 < 2^{-1}(\lambda_0+\lambda_1) < \theta_{\mathrm{Alt}} \cdot \sqrt{(1-\rho)/2}$. Accordingly, when $\rho = 0.3, 0.5, 0.7$, we need to have $0 < 2^{-1}(\lambda_0+\lambda_1) < 0.41\theta_{\mathrm{Alt}}, 0.50\theta_{\mathrm{Alt}}, 0.61\theta_{\mathrm{Alt}}$ respectively so that there is no type I error inflation and the crossover design is still more powerful than the parallel one. Such small carry-over effect condition may be plausible in many scenarios as carry-over effects are usually relatively small compared to the treatment effects and thus, a crossover design can still be more efficient than a parallel one in many cases even when carry-over effects exist.

In Appendix A.3, we use a sensitivity analysis approach[17] to discuss how to control the type I error when $\lambda_0 + \lambda_1 < 0$.

2.3 Covariate Adjustment for Crossover Trials

Covariate adjustment has high potential to improve the precision for many clinical trials.[18] It often uses a *working* model between the outcomes and covariates; however, the estimand of the treatment effect under covariate adjustment is the same as when using the unadjusted method and the inference of the treatment effect does not rely on whether the working model is correctly specified. Covariate adjustment for parallel trials has been extensively studied recently: in particular, it has been established that covariate adjustment using an analysis of heterogeneous covariance (ANHECOVA) working model that includes all treatment-by-covariate interaction terms can lead to *guaranteed efficiency gain* regardless of the model is mis-specified or

not.[19,20,21,22] These recent results have not been extended to crossover trials, although covariate adjustment is broadly recommended for crossover trials.[23,24,25]

To use the baseline covariate vector $\boldsymbol{X}_i$ in crossover trials, we propose the following covariate-adjusted ANHECOVA estimator:

$$\hat{\theta}_{\text{cr,adj}} = \frac{1}{2}\left[\left\{\overline{\Delta}_1 - \widehat{\boldsymbol{\beta}}_1^T(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}})\right\} - \left\{\overline{\Delta}_0 - \widehat{\boldsymbol{\beta}}_0^T(\overline{\boldsymbol{X}}_0 - \overline{\boldsymbol{X}})\right\}\right]$$

where $\overline{\boldsymbol{X}}$ is the sample mean of all $\boldsymbol{X}_i$'s, $\overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_0$ are the sample mean of $\boldsymbol{X}_i$'s from subjects under $A_i = 1,0$ respectively, $\Delta_i, \overline{\Delta}_1, \overline{\Delta}_0$ are defined in (1), and

$$\widehat{\boldsymbol{\beta}}_a = \left\{\sum\nolimits_{i:A_i=a} (\boldsymbol{X}_i - \overline{\boldsymbol{X}}_a)(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_a)^T\right\}^{-1}\left\{\sum\nolimits_{i:A_i=a} (\boldsymbol{X}_i - \overline{\boldsymbol{X}}_a)\right\}\overline{\Delta}_a$$

is the least squares estimator of $\boldsymbol{\beta}_a$ from fitting the linear working model $E(\Delta_i|A_i = a, \boldsymbol{X}_i) = \mu_a + \boldsymbol{\beta}_a^T\boldsymbol{X}_i$ using subjects with $A_i = 1, 0$.

The following heuristics reveal why ANHECOVA does not change the estimand, often gains but never hurts efficiency even when the linear working model is wrong. As randomization balances the covariate distribution, both $\overline{\boldsymbol{X}}_a$ and $\overline{\boldsymbol{X}}$ estimate the same quantity and thus, $\widehat{\boldsymbol{\beta}}_a^T(\overline{\boldsymbol{X}}_a - \overline{\boldsymbol{X}})$ is an "estimator" of zero. Therefore, $\overline{\Delta}_a - \widehat{\boldsymbol{\beta}}_a^T(\overline{\boldsymbol{X}}_a - \overline{\boldsymbol{X}})$ and $\overline{\Delta}_a$ correspond to the same estimand. Additionally, as $n \to \infty, \widehat{\boldsymbol{\beta}}_a \xrightarrow{\text{p}} \text{Var}(\boldsymbol{X}_i)^{-1}\text{Cov}(\boldsymbol{X}_i, \Delta_i|A_i = a)$, regardless of the linear working model is correct or not. Hence, $\overline{\Delta}_a - \widehat{\boldsymbol{\beta}}_a^T(\overline{\boldsymbol{X}}_a - \overline{\boldsymbol{X}})$ is asymptotically equivalent to $\overline{\Delta}_a - \boldsymbol{\beta}_a^T(\overline{\boldsymbol{X}}_a - \overline{\boldsymbol{X}})$, whose variance is:

$$\begin{aligned}\text{Var}\left\{\overline{\Delta}_a - \widehat{\boldsymbol{\beta}}_a^T(\overline{\boldsymbol{X}}_a - \overline{\boldsymbol{X}})\right\} &= \text{Var}(\overline{\Delta}_a) + \text{Var}\left\{\widehat{\boldsymbol{\beta}}_a^T(\overline{\boldsymbol{X}}_a - \overline{\boldsymbol{X}})\right\} - 2\text{Cov}\left\{\overline{\Delta}_a, \widehat{\boldsymbol{\beta}}_a^T(\overline{\boldsymbol{X}}_a - \overline{\boldsymbol{X}})\right\} \\ &= \text{Var}(\overline{\Delta}_a) - \text{Var}\left\{\widehat{\boldsymbol{\beta}}_a^T(\overline{\boldsymbol{X}}_a - \overline{\boldsymbol{X}})\right\}.\end{aligned}$$

Consequently, we can have $\text{Var}\left\{\overline{\Delta}_a - \widehat{\boldsymbol{\beta}}_a^T(\overline{\boldsymbol{X}}_a - \overline{\boldsymbol{X}})\right\} < \text{Var}(\overline{\Delta}_a)$. These results are formally stated in Theorem 3.

- *Theorem 3:* Under Assumption 1,
  - *(a)* $\sqrt{n}\{\hat{\theta}_{\mathrm{cr,adj}} - 2^{-1}(\theta_1 + \tilde{\theta}_2 - \lambda_0 - \lambda_1)\} \xrightarrow{\mathrm{d}} N(0, \tilde{\sigma}^2_{\mathrm{cr,adj}})$, where $\tilde{\sigma}^2_{\mathrm{cr,adj}} = (4\pi_1)^{-1}\mathrm{Var}\left(Y_{i1}^{(1)} - Y_{i2}^{(10)} - \boldsymbol{\beta}_1^T \boldsymbol{X}_i\right) + (4\pi_0)^{-1}\mathrm{Var}\left(Y_{i1}^{(0)} - Y_{i2}^{(01)} - \boldsymbol{\beta}_0^T \boldsymbol{X}_i\right) + 4^{-1}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)^T \mathrm{Var}(\boldsymbol{X})(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)$ where $\boldsymbol{\beta}_a = \mathrm{Var}(\boldsymbol{X}_i)^{-1}\mathrm{Cov}(\boldsymbol{X}_i, \Delta_i | A_i = a)$.
  - *(b)* Moreover, $\tilde{\sigma}^2_{\mathrm{cr,adj}} - \tilde{\sigma}^2_{\mathrm{cr}} = -(4\pi_1\pi_0)^{-1}(\pi_0\boldsymbol{\beta}_1 + \pi_1\boldsymbol{\beta}_0)^T \mathrm{Var}(\boldsymbol{X})(\pi_0\boldsymbol{\beta}_1 + \pi_1\boldsymbol{\beta}_0) \leq 0$.

Theorem 3 is proved by applying Theorem 1 and Corollary 1 in Ye et al. (2023)[22] with Δ_i as the outcome; see Appendix A1.3 for details. From Theorem 3(b), we see that the asymptotic variance of $\hat{\theta}_{\mathrm{cr,adj}}$ is no larger than that of $\hat{\theta}_{\mathrm{cr}}$, where the equality holds if and only if $\pi_0\boldsymbol{\beta}_1 + \pi_1\boldsymbol{\beta}_0 = 0$. This occurs, for instance, when the covariates are uncorrelated with the change in outcome Δ_i. In fact, Theorem 1 of Ye et al. (2023)[22] implies a stronger result that $\hat{\theta}_{\mathrm{cr,adj}}$ has the smallest asymptotic variance among all linearly adjusted estimators of the form $2^{-1}[\{\overline{\Delta}_1 - \boldsymbol{b}_1^T(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}})\} - \{\overline{\Delta}_0 - \boldsymbol{b}_0^T(\overline{\boldsymbol{X}}_0 - \overline{\boldsymbol{X}})\}]$, where $\boldsymbol{b}_1, \boldsymbol{b}_0$ are any fixed or random vectors that have the same dimension as $\boldsymbol{X}_i$. A consistent estimator of $\tilde{\sigma}^2_{\mathrm{cr,adj}}$ is

$$\hat{\sigma}^2_{\mathrm{cr,adj}} = \frac{1}{4\pi_1} S^2_{\Delta_{1,\mathrm{adj}}} + \frac{1}{4\pi_0} S^2_{\Delta_{0,\mathrm{adj}}} + \frac{1}{4}\left(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_0\right)^T \hat{\Sigma}_X \left(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_0\right),$$

where $S^2_{\Delta_{a,\mathrm{adj}}}$ is the sample variance of $Y_{i1} - Y_{i2} - \hat{\boldsymbol{\beta}}_a^T \boldsymbol{X}_i$ based on subjects under $A_i = a$ for $a = 1,0$, $\hat{\Sigma}_X$ is the sample covariance matrix of $\boldsymbol{X}_i$ based on the entire sample. One can easily construct a *Z*-test based on the covariate-adjusted estimator $\hat{\theta}_{\mathrm{cr,adj}}$, which from Theorem 3 is guaranteed to be more powerful than the unadjusted counterpart T_{cr}.

Correspondingly, for a parallel trial, the covariate-adjusted ANHECOVA estimator is

$$\hat{\theta}_{\text{pr,adj}} = \left\{\frac{1}{n_1}\sum_{i=1}^{n} A_i Y_{i1} - \hat{\boldsymbol{\gamma}}_1^T(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}})\right\} - \left\{\frac{1}{n_0}\sum_{i=1}^{n} A_i Y_{i1} - \hat{\boldsymbol{\gamma}}_0^T(\bar{\boldsymbol{X}}_0 - \bar{\boldsymbol{X}})\right\},$$

where for $A_i = a$ and $a = 1,0$, $\hat{\boldsymbol{\gamma}}_a$ is

$$\hat{\boldsymbol{\gamma}}_a = \left\{\sum\nolimits_{i:A_i=a} (\boldsymbol{X}_i - \bar{\boldsymbol{X}}_a)(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_a)^T\right\}^{-1}\left\{\sum\nolimits_{i:A_i=a} (\boldsymbol{X}_i - \bar{\boldsymbol{X}}_a)\right\}\left(\frac{1}{n_a}\sum_{A_i=a} Y_{i1}\right).$$

We can also derive asymptotic properties for $\hat{\theta}_{\text{pr,adj}}$. $\sqrt{n}\left(\hat{\theta}_{\text{pr,adj}} - 2^{-1}\theta_1\right) \xrightarrow{\text{d}} N\left(0, \tilde{\sigma}^2_{\text{pr,adj}}\right)$, where $\tilde{\sigma}^2_{\text{pr,adj}} = \pi_1^{-1}\text{Var}\left(Y_{i1}^{(1)} - \boldsymbol{\gamma}_1^T\boldsymbol{X}_i\right) + \pi_0^{-1}\text{Var}\left(Y_{i1}^{(0)} - \boldsymbol{\gamma}_0^T\boldsymbol{X}_i\right) + (\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_0)^T\text{Var}(\boldsymbol{X})(\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_0)$, and $\boldsymbol{\gamma}_a = \text{Var}(\boldsymbol{X}_i)^{-1}\text{Cov}(\boldsymbol{X}_i, Y_{i1}|A_i = a)$. A consistent estimator of $\tilde{\sigma}^2_{\text{pr,adj}}$ is $\hat{\sigma}_{\text{pr,adj}} = \pi_1^{-1}S^2_{1,\text{adj}} + \pi_0^{-1}S^2_{0,\text{adj}} + (\hat{\boldsymbol{\gamma}}_1 - \hat{\boldsymbol{\gamma}}_0)^T\ \hat{\Sigma}_X(\hat{\boldsymbol{\gamma}}_1 - \hat{\boldsymbol{\gamma}}_0)$.

Chapter 3. RESULTS

3.1 POWER CALCULATIONS

To compare the power of crossover and parallel design, we consider the following data-generating process based on a set of parametric models:

$$Y_{i1}^{(0)} = X_{i1} + X_{i2} + X_{i3} + \epsilon_{i1}$$
$$Y_{i1}^{(1)} = \theta_1 + X_{i1} + X_{i2} + X_{i3} + \epsilon_{i2}$$
$$Y_{i2}^{(10)} = \tilde{\tau} + \lambda_0 + X_{i1} + bX_{i2} + X_{i3} + \epsilon_{i3}$$
$$Y_{i2}^{(01)} = \tilde{\tau} + \tilde{\theta}_2 - \lambda_1 + X_{i1} + bX_{i2} + X_{i3} + \epsilon_{i4},$$

where $X_{ij}, \epsilon_{ik} \sim N(0,1)$ for $j = 1,2$ and $k = 1,2,3,4$ and X_3 is Bernoulli with $p = 1/2$. The treatment arm indicator A_i is Bernoulli with $\pi_1 = 1/2$. Following the consistency assumption, the

observed outcomes for subject i are $Y_{i1} = Y_{i1}^{(1)}, Y_{i2} = Y_{i2}^{(10)}$ if $A_i = 1$, and $Y_{i1} = Y_{i1}^{(0)}, Y_{i2} = Y_{i2}^{(01)}$ if $A_i = 0$, and then the observed data are $(\boldsymbol{X}_i, A_i, Y_{i1}, Y_{i2}), i = 1, \dots, n$. In this data-generating process, we can have $\mathrm{Var}\left(Y_{i1}^{(1)}\right) = \mathrm{Var}\left(Y_{i1}^{(0)}\right) = 3 + \frac{1}{4}$, $\mathrm{Var}\left(Y_{i2}^{(10)}\right) = \mathrm{Var}\left(Y_{i2}^{(01)}\right) = b^2 + 2 + \frac{1}{4}$, and $\mathrm{Cov}\left(Y_{i1}^{(1)}, Y_{i2}^{(10)}\right) = \mathrm{Cov}\left(Y_{i1}^{(0)}, Y_{i2}^{(01)}\right) = b + 1 + \frac{1}{4}$. So $\sigma_{\mathrm{pr}}^2 = 13$ and $\sigma_{\mathrm{cr}}^2 = (1-b)^2 + 2$. For the covariate-adjusted estimators, we adjust for $\boldsymbol{X}_i = (X_{i1}, X_{i2}, X_{i3})^T$. In this data-generating example, $\mathrm{Var}(\boldsymbol{X}_i) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/4 \end{pmatrix}$ and for the covariate-adjusted crossover estimator, $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_0 = (0, 1-b, 0)^T$, so $\tilde{\sigma}_{\mathrm{cr,adj}}^2 = 2$. For the covariate-adjusted parallel estimator, $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_0 = (1,1,1)^T$ and $\tilde{\sigma}_{\mathrm{pr,adj}}^2 = 4$.

We set $n = 550$ so that the power of T_{pr} is approximately 90% when $\theta_1 = 0.5$; we set $\tilde{\tau} = -0.05, \alpha = 0.025, \theta^* = 0$ and $b \in \{0, 1/3\}$. Figure 2 shows the type I error rate and power for four one-sided tests of superiority $T_{\mathrm{pr}}, T_{\mathrm{pr,adj}}, T_{\mathrm{cr}}, T_{\mathrm{cr,adj}}$ when $\theta_1 = \tilde{\theta}_2 = \theta = \{0, 0.05, 0.1, 0.15, \dots, 0.45, 0.5\}$ and $\lambda_1 = \lambda_0 = \lambda \in \{-0.1, 0, 0.1, 0.3\}$ based on the formula in Equations (4), (6) and their covariate-adjusted counterparts. The type I error rate and power obtained by simulation are also plotted in Figure 2. The R codes for the data generating process here and the R codes for all data analyses and data simulation in Section 3 can be found in Appendix A1.4. The powers of T_{pr} and $T_{\mathrm{pr,adj}}$ are not affected by λ, so they can be used to benchmark the performance of the other tests. The value of b only affects the power of T_{cr} but not the other three tests, so the formula powers of $T_{\mathrm{pr}}, T_{\mathrm{pr,adj}}, T_{\mathrm{cr,adj}}$ in Figure 2(a) and 2(b) are identical. All the simulation power curves well agree with the corresponding formula ones in Figure 2, and the simulation powers of $T_{\mathrm{pr}}, T_{\mathrm{pr,adj}}, T_{\mathrm{cr,adj}}$ seem unaffected by the values of b.

When $\theta = 0$, under the null hypothesis that the treatment effect is 0, the theoretical type I error rate of $T_{\text{pr}}, T_{\text{pr,adj}}$ are both equal to $\alpha = 0.025$. When $\lambda < 0$, the type I error rates of $T_{\text{cr}}, T_{\text{cr,adj}}$ are greater than α; the type I error rates are equal to α when $\lambda = 0$ and smaller than α when $\lambda > 0$, which means that T_{cr} and $T_{\text{cr,adj}}$ can control the type I error rate when $\lambda \geq 0$, and furthermore, T_{cr} and $T_{\text{cr,adj}}$ become more and more conservative as λ increases.

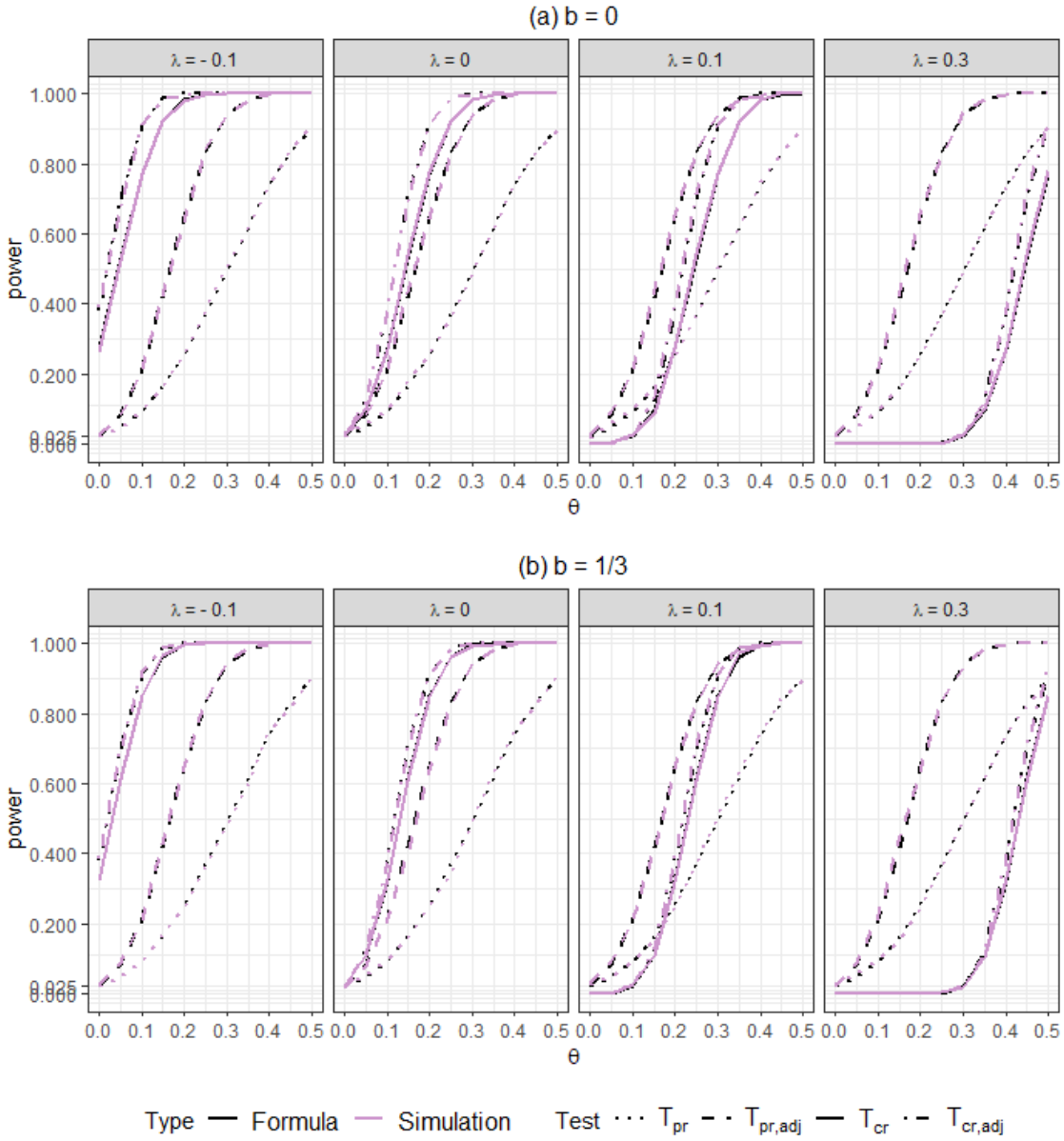

Figure 3: Power curves for four tests calculated by formulas and simulations (with 2,500 repetitions) when $\lambda = -0.1, 0, 0.1, 0.3$ and (a) $b = 0$, (b) $b = 1/3$.

When $\theta > 0$, under the alternative hypothesis, from (7) we have that the power comparison of T_{cr} and T_{pr} depends on the sign of $\lambda - \theta\left(1 - \sqrt{\frac{(1-b)^2+2}{13}}\right)$. Specifically, when $b = 0$, T_{cr} is more powerful than T_{pr} when $\lambda < 0.52\theta$. Hence, when $\lambda = 0$, we see that T_{cr} has substantially greater power than T_{pr} does; when $\lambda = 0.3$, the carry-over effect is large and greater than 0.52θ for all $\theta \in (0, 0.5]$, and T_{pr} has substantially greater power than T_{cr} does. Moreover, since a larger b leads to a larger ICC, it also leads to the slightly higher power of T_{cr}.

Lastly, covariate adjustment can result in greater power, regardless of whether the model is correct or not. The covariate-adjusted tests are always more powerful than the unadjusted counterparts. When $\lambda = 0$, the power of $T_{\mathrm{cr,adj}} > T_{\mathrm{cr}} > T_{\mathrm{pr,adj}} > T_{\mathrm{pr}}$; when $\lambda = 0.1$, the power of $T_{\mathrm{cr,adj}} > T_{\mathrm{pr,adj}} > T_{\mathrm{cr}}$ when $\theta \geq 0.35$.

3.2 Real Data Analysis Example

Laird et al. (1992)[10] discussed the analysis of a randomized two-treatment two-period crossover pharmacological study comparing two active analgesic drugs and placebo for relief of tension headachaes. Two active drugs (treatment A and B) have identical active ingredients of 1,000 mg of acetaminophen and treatment A contained caffeine. Caffeine alone is ineffective as an analgesic. In this study, researchers were interested in whether caffeine could work as na adjuvant medication. The placebo (P) was included for regulatory purposes. Patients were randomly allocated to 6 treatment sequences (AB, BA; AP, PA; BP, PB) and participated the study in 14 different clinical centers; see Table 3 for the number of patients per sequence at each clinical center. In each period,

two types of headache were treated, and the response variables were the average of each patient's two scores that measured the relief of pain for the two treat headaches; see Table 4 for the descriptive statistics by treatment sequences for total pain relief in each period.

Table 3: Sample sizes per treatment sequence at 14 clinical centers

	Treatment Sequence						
Clinical Center	Total (*n=423*)	AB (*n=126*)	BA (*n=127*)	AP (*n=43*)	PA (*n=43*)	BP (*n=42*)	PB (*n=42*)
1	30	9	9	3	3	3	3
2	29	9	8	3	3	3	3
3	31	9	10	3	3	3	3
4	31	9	9	4	3	3	3
5	29	9	8	3	3	3	3
6	30	9	9	3	3	3	3
7	30	9	9	3	3	3	3
8	31	9	9	3	4	3	3
9	31	9	10	3	3	3	3
10	30	9	9	3	3	3	3
11	30	9	9	3	3	3	3
12	30	9	9	3	3	3	3
13	30	9	9	3	3	3	3
14	31	9	10	3	3	3	3

Table 4: Descriptive statistics by treatment sequence (mean and standard deviation of the relief scores and treatment effects and intraclass correlation ρ between relief scores at two periods) for total pain relief scores of headache in period 1 and 2.

		Period 1		Period 2		
Sequence	*N*	Mean	SD	Mean	SD	ρ
AB	126	10.196	3.347	9.153	3.429	0.20
BA	127	9.581	3.881	10.791	3.530	0.30
AP	43	10.477	3.546	7.273	4.451	0.31
PA	43	8.366	3.777	10.855	3.204	0.47
BP	42	10.333	3.306	8.357	3.944	0.72
PB	42	7.464	4.265	9.911	4.183	0.67

The treatment outcomes at period in PB and PA sequences were unlikely to be affected by carry-over effects since the placebo were provided during the previous period. The 95% confidence intervals for the average treatment outcomes at period 2 of the AB and PB sequences are (8.552, 9.751) and (8.647, 11.176) respectively, which largely overlap with each other. Similarly, the 95% confidence intervals for the average treatment outcomes at period 2 of the BA and PA sequences are (10.178, 11.404) and (9.897, 11.813) respectively. Given the overlapping of the confidence intervals, we make the assumption that treatment A and B do not have carry-over effects. When comparing the treatment effects of A and B, we consider data from AB and BA sequences only.

We adjust on the clinical centers and for the treatment effect of treatment A from AB and BA sequences, we compare the parallel estimator, the covariate-adjusted parallel estimator, the crossover estimator and the covariate-adjusted crossover estimator, and the latter two are estimated from data of period 1 only. In Table 5, neither of the two 95% confidence intervals for the average treatment effects based on the two crossover estimators cover 0, while the ones based on the parallel estimators do. The crossover estimators have smaller standard errors than the parallel estimators do. The covariate-adjusted estimators have smaller standard errors than their unadjusted counterparts do.

Table 5: Clinical-center-adjusted and unadjusted parallel and crossover estimators for the average treatment effect of treatment A, with standard errors and 95% confidence intervals.

Type	Mean	SE	95% CI
Parallel, unadjusted	0.616	0.456	(-0.277, 1.509)
Parallel, adjusted	0.627	0.446	(-0.250, 1.504)
Crossover, unadjusted	1.127	0.273	(0.592, 1.663)
Crossover, adjusted	1.139	0.265	(0.621, 1.658)

3.3 A Simulation Study Based on Data from HPTN 084

In this section, we revisit the HIV prevention trial described in Section 1. In this example, we expect the carry-over effect to be nonnegative because if participants taking DPP at period 1 are more likely to develop adherence habits that affect their adherence at period 2, then we expect this persisting carry-over effect to result in better adherence at period 2. Such carry-over effect is behavioral instead of biological because it is due to the altering of the participant's drug-taking behavior. From the results in Section 2.2.3, we expect that a crossover design with nonnegative carry-over effects does not result in the inflation of type I error rate.

To evaluate the power trade-off between the crossover and parallel designs in the HIV prevention trial described in Section 1 and to demonstrate the potential efficiency gain from covariate adjustment, we simulate a randomized two-treatment two-period crossover trial comparing adherence to DPP versus 2PR based on data from HPTN 084 (ClinicalTrials.gov number, NCT03164564).[26] HPTN 084 is a randomized controlled trial comparing cabotegravir and TDF-FTC for HIV prevention among HIV-uninfected cisgender women. In the TDF-FTC arm, a randomly selected cohort of 405 participants were evaluated for their adherence to TDF-FTC during the course of the trial. Among women in this group, 336 participants had their adherence evaluated at week 12. Our simulations are bsed on the data of these 336 participants; see Table 5 for their baseline characteristics. The adherence variable is binary and defined as intraerythrocytic tenofovir-diphosphate (TFV-DP) in dried blood spots (DBS) ≥ 700 fmol/punch, roughly corresponding to ≥ 4 pills/week. We obtain their baseline covariates and their adherence to TDF-FTC at week 12. Four participants had missing values for both gonorrhea and chlamydia testing results. Since single imputation is valid for handling missing baseline covariate values[27] and the

proportion of positive gonorrhea or chlamydia testing results were low, we impute these four individuals' gonorrhea and chlamydia testing results as negative.

Table 6: Baseline characteristics of the 336 participants who had measures of adherence at week 12. Data are *n(%)* or mean (SD).

Total	336 (100%)
Country	
Botswana	8 (2.4%)
Eswatini	17 (5.1%)
Kenya	8 (2.4%)
Malawi	26 (7.7%)
South Africa	126 (37.5%)
Uganda	63 (18.8%)
Zimbabwe	88 (26.2%)
Age (in years)	26.5 (5.89)
Marital status	
Married / civil union / legal partnership	39 (11.6%)
Living with primary or main partner	30 (8.9%)
Having primary or main partner, not living together	171 (50.9%)
Single / divorced / widowed / other	96 (28.6%)
Gonorrhea	
Positive	20 (6.0%)
Negative	213 (92.9%)
Missing	4 (1.2%)
Chlamydia	
Positive	54 (16.1%)
Negative	278 (82.7%)
Missing	4 (1.2%)

In the simulation study, we set $\pi_1 = \pi_0 = \frac{1}{2}, \theta_1 = \tilde{\theta}_2 = \theta$ and $\lambda_1 = \lambda_0 = \lambda$, where $\theta \in \{0.05, 0.08, 0.1, 0.15\}$ for different treatment effects and $\lambda \in \left\{0, \frac{1}{4}\theta, \frac{1}{2}\theta\right\}$ for varying carry-over effects. We expect that fewer participants would adhere to the assigned treatments at period 2 for both treatment arms and thus set $\tilde{\tau} = -0.05$. We set $\text{ICC} = 0.33$, which is estimated from

participants' adherence rates to TDF-FTC at the first two visits in HPTN 084. Consider the four test statistics: $T_{\text{pr}}, T_{\text{pr,adj}}, T_{\text{cr}}, T_{\text{cr,adj}}$. The simulation process involves the following steps:

1) Denote the baseline covariates and the adherence variable at week 12 from HPTN 084 as $\left(\boldsymbol{X}_i, Y_{i1}^{(0)}\right), i = 1, \ldots, N$ with $N = 336$, and the mean of $Y_{i1}^{(0)}$ is 0.185. Fit a logistic regression model for the probability of adherence to TDF-FTC with age, current marital status and baseline gonorrhea and chlamydia testing results as covariates. The country variable is excluded from the covariates due to the low frequency of some of its categories, which may cause problematic issues during the resampling process later. Denote the fitted model as $\hat{\mu}_1(\boldsymbol{x}) = \text{expit}\left(\hat{\alpha}_0 + \widehat{\boldsymbol{\beta}}_1^T(\boldsymbol{x})\right)$, where $\text{expit}(x) = \exp(x)/\{1 + \exp(x)\}$.

2) Generate $Y_{i1}^{(1)} \sim \text{Bernoulli}\left(\text{expit}\left(\hat{\alpha}_1 + \widehat{\boldsymbol{\beta}}_1^T \boldsymbol{X}_i\right)\right)$, where $\hat{\alpha}_1$ satisfies $n^{-1}\sum_{i=1}^n \text{expit}\left(\hat{\alpha}_1 + \widehat{\boldsymbol{\beta}}_1^T \boldsymbol{X}_i\right) - n^{-1}\sum_{i=1}^n \text{expit}\left(\hat{\alpha}_0 + \widehat{\boldsymbol{\beta}}_1^T \boldsymbol{X}_i\right) = \theta$.

3) Generate $Y_{i2}^{(00)}$ from $Y_{i2}^{(00)}|Y_{i1}^{(0)} = 1 \sim \text{Bernoulli}\left(\frac{s}{0.185}\right)$ and $Y_{i2}^{(00)}|Y_{i1}^{(0)} = 0 \sim \text{Bernoulli}\left(\frac{0.185 + \tilde{\tau} - s}{1 - 0.185}\right)$ with $s = E\left(Y_{i2}^{(00)} Y_{i1}^{(0)}\right) = 0.185(0.185 + \tilde{\tau}) + \rho \cdot \sqrt{0.185(1 - 0.185)(0.185 + \tilde{\tau})(1 - 0.185 - \tilde{\tau})}$; a proof can be found in Appendix A1.4. Based on the generated $Y_{i2}^{(00)}$, fit a logistic regression model and denote the fitted model as $\hat{\mu}_2(\boldsymbol{x}) = \text{expit}\left(\hat{\alpha}_{00} + \widehat{\boldsymbol{\beta}}_2^T(\boldsymbol{x})\right)$.

4) Generate $Y_{i2}^{(10)} \sim \text{Bernoulli}\left(\text{expit}\left(\hat{\alpha}_{10} + \widehat{\boldsymbol{\beta}}_2^T \boldsymbol{X}_i\right)\right)$, $Y_{i2}^{(01)} \sim \text{Bernoulli}\left(\text{expit}\left(\hat{\alpha}_{01} + \widehat{\boldsymbol{\beta}}_2^T \boldsymbol{X}_i\right)\right)$ and $Y_{i2}^{(11)} \sim \text{Bernoulli}\left(\text{expit}\left(\hat{\alpha}_{11} + \widehat{\boldsymbol{\beta}}_2^T \boldsymbol{X}_i\right)\right)$, where $\hat{\alpha}_{10}, \hat{\alpha}_{01}, \hat{\alpha}_{11}$ satisfy $n^{-1}\sum_{i=1}^n \text{expit}\left(\hat{\alpha}_{10} + \widehat{\boldsymbol{\beta}}_2^T \boldsymbol{X}_i\right) - n^{-1}\sum_{i=1}^n \hat{\mu}_2(\boldsymbol{X}_i) = \lambda$, $n^{-1}\sum_{i=1}^n \text{expit}\left(\hat{\alpha}_{01} + \widehat{\boldsymbol{\beta}}_2^T \boldsymbol{X}_i\right) - n^{-1}\sum_{i=1}^n \hat{\mu}_2(\boldsymbol{X}_i) = \theta - \lambda$ and $n^{-1}\sum_{i=1}^n \text{expit}\left(\hat{\alpha}_{11} + \widehat{\boldsymbol{\beta}}_2^T \boldsymbol{X}_i\right) - n^{-1}\sum_{i=1}^n \hat{\mu}_2(\boldsymbol{X}_i) = \theta$. Byy

now, we have obtained the covariates and all 6 potential outcomes $\left(\boldsymbol{X}_i, Y_{i1}^{(0)}, Y_{i1}^{(1)}, Y_{i2}^{(00)}, Y_{i2}^{(01)}, Y_{i2}^{(11)}, Y_{i2}^{(10)}\right)$ for 336 individuals.

5) A random sample of size n is drawn from the 336 individuals' covariates and potential outcomes with replacement. Each of the obtained n subjects is assigned randomly to $A_i = 0$ or 1 with equal probability. For $A_i = 1$, under the consistency assumption, $Y_{i1} = Y_{i1}^{(1)}$ and $Y_{i2} = Y_{i2}^{(10)}$; for $A_i = 0$, $Y_{i1} = Y_{i1}^{(0)}$ and $Y_{i2} = Y_{i2}^{(01)}$. Hence, we obtain the observed data $(A_i, \boldsymbol{X}_i, Y_{i1}, Y_{i2}), i = 1, \dots, n$ and calculate the test statistics $T_{\text{pr}}, T_{\text{pr,adj}}, T_{\text{cr}}, T_{\text{cr,adj}}$.

6) Steps (2)-(5) are repeated 1,000 times, from which we obtain the empirical powers of the test statistics.

Table 6 shows the empirical powers for $T_{\text{pr}}, T_{\text{pr,adj}}, T_{\text{cr}}, T_{\text{cr,adj}}$ for different values of θ and λ based on 1,000 simulations. The sample sizes $n = 380, 770, 2780$ are determined such that the power of T_{cr} is close to 0.9 when $\theta = 0.10$. First, in Table 6 wen can tell that the empirical power increases as θ increases. Second, under both parallel and crossover designs, covariate-adjustment always results in a larger power. Third, when $\lambda = 0$, the crossover tests have considerably larger powers than the parallel tests do. When $\lambda = \theta/4$, the crossover tests still obtain larger powers but the differences in powers are not as pronounced as when $\lambda = 0$. When $\lambda = \theta/2$, the crossover tests are less powerful.

Table 7: Empirical powers for $T_{\text{pr}}, T_{\text{pr,adj}}, T_{\text{cr}}, T_{\text{cr,adj}}$

n	λ	θ	power_{pr}	$\text{power}_{\text{pr,adj}}$	power_{cr}	$\text{power}_{\text{cr,adj}}$
380	0	0.05	0.232	0.254	0.443	0.455
		0.08	0.457	0.495	0.762	0.778
		0.10	0.639	0.664	**0.901**	0.906
		0.15	0.895	0.916	0.994	0.996
770	$\theta/4$	0.05	0.451	0.484	0.495	0.516
		0.08	0.698	0.731	0.796	0.799
		0.10	0.853	0.873	**0.901**	0.907
		0.15	0.988	0.991	0.991	0.992
2780	$\theta/2$	0.05	0.770	0.797	0.697	0.613
		0.08	0.964	0.971	0.798	0.800
		0.10	0.995	0.996	**0.909**	0.915
		0.15	1.000	1.000	0.992	0.992

Chapter 4. DISCUSSION

4.1 TIME CONSIDERATIONS

In previous sections, the crossover trial is preferred when it requires smaller sample sizes to achieve the desired statistical power. Meanwhile, time is another important factor contributing to the cost of a trial.[27] Although we show that a crossover trial is more powerful than a parallel one that recruits the same number of participants when the carry-over effect is relatively small, the total observation time in a crossover trial is often twice as long as the one in a parallel trial. Consider the following models for various times.[28] Let t denote the length of each treatment period, v denote the run-in period and let w denote the washout period. Let r_1 and r_2 denote the average time between recruitment of individual patients for the parallel and crossover trials respectively. Let p and c denote the number of participants recruited for the parallel and the crossover designs. A simple expression for the time that each participant spends in the trial is $v +$

t under the parallel design and $v + 2t + w$ under the crossover design, and the extra time each participant enrolled in the crossover trial is one treatment period plus the washout period. The total investigation time that investigators have to dedicate are $2p(v + t)$ and $2c(v + 2t + w)$ for the parallel and crossover designs respectively, and the total trial time are $2pr_1 + v + t$ and $2cr_2 + v + 2t + w$ respectively.

When the crossover trial has shorter total investigation time and $2p(v + t) \geq 2c(v + 2t + w)$, we have $\frac{p}{c} \geq 1 + \frac{w+t}{v+t}$. When a washout period is designed, it might be able to sufficiently diminish the carry-over effect, but it may increase the total investigation time for the crossover trial; when there is no washout period, the carry-over effect is likely to exist, which requires larger sample sizes and increases the total investigation time. Consider the situation similar to the analgesic example discussed in Section 3.2 where it is eligible to assume that carry-over effects could be diminished via the washout period. Under the setting of the simple example at the end of Section 2.1.3, if we set $\frac{p}{c} = \frac{n_{\mathrm{pr}}}{n_{\mathrm{cr}}} = \frac{\sigma_{\mathrm{pr}}^2}{\sigma_{\mathrm{cr}}^2} = \frac{2}{1-\rho}$, then ideally we need to have the washout period w to be long enough to sufficiently diminish the carry-over effect and $w \leq \frac{2}{1-\rho}(v + t) - (v + 2t)$ as well; when $\rho = 0$, $w \leq v$. Consider a scenario similar to the hypothetical HIV prevention trial discussed in Section 3.3 where there is no washout period and $w = 0$, but there exist carry-over effects. Suppose $\lambda_1 = \lambda_0 = \lambda$ and suppose $\frac{p}{c} = \frac{n_{\mathrm{pr}}}{\tilde{n}_{\mathrm{cr}}} = \frac{2}{1-\rho}\left(1 - \frac{\lambda}{\theta}\right)^2$ as defined at the end of Section 2.2.3. When the crossover trial has shorter total investigation time, we need the ratio of carry-over effect and treatment effect to be $0 \leq \frac{\lambda}{\theta} \leq 1 - \sqrt{\frac{v+2t}{v+t} \cdot \frac{1-\rho}{2}}$; when $\rho = 0$, $0 \leq \frac{\lambda}{\theta} \leq 1 - \sqrt{\frac{v+2t}{2v+2t}}$.

When the crossover trial has shorter total trial time and $2pr_1 + v + t \geq 2cr_2 + v + 2t + w$, we need $p \geq c \cdot \frac{r_2}{r_1} + \frac{t+w}{2r_1}$. Similarly, the washout period and the carry-over effect could also affect the total trial time. Consider a simple case with $r_1 = r_2 = r$. When it is eligible to assume that the carry-over effect can be eliminated by the washout period, if we set $\frac{p}{c} = \frac{2}{1-\rho}$, then we need to have w to be long enough to sufficiently diminish the carry-over effect and $w \leq \frac{2rc(1+\rho)}{t(1-\rho)}$; when $\rho = 0$, $w \leq \frac{2rc}{t}$. When there is no washout period but exist carry-over effects, if we set $\frac{p}{c} = \frac{2}{1-\rho}\left(1 - \frac{\lambda}{\theta}\right)^2$, then $0 \leq \frac{\lambda}{\theta} \leq 1 - \sqrt{\left(\frac{t}{2rc} + 1\right) \cdot \frac{1-\rho}{2}}$; when $\rho = 0$, $0 \leq \frac{\lambda}{\theta} \leq 1 - \sqrt{\frac{t}{4rc} + \frac{1}{2}} < \frac{1}{2}$.

However, the estimation for recruitment rate is crucial in clinical trial design[29] and often whether $\frac{r_2}{r_1} \geq 1$ remains uncertain. Participants may find the crossover trial more inconvenient as they have to spend longer time under observation and submit to a number of various treatments and would thus be less willing to enter the study, which may result in a lower recruitment rate and $\frac{r_2}{r_1} > 1$. However, there are occassions when, for instance in some other studies of analgesics for migraine,[30] the guaranteed opportunity to try each treatment allows participants to assess their benefit/tolerability ratios of different treatments or doses and therefore the crossover trial is preferred.

4.2 CONCLUSION

The crossover trial is an efficient design as it uses participants as their own controls. The carry-over effect, especially the behavioral carry-over effect, is an outstanding concern because it can

bias the estimation of the treatment effect and affect the efficiency of the trial. Under a potential outcome framework, we investigate the impact of the carry-over effect for the two-treatment two-period crossover trial. We find that when the carry-over effect $\lambda_1 + \lambda_0 > 0$, the classic estimator is biased and underestimates the treatment effect, which does not inflate the type I error rate of one-sided tests of superiority or non-inferiority but decreases the power. Meanwhile, when $\lambda_1 + \lambda_0$ satisfies some sign condition and is relatively small compared to the treatment effect, the crossover design can still be more powerful than the parallel design and require less sample sizes. We can further apply covariate adjustment in crossover trials for guaranteed efficiency gain. Meanwhile, although the crossover trial requires less sample sizes, it requires longer observation time, which may result in higher cost. We show that carry-over effects and washout periods can affect the total investigation time and the total trial time; recruitment rate is another consideration in choosing the design and it could affect the total trial time as well. All methods in this thesis can be implemented via the R package RobinCar, which is available at https://github.com/tye27/RobinCar.

REFERENCES

1. Hills, M. and Armitage, P. (1979). The two-period cross-over clinical trial. *British Journal of Clinical Pharmacology* **8**, 7.
2. Senn, S. (1994). The AB/BA crossover: past, present and future? *Statistical methods in medical research* **3**, 303–324.
3. Kim, K., Bretz, F., Cheung, Y. K. K., and Hampson, L. V. (2021). *Handbook of statistical methods for randomized controlled trials*. CRC Press.
4. Jones, B. and Lewis, J. (1995). The case for cross-over trials in phase III. *Statistics in Medicine* **14,** 1025–1038.
5. Minnis, A. M., Roberts, S. T., Agot, K., Weinrib, R., Ahmed, K., Manenzhe, K., Owino, F., and van der Straten, A. (2018). Young women's ratings of three placebo multipurpose prevention technologies for HIV and pregnancy prevention in a randomized, cross-over study in kenya and south africa. *AIDS and Behavior* **22,** 2662–2673.
6. Hemming, K., Taljaard, M., Weijer, C., and Forbes, A. B. (2020). Use of multiple period, cluster randomised, crossover trial designs for comparative effectiveness research. *bmj* 371,.
7. Grizzle, J. E. (1965). The two-period change-over design and its use in clinical trials. *Biometrics* pages 467–480.
8. Freeman, P. (1989). The performance of the two-stage analysis of two-treatment, two-period crossover trials. *Statistics in medicine* **8,** 1421–1432.
9. Brown Jr, B. W. (1980). The crossover experiment for clinical trials. *Biometrics* pages 69-79.
10. Laird, N. M., Skinner, J., and Kenward, M. (1992). An analysis of two-period crossover designs with carry-over effects. *Statistics in Medicine* **11,** 1967–1979.
11. Jones, B. and Donev, A. (1996). Modelling and design of cross-over trials. *Statistics in Medicine* **15,** 1435–1446.
12. Kunert, J. and Stufken, J. (2002). Optimal crossover designs in a model with self and mixed carryover effects. *Journal of the American Statistical Association* **97,** 898–906.
13. Bailey, R. and Kunert, J. (2006). On optimal crossover designs when carryover effects are proportional to direct effects. *Biometrika* **93,** 613–625.
14. Neyman, J. (1923). On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science* **5,** 465–472. Trans. Dorota M. Dabrowska and Terence P. Speed (1990).
15. Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **6,** 688–701.
16. Jones, B. and Kenward, M. G. (1989). *Design and analysis of cross-over trials*. Chapman and Hall/CRC.
17. Rosenbaum, P. R. (2020). *Design of observational studies (2nd ed.)*. Springer.

18. FDA (2021). Adjusting for covariates in randomized clinical trials for drugs and biological products. Draft Guidance for Industry. Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research, Food and Drug Administration (FDA), U.S. Department of Health and Human Services. May 2021.
19. Yang, L. and Tsiatis, A. A. (2001). Efficiency study of estimators for a treatment effect in a pretest–posttest trial. *The American Statistician* **55,** 314–321.
20. Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining freedman's critique. *Annals of Applied Statistics* **7,** 295–318.
21. Ye, T., Yi, Y., and Shao, J. (2022). Inference on the average treatment effect under minimization and other covariate-adaptive randomization methods. *Biometrika* **109,** 33–47.
22. Ye, T., Shao, J., Yi, Y., and Zhao, Q. (2023). Toward better practice of covariate adjustment in analyzing randomized clinical trials. *Journal of the American Statistical Association* pages 1–13.
23. Metcalfe, C. (2010). The analysis of cross-over trials with baseline measurements. *Statistics in medicine* **29,** 3211–3218.
24. Mehrotra, D. V. (2014). A recommended analysis for 2× 2 crossover trials with baseline measurements. *Pharmaceutical statistics* **13,** 376–387.
25. Jemielita, T., Putt, M., and Mehrotra, D. (2016). Improved power in crossover designs through linear combinations of baselines. *Statistics in Medicine* **35,** 5625–5641.
26. Delany-Moretlwe, S., Hughes, J. P., Bock, P., Hunidzarira, P., Kalonji, D., Kayange, N., Makhema, J., Mandima, P., Mathew, C., Spooner, E., Mpendo, J., Mukwekwerere, P., Mgodi, N., Ntege, P. N., Nakabiito, C., Nuwagaba-Biribonwoha, H., Panchia, R., Singh, N., Farrior, J., Rose, S., Anderson, P. L., Eshleman, S. H., Marzinke, M. A., Hendrix, C. W., Beigel-Orme, S., Hosek, S., Tolley, E., Sista, N., Adeyeye, A., Rooney, J. F., Rinehart, A., Spreen, W. R., Hanscom, B., Cohen, M. S., and Hosseinipour, M. C. (2022). Cabotegravir for the prevention of HIV-1 in women: results from HPTN 084, a phase 3, randomised clinical trial. *The Lancet (British Edition)* **399,** 1779–1789.
27. Bentley, C., Cressman, S., van der Hoek, K., Arts, K., Dancey, J., & Peacock, S. (2019). Conducting clinical trials—costs, impacts, and the value of clinical trials networks: A scoping review. *Clinical Trials*, **16**(2), 183–193.
28. Senn, S. (2002). *Cross-over Trials in Clinical Research*, John Wiley & Sons.
29. Johnson, A. L. (1989) Methodology of clinical trials in psychiatry, in G. Freeman & P. Tyrer (eds), *Research Methods in Psychiatry: A Beginner's Guide*. Royal College of Psychiatrists, London.
30. Tfelt-Hansen, P., Pascual, J., Ramadan, N., Dahlöf, C., D'Amico, D., Diener, H.-C., Hansen, J. M., Lanteri-Minet, M., Loder, E., McCrory, D., Plancade, S., & Schwedt, T. (2012). Guidelines for controlled trials of drugs in migraine: Third edition. A guide for investigators. *Cephalalgia*, **32**(1), 6–38.

APPENDIX

A.1 TECHNICAL PROOFS

A.1.1 Proof of Theorem 1

Proof. (a) Let $\tau = E\left(Y_{i2}^{(0)} - Y_{i1}^{(0)}\right)$ denote the expected temporal trend in the absence of treatment. Then,

$$E\left(\hat{\theta}_{\text{cr}}\right) = \frac{1}{2}E\left\{\frac{1}{n_1}\sum_{i=1}^{n} A_i(Y_{i1} - Y_{i2}) + \frac{1}{n_0}\sum_{i=1}^{n}(1 - A_i)(Y_{i2} - Y_{i1})\right\}$$

$$= \frac{1}{2}E\left\{\frac{1}{n_1}\sum_{i=1}^{n} A_i\left(Y_{i1}^{(1)} - Y_{i2}^{(0)}\right) + \frac{1}{n_0}\sum_{i=1}^{n}(1 - A_i)\left(Y_{i2}^{(1)} - Y_{i1}^{(0)}\right)\right\}$$

$$= \frac{1}{2}E\left\{\frac{1}{n_1}\sum_{i=1}^{n} A_i\left(Y_{i1}^{(1)} - Y_{i1}^{(0)} + Y_{i1}^{(0)} - Y_{i2}^{(0)}\right) + \frac{1}{n_0}\sum_{i=1}^{n}(1 - A_i)\left(Y_{i2}^{(1)} - Y_{i2}^{(0)} + Y_{i2}^{(0)} - Y_{i1}^{(0)}\right)\right\}$$

$$= \frac{1}{2}(\theta_1 - \tau) + \frac{1}{2}(\theta_2 + \tau) = \frac{\theta_1 + \theta_2}{2}$$

(b) Asymptotic normality is straightforward from the central limit theorem. In what follows, we derive the asymptotic variance of $\sqrt{n}\left(\hat{\theta}_{\text{cr}} - \frac{\theta_1+\theta_2}{2}\right)$. Let $M_1 \coloneqq \hat{\theta}_{\text{cr}} - \frac{\theta_1+\theta_2}{2}$. Note that:

$$\hat{\theta}_{\text{cr}} - \frac{\theta_1 + \theta_2}{2} = \underbrace{\frac{1}{2}\left\{\frac{1}{n_1}\sum_{i=1}^{n} A_i\left(Y_{i1}^{(1)} - Y_{i2}^{(0)} - \theta_1 + \tau\right) + \frac{1}{n_0}\sum_{i=1}^{n}(1 - A_i)\left(Y_{i2}^{(1)} - Y_{i1}^{(0)} - \theta_2 - \tau\right)\right\}}_{M_1}$$

We can show that:

$$\sqrt{n}M_1 | A_1, \ldots, A_n \xrightarrow{\text{d}} N\left(0, \frac{\text{Var}\left(Y_{i1}^{(1)} - Y_{i2}^{(0)}\right)}{4\pi_1} + \frac{\text{Var}\left(Y_{i2}^{(1)} - Y_{i1}^{(0)}\right)}{4\pi_0}\right).$$

From the bounded convergence theorem, this result still holds unconditionally, i.e.,

$$\sqrt{n}\left(\hat{\theta}_{\mathrm{cr}}-\frac{\theta_1+\theta_2}{2}\right)\xrightarrow{\mathrm{d}}N\left(0,\frac{\mathrm{Var}\left(Y_{i1}^{(1)}-Y_{i2}^{(0)}\right)}{4\pi_1}+\frac{\mathrm{Var}\left(Y_{i2}^{(1)}-Y_{i1}^{(0)}\right)}{4\pi_0}\right).$$

A.1.2 Proof of Theorem 2

Proof. (a) We calculate the expectation as follows:

$$E\left(\hat{\theta}_{\mathrm{cr}}\right)=\frac{1}{2}E\left\{\frac{1}{n_1}\sum_{i=1}^{n}A_i\left(Y_{i1}^{(1)}-Y_{i2}^{(10)}\right)+\frac{1}{n_0}\sum_{i=1}^{n}(1-A_i)\left(Y_{i2}^{(01)}-Y_{i1}^{(0)}\right)\right\}$$

$$=\frac{1}{2}E\left\{\frac{1}{n_1}\sum_{i=1}^{n}A_i\left(Y_{i1}^{(1)}-Y_{i1}^{(0)}+Y_{i1}^{(0)}-Y_{i2}^{(00)}+Y_{i2}^{(00)}-Y_{i2}^{(10)}\right)\right.$$

$$\left.+\frac{1}{n_0}\sum_{i=1}^{n}(1-A_i)\left(Y_{i2}^{(01)}-Y_{i2}^{(11)}+Y_{i2}^{(11)}-Y_{i2}^{(00)}+Y_{i2}^{(00)}-Y_{i1}^{(0)}\right)\right\}$$

$$=\frac{1}{2}(\theta_1-\tilde{\tau}-\lambda_0)+\frac{1}{2}\left(\tilde{\theta}_2-\lambda_1+\tilde{\tau}\right)=\frac{1}{2}\left(\theta_1+\tilde{\theta}_2-\lambda_0-\lambda_1\right).$$

(b) Asymptotic normality is straightforward from the central limit theorem. In what follows, we derive the asymptotic variance of $\sqrt{n}\left\{\hat{\theta}_{\mathrm{cr}}-2^{-1}\left(\theta_1+\tilde{\theta}_2-\lambda_0-\lambda_1\right)\right\}$. Similar to the proof of Theorem 1(b), we have:

$$\sqrt{n}\left(\hat{\theta}_{\mathrm{cr}}-2^{-1}\left(\theta_1+\tilde{\theta}_2-\lambda_0-\lambda_1\right)\right)\xrightarrow{\mathrm{d}}N\left(0,\frac{\mathrm{Var}\left(Y_{i1}^{(1)}-Y_{i2}^{(10)}\right)}{4\pi_1}+\frac{\mathrm{Var}\left(Y_{i2}^{(01)}-Y_{i1}^{(0)}\right)}{4\pi_0}\right).$$

A.1.3 Proof of Theorem 3

(a) Theorem 3 is proved via applying Theorem 1 and Corollary 1 in Ye et al. (2023) with Δ_i being the outcome. From $\frac{1}{n}\sum_{i=1}^{n} A_i(X_i - \bar{X}) = \frac{n_1}{n}(X_i - \bar{X}) = O_p\left(n^{-\frac{1}{2}}\right)$ and $\widehat{\boldsymbol{\beta}}_a = \boldsymbol{\beta}_a + o_p(1)$ from Lemma 3 in Ye et al. (2023), we have that

$$\hat{\theta}_{\text{cr}} = \frac{1}{2}\left[\frac{1}{n_1}\sum_{i=1}^{n} A_i\{Y_{i1} - Y_{i2} - \boldsymbol{\beta}_1^T(X_i - \bar{X})\} - \frac{1}{n_1}\sum_{i=1}^{n}(1 - A_i)\{Y_{i1} - Y_{i2} - \boldsymbol{\beta}_0^T(X_i - \bar{X})\}\right] + o_p\left(n^{-\frac{1}{2}}\right).$$

Then,

$$\begin{aligned}
&\hat{\theta}_{\text{cr}} - \left\{\frac{1}{2}\left(\theta_1 + \tilde{\theta}_2 - \lambda_0 - \lambda_1\right)\right\} \\
&= \frac{1}{2n_1}\sum_{i=1}^{n} A_i\{Y_{i1}^{(1)} - Y_{i2}^{(10)} - (\theta_1 - \tilde{\tau} - \lambda_0) - \boldsymbol{\beta}_1^T(X_i - \bar{X})\} \\
&- \frac{1}{2n_0}\sum_{i=1}^{n}(1 - A_i)\{Y_{i1}^{(0)} - Y_{i2}^{(01)} - \left(\tilde{\theta}_2 + \tilde{\tau} - \lambda_1\right) - \boldsymbol{\beta}_0^T(X_i - \bar{X})\} + o_p\left(n^{-\frac{1}{2}}\right) \\
&= \underbrace{\frac{1}{2n_1}\sum_{i=1}^{n} A_i\{Y_{i1}^{(1)} - Y_{i2}^{(10)} - (\theta_1 - \tilde{\tau} - \lambda_0) - \boldsymbol{\beta}_1^T(X_i - \mu_X)\}}_{M_{11}} \\
&- \underbrace{\frac{1}{2n_0}\sum_{i=1}^{n}(1 - A_i)\{Y_{i1}^{(0)} - Y_{i2}^{(01)} - \left(\tilde{\theta}_2 + \tilde{\tau} - \lambda_1\right) - \boldsymbol{\beta}_0^T(X_i - \mu_X)\}}_{M_{12}} \\
&+ \underbrace{\frac{1}{2}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)^T(\bar{X} - \mu_X)}_{M_2} + o_p\left(n^{-\frac{1}{2}}\right)
\end{aligned}$$

Consider the random vector

$$\sqrt{n}\begin{pmatrix} E_n\left[A_i\left\{Y_{i1}^{(1)} - Y_{i2}^{(10)} - (\theta_1 - \tilde{\tau} - \lambda_0) - \boldsymbol{\beta}_1^T(X_i - \mu_X)\right\}\right] \\ E_n\left[(1-A_i)\left\{Y_{i1}^{(0)} - Y_{i2}^{(01)} - \left(\tilde{\theta}_2 + \tilde{\tau} - \lambda_1\right) - \boldsymbol{\beta}_0^T(X_i - \mu_X)\right\}\right] \\ E_n[(X_i - \mu_X)] \end{pmatrix},$$

where $E_n[K_i] \coloneqq \frac{1}{n}\sum_{i=1}^n K_i$. Since every component in the above vector is an average of independent terms, by the assumption that all second moments are finite, the central limit theorem justifies that this vector is asymptotically normal with mean 0 as $n \to \infty$. This implies that $\sqrt{n}(M_{11} + M_{12} + M_2)$ is asymptotically normal.

In what follows, we derive the asymptotic variance of $\hat{\theta}_{\text{cr}} - \left\{\frac{1}{2}\left(\theta_1 + \tilde{\theta}_2 - \lambda_0 - \lambda_1\right)\right\}$. Note that:

$$\sqrt{n}(M_{11} - M_{12})|A_1, \dots, A_n \xrightarrow{\text{d}} N\left(0, \frac{\text{Var}\left(Y_{i1}^{(1)} - Y_{i2}^{(10)} - \boldsymbol{\beta}_1^T\boldsymbol{X}_i\right)}{4\pi_1} + \frac{\text{Var}\left(Y_{i1}^{(0)} - Y_{i2}^{(01)} - \boldsymbol{\beta}_0^T\boldsymbol{X}_i\right)}{4\pi_0}\right),$$

$$\sqrt{n}M_2|A_1, \dots, A_n \xrightarrow{\text{d}} N\left(0, \frac{1}{4}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)^T\text{Var}(X)(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)\right).$$

Therefore,

$$\sqrt{n}\left\{\hat{\theta}_{\text{cr,adj}} - 2^{-1}\left(\theta_1 + \tilde{\theta}_2 - \lambda_0 - \lambda_1\right)\right\} \xrightarrow{\text{d}} N\left(0, \tilde{\sigma}_{\text{cr,adj}}^2\right),$$

where $\tilde{\sigma}_{\text{cr,adj}}^2 = (4\pi_1)^{-1}\text{Var}\left(Y_{i1}^{(1)} - Y_{i2}^{(10)} - \boldsymbol{\beta}_1^T\boldsymbol{X}_i\right) + (4\pi_0)^{-1}\text{Var}\left(Y_{i1}^{(0)} - Y_{i2}^{(01)} - \boldsymbol{\beta}_0^T\boldsymbol{X}_i\right) + 4^{-1}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)^T\text{Var}(\boldsymbol{X})(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)$.

(b) This part is a direct consequence of

$$(4\pi_1)^{-1}\mathrm{Var}\left(Y_{i1}^{(1)} - Y_{i2}^{(10)} - \boldsymbol{\beta}_1^T\boldsymbol{X}_i\right) + (4\pi_0)^{-1}\mathrm{Var}\left(Y_{i1}^{(0)} - Y_{i2}^{(01)} - \boldsymbol{\beta}_0^T\boldsymbol{X}_i\right)$$
$$+ 4^{-1}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)^T\mathrm{Var}(\boldsymbol{X})(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)$$
$$= (4\pi_1)^{-1}\mathrm{Var}\left(Y_{i1}^{(1)} - Y_{i2}^{(10)}\right) + (4\pi_0)^{-1}\mathrm{Var}\left(Y_{i1}^{(0)} - Y_{i2}^{(01)}\right)$$
$$- 4^{-1}(\boldsymbol{\beta}_1 + \boldsymbol{\beta}_0)^T\mathrm{Var}(\boldsymbol{X})(\boldsymbol{\beta}_1 + \boldsymbol{\beta}_0)$$

from applying the definitions of $\boldsymbol{\beta}_1, \boldsymbol{\beta}_0$.

A.1.4 Generating correlated binary variables

Lemma A1: Consider a binary variable $Z_1 \sim \mathrm{Bernoulli}(p_1)$ where $\Pr(Z_1 = 1) = p_1$ and $p_1 \in (0,1)$. Given some $p_2 \in (0,1)$ and

$$-1 \le \max\left\{-\sqrt{\frac{p_1p_2}{(1-p_1)(1-p_2)}}, -\sqrt{\frac{(1-p_1)(1-p_2)}{p_1p_2}}\right\} < \rho < \min\left\{\sqrt{\frac{p_1(1-p_2)}{p_2(1-p_1)}}, \sqrt{\frac{p_2(1-p_1)}{p_1(1-p_2)}}\right\} \le 1,$$

if $Z_2|Z_1 = 1 \sim \mathrm{Bernoulli}\left(\frac{s}{p_1}\right)$ and $Z_2|Z_1 = 0 \sim \mathrm{Bernoulli}\left(\frac{p_2 - s}{1-p_1}\right)$ where $s = \rho\sqrt{p_1(1-p_1)} \cdot \sqrt{p_2(1-p_2)} + p_1p_2$, then $E(Z_2) = p_2$ and $\mathrm{Cor}(Z_1, Z_2) = \rho$.

Proof: First, it is straightforward to show that when ρ satisfies the aforementioned condition, it implies that:

$$\begin{cases} -1 \le \rho \le 1 \\ 0 < \frac{s}{p_1} < 1 \\ 0 < \frac{p_2 - s}{1-p_1} < 1 \end{cases}.$$

Then, we could also show that:

$$\Pr(Z_2 = 1|Z_1 = 1) = \frac{\Pr(Z_2 = 1, Z_1 = 1)}{\Pr(Z_1 = 1)} = \frac{s}{p_1} \Rightarrow \Pr(Z_2 = 1, Z_1 = 1) = s;$$

$$\Pr(Z_2 = 1|Z_1 = 0) = \frac{\Pr(Z_2 = 1, Z_1 = 0)}{\Pr(Z_1 = 0)} = \frac{p_2 - s}{1 - p_1} \Rightarrow \Pr(Z_2 = 1, Z_1 = 0) = p_2 - s.$$

This implies that:

$\Pr(Z_2 = 1) = \Pr(Z_2 = 1, Z_1 = 1) + \Pr(Z_2 = 1, Z_1 = 0) = p_2 = E(Z_2)$.

Then we have $\mathrm{Var}(Z_2) = p_2(1 - p_2)$. Given that $\mathrm{Var}(Z_1) = p_1(1 - p_1)$, we have:

$$\mathrm{Cor}(Z_1, Z_2) = \frac{\mathrm{Cov}(Z_1, Z_2)}{\sqrt{\mathrm{Var}(Z_1)\mathrm{Var}(Z_2)}} = \frac{E(Z_1 Z_2) - p_1 p_2}{\sqrt{p_1(1 - p_1)p_2(1 - p_2)}} = \frac{\Pr(Z_1 = 1, Z_2 = 1) - p_1 p_2}{\sqrt{p_1(1 - p_1)p_2(1 - p_2)}}$$

$$= \frac{s - p_1 p_2}{\sqrt{p_1(1 - p_1)p_2(1 - p_2)}} = \rho.$$

A.2 Two Sample T-Tests for Treatment Effects

To test $H_0: 2^{-1}(\theta_1 + \theta_2) = \theta^*$ versus $H_A: 2^{-1}(\theta_1 + \theta_2) > \theta^*$, under assumptions of normality and independence, when it is eligible to assume $\mathrm{Var}\left(Y_{i1}^{(1)} - Y_{i1}^{(0)}\right) = \mathrm{Var}\left(Y_{i2}^{(01)} - Y_{i2}^{(10)}\right)$, the test statistic based on the classic estimator is

$$T_{\mathrm{T,cr}} = \frac{\hat{\theta}_{\mathrm{cr}} - \theta^*}{\sqrt{\frac{(n_1 - 1)S_{\Delta_1}^2 + (n_0 - 1)S_{\Delta_0}^2}{n_1 + n_0 - 2}} \cdot \sqrt{\frac{1}{4n_1} + \frac{1}{4n_0}}},$$

and $T_{\mathrm{T,cr}}$ follows a t-distribution with $n_1 + n_0 - 2$ degrees of freedom.

When $\mathrm{Var}\left(Y_{i1}^{(1)} - Y_{i1}^{(0)}\right) \neq \mathrm{Var}\left(Y_{i2}^{(01)} - Y_{i2}^{(10)}\right)$, we can perform a Welch's t-test and the test statistic is

$$T_{\mathrm{T}^*,\mathrm{cr}} = \frac{\hat{\theta}_{\mathrm{cr}} - \theta^*}{\sqrt{\frac{S_{\Delta_1}^2}{4n_1} + \frac{S_{\Delta_0}^2}{4n_0}}}$$

and $T_{\mathrm{T}^*,\mathrm{cr}}$ follows a t-distribution with the degrees of freedom being

$$df = \frac{\left(\frac{S_{\Delta_1}^2}{n_1} + \frac{S_{\Delta_0}^2}{n_0}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{S_{\Delta_1}^2}{n_1}\right)^2 + \frac{1}{n_0 - 1}\left(\frac{S_{\Delta_0}^2}{n_0}\right)^2}.$$

A.3 Controlling for Type I Error with Negative Carry-over Effects

Similar to Yi et al. (2022), we consider a sensitivity parameter $\Lambda < 0$ that bounds the bias, i.e., $2^{-1}(\lambda_0 + \lambda_1) \geq \Lambda$. Then the rejection region

$$\frac{\sqrt{n}\left(\hat{\theta}_{\text{cr}} - \theta^* + \Lambda\right)}{\hat{\sigma}_{\text{cr}}} > z_{1-\alpha}$$

can control the type I error at level α.

For the implementation of the sensitivity analysis, practitioners are not required to specify the value of the sensitivity parameter Λ. Results from the sensitivity parameter can be summarized by the "tipping point", i.e., the magnitude of Λ that would be needed such that the null hypothesis can no longer be rejected. [A1, A2] If such a value of Λ is deemed implausible, then we still have evidence to reject the null hypothesis.

A.4 R Codes

All R codes for Section 3 are available at https://github.com/dshidanni/UW-MasterThesis, including R codes for calculating formula powers and simulation powers and the corresponding

simulation data for Section 3.1, R codes for the data analysis of the analgesic example and the real data, and R codes for the power calculations for Section 3.3.

References

A1. Cornfield, J., Haenszel, W., Hammond, E., Lilienfeld, A., Shimkin, M., and Wynder, E. (1959). Smoking and lung cancer. *Journal of the National Cancer Institute* **22**, 173 – 203. Rosenbaum, P. R. (2020). Design of observational studies (2nd ed.). Springer.

A2. Yi, Y., Zhang, Y., Du, Y., and Ye, T. (2022). Testing for treatment effect twice using internal and external controls in clinical trials. *arXiv preprint arXiv:2203.04194.*