axioms

MDPI

Article

Equivalence Test and Sample Size Determination Based on Odds Ratio in an AB/BA Crossover Study with Binary Outcomes

Shi-Fang Qiu [1,*], Xue-Qin Yu [1] and Wai-Yin Poon [2]

[1] Department of Statistics and Data Science, Chongqing University of Technology, Chongqing 400054, China; yuxq0726@163.com
[2] Department of Statistics, The Chinese University of Hong Kong, Hong Kong, China; wypoon@cuhk.edu.hk
* Correspondence: sfqiu@cqut.edu.cn

Abstract

Crossover trials are specifically designed to evaluate treatment effects within individual participants through within-subject comparisons. In a standard AB/BA crossover trial, participants are randomly allocated to one of two treatment sequences: either the AB sequence (where patients receive treatment A first and then cross over to treatment B after a washout period) or the BA sequence (where patients receive B first and then cross over to A after a washout period). Asymptotic and approximate unconditional test procedures, based on two Wald-type statistics, the likelihood ratio statistic, and the score test statistic for the odds ratio (OR), are developed to evaluate the equality of treatment effects in this trial design. Additionally, confidence intervals for OR are constructed, accompanied by an approximate sample size calculation methodology to control the interval width at a pre-specified precision. Empirical analyses demonstrate that asymptotic test procedures exhibit robust performance in moderate to large sample sizes, though they occasionally yield unsatisfactory type I error rates when the sample size is small. In such cases, approximate unconditional test procedures emerge as a rigorous alternative. All proposed confidence intervals achieve satisfactory coverage probabilities, and the approximate sample size estimation method demonstrates high accuracy, as evidenced by empirical coverage probabilities aligning closely with pre-specified confidence levels under estimated sample sizes. To validate practical utility, two real examples are used to illustrate the proposed methodologies.

Keywords: AB/BA crossover trial; approximate unconditional method; confidence interval; odds ratio; sample size determination

MSC: 62F03; 62F25

check for updates

Academic Editor: Hari Mohan Srivastava

Received: 30 June 2025
Revised: 20 July 2025
Accepted: 25 July 2025
Published: 27 July 2025

Citation: Qiu, S.-F.; Yu, X.-Q.; Poon, W.-Y. Equivalence Test and Sample Size Determination Based on Odds Ratio in an AB/BA Crossover Study with Binary Outcomes. *Axioms* **2025**, *14*, 582. https://doi.org/10.3390/axioms14080582

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In a parallel-group trial design, each participant is randomly assigned to receive a single experimental treatment. This methodology serves as a cornerstone of clinical research, particularly in medical and health science investigations, due to its ability to isolate treatment effects through randomization. By contrast, crossover trials employ a distinct approach wherein each participant sequentially receives multiple experimental treatments under controlled conditions, enabling within-subject comparisons while minimizing inter-individual variability. The design's methodological strength stems from its ability to isolate treatment effects through paired within-subject comparisons, enhancing both statistical

precision and resource efficiency in clinical investigations (Hills and Armitage [1]; Fleiss [2]; Senn [3]). It can significantly reduce both costs and required sample sizes due to within-patient comparison (Sever et al. [4]). For example, in a study by Ménard et al. [5], a crossover design with alternating 2-week active treatment phases and 2-week placebo washout periods was employed to assess antihypertensive efficacy. Their findings indicated that a reduction of 5 mmHg in diastolic blood pressure between active treatment and placebo could be detected with 27 clinic patients and 20 self-monitored home patients under usual statistical risk ($\alpha = 0.05$, $\beta = 0.10$). At the conclusion of the 3-month follow-up period, each participant could continue receiving the treatment that was the most effective and the best tolerated. This demonstrates that implementing a crossover trial design (with 15-day washout periods between active treatments) combined with precise blood pressure recording not only minimizes the number of patients required for hypertension trials but also enables personalized treatment optimization for individual patients. Recent work by Grenet et al. [6] established that when within-patient correlation ranges from 0.5 to 0.9, crossover trials require only 5–25% as many participants as parallel-group trials to achieve equivalent statistical power for detecting interaction effects.

Crossover trials are commonly utilized in clinical research to compare treatments for chronic diseases. This design is particularly valuable when evaluating new versus existing therapies, as seen in conditions such as asthma and hypertension. The choice of sequences in crossover designs depends on treatment number, sequence length, and trial objectives. The simplest design is the AB/BA crossover, where participants are randomized to either AB or BA sequences to receive two treatments. For the AB treatment group, patients first receive treatment A and then cross over to treatment B after a washout period; for the BA treatment group, patients first receive treatment B and then cross over to treatment A after a washout period. The AB/BA crossover design is also known as a simple crossover or 2×2 design (Jones and Kenward [7]). Due to its simplicity, the AB/BA crossover design constitutes a substantial portion of crossover trials in practice (Hills and Armitage [1]; Senn [3,8]; Mills et al. [9]). For example, Fava and Patel [10] reported that over half of 72 crossover trials in a survey of 12 major U.S. pharmaceutical companies used this design (Jones and Kenward [11]).

Significant and productive research has focused on AB/BA crossover trial designs. Under the assumption that carryover effects are absent (typically ensured by an adequate washout period eliminating the impact of the first-stage treatment on the second stage), Kershner and Federer [12] compared the variances for direct, residual, and cumulative treatment effects in two-treatment crossover designs. They considered variance estimates for the second-order residual effects and the interaction effects between period and first-order residuals. To analyze ordinal categorical data from AB/BA crossover trials, Ezzet and Whitehead [13] fitted a random effects model and used the Newton–Raphson algorithm for maximum likelihood estimation. For binomial data, Becker and Balagtas [14] developed likelihood-based methods for hypothesis testing and parameter estimation using two models: a log-linear model for marginal response probabilities and a linear model for log-odds ratios. For AB/BA crossover trials with continuous data, Jaki, Pallmann, and Wolfsegger [15] developed three parameter estimation methods and derived corresponding confidence intervals assuming a normal distribution. Lui and Chang [16] addressed methodological issues in hypothesis testing and parameter estimation for AB/BA crossover trials with ordinal outcomes.

Recently, Lui [17] investigated methods for testing inequality, non-inferiority, and equivalence of treatment effects across continuous, ordinal, binomial, and count data types. The author also provided interval estimation methods for the mean difference between two treatments and explored sample size estimation formulas using the power of the tests.

Furthermore, the findings on AB/BA crossover designs were extended to crossover designs involving multiple treatments. Li et al. [18] developed a likelihood ratio test and score test for assessing non-inferiority or equivalence specifically based on the square root of the odds ratio in binomial data in AB/BA crossover trials. Lui [19,20], Lui and Chang [21] developed large-sample and exact small-sample testing procedures for equivalence and non-inferiority assessments of binomial and ordinal categorical data in incomplete block crossover designs. Using Monte Carlo simulations with maximum likelihood estimation, Zhu and Lui [22] quantified the effects of violated normality assumptions for random effects on hypothesis testing and parameter estimation in logistic regression models analyzing binomial data in AB/BA crossover designs. Although extensive research exists on treatment effects in AB/BA crossover trials, there remains a need to investigate equivalence testing based on odds ratios (OR), develop corresponding confidence intervals, and establish sample size determination methods based on interval precision. Focusing on binomial outcome data without carryover effects, this paper introduces OR-based equivalence test statistics and procedures; specifically, approximate unconditional test procedures are developed for small-sample applications. We also develop confidence interval methods for the OR and establish sample size determination techniques with controlled interval precision.

This paper is organized as follows. Model and parameter estimations are presented in Section 2, and the hypothesis test statistics and test procedures for the equivalence test based on OR are provided in Section 3. Confidence interval construction and sample size, which control the interval width, are given in Sections 4 and 5, respectively. The performance of the proposed methods is evaluated by simulation studies in Section 6. Two examples based on real data about two new devices delivering salbutamol and relieving heartburn are analyzed in Section 7. We summarize our conclusions and a brief discussion in Section 8.

2. Model and Parameter Estimation

In this article, we consider the two-period two-treatment or AB/BA crossover trial design for comparing two treatments. Suppose that n_1 patients are randomly assigned to receive the treatment with the order AB (i.e., Group 1 ($g = 1$)), in which patients receive treatment A at the first period and then receive treatment B at the second period, and n_2 patients are randomly assigned to receive the treatment with the order BA (i.e., Group 2 ($g = 2$)), in which patients receive treatment B at the first period and then receive treatment A at the second period. After an adequate wash-out period, we assume that there is no carry-out effect. Let $Y_{iz}^{(g)} = 1$ if the ith patient in the zth period in Group g has the positive response of interest, and $Y_{iz}^{(g)} = 0$; otherwise, ($i = 1, 2, \cdots, n_g$, $z, g = 1, 2$), $n_{rc}^{(g)}$ denote the number of patients among $n_{(g)}$ patients in Group g with the response vector $(Y_{i1}^{(g)} = r, Y_{i2}^{(g)} = c)$ $(r, c = 0, 1)$. Then, the random frequencies $\{n_{rc}^{(g)}\}_{r,c=0,1}$ follow the multinomial distribution with parameters n_g and $\{\pi_{rc}^{(g)}\}_{r,c=0,1}$, where $\pi_{rc}^{(g)}$ denote the cell probability that a randomly selected patient from the gth group has the response vector $(Y_{i1}^{(g)} = r, Y_{i2}^{(g)} = c)$. The data structure is given in Table 1.

The results (0, 0) and (1, 1) imply that there is no significant difference between the two groups. Therefore, the results (0, 1) or (1, 0) are of greater interest. The ratio of probabilities in the cell (0, 1) to the cell (1, 0) in Group g is $\pi_{01}^{(g)}/\pi_{10}^{(g)}$ $(g = 1, 2)$. Obviously, the two probability ratios, i.e., $\pi_{01}^{(1)}/\pi_{10}^{(1)}$ and $\pi_{01}^{(2)}/\pi_{10}^{(2)}$ should be equal if there is no significant difference between the two treatments; otherwise, it means that the treatment effects of the two treatments are different. Let

$$\phi = \frac{\pi_{01}^{(1)}/\pi_{10}^{(1)}}{\pi_{01}^{(2)}/\pi_{10}^{(2)}} = \frac{\pi_{01}^{(1)}\pi_{10}^{(2)}}{\pi_{10}^{(1)}\pi_{01}^{(2)}}. \quad (1)$$

Table 1. Data structure of the AB/BA crossover trial.

AB Sequence				
		Period 2		
		1	0	Total
Period 1	1	$n_{11}^{(1)}(\pi_{11}^{(1)})$	$n_{10}^{(1)}(\pi_{10}^{(1)})$	$n_{1\cdot}^{(1)}$
	0	$n_{01}^{(1)}(\pi_{01}^{(1)})$	$n_{00}^{(1)}(\pi_{00}^{(1)})$	$n_{0\cdot}^{(1)}$
	Total	$n_{\cdot 1}^{(1)}$	$n_{\cdot 0}^{(1)}$	n_1
BA Sequence				
		Period 2		
		1	0	Total
Period 1	1	$n_{11}^{(2)}(\pi_{11}^{(2)})$	$n_{10}^{(2)}(\pi_{10}^{(2)})$	$n_{1\cdot}^{(2)}$
	0	$n_{01}^{(2)}(\pi_{01}^{(2)})$	$n_{00}^{(2)}(\pi_{00}^{(2)})$	$n_{0\cdot}^{(2)}$
	Total	$n_{\cdot 1}^{(2)}$	$n_{\cdot 0}^{(2)}$	n_2

Obviously, $\phi = 1$ denotes that there is no significant difference between the two treatment effects. Therefore, we are interested in the following hypothesis testing:

$$H_0 : \phi = 1 \text{ versus } H_1 : \phi \neq 1.$$

Let $\boldsymbol{m} = (n_{00}^{(1)}, n_{01}^{(1)}, n_{10}^{(1)}, n_{11}^{(1)}, n_{00}^{(2)}, n_{01}^{(2)}, n_{10}^{(2)}, n_{11}^{(2)})$ be the collection of the observed frequencies in Table 1, with $(n_{00}^{(1)}, n_{10}^{(1)}, n_{01}^{(1)}, n_{11}^{(1)}) \sim M(n_1; \pi_{00}^{(1)}, \pi_{01}^{(1)}, \pi_{10}^{(1)}, \pi_{11}^{(1)})$, $(n_{00}^{(2)}, n_{10}^{(2)}, n_{01}^{(2)}, n_{11}^{(2)}) \sim M(n_2; \pi_{00}^{(2)}, \pi_{01}^{(2)}, \pi_{10}^{(2)}, \pi_{11}^{(2)})$, where $\sum_{r,c\in\{0,1\}} \pi_{rc}^{(g)} = 1, g = 1, 2, r, c = 0, 1$, M denotes the multinomial distribution. The likelihood function based on the observed data $\boldsymbol{m}$ is given by

$$L(\boldsymbol{m}; \pi_{rc}^{(g)}) = \Pi_{k\in\{1,2\}} \Pi_{r,c\in\{0,1\}} \frac{n_1! n_2!}{n_{rc}^{(g)}!} {\pi_{rc}^{(g)}}^{n_{rc}^{(g)}},$$

where $\sum\limits_{r,c\in\{0,1\}} \pi_{rc}^{(g)} = 1, g = 1, 2$ and $\sum\limits_{r,c\in\{0,1\}} n_{rc}^{(g)} = n_g, g = 1, 2$ are the two constraints of the parameter. And then, the log-likelihood function based on the observed data $\boldsymbol{m}$ is given by

$$\begin{aligned} l_1 = C_1 &+ n_{00}^{(1)} \log \pi_{00}^{(1)} + n_{01}^{(1)} \log \pi_{01}^{(1)} + n_{10}^{(1)} \log \pi_{10}^{(1)} + n_{11}^{(1)} \log \pi_{11}^{(1)} + n_{00}^{(2)} \log \pi_{00}^{(2)} + \\ &+ n_{01}^{(2)} \log \pi_{01}^{(2)} + n_{10}^{(2)} \log \pi_{10}^{(2)} + n_{11}^{(2)} \log \pi_{11}^{(2)}, \end{aligned} \tag{2}$$

where C_1 is a constant that does not involve the unknown parameters. The maximum likelihood estimators (MLEs) $\{\hat{\pi}_{rc}^{(g)} : r, c = 0, 1, g = 1, 2\}$ and $\hat{\phi}$ of the parameters $\{\pi_{rc}^{(g)} : r, c = 0, 1, g = 1, 2\}$ and ϕ are given by the following:

$$\hat{\pi}_{rc}^{(g)} = \frac{n_{rc}^{(g)}}{n_g} \quad \text{and} \quad \hat{\phi} = \frac{\hat{\pi}_{01}^{(1)} \hat{\pi}_{10}^{(2)}}{\hat{\pi}_{10}^{(1)} \hat{\pi}_{01}^{(2)}} = \frac{n_{01}^{(1)} n_{10}^{(2)}}{n_{10}^{(1)} n_{01}^{(2)}}. \tag{3}$$

Similar to Li et al. [18], let $M_1 = 1 - \pi_{00}^{(1)} - \pi_{11}^{(1)}$ and $M_2 = 1 - \pi_{00}^{(2)} - \pi_{11}^{(2)}$, then we can re-parameterize the log-likelihood function as

$$\begin{aligned} l_2(\boldsymbol{m};\phi,\pi_{10}^{(2)},\pi_{00}^{(1)},\pi_{11}^{(1)},\pi_{00}^{(2)},\pi_{11}^{(2)}) = {} & n_{01}^{(1)}\log(\frac{\phi(M_2-\pi_{10}^{(2)})M_1}{\pi_{10}^{(2)}+\phi(M_2-\pi_{10}^{(2)})}) + n_{00}^{(1)}\log\pi_{00}^{(1)} \\ & + n_{10}^{(1)}\log(\frac{\pi_{10}^{(2)}M_1}{\pi_{10}^{(2)}+\phi(M_2-\pi_{10}^{(2)})}) + n_{11}^{(1)}\log\pi_{11}^{(1)} \\ & + n_{00}^{(2)}\log\pi_{00}^{(2)} + n_{10}^{(2)}\log\pi_{10}^{(2)} + n_{11}^{(2)}\log\pi_{11}^{(2)} \\ & + n_{01}^{(2)}\log(M_2-\pi_{10}^{(2)}) + C_1. \end{aligned} \tag{4}$$

This log-likelihood function contains six parameters (ϕ, $\pi_{10}^{(2)}$, $\pi_{00}^{(1)}$, $\pi_{11}^{(1)}$, $\pi_{00}^{(2)}$, $\pi_{11}^{(2)}$), where ϕ is the interest parameter. Given the value of ϕ, the constrained maximum likelihood estimator (CMLE) of $\pi_{10}^{(2)}$ can be obtained by solving the following equation:

$$\frac{n_{01}^{(1)}+n_{01}^{(2)}}{m_2-\pi_{10}^{(2)}} - \frac{n_{10}^{(1)}+n_{10}^{(2)}}{\pi_{10}^{(2)}} + \frac{(n_{10}^{(1)}+n_{01}^{(1)})(1-\phi)}{\pi_{10}^{(2)}+\phi(m_2-\pi_{10}^{(2)})} = 0, \tag{5}$$

where $m_1 = (n_{10}^{(1)}+n_{01}^{(1)})/n_1$, $m_2 = (n_{10}^{(2)}+n_{01}^{(2)})/n_2$. Let $A = (\phi-1)(n_{01}^{(2)}+n_{10}^{(2)})$, $B = m_2(n_{10}^{(2)}-n_{01}^{(1)}) - m_2(n_{01}^{(2)}+n_{10}^{(1)}+2n_{10}^{(2)})\phi$, and $C = \phi(n_{10}^{(1)}+n_{10}^{(2)})m_2^2$, then the above equation is given by $A(\pi_{10}^{(2)})^2 + B(\pi_{10}^{(2)}) + C = 0$. Therefore, when $A \neq 0$, the CMLE $\tilde{\pi}_{10}^{(2)}$ of $\pi_{10}^{(2)}$ given the value of ϕ can be given by $\tilde{\pi}_{10}^{(2)} = (-B-\sqrt{B^2-4AC})/(2A)$ (please refer to Appendix A for details); otherwise, $\tilde{\pi}_{10}^{(2)} = -C/B$, and then the CMLEs of the other parameters given ϕ are given by

$$\begin{aligned} & \tilde{\pi}_{01}^{(1)}(\phi) = \frac{\phi(m_2-\tilde{\pi}_{10}^{(2)}(\phi))m_1}{\tilde{\pi}_{10}^{(2)}(\phi)+\phi(m_2-\tilde{\pi}_{10}^{(2)}(\phi))}, \tilde{\pi}_{10}^{(1)}(\phi) = \frac{\tilde{\pi}_{10}^{(2)}(\phi)m_1}{\tilde{\pi}_{10}^{(2)}(\phi)+\phi(m_2-\tilde{\pi}_{10}^{(2)}(\phi))}, \\ & \tilde{\pi}_{01}^{(2)}(\phi) = m_2-\tilde{\pi}_{10}^{(2)}(\phi), \tilde{\pi}_{00}^{(1)}(\phi) = \frac{n_{00}^{(1)}}{n_1}, \tilde{\pi}_{11}^{(1)}(\phi) = \frac{n_{11}^{(1)}}{n_1}, \tilde{\pi}_{00}^{(2)}(\phi) = \frac{n_{00}^{(2)}}{n_2}, \\ & \tilde{\pi}_{11}^{(2)}(\phi) = \frac{n_{11}^{(2)}}{n_2}, \end{aligned} \tag{6}$$

respectively. Specially, when $\phi = 1$, the CMLEs of the parameters $\pi_{01}^{(1)}$, $\pi_{10}^{(1)}$, $\pi_{01}^{(2)}$, $\pi_{10}^{(2)}$ are given by

$$\begin{aligned} & \tilde{\pi}_{10}^{(2)} = \frac{(n_{10}^{(2)}+n_{10}^{(1)})m_2}{n_{10}^{(2)}+n_{10}^{(1)}+n_{01}^{(1)}+n_{01}^{(2)}}, \ \tilde{\pi}_{01}^{(1)} = \frac{(m_2-\tilde{\pi}_{10}^{(2)})m_1}{m_2}, \\ & \tilde{\pi}_{10}^{(1)} = \frac{\tilde{\pi}_{10}^{(2)}m_1}{m_2}, \ \tilde{\pi}_{01}^{(2)} = m_2-\tilde{\pi}_{10}^{(2)}. \end{aligned} \tag{7}$$

3. Hypothesis Testing

3.1. Test Statistics

For the hypothesis testing $H_0 : \phi = 1$ versus $H_1 : \phi \neq 1$, we consider the following test statistics.

(i) Wald-type test statistic (T_{w1})

Let $\boldsymbol{c} = (1,-1,-1,1)'$, and $\boldsymbol{\beta} = (\log(\pi_{01}^{(1)}), \log(\pi_{10}^{(1)}), \log(\pi_{01}^{(2)}), \log(\pi_{10}^{(2)}))'$, then the above hypothesis is equivalent to the following hypothesis test:

$$H_0 : \boldsymbol{c}'\boldsymbol{\beta} = 0 \text{ versus } H_1 : \boldsymbol{c}'\boldsymbol{\beta} \neq 0.$$

The estimator of $\boldsymbol{\beta}$ is given by $\hat{\boldsymbol{\beta}} = (\log(\hat{\pi}_{01}^{(1)}), \log(\hat{\pi}_{10}^{(1)}), \log(\hat{\pi}_{01}^{(2)}), \log(\hat{\pi}_{10}^{(2)}))'$. Using the delta method, it is easily shown that the covariance matrix of $\hat{\boldsymbol{\beta}}$ is given by

$$\boldsymbol{\Sigma} = Var(\hat{\boldsymbol{\beta}}) = \begin{pmatrix} \frac{1-\pi_{01}^{(1)}}{n_1\pi_{01}^{(1)}} & -\frac{1}{n_1} & 0 & 0 \\ -\frac{1}{n_1} & \frac{1-\pi_{10}^{(1)}}{n_1\pi_{10}^{(1)}} & 0 & 0 \\ 0 & 0 & \frac{1-\pi_{01}^{(2)}}{n_2\pi_{01}^{(2)}} & -\frac{1}{n_2} \\ 0 & 0 & -\frac{1}{n_2} & \frac{1-\pi_{10}^{(2)}}{n_2\pi_{10}^{(2)}} \end{pmatrix}. \tag{8}$$

$Var(\hat{\boldsymbol{\beta}})$ can be estimated by replacing the parameters with their MLEs given by Equation (3), which is given by

$$\hat{\boldsymbol{\Sigma}} = \widehat{Var}(\hat{\boldsymbol{\beta}}) = \begin{pmatrix} \frac{1}{n_{01}^{(1)}} - \frac{1}{n_1} & -\frac{1}{n_1} & 0 & 0 \\ -\frac{1}{n_1} & \frac{1}{n_{10}^{(1)}} - \frac{1}{n_1} & 0 & 0 \\ 0 & 0 & \frac{1}{n_{01}^{(2)}} - \frac{1}{n_2} & -\frac{1}{n_2} \\ 0 & 0 & -\frac{1}{n_2} & \frac{1}{n_{10}^{(2)}} - \frac{1}{n_2} \end{pmatrix}. \tag{9}$$

Therefore, the Wald test statistic for testing $H_0 : \boldsymbol{c'\beta} = 0$ is given by

$$T_{w1} = \frac{(\boldsymbol{c'}\hat{\boldsymbol{\beta}})^2}{\boldsymbol{c'}\widehat{Var}(\hat{\boldsymbol{\beta}})\boldsymbol{c}} = \frac{(\boldsymbol{c'}\hat{\boldsymbol{\beta}})^2}{\boldsymbol{c'}\hat{\boldsymbol{\Sigma}}\boldsymbol{c}}.$$

Under the null hypothesis $H_0 : \boldsymbol{c'\beta} = 0$, T_{w1} is asymptotically distributed as the χ^2 distribution with one degree of freedom when $\min\{n_1, n_2\} \to \infty$ (please refer to Appendix B for details).

(ii) Wald-type test statistic (T_{w2})

Let $\tilde{\pi}_{01}^{(1)}$, $\tilde{\pi}_{10}^{(1)}$, $\tilde{\pi}_{01}^{(2)}$, $\tilde{\pi}_{10}^{(2)}$ be the constrained MLEs of the parameters $\pi_{01}^{(1)}$, $\pi_{10}^{(1)}$, $\pi_{01}^{(2)}$, $\pi_{10}^{(2)}$ under $H_0 : \boldsymbol{c'\beta} = 0$. The variance of $\hat{\boldsymbol{\beta}}$ can be estimated by replacing the parameters with their CMLEs given in Equation (7), i.e., $\tilde{\boldsymbol{\Sigma}} = \widehat{Var}(\hat{\boldsymbol{\beta}}|H_0)$. Therefore, another Wald-type statistic for testing $H_0 : \boldsymbol{c'\beta} = 0$ is given by

$$T_{w2} = \frac{(\boldsymbol{c'}\hat{\boldsymbol{\beta}})^2}{\boldsymbol{c'}\widehat{Var}(\hat{\boldsymbol{\beta}}|H_0)\boldsymbol{c}} = \frac{(\boldsymbol{c'}\hat{\boldsymbol{\beta}})^2}{\boldsymbol{c'}\tilde{\boldsymbol{\Sigma}}\boldsymbol{c}},$$

where

$$\tilde{\boldsymbol{\Sigma}} = \widehat{Var}(\hat{\boldsymbol{\beta}}|H_0) = \begin{pmatrix} \frac{1-\tilde{\pi}_{01}^{(1)}}{n_1\tilde{\pi}_{01}^{(1)}} & -\frac{1}{n_1} & 0 & 0 \\ -\frac{1}{n_1} & \frac{1-\tilde{\pi}_{10}^{(1)}}{n_1\tilde{\pi}_{10}^{(1)}} & 0 & 0 \\ 0 & 0 & \frac{1-\tilde{\pi}_{01}^{(2)}}{n_2\tilde{\pi}_{01}^{(2)}} & -\frac{1}{n_2} \\ 0 & 0 & -\frac{1}{n_2} & \frac{1-\tilde{\pi}_{10}^{(2)}}{n_2\tilde{\pi}_{10}^{(2)}} \end{pmatrix}.$$

Under the null hypothesis, T_{w2} is asymptotically distributed as the χ^2 distribution with one degree of freedom when $\min\{n_1, n_2\} \to \infty$ (please refer to Appendix B for details).

(iii) Likelihood ratio test statistic (T_l)

The likelihood ratio statistic for testing the hypothesis $H_0 : \phi = 1$ versus $H_1 : \phi \neq 1$ can be given by

$$\begin{aligned} T_l &= 2\{l_2(\boldsymbol{m}; \hat{\phi}, \hat{\pi}_{10}^{(2)}, \hat{\pi}_{00}^{(1)}, \hat{\pi}_{11}^{(1)}, \hat{\pi}_{00}^{(2)}, \hat{\pi}_{11}^{(2)}) - l_2(\boldsymbol{m}; 1, \tilde{\pi}_{10}^{(2)}, \tilde{\pi}_{00}^{(1)}, \tilde{\pi}_{11}^{(1)}, \tilde{\pi}_{00}^{(2)}, \tilde{\pi}_{11}^{(2)})\} \\ &= 2\{n_{10}^{(1)} \log(\hat{\pi}_{10}^{(1)}) + n_{01}^{(1)} \log(\hat{\pi}_{01}^{(1)}) + n_{10}^{(2)} \log(\hat{\pi}_{10}^{(2)}) + n_{01}^{(2)} \log(\hat{\pi}_{01}^{(2)}) \\ &\quad - n_{10}^{(1)} \log(\tilde{\pi}_{10}^{(1)}) - n_{01}^{(1)} \log(\tilde{\pi}_{01}^{(1)}) - n_{10}^{(2)} \log(\tilde{\pi}_{10}^{(2)}) - n_{01}^{(2)} \log(\tilde{\pi}_{01}^{(2)})\}. \end{aligned}$$

Under the null hypothesis $H_0 : \phi = 1$, T_l is asymptotically distributed as a chi-square distribution with one degree of freedom when $\min\{n_1, n_2\} \to \infty$.

(iv) Score test statistic (T_{sc})

Taking the partial derivative of the log-likelihood function given in Equation (4) with respect to ϕ, we can obtain the following score function:

$$S_\phi = \frac{\partial l_2(\phi; \pi_{10}^{(2)}, \pi_{00}^{(1)}, \pi_{11}^{(1)}, \pi_{00}^{(2)}, \pi_{11}^{(2)})}{\partial \phi} = \frac{n_{01}^{(2)}}{\phi} - \frac{(n_{01}^{(1)} + n_{10}^{(1)})(M_2 - \pi_{10}^{(2)})}{\pi_{10}^{(2)} + \phi(M_2 - \pi_{10}^{(2)})}.$$

According to the general theory of score test (Rao [23]), the score test statistic for testing $H_0 : \phi = 1$ can be given by

$$T_{sc} = S_\phi \sqrt{I^{11}}|_{\phi=1, \pi_{10}^{(2)}=\tilde{\pi}_{10}^{(2)}, \pi_{00}^{(1)}=\tilde{\pi}_{00}^{(1)}, \pi_{11}^{(1)}=\tilde{\pi}_{11}^{(1)}, \pi_{00}^{(2)}=\tilde{\pi}_{00}^{(2)}, \pi_{11}^{(2)}=\tilde{\pi}_{11}^{(2)}},$$

where I^{11} is the first main diagonal element of the inverse of the Fisher information matrix (please refer to Appendix C for details). Under the null hypothesis $H_0 : \phi = 1$, T_{sc} is asymptotically distributed as a standard normal distribution when $\min\{n_1, n_2\} \to \infty$.

3.2. Test Procedures

3.2.1. Asymptotic Test Procedure

Assume that $\chi^2_{1,\alpha}$ is the upper α quantile of the chi-square distribution with one degree of freedom, and $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution. For the asymptotical test procedure, we reject the null hypothesis $H_0 : \phi = 1$ at the significance level α when $t_i > \chi^2_{1,\alpha} (i = w1, w2, l)$, or $|t_{sc}| > z_{\alpha/2}$. It is well known that the asymptotic test procedure is easily implemented; however, it usually does not perform well for small sample size designs or sparse data structures. Therefore, we consider the following approximate unconditional test procedure.

3.2.2. Approximate Unconditional Test Procedure

Similar to Tang et al. [24], we consider the approximate unconditional test procedure when sample size is small. Let

$$\begin{aligned} \Omega = \{\mathbf{m} = (&n_{00}^{(1)}, n_{01}^{(1)}, n_{10}^{(1)}, n_{11}^{(1)}, n_{00}^{(2)}, n_{01}^{(2)}, n_{10}^{(2)}, n_{11}^{(2)}) : \\ &0 \leq n_{00}^{(1)}, n_{10}^{(1)}, n_{01}^{(1)}, n_{11}^{(1)} \leq n_1, 0 \leq n_{00}^{(2)}, n_{10}^{(2)}, n_{01}^{(2)}, n_{11}^{(2)} \leq n_2\}, \end{aligned}$$

where Ω is the set of all possible observations given N and n_1. Given the observed value $t_i (i = w1, w2, l, sc)$, we find all possible $\boldsymbol{m}$ such that the test statistic T_i satisfies $T_i \geq t_i$ for $i = w1, w2, l$ or $|T_{sc}| \geq |t_{sc}|$. For each $\boldsymbol{m}$ satisfying $T_i \geq t_i$ $(i = w1, w2, l)$ or $|T_{sc}| \geq |t_{sc}|$, we calculate the probability under $H_0 : \phi = 1$ by using the corresponding likelihood function value. The approximate unconditional p-value is computed by

$$p_i^{AU} = \Pr(T_i \geq t_i | \phi = 1, \pi_{rc}^{(g)} = \tilde{\pi}_{rc}^{(g)})(i = w1, w2, l)$$
$$= \sum_{\boldsymbol{m} \in \Omega(T_i \geq t_i)} L(\boldsymbol{m}; \phi = 1, \tilde{\pi}_{rc}^{(g)})$$

or

$$p_{sc}^{AU} = \Pr(|T_{sc}| \geq |t_{sc}| | \phi = 1, \pi_{rc}^{(g)} = \tilde{\pi}_{rc}^{(g)})$$
$$= \sum_{\boldsymbol{m} \in \Omega(|T_{sc}| \geq |t_{sc}|)} L(\boldsymbol{m}; \phi = 1, \tilde{\pi}_{rc}^{(g)}),$$

If $p_i^{AU} \leq \alpha$ $(i = w1, w2, l, sc)$, then the null hypothesis $H_0 : \phi = 1$ is rejected at the nominal level α.

4. Confidence Interval

As pointed in several articles, for example, Alhija and Levy [25], Odgaard and Fowler [26], Sun et al. [27], Dunst and Hamby [28], and Fritz et al. [29], it is necessary to include some measures of effect size and confidence intervals for all primary outcomes, which is recommended in editorial guidelines and methodological recommendations of several prominent educational and psychological journals. Indeed, based on the Publication Manual of the American Psychological Association [30], it is well documented that confidence intervals are more informative than simple hypothesis tests, and are the best reporting strategy for the description of location and precision of the statistic. Therefore, we investigate confidence interval construction for ϕ in this section.

4.1. Wald CIs

Let $n_1 = \frac{N}{1+r}$ and $n_2 = \frac{rN}{1+r}$, and the variance in $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ can be given by $Var(\boldsymbol{c}'\hat{\boldsymbol{\beta}}) = \boldsymbol{c}' Var(\hat{\boldsymbol{\beta}})\boldsymbol{c}$, where $\boldsymbol{c} = (1, -1, -1, 1)'$ and $\hat{\boldsymbol{\beta}} = (\log(\hat{\pi}_{01}^{(1)}), \log(\hat{\pi}_{10}^{(1)}), \log(\hat{\pi}_{01}^{(2)}), \log(\hat{\pi}_{10}^{(2)}))'$. If we estimate the covariance matrix $Var(\hat{\boldsymbol{\beta}})$ by the unconstrained MLEs given in (3), then the variance in $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ can be estimated by

$$\widehat{Var}(\boldsymbol{c}'\hat{\boldsymbol{\beta}}) = \boldsymbol{c}' \widehat{Var}(\hat{\boldsymbol{\beta}})\boldsymbol{c} = \frac{(1+r)}{N}\left(\frac{1}{\hat{\pi}_{01}^{(1)}} + \frac{1}{\hat{\pi}_{10}^{(1)}} + \frac{1}{r\hat{\pi}_{01}^{(2)}} + \frac{1}{r\hat{\pi}_{10}^{(2)}}\right)$$

According to the Central Limits Theorem,

$$\frac{(\boldsymbol{c}'\hat{\boldsymbol{\beta}} - \log\phi)}{\sqrt{\boldsymbol{c}' \widehat{Var}(\hat{\boldsymbol{\beta}})\boldsymbol{c}}} = \frac{\sqrt{N}(\boldsymbol{c}'\hat{\boldsymbol{\beta}} - \log\phi)}{\sqrt{(1+r)\left(\frac{1}{\hat{\pi}_{01}^{(1)}} + \frac{1}{\hat{\pi}_{10}^{(1)}} + \frac{1}{r\hat{\pi}_{01}^{(2)}} + \frac{1}{r\hat{\pi}_{10}^{(2)}}\right)}}$$

is asymptotically followed as the standard normal distribution. Therefore, the $100(1-\alpha)\%$ confidence interval for $\log\phi$ can be given by

$$[l, u] = \left[\boldsymbol{c}'\hat{\boldsymbol{\beta}} - z_{\alpha/2}\sqrt{\boldsymbol{c}' \widehat{Var}(\hat{\boldsymbol{\beta}})\boldsymbol{c}}, \boldsymbol{c}'\hat{\boldsymbol{\beta}} + z_{\alpha/2}\sqrt{\boldsymbol{c}' \widehat{Var}(\hat{\boldsymbol{\beta}})\boldsymbol{c}}\right].$$

Then, the $100(1-\alpha)\%$ confidence interval for ϕ is given by $[\phi_l, \phi_u] = [exp(l), exp(u)]$, which is denoted as CI_{w1}.

Let $\tilde{\pi}_{10}^{(1)}(\phi)$, $\tilde{\pi}_{01}^{(1)}(\phi)$, $\tilde{\pi}_{01}^{(2)}(\phi)$, and $\tilde{\pi}_{10}^{(2)}(\phi)$ are the constrained MLEs of parameters $\pi_{10}^{(1)}$, $\pi_{01}^{(1)}$, $\pi_{01}^{(2)}$, and $\pi_{10}^{(2)}$ given ϕ. The variance of $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ given ϕ can be estimated by

$$\widehat{Var}(\boldsymbol{c}'\hat{\boldsymbol{\beta}}|\phi) = \boldsymbol{c}' \widehat{Var}(\hat{\boldsymbol{\beta}}|\phi)\boldsymbol{c}$$

where

$$\widehat{Var}(\hat{\boldsymbol{\beta}}|\phi) = \begin{pmatrix} \frac{1-\tilde{\pi}_{01}^{(1)}(\phi)}{n_1\tilde{\pi}_{01}^{(1)}(\phi)} & -\frac{1}{n_1} & 0 & 0 \\ -\frac{1}{n_1} & \frac{1-\tilde{\pi}_{10}^{(1)}(\phi)}{n_1\tilde{\pi}_{10}^{(1)}(\phi)} & 0 & 0 \\ 0 & 0 & \frac{1-\tilde{\pi}_{01}^{(2)}(\phi)}{n_2\tilde{\pi}_{01}^{(2)}(\phi)} & -\frac{1}{n_2} \\ 0 & 0 & -\frac{1}{n_2} & \frac{1-\tilde{\pi}_{10}^{(2)}(\phi)}{n_2\tilde{\pi}_{10}^{(2)}(\phi)} \end{pmatrix}.$$

According to the Central Limits Theorem,

$$\frac{\boldsymbol{c}'\hat{\boldsymbol{\beta}} - \log\phi}{\sqrt{\boldsymbol{c}'\widehat{Var}(\hat{\boldsymbol{\beta}}|\phi)\boldsymbol{c}}} = \frac{\sqrt{N}(\boldsymbol{c}'\hat{\boldsymbol{\beta}} - \log\phi)}{\sqrt{(1+r)(\frac{1}{\tilde{\pi}_{01}^{(1)}(\phi)} + \frac{1}{\tilde{\pi}_{10}^{(1)}(\phi)} + \frac{1}{r\tilde{\pi}_{01}^{(2)}(\phi)} + \frac{1}{r\tilde{\pi}_{10}^{(2)}(\phi)})}}$$

is asymptotically followed as the standard normal distribution. Therefore, the lower and upper confidence limits for ϕ can be calculated by solving the equation

$$\frac{N(\boldsymbol{c}'\hat{\boldsymbol{\beta}} - \log\phi)^2}{(1+r)(\frac{1}{\tilde{\pi}_{01}^{(1)}(\phi)} + \frac{1}{\tilde{\pi}_{10}^{(1)}(\phi)} + \frac{1}{r\tilde{\pi}_{01}^{(2)}(\phi)} + \frac{1}{r\tilde{\pi}_{10}^{(2)}(\phi)})} \leq \chi^2_{1,\alpha}, \tag{10}$$

where $\tilde{\pi}_{01}^{(1)}(\phi)$, $\tilde{\pi}_{10}^{(1)}(\phi)$, $\tilde{\pi}_{01}^{(2)}(\phi)$ and $\tilde{\pi}_{10}^{(2)}(\phi)$ are the constrained MLEs of $\pi_{01}^{(1)}$, $\pi_{10}^{(1)}$, $\pi_{01}^{(2)}$ and $\pi_{10}^{(2)}$ given by (6) and (7). The $100(1-\alpha)\%$ confidence interval for ϕ is denoted as $[\phi_l, \phi_u]$, where $0 < \phi_l < \phi_u < +\infty$. No closed form exists; an iterative algorithm, for example, the Newton–Raphson iterative algorithm can be used to find the solutions. This CI is denoted as CI_{w2}.

4.2. CI Based on Likelihood Ratio Test T_l

The likelihood ratio statistic for testing the null hypothesis $H_0 : \phi = \phi_0$ is given by

$$\begin{aligned} T_l = 2\{&l_2(m; \hat{\phi}, \hat{\pi}_{10}^{(2)}, \hat{\pi}_{00}^{(1)}, \hat{\pi}_{11}^{(1)}, \hat{\pi}_{00}^{(2)}, \hat{\pi}_{11}^{(2)}) \\ &- l_2(m; \phi_0, \tilde{\pi}_{10}^{(2)}(\phi_0), \tilde{\pi}_{00}^{(1)}(\phi_0), \tilde{\pi}_{11}^{(1)}(\phi_0), \tilde{\pi}_{00}^{(2)}(\phi_0), \tilde{\pi}_{11}^{(2)}(\phi_0))\}. \end{aligned}$$

Since T_l asymptotically follows the chi-square distribution with one degree of freedom when $N \to \infty$ under H_0, then the $100(1-\alpha)\%$ confidence interval for ϕ is given by $[\phi_l, \phi_u]$, where $0 < \phi_l < \phi_u < +\infty$ are the smaller and the larger roots of the following equation with respect to ϕ:

$$\begin{aligned} &2\{l_2(m; \hat{\phi}, \hat{\pi}_{10}^{(2)}, \hat{\pi}_{00}^{(1)}, \hat{\pi}_{11}^{(1)}, \hat{\pi}_{00}^{(2)}, \hat{\pi}_{11}^{(2)}) \\ &- l_2(m; \phi_0, \tilde{\pi}_{10}^{(2)}(\phi_0), \tilde{\pi}_{00}^{(1)}(\phi_0), \tilde{\pi}_{11}^{(1)}(\phi_0), \tilde{\pi}_{00}^{(2)}(\phi_0), \tilde{\pi}_{11}^{(2)}(\phi_0))\} \leq \chi^2_{1,\alpha}, \end{aligned} \tag{11}$$

Similar to that of Wald CIs, no closed form exists; an iterative algorithm (e.g., Newton–Raphson algorithm) can be used to find the solutions. This CI is denoted as CI_l.

4.3. CI Based on Score Test T_{sc}

Differentiating the log-likelihood function given in (4) with respect to ϕ, we obtain the following score function:

$$S_\phi = \frac{\partial l_2(\phi; \pi_{10}^{(2)}, \pi_{00}^{(1)}, \pi_{11}^{(1)}, \pi_{00}^{(2)}, \pi_{11}^{(2)})}{\partial\phi} = \frac{n_{01}^{(2)}}{\phi} - \frac{(n_{01}^{(1)} + n_{10}^{(1)})(M_2 - \pi_{10}^{(2)})}{\pi_{10}^{(2)} + \phi(M_2 - \pi_{10}^{(2)})}. \tag{12}$$

Let $I(\phi, \pi_{10}^{(2)}, \pi_{00}^{(1)}, \pi_{11}^{(1)}, \pi_{00}^{(2)}, \pi_{11}^{(2)})$ be the Fisher information matrix, which can be given by

$$I(\phi, \pi_{10}^{(2)}, \pi_{00}^{(1)}, \pi_{11}^{(1)}, \pi_{00}^{(2)}, \pi_{11}^{(2)}) = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}, \tag{13}$$

where $I_{11} = E(-\frac{\partial^2 l_2}{\partial \phi^2}) = -\frac{n_{01}^{(1)}}{\phi^2} + \frac{(n_{01}^{(1)}+n_{10}^{(1)})(M_2-\pi_{10}^{(2)})^2}{(\pi_{10}^{(2)}+\phi(M_2-\pi_{10}^{(2)}))^2}$, $I_{12} = I_{21}'$ is a 5×1 matrix and I_{22} is a 5×5 symmetric matrix.

According to the general theory of efficient scores (Rao [23]), the score statistic for testing $H_0 : \phi = \phi_0$ is given by

$$T_{sc}(\phi_0) = S_\phi \sqrt{I^{11}}|_{\phi=\phi_0, \tilde{\pi}_{10}^{(2)}(\phi_0), \tilde{\pi}_{00}^{(1)}(\phi_0), \tilde{\pi}_{11}^{(1)}(\phi_0), \tilde{\pi}_{00}^{(2)}(\phi_0), \tilde{\pi}_{11}^{(2)}(\phi_0)},$$

where $I^{11} = (I_{11} - I_{12} I_{22}^{-1} I_{21})^{-1}$ is the first main diagonal element of the inverse of the Fisher information matrix, T_{sc} is asymptotically distributed as the standard normal distribution under $H_0 : \phi = \phi_0$ when $N \to \infty$. Therefore, the $100(1-\alpha)\%$ confidence interval for ϕ is given by $[\phi_l, \phi_u]$, with the lower limit ϕ_l and the upper limit ϕ_u can be obtained by solving the following equation:

$$T_{sc}(\phi_0) = z_{\alpha/2}, \quad T_{sc}(\phi_0) = -z_{\alpha/2}, \tag{14}$$

respectively. The solutions for the equations in (14) can be obtained by the secant method (Traub [31]), and this CI is denoted as CI_{sc}.

5. Sample Size Determination

It is an important step in clinical trials to determine the required number of participants. In this section, we investigate the determination of sample sizes for equivalence evaluation based on ϕ in AB/BA crossover trials. Lui and Chang [32]; Li et al. [18] have investigated the determination of the sample size from the perspective of testing hypothesis, so we find the sample size that can control the width of a confidence interval with a pre-specified confidence level. Let $n_1 : n_2 = 1 : r$ and $n_1 + n_2 = N$, i.e., $n_1 = \frac{N}{1+r}$ and $n_2 = \frac{rN}{1+r}$. We investigate sample sizes that can control the width of the confidence interval (ϕ_l, ϕ_u) within 2ω with a pre-specified confidence level $1-\alpha$, i.e., the sample size N satisfies the following condition:

$$\{N : \phi_u - \phi_l \leq 2\omega\}.$$

No closed form exists since confidence intervals CI_{w2}, CI_l and CI_{sc} would be obtained via iterative algorithms. The following search algorithm can be used to find the approximate solution.

Step 1. For given $\pi_{00}^{(1)}$, $\pi_{11}^{(1)}$, $\pi_{00}^{(2)}$, $\pi_{10}^{(2)}$, $\pi_{11}^{(2)}$, r, ϕ and N, generate K random samples $\mathbf{m} = (n_{00}^{(1)}, n_{10}^{(1)}, n_{01}^{(1)}, n_{11}^{(1)}, n_{00}^{(2)}, n_{10}^{(2)}, n_{01}^{(2)}, n_{11}^{(2)})$, where $(n_{00}^{(1)}, n_{10}^{(1)}, n_{01}^{(1)}, n_{11}^{(1)}) \sim M(\frac{1}{r+1}N; \pi_{00}^{(1)}, \pi_{01}^{(1)}, \pi_{10}^{(1)}, \pi_{11}^{(1)})$ and $(n_{00}^{(2)}, n_{10}^{(2)}, n_{01}^{(2)}, n_{11}^{(2)}) \sim M(\frac{r}{1+r}N; \pi_{00}^{(2)}, \pi_{01}^{(2)}, \pi_{10}^{(2)}, \pi_{11}^{(2)})$ with $\pi_{01}^{(2)} = 1.0 - \pi_{00}^{(2)} - \pi_{11}^{(2)} - \pi_{10}^{(2)}$, $\pi_{10}^{(1)} = \frac{\pi_{10}^{(2)}(1-\pi_{00}^{(1)}-\pi_{11}^{(1)})}{\pi_{10}^{(2)}+\phi\pi_{01}^{(2)}}$ and $\pi_{01}^{(1)} = \frac{\phi\pi_{10}^{(1)}\pi_{01}^{(2)}}{\pi_{10}^{(2)}}$.

Step 2. Based on each sample generated in Step 1, compute the confidence intervals using equations given in Section 4 and approximate the interval width by averaging the widths obtained from the K samples, which is denoted as $2\omega^*(N)$.

Step 3. Repeat Steps 1 and 2 via smaller (or larger) N if the $2\omega^*(N)$ is less (or greater) than 2ω.

Step 4. Repeat Step 3 until the approximate half-interval-width $\omega^*(N)$ is close to ω, i.e., $N = \min\{N : |\omega^*(N) - \omega| \leq 0.001\}$. The resulting N is the approximate sample size.

The sample sizes of Wald CIs, CI based on T_l, and CI based on T_{sc} obtained by the above algorithm are denoted as N_{w1}, N_{w2}, N_l and N_{sc}, respectively.

6. Simulation Studies

6.1. Empirical Study for Hypothesis Testing

In this section, we investigate the performance of various test procedures for the hypothesis testing $H_0 : \phi = 1$ versus $H_1 : \phi \neq 1$ proposed in Section 3. We consider the following sample size designs: three balanced sample size designs, i.e., (i) moderate sample size $n_1 = n_2 = 100$; (ii) large sample size $n_1 = n_2 = 150$; (iii) large sample size $n_1 = n_2 = 200$ and three unbalanced sample size designs, i.e., (iv) moderate sample size $n_1 = 50$ and $n_2 = 100$; (v) large sample size $n_1 = 150$ and $n_2 = 100$; (vi) large sample size $n_1 = 200$ and $n_2 = 250$. We do not consider the approximate unconditional methods for moderate to large sample sizes due to the extremely large number of possible values for $\mathbf{m} = (n_{00}^{(1)}, n_{10}^{(1)}, n_{01}^{(1)}, n_{11}^{(1)}, n_{00}^{(2)}, n_{10}^{(2)}, n_{01}^{(2)}, n_{11}^{(2)})$. However, the comparisons between the asymptotic (AS) and approximate unconditional (AU) methods are considered for small sample settings, including (vii) $n_1 = n_2 = 10$ and (viii) $n_1 = n_2 = 15$. For each sample size design, we consider twelve different settings for the nuisance parameters $(\pi_{00}^{(1)}, \pi_{11}^{(1)}, \pi_{00}^{(2)}, \pi_{11}^{(2)}, \pi_{10}^{(2)})$ to investigate the performance of the various test procedures, and it is given in Table 2.

Table 2. Parameter settings for the hypothesis testing $H_0 : \phi = 1$ versus $H_0 : \phi \neq 1$.

Par.	$(\pi_{00}^{(1)}, \pi_{11}^{(1)}, \pi_{00}^{(2)}, \pi_{11}^{(2)}, \pi_{10}^{(2)})$	Par.	$(\pi_{00}^{(1)}, \pi_{11}^{(1)}, \pi_{00}^{(2)}, \pi_{11}^{(2)}, \pi_{10}^{(2)})$
A_1	(0.5, 0.25, 0.4, 0.35, 0.10)	A_7	(0.5, 0.20, 0.4, 0.10, 0.25)
A_2	(0.5, 0.25, 0.4, 0.35, 0.15)	A_8	(0.5, 0.20, 0.4, 0.10, 0.30)
A_3	(0.4, 0.25, 0.4, 0.25, 0.15)	A_9	(0.5, 0.20, 0.2, 0.30, 0.30)
A_4	(0.4, 0.25, 0.4, 0.25, 0.20)	A_{10}	(0.5, 0.20, 0.2, 0.20, 0.30)
A_5	(0.5, 0.20, 0.5, 0.20, 0.15)	A_{11}	(0.5, 0.20, 0.2, 0.20, 0.40)
A_6	(0.5, 0.20, 0.4, 0.10, 0.20)	A_{12}	(0.3, 0.25, 0.4, 0.30, 0.15)

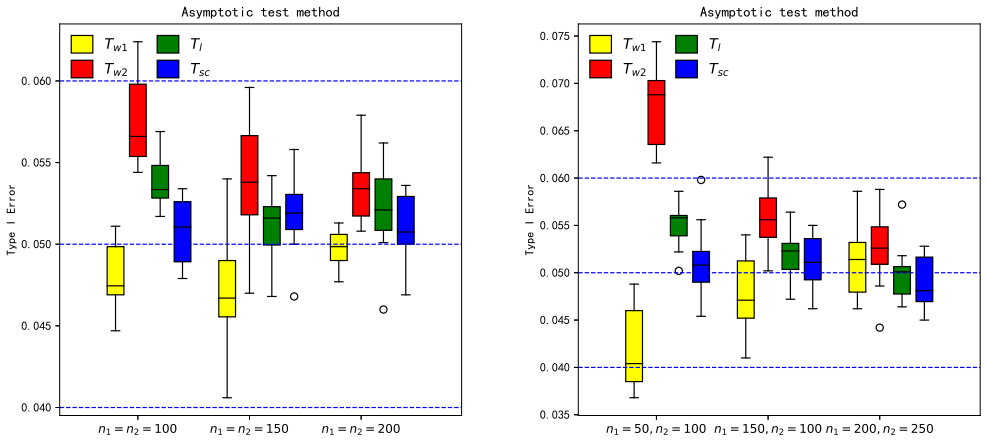
First, we examine the behavior of the actual test size, that is, the actual type I error rate of the testing procedure. For each sample size design and parameter setting, we generate $M = 5000$ observed samples for all test statistics. For the asymptotical test procedures, the empirical type I error rate for a given test $T_i(i = w1, w2, l, sc)$, at significance level $\alpha = 0.05$ for the settings under consideration, is simply estimated by (the number of rejections of H_0 by test T_i at the α level)$/M$ when $\phi = 1$. The results for moderate to large sample sizes are displayed in Figure 1. According to the Reviewer's suggestion, we report the CMLEs and the MLEs of parameters $\pi_{00}^{(1)}, \pi_{11}^{(1)}, \pi_{01}^{(1)}, \pi_{10}^{(1)}, \pi_{00}^{(2)}, \pi_{11}^{(2)}, \pi_{01}^{(2)}, \pi_{10}^{(2)}$ under small, moderate and large sample size designs via the online Supplemental Materials, respectively. Note that the CMLEs of parameters $\pi_{00}^{(1)}$, $\pi_{11}^{(1)}$, $\pi_{00}^{(2)}$ and $\pi_{11}^{(2)}$ are the same as their MLEs; we only report the results of parameters $\pi_{01}^{(1)}$, $\pi_{10}^{(1)}$, $\pi_{01}^{(2)}$ and $\pi_{10}^{(2)}$. For the approximate unconditional test procedures, the actual type I error rate at the α level can be obtained by

$$\sum_{m} I\{T_i(m); \alpha\}[\exp(l_2(\mathbf{m}; \phi = 1, \pi_{10}^{(2)}, \pi_{00}^{(1)}, \pi_{11}^{(1)}, \pi_{00}^{(2)}, \pi_{11}^{(2)}))].$$

where $I\{T_i(m); \alpha\} = 1$ when the null hypothesis is rejected at α level, and $I\{T_i(m); \alpha\} = 0$ otherwise. According to Tang et al. [24], a test is liberal if its actual type I error rate exceeds 0.06, conservative if it falls below 0.04, and robust if it lies within the interval [0.04, 0.06] at 0.05 significant level. We summarize the results for $n_1 = n_2 = 10$ and are $n_1 = n_2 = 15$ in Table 3.

Table 3. The actual type I error rates (percent) of various test procedures for testing $H_0 : \phi = 1$ at $\alpha = 0.05$ under $n_1 = n_2 = n$.

		T_{w1}		T_{w2}		T_l		T_{sc}	
n	**Par.**	**AS**	**AU**	**AS**	**AU**	**AS**	**AU**	**AS**	**AU**
10	A_1	0.06	3.02	3.38	4.04	1.90	4.93	0.64	4.97
	A_2	0.12	3.02	3.32	4.04	1.50	4.93	0.90	5.11
	A_3	0.24	4.14	5.04	5.45	2.90	5.51	1.44	5.79
	A_4	0.44	4.14	5.52	5.45	2.36	5.51	1.28	5.89
	A_5	0.10	3.94	4.98	5.06	2.50	5.48	1.58	5.74
	A_6	0.62	3.67	5.24	4.59	2.96	5.03	2.00	5.33
	A_7	0.60	4.51	4.86	5.26	3.62	5.25	1.92	5.55
	A_8	0.32	4.91	4.84	5.42	2.74	5.06	1.96	5.38
	A_9	0.32	4.81	5.42	5.49	2.74	5.13	1.98	5.46
	A_{10}	0.40	4.42	6.08	5.12	3.20	5.18	3.00	5.42
	A_{11}	0.40	5.24	4.68	5.34	2.94	4.83	2.36	5.12
	A_{12}	0.24	4.23	5.32	5.44	3.16	5.45	1.86	5.74
15	A_1	0.22	5.06	5.16	6.51	3.30	4.72	1.84	5.03
	A_2	0.40	4.05	5.80	5.30	2.92	4.72	2.14	5.08
	A_3	1.20	5.26	8.26	5.85	4.92	5.99	3.72	6.06
	A_4	1.22	5.26	7.92	5.85	4.46	5.99	2.84	6.08
	A_5	0.82	4.99	7.34	5.88	4.08	5.67	2.84	5.91
	A_6	1.42	4.64	7.26	5.33	4.44	5.67	4.04	5.69
	A_7	1.44	5.44	7.14	5.69	4.46	5.85	3.32	5.85
	A_8	1.24	5.82	6.76	5.75	4.38	5.66	3.68	5.70
	A_9	1.40	5.78	7.10	5.75	4.44	5.74	3.82	5.72
	A_{10}	1.94	5.33	7.10	5.68	4.56	5.70	3.74	5.49
	A_{11}	1.12	5.91	5.94	5.61	4.16	5.36	3.50	5.28
	A_{12}	1.38	5.29	7.54	5.75	5.04	5.98	3.54	6.00



(**a**) Balanced sample size design (**b**) Unbalanced sample size design

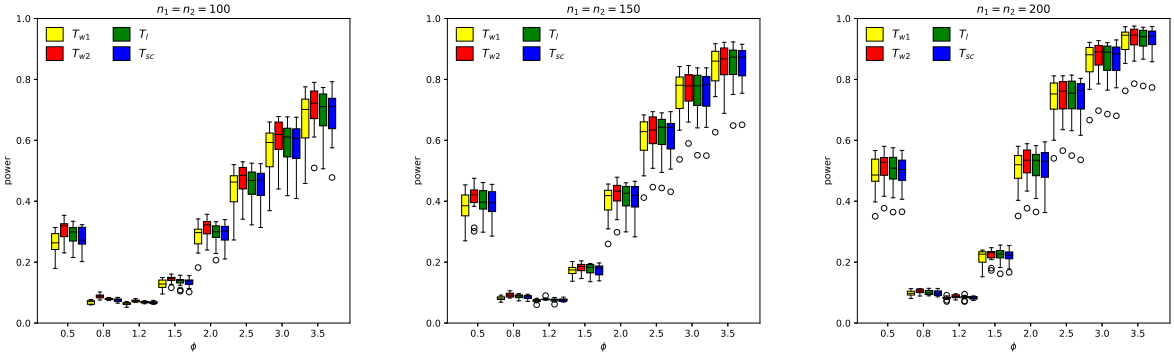
Figure 1. Boxplots of actual Type I errors for various test procedures under (**a**) balanced sample size designs: $n_1 = n_2 = 100, 150, 200$ and (**b**) unbalanced sample size designs: $n_1 = 50, n_2 = 100$; $n_1 = 150, n_2 = 100$; $n_1 = 200, n_2 = 250$.

For the actual power performance of various test procedures, we consider the above sample size designs and the above nuisance parameter settings and $\phi_1 = 0.5$, 0.8, 1.2, $1.5(0.5)3.5$, where $a(b)c$ denotes the value is from a to c with step size b. For the asymptotical test procedures, the empirical power for a given test $T_i (i = w1, w2, l, sc)$, at significance level $\alpha = 0.05$ for the settings under consideration, is estimated by (the number of rejections

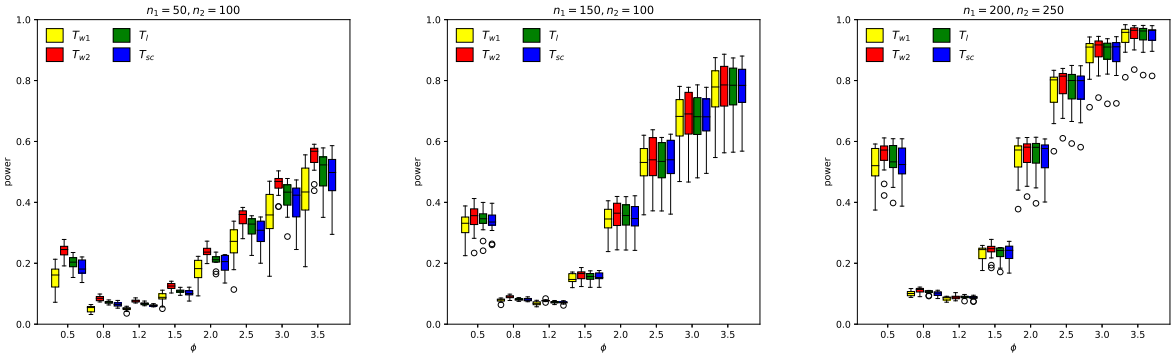
of H_0 by test T_i at the α level)/M when $\phi = \phi_1$. The simulation results for moderate to large sample sizes are displayed in Figure 2. For the approximate unconditional test procedures, the actual power at the α level can be obtained by

$$\sum_{m} I\{T_i(m); \alpha\}[\exp(l_2(\mathbf{m}; \phi = \phi_1, \pi_{10}^{(2)}, \pi_{00}^{(1)}, \pi_{11}^{(1)}, \pi_{00}^{(2)}, \pi_{11}^{(2)}))].$$

We only reported the simulation results for the sample size design $n_1 = n_2 = 15$ by Figure 3 due to the space limitation. According to Figures 1 and 2, we can observe that the asymptotical test procedures based on all test statistics perform well in the sense that the empirical type I errors are close to the nominal level under large sample sizes, and T_{w2} has a slightly inflated type I error under some moderate sample sizes (e.g., $n_1 = 50, n_2 = 100$). All test procedures have similar empirical powers under moderate to large sample sizes, except that T_{w2} has slightly more power than the other test procedures. As expected, the empirical powers increase in all test procedures with the increase in the absolute difference between $\phi = 1$ and ϕ_1. When the sample size is large (e.g., sample sizes $n_1 = 200$ and $n_2 = 250$), four test statistics have similar performance; when the sample size increases, as expected, these four testing procedures are asymptotically equivalent. According to Table 3 and Figure 3, since the sample size is very small, the power performance was not satisfactory in some cases for both asymptotical and approximate unconditional test procedures. However, approximate unconditional procedures usually outperform the asymptotic test procedures in the sense that their empirical type I errors are closer to the nominal level and have higher powers under these small sample sizes. Therefore, when the sample size is small, the approximate unconditional test procedures based on these four test statistics are recommended for practical applications.



(**a**) Balanced sample size design



(**b**) Unbalanced sample size design

Figure 2. Boxplots of the actual powers for various test procedures for testing $H_0 : \phi = 1$ versus $H_1 : \phi = \phi_1 \neq 1$ at $\alpha = 0.05$ under (**a**) balanced sample size designs: $n_1 = n_2 = 100, 150, 200$ and (**b**) unbalanced sample size designs: $n_1 = 50, n_2 = 100$; $n_1 = 150, n_2 = 100$; $n_1 = 200, n_2 = 250$.

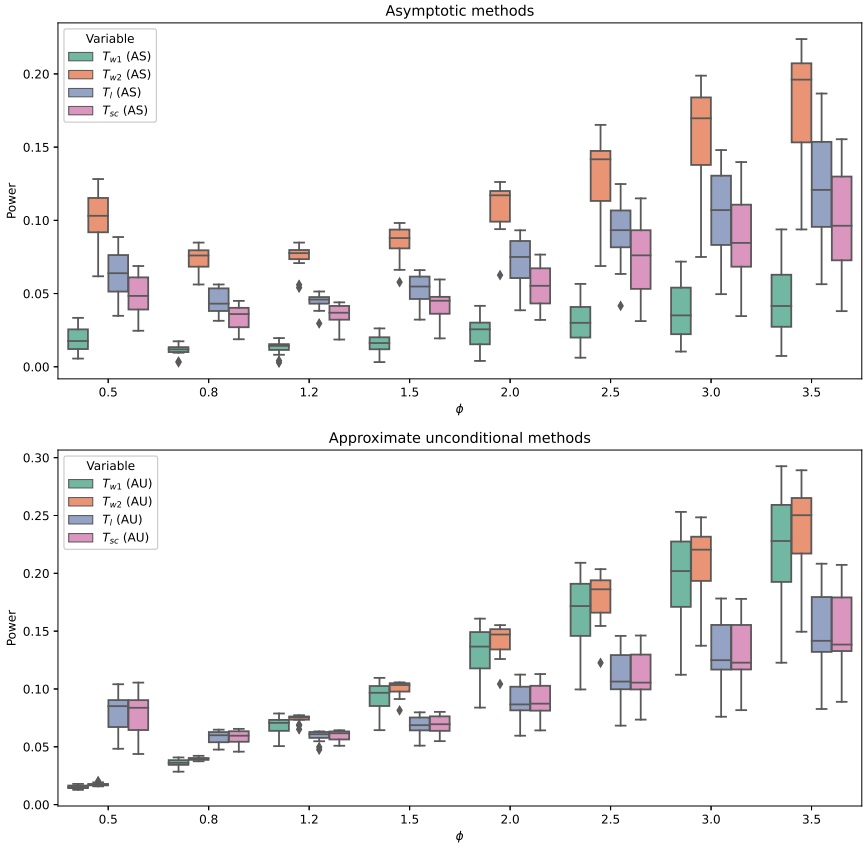


Figure 3. Comparison of empirical powers between asymptotic methods and approximate unconditional methods under small sample size design $n_1 = n_2 = 15$.

6.2. Empirical Study for Confidence Interval

In this section, we investigate the performance of these CIs for ϕ under various settings. The same parameter settings as those in Section 6.1 are considered. Four sample size designs were also considered, i.e., (i) $n_1 = n_2 = 100$; (ii) $n_1 = n_2 = 200$; (iii) $n_1 = 150$ and $n_2 = 100$; (iv) $n_1 = 200$ and $n_2 = 250$. Under each sample size setting and each parameter combination, the empirical coverage probability, the empirical coverage width and the left and right non-coverage probability of the confidence interval are calculated by repeated simulation $K = 5000$ times. The three indices are calculated as follows:

(i) Empirical coverage probability (ECP)

$$\text{ECP} = \frac{1}{K}\sum_{k=1}^{K} I\{\phi \in [\phi_l(\mathbf{m}^{(k)}), \phi_u(\mathbf{m}^{(k)})]\},$$

where $[\phi_l(\mathbf{m}^{(k)}), \phi_u(\mathbf{m}^{(k)})]$ is the confidence interval of ϕ at the kth replication, $I(\cdot)$ is the indicator function, $\mathbf{m}^{(k)} = (n_{00}^{(1)}, n_{10}^{(1)}, n_{01}^{(1)}, n_{11}^{(1)}, n_{00}^{(2)}, n_{10}^{(2)}, n_{01}^{(2)}, n_{11}^{(2)})^{(k)}$. Since the empirical coverage probability is defined as the proportion that the K confidence interval includes the true value of the interest parameter, then the closer the empirical coverage probability is to the nominal confidence level (e.g., 95%), the better the performance of the proposed method.

(ii) Empirical coverage width (ECW)

$$\text{ECW} = \frac{1}{K}\sum_{k=1}^{K}\{\phi \in [\phi_u(\mathbf{m}^{(k)}) - \phi_l(\mathbf{m}^{(k)})]\}.$$

It is obvious that a small empirical coverage width suggests a better confidence interval procedure if coverage can be maintained.

(iii) Left and right non-coverage probability (LNCP, RNCP)

$$\text{LNCP} = \frac{1}{K}\sum_{k=1}^{K} I\{\phi < \phi_l(\mathbf{m}^{(k)})\}, \text{RNCP} = \frac{1}{K}\sum_{k=1}^{K} I\{\phi > \phi_u(\mathbf{m}^{(k)})\}.$$

Since left and right non-coverage probabilities are calculated as the proportions that the lower limit is above and the upper limit is below the true parameter value, respectively, so the confidence interval is regarded as having satisfactory interval location if the confidence interval procedure has equal left and right non-coverage probabilities.

Note that when $n_{01}^{(1)} = 0$ or $n_{10}^{(1)} = 0$ or $n_{01}^{(2)} = 0$ or $n_{10}^{(2)} = 0$, the variance $Var(\boldsymbol{c}'\hat{\boldsymbol{\beta}})$ is undefined and its corresponding test statistics are undefined, so all Wald CIs are undefined. In this case, we use the commonly used adjustment for sparse data structures in contingency table analysis by adding 0.5 to each cell. For space limitations, we only present the simulation results for balanced sample sizes $(n_1, n_2) = (100, 100), (200, 200)$ and unbalanced sample sizes $(n_1, n_2) = (150, 100)$ in Tables 4–6, respectively. When the sample size is large, we can observe that (i) all CIs perform well in the sense that their empirical coverage probabilities are very close to the nominal confidence level; (ii) all CIs have satisfactory interval locations due to producing approximately symmetrical left-right non-coverage probabilities; (iii) similar interval widths for all CIs. However, when the sample size is not large (e.g., $n_1 = n_2 = 100$), CI_{w2} based on the Wald statistic appears to perform poorly in some parameter settings, for example, A_2, A_8, A_9 and A_{11}. Also, there is a slightly unbalance is observed in terms of left and right non-coverage probabilities and their expected confidence width is relatively wider than other CIs under those parameter sittings.

Table 4. The empirical coverage probability (percent), empirical coverage width, left and right non coverage probabilities (percent) of 95% confidence interval under sample size $(n_1, n_2) = (100, 100)$.

Par.	CI_{w1} $ECP(L,R)ECW$	CI_{w2} $ECP(L,R)ECW$	CI_l $ECP(L,R)ECW$	CI_{sc} $ECP(L,R)ECW$
		$\phi = 0.5$		
A1	95.18 (2.50, 2.32) 1.68	94.02 (2.40, 3.58) 4.66	94.48 (2.56, 2.96) 1.68	94.66 (2.72, 2.62) 1.64
A2	95.98 (2.40, 1.62) 1.93	93.08 (1.84, 5.08) 2.07	94.56 (2.52, 2.92) 1.94	94.92 (2.76, 2.32) 1.85
A3	95.18 (2.44, 2.38) 1.25	94.14 (2.32, 3.54) 3.97	94.60 (2.44, 2.96) 1.25	94.88 (2.52, 2.60) 1.23
A4	95.88 (2.02, 2.10) 1.35	94.64 (1.90, 3.46) 1.39	95.08 (2.08, 2.84) 1.34	95.36 (2.22, 2.42) 1.32
A5	95.26 (2.36, 2.38) 1.45	94.20 (2.16, 3.64) 1.50	94.70 (2.42, 2.88) 1.45	94.78 (2.62, 2.60) 1.43
A6	95.34 (2.66, 2.00) 1.19	94.46 (2.56, 2.98) 1.23	94.98 (2.64, 2.38) 1.19	95.06 (2.70, 2.24) 1.18
A7	95.88 (2.22, 1.90) 1.21	94.04 (1.96, 4.00) 1.21	95.08 (2.14, 2.78) 1.20	95.38 (2.40, 2.22) 1.19
A8	96.04 (2.26, 1.70) 1.31	92.22 (1.66, 6.12) 1.28	94.62 (2.06, 3.32) 1.28	95.20 (2.44, 2.36) 1.28
A9	95.56 (2.42, 2.02) 1.30	92.20 (1.82, 5.98) 1.26	94.46 (2.30, 3.24) 1.27	94.92 (2.58, 2.50) 1.27
A10	95.90 (2.16, 1.94) 1.16	94.52 (1.68, 3.80) 1.16	95.44 (2.08, 2.48) 1.14	95.48 (2.32, 2.20) 1.14
A11	96.26 (2.46, 1.28) 1.38	90.78 (1.36, 7.86) 1.30	94.48 (2.16, 3.36) 1.33	95.04 (2.68, 2.28) 1.34
A12	95.46 (2.16, 2.38) 1.25	94.58 (2.10, 3.32) 1.55	94.90 (2.24, 2.86) 1.26	95.12 (2.30, 2.58) 1.23
		$\phi = 0.8$		
A1	95.24 (2.52, 2.24) 2.80	94.14 (2.52, 3.34) 3.09	94.64 (2.72, 2.64) 2.86	94.76 (2.72, 2.52) 2.73
A2	95.36 (2.48, 2.16) 2.93	93.34 (2.50, 4.16) 4.33	94.52 (2.76, 2.72) 3.01	94.88 (2.70, 2.42) 2.83
A3	95.04 (2.46, 2.50) 2.02	94.38 (2.66, 2.96) 2.13	94.64 (2.62, 2.74) 2.04	94.78 (2.58, 2.64) 1.99
A4	95.26 (2.40, 2.34) 2.08	94.46 (2.64, 2.90) 2.21	94.84 (2.60, 2.56) 2.10	94.94 (2.60, 2.46) 2.04
A5	95.30 (2.38, 2.32) 2.29	94.44 (2.64, 2.92) 2.46	94.78 (2.56, 2.66) 2.32	95.02 (2.54, 2.44) 2.25
A6	95.24 (2.68, 2.08) 1.97	94.54 (2.98, 2.48) 2.19	94.88 (2.86, 2.26) 2.01	94.96 (2.80, 2.24) 1.93
A7	95.58 (2.28, 2.14) 1.90	94.56 (2.40, 3.04) 2.00	95.22 (2.34, 2.44) 1.91	95.34 (2.44, 2.22) 1.87
A8	95.24 (2.40, 2.36) 1.98	93.86 (2.32, 3.82) 2.03	94.64 (2.50, 2.86) 1.97	94.86 (2.54, 2.60) 1.95
A9	95.56 (2.22, 2.22) 1.96	94.28 (2.14, 3.58) 2.01	95.10 (2.30, 2.60) 1.95	95.26 (2.36, 2.38) 1.93
A10	95.66 (2.36, 1.98) 1.81	94.50 (2.54, 2.96) 1.91	95.16 (2.50, 2.34) 1.82	95.38 (2.44, 2.18) 1.78
A11	95.70 (2.28, 2.02) 2.00	92.30 (1.84, 5.86) 1.99	94.68 (2.24, 3.08) 1.98	94.94 (2.54, 2.52) 1.96
A12	95.24 (2.04, 2.72) 1.97	94.58 (2.26, 3.16) 2.10	94.90 (2.22, 2.88) 1.99	95.04 (2.20, 2.76) 1.94

Table 4. *Cont.*

Par.	CI_{w1} $ECP(L,R)ECW$	CI_{w2} $ECP(L,R)ECW$	CI_{l} $ECP(L,R)ECW$	CI_{sc} $ECP(L,R)ECW$
		$\phi = 1.2$		
A1	95.10 (2.54, 2.36) 4.69	94.14 (0.12, 5.64) 5.03	94.40 (3.00, 2.60) 5.04	94.72 (2.80, 2.48) 4.50
A2	95.18 (2.56, 2.26) 4.37	93.68 (2.70, 3.62) 5.11	94.52 (2.94, 2.54) 4.55	94.78 (2.76, 2.46) 4.24
A3	95.12 (2.40, 2.48) 3.19	94.44 (2.72, 2.84) 3.58	94.74 (2.60, 2.66) 3.26	94.84 (2.54, 2.62) 3.14
A4	95.22 (2.52, 2.26) 3.11	94.48 (2.88, 2.64) 3.40	94.70 (2.78, 2.52) 3.16	94.86 (2.62, 2.52) 3.06
A5	95.36 (2.36, 2.28) 3.54	94.46 (2.84, 2.70) 3.96	94.92 (2.66, 2.42) 3.63	95.04 (2.54, 2.42) 3.48
A6	95.38 (2.44, 2.18) 3.23	94.62 (2.68, 2.70) 3.85	94.82 (2.94, 2.24) 3.39	94.96 (2.72, 2.32) 3.15
A7	95.64 (2.20, 2.16) 2.94	94.80 (2.76, 2.44) 3.30	95.16 (2.54, 2.30) 3.01	95.28 (2.38, 2.34) 2.89
A8	95.22 (2.38, 2.40) 2.93	94.44 (2.60, 2.96) 3.13	94.80 (2.52, 2.68) 2.96	94.84 (2.54, 2.62) 2.88
A9	95.64 (2.14, 2.22) 2.90	94.82 (2.40, 2.78) 3.10	95.20 (2.34, 2.46) 2.93	95.32 (2.32, 2.36) 2.86
A10	95.52 (2.26, 2.22) 2.80	94.88 (2.72, 2.40) 3.17	95.10 (2.56, 2.34) 2.87	95.30 (2.38, 2.32) 2.75
A11	95.50 (2.44, 2.06) 2.89	94.24 (2.44, 3.32) 3.00	95.08 (2.56, 2.36) 2.90	95.14 (2.58, 2.28) 2.84
A12	94.98 (2.64, 2.38) 3.03	94.38 (2.98, 2.64) 3.31	94.68 (2.84, 2.48) 3.08	94.78 (2.72, 2.50) 2.98

Table 5. The empirical coverage probability (percent), empirical coverage width, left and right non-coverage probabilities (percent) of a 95% confidence interval under sample size $(n_1, n_2) = (200, 200)$.

Par.	CI_{w1} $ECP(L,R)ECW$	CI_{w2} $ECP(L,R)ECW$	CI_{l} $ECP(L,R)ECW$	CI_{sc} $ECP(L,R)ECW$
		$\phi = 0.5$		
A1	94.90 (2.70, 2.40) 0.97	94.42 (2.56, 3.02) 0.97	94.64 (2.72, 2.64) 0.97	94.72 (2.78, 2.50) 0.96
A2	95.48 (2.14, 2.38) 1.07	94.54 (2.00, 3.46) 1.08	95.14 (2.12, 2.74) 1.06	95.18 (2.26, 2.56) 1.05
A3	94.48 (2.90, 2.62) 0.78	94.12 (2.76, 3.12) 0.78	94.22 (2.90, 2.88) 0.77	94.34 (2.98, 2.68) 0.77
A4	95.38 (2.28, 2.34) 0.81	94.80 (2.08, 3.12) 0.81	95.20 (2.28, 2.52) 0.81	95.20 (2.36, 2.44) 0.81
A5	95.58 (2.42, 2.00) 0.87	95.00 (2.20, 2.80) 0.87	95.12 (2.38, 2.50) 0.86	95.44 (2.44, 2.12) 0.86
A6	95.14 (2.44, 2.42) 0.74	94.72 (2.34, 2.94) 0.75	94.98 (2.46, 2.56) 0.74	95.02 (2.52, 2.46) 0.74
A7	95.32 (2.64, 2.04) 0.76	94.78 (2.32, 2.90) 0.76	95.16 (2.48, 2.36) 0.76	95.18 (2.70, 2.12) 0.75
A8	95.34 (2.74, 1.92) 0.82	94.86 (2.28, 2.86) 0.81	95.08 (2.66, 2.26) 0.81	95.06 (2.84, 2.10) 0.81
A9	95.42 (2.52, 2.06) 0.82	94.68 (2.02, 3.30) 0.81	94.96 (2.40, 2.64) 0.81	95.08 (2.66, 2.26) 0.81
A10	95.32 (2.52, 2.16) 0.73	94.62 (2.24, 3.14) 0.73	94.98 (2.44, 2.58) 0.73	95.04 (2.58, 2.38) 0.73
A11	95.94 (2.22, 1.84) 0.85	94.88 (1.48, 3.64) 0.83	95.52 (2.06, 2.42) 0.84	95.52 (2.44, 2.04) 0.84
A12	94.88 (2.58, 2.54) 0.77	94.20 (2.56, 3.24) 0.78	94.52 (2.58, 2.90) 0.77	94.66 (2.62, 2.72) 0.76
		$\phi = 0.8$		
A1	94.98 (2.76, 2.26) 1.57	94.26 (2.92, 2.82) 1.62	94.70 (2.80, 2.50) 1.58	94.86 (2.80, 2.34) 1.56
A2	95.36 (2.18, 2.46) 1.60	94.56 (2.36, 3.08) 1.66	94.92 (2.34, 2.74) 1.61	95.06 (2.36, 2.58) 1.59
A3	94.90 (2.74, 2.36) 1.23	94.56 (2.84, 2.60) 1.25	94.70 (2.82, 2.48) 1.24	94.70 (2.88, 2.42) 1.23
A4	95.84 (2.08, 2.08) 1.24	95.54 (2.18, 2.28) 1.27	95.66 (2.20, 2.14) 1.25	95.70 (2.20, 2.10) 1.24
A5	95.58 (2.24, 2.18) 1.36	95.22 (2.38, 2.40) 1.38	95.38 (2.36, 2.26) 1.36	95.38 (2.38, 2.24) 1.35
A6	94.88 (2.56, 2.56) 1.19	94.60 (2.72, 2.68) 1.23	94.74 (2.62, 2.64) 1.20	94.76 (2.60, 2.64) 1.19
A7	94.90 (2.40, 2.70) 1.18	94.42 (2.40, 3.18) 1.19	94.70 (2.46, 2.84) 1.18	94.76 (2.46, 2.78) 1.17
A8	95.08 (2.56, 2.36) 1.23	94.56 (2.42, 3.02) 1.23	94.66 (2.58, 2.76) 1.22	94.88 (2.60, 2.52) 1.22
A9	95.04 (2.66, 2.30) 1.23	94.66 (2.50, 2.84) 1.24	94.82 (2.64, 2.54) 1.23	94.82 (2.78, 2.40) 1.22
A10	95.28 (2.52, 2.20) 1.13	94.84 (2.56, 2.60) 1.15	95.14 (2.58, 2.28) 1.14	95.16 (2.62, 2.22) 1.13
A11	95.04 (2.76, 2.20) 1.25	94.28 (2.32, 3.40) 1.24	94.60 (2.68, 2.72) 1.24	94.84 (2.78, 2.38) 1.23
A12	94.62 (2.66, 2.72) 1.21	94.16 (2.84, 3.00) 1.24	94.38 (2.78, 2.84) 1.22	94.46 (2.74, 2.80) 1.21
		$\phi = 1.2$		
A1	94.78 (2.70, 2.52) 2.48	94.00 (3.26, 2.74) 2.64	94.36 (2.98, 2.66) 2.51	94.46 (2.88, 2.66) 2.45
A2	95.12 (2.42, 2.46) 2.40	94.80 (2.66, 2.54) 2.53	94.88 (2.62, 2.50) 2.43	95.00 (2.50, 2.50) 2.38
A3	94.80 (2.48, 2.72) 1.90	94.42 (2.74, 2.84) 1.97	94.60 (2.58, 2.82) 1.91	94.64 (2.52, 2.84) 1.89
A4	95.20 (2.44, 2.36) 1.85	94.88 (2.70, 2.42) 1.90	94.98 (2.58, 2.44) 1.86	95.02 (2.54, 2.44) 1.84
A5	95.30 (2.32, 2.38) 2.05	94.82 (2.64, 2.54) 2.12	95.00 (2.52, 2.48) 2.07	95.08 (2.44, 2.48) 2.04
A6	95.18 (2.54, 2.28) 1.87	94.76 (3.12, 2.12) 2.00	94.90 (2.84, 2.26) 1.90	94.92 (2.70, 2.38) 1.86
A7	95.36 (2.52, 2.12) 1.78	94.98 (2.92, 2.10) 1.85	95.18 (2.70, 2.12) 1.80	95.30 (2.58, 2.12) 1.77
A8	95.14 (2.62, 2.24) 1.80	94.82 (2.78, 2.40) 1.84	94.96 (2.70, 2.34) 1.80	95.06 (2.66, 2.28) 1.79
A9	94.94 (2.54, 2.52) 1.81	94.52 (2.72, 2.76) 1.85	94.80 (2.60, 2.60) 1.82	94.84 (2.62, 2.54) 1.80
A10	95.32 (2.24, 2.44) 1.71	94.74 (2.74, 2.52) 1.77	95.08 (2.42, 2.50) 1.72	95.16 (2.34, 2.50) 1.70
A11	95.00 (2.46, 2.54) 1.78	94.54 (2.48, 2.98) 1.80	94.78 (2.54, 2.68) 1.78	94.88 (2.56, 2.56) 1.76
A12	95.00 (2.44, 2.56) 1.83	94.54 (2.72, 2.74) 1.88	94.82 (2.56, 2.62) 1.84	94.84 (2.52, 2.64) 1.82

Table 6. The empirical coverage probability (percent), empirical coverage width, left and right non-coverage probabilities (percent) of a 95% confidence interval under sample size $(n_1, n_2) = (150, 100)$.

Par.	CI_{w1} $ECP(L,R)ECW$	CI_{w2} $ECP(L,R)ECW$	CI_l $ECP(L,R)ECW$	CI_{sc} $ECP(L,R)ECW$
		$\phi = 0.5$		
A1	95.36 (2.16, 2.48) 1.41	94.04 (2.10, 3.86)1.47	94.62 (2.24, 3.14) 1.41	94.94 (2.34, 2.72) 1.39
A2	95.38 (2.12, 2.50) 1.66	93.86 (1.74, 4.40)1.78	94.78 (2.38, 2.84) 1.73	94.98 (2.34, 2.68) 1.60
A3	95.20 (2.08, 2.72) 1.08	94.44 (2.02, 3.54)1.09	94.84 (2.10, 3.06) 1.07	94.90 (2.18, 2.92) 1.06
A4	95.06 (2.34, 2.60) 1.17	94.02 (2.46, 3.52)1.25	94.50 (2.48, 3.02) 1.18	94.56 (2.54, 2.90) 1.15
A5	95.24 (2.40, 2.36) 1.26	94.32 (2.48, 3.20)1.32	94.78 (2.58, 2.64) 1.26	94.90 (2.58, 2.52) 1.23
A6	95.14 (2.64, 2.22) 1.01	94.52 (2.52, 2.96)1.01	94.82 (2.66, 2.52) 1.00	94.82 (2.84, 2.34) 1.00
A7	95.36 (2.46, 2.18) 1.02	94.80 (2.32, 2.88)1.03	95.14 (2.44, 2.42) 1.02	95.22 (2.54, 2.24) 1.01
A8	95.06 (2.54, 2.40) 1.11	94.20 (2.18, 3.62)1.12	94.60 (2.50, 2.90) 1.10	94.72 (2.70, 2.58) 1.09
A9	96.00 (2.08, 1.92) 1.09	94.82 (1.78, 3.40)1.10	95.22 (2.04, 2.74) 1.09	95.50 (2.28, 2.22) 1.08
A10	95.52 (2.50, 1.98) 0.97	94.68 (2.22, 3.10)0.97	95.08 (2.48, 2.44) 0.97	95.24 (2.58, 2.18) 0.96
A11	95.78 (2.60, 1.62) 1.14	94.26 (1.94, 3.80)1.14	95.02 (2.50, 2.48) 1.12	95.20 (2.78, 2.02) 1.12
A12	95.16 (2.40, 2.44) 1.13	94.22 (2.64, 3.14)1.21	94.64 (2.58, 2.78) 1.14	94.80 (2.64, 2.56) 1.11
		$\phi = 0.8$		
A1	95.40 (2.22, 2.38) 2.32	93.96 (2.54, 3.50) 2.47	94.82 (2.42, 2.76) 2.34	94.86 (2.80, 2.34) 1.56
A2	95.48 (2.26, 2.26) 2.52	94.24 (2.02, 3.74) 2.79	94.80 (2.68, 2.52) 2.63	95.06 (2.36, 2.58) 1.59
A3	95.58 (2.22, 2.20) 1.75	94.66 (2.34, 3.00) 1.80	95.18 (2.28, 2.54) 1.75	94.70 (2.88, 2.42) 1.23
A4	94.52 (2.32, 3.16) 1.80	93.48 (2.90, 3.62) 1.97	94.14 (2.54, 3.32) 1.83	95.70 (2.20, 2.10) 1.24
A5	95.02 (2.42, 2.56) 1.98	94.02 (2.70, 3.28) 2.14	94.52 (2.56, 2.92) 2.01	95.38 (2.38, 2.24) 1.35
A6	94.70 (2.46, 2.84) 1.60	94.04 (2.66, 3.30) 1.66	94.40 (2.56, 3.04) 1.61	94.76 (2.60, 2.64) 1.19
A7	95.22 (2.30, 2.48) 1.59	94.50 (2.42, 3.08) 1.63	94.92 (2.46, 2.62) 1.59	94.76 (2.46, 2.78) 1.17
A8	95.04 (2.46, 2.50) 1.67	94.44 (2.56, 3.00) 1.72	94.74 (2.52, 2.74) 1.67	94.88 (2.60, 2.52) 1.22
A9	94.70 (2.50, 2.80) 1.65	94.02 (2.52, 3.46) 1.71	94.40 (2.56, 3.04) 1.66	94.82 (2.78, 2.40) 1.22
A10	95.46 (2.06, 2.48) 1.49	95.08 (2.08, 2.84) 1.53	95.26 (2.06, 2.68) 1.50	95.16 (2.62, 2.22) 1.13
A11	95.08 (2.52, 2.40) 1.67	94.18 (2.50, 3.32) 1.71	94.68 (2.54, 2.78) 1.67	94.84 (2.78, 2.38) 1.23
A12	95.22 (2.30, 2.48) 1.78	94.48 (2.68, 2.84) 1.94	94.72 (2.64, 2.64) 1.80	94.46 (2.74, 2.80) 1.21
		$\phi = 1.2$		
A1	95.36 (2.40, 2.24) 3.65	93.94 (2.72, 3.34) 4.03	94.78 (2.62, 2.60) 3.71	94.46 (2.88, 2.66) 2.45
A2	95.12 (2.18, 2.70) 3.74	93.72 (2.18, 4.10) 4.22	94.60 (2.52, 2.88) 3.93	95.00 (2.50, 2.50) 2.38
A3	95.18 (2.52, 2.30) 2.69	94.42 (2.68, 2.90) 2.84	94.74 (2.60, 2.66) 2.71	94.64 (2.52, 2.84) 1.89
A4	94.80 (2.30, 2.90) 2.72	94.18 (2.84, 2.98) 2.99	94.56 (2.50, 2.94) 2.77	95.02 (2.54, 2.44) 1.84
A5	95.40 (2.02, 2.58) 2.95	94.76 (2.40, 2.84) 3.21	95.10 (2.16, 2.74) 3.00	95.08 (2.44, 2.48) 2.04
A6	94.88 (2.58, 2.54) 2.59	94.28 (3.08, 2.64) 2.84	94.56 (2.78, 2.66) 2.64	94.92 (2.70, 2.38) 1.86
A7	94.66 (2.10, 3.24) 2.39	94.26 (2.36, 3.38) 2.51	94.42 (2.28, 3.30) 2.42	95.30 (2.58, 2.12) 1.77
A8	94.78 (2.52, 2.70) 2.47	94.28 (2.72, 3.00) 2.60	94.44 (2.66, 2.90) 2.49	95.06 (2.66, 2.28) 1.79
A9	95.54 (1.80, 2.66) 2.42	95.06 (2.08, 2.86) 2.54	95.38 (1.92, 2.70) 2.44	94.84 (2.62, 2.54) 1.80
A10	95.18 (1.86, 2.96) 2.25	94.62 (2.18, 3.20) 2.37	94.90 (2.00, 3.10) 2.28	95.16 (2.34, 2.50) 1.70
A11	95.14 (2.50, 2.36) 2.43	94.56 (2.84, 2.60) 2.54	94.86 (2.68, 2.46) 2.44	94.88 (2.56, 2.56) 1.76
A12	95.48 (2.24, 2.28) 2.70	94.32 (2.84, 2.84) 2.96	94.80 (2.54, 2.66) 2.74	94.84 (2.52, 2.64) 1.82

6.3. Empirical Study for Sample Size Determination

In this section, we investigate the performance of approximate sample size estimation methods that can control the width of a confidence interval within a pre-specified width at a given confidence level. For each setting of ϕ, $\pi_{10}^{(2)}$, $\pi_{00}^{(1)}$, $\pi_{11}^{(1)}$, $\pi_{00}^{(2)}$, $\pi_{11}^{(2)}$ and r values, the approximate sample sizes N_{w1}, N_{w2}, N_l and N_{sc} are obtained via the algorithm proposed in Section 5. $K = 5000$ random samples are generated in the numerical algorithm. In order to investigate the accuracy of the estimated sample sizes, ECP and ECW of various CIs are calculated based on the estimated sample sizes. For given combinations of the parameters and a confidence interval CI_i, sample size N_i, where $i = w1, w2, l, sc$ is determined. Based on the sample size, $M = 5000$ random samples for $\mathbf{m}$ are generated and confidence intervals are computed. The empirical coverage probabilities and empirical coverage widths based on the $M = 5000$ samples are obtained.

We consider different settings in our simulation study to study the effects of various factors. (i) To investigate the impact of ϕ, we consider various values of ϕ ranging from 0.5 to 1.5 with step size 0.1, with $\pi_{00}^{(1)} = 0.5$, $\pi_{11}^{(1)} = 0.25$, $\pi_{00}^{(2)} = 0.4$, $\pi_{11}^{(2)} = 0.35$, half-width $\omega = 0.3$, and (a) $r = 0.9$, $\pi_{10}^{(2)} = 0.1$; (b) $r = 0.9$, $\pi_{10}^{(2)} = 0.05$; (c) $r = 1$, $\pi_{10}^{(2)} = 0.1$; (d) $r = 1$, $\pi_{10}^{(2)} = 0.05$. The simulation results are available in Figure 4. (ii) To investigate the effect of r, we considered various r values ranging from 0.1 to 1 with step sizes 0.1, with $\pi_{00}^{(1)} = 0.5$, $\pi_{11}^{(1)} = 0.2$, $\pi_{00}^{(2)} = 0.4$, $\pi_{11}^{(2)} = 0.3$, half-width $\omega = 0.15$, and (a) $\phi = 0.2$, $\pi_{10}^{(2)} = 0.15$; (b) $\phi = 0.3$, $\pi_{10}^{(2)} = 0.15$; (c) $\phi = 0.2$, $\pi_{10}^{(2)} = 0.1$; (d) $\phi = 0.3$, $\pi_{10}^{(2)} = 0.1$. The simulation results are available in Figure 5.

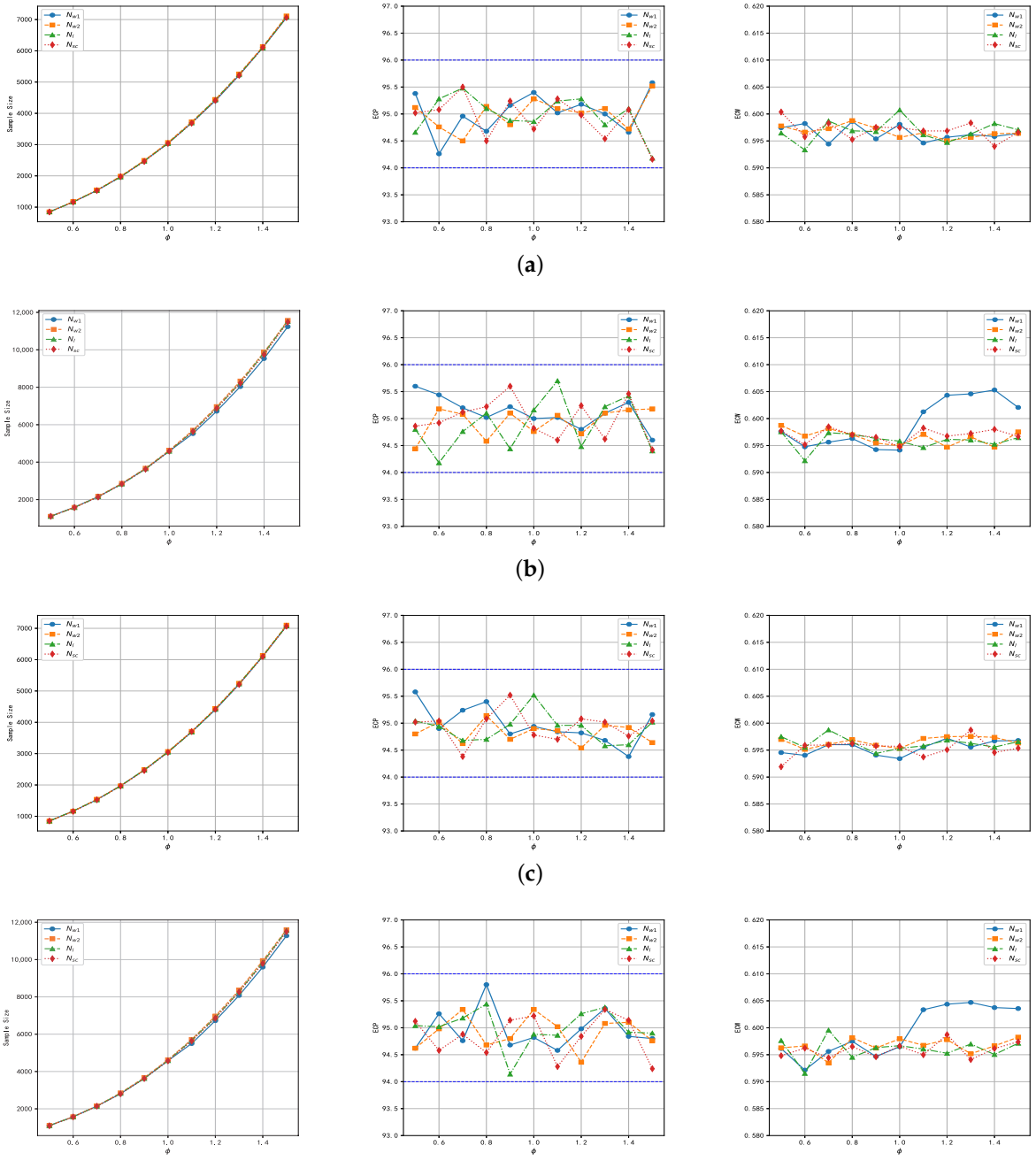
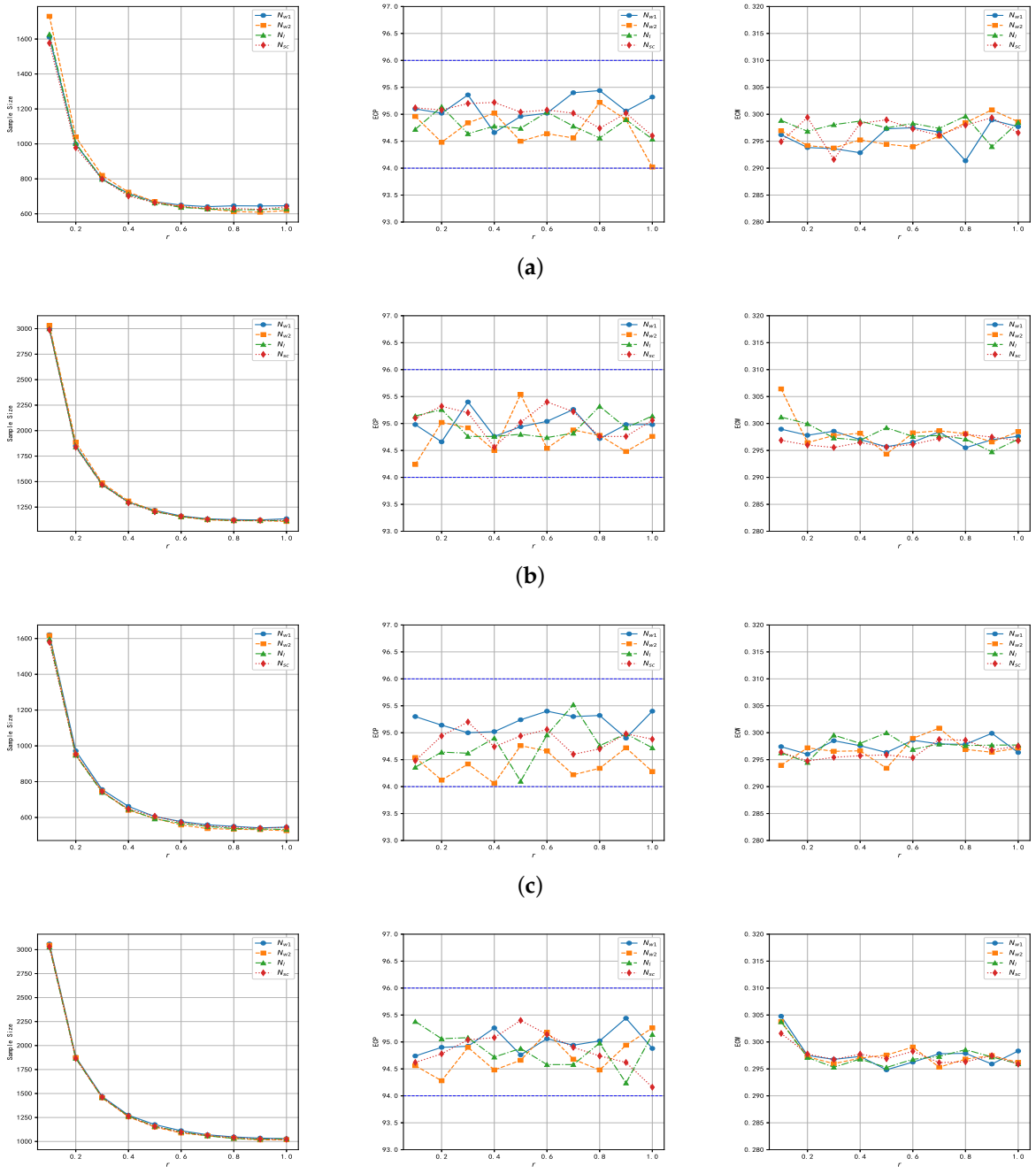


Figure 4. Plot of sample sizes, ECPs(%) and ECWs of CIs against ϕ with $\pi_{00}^{(1)} = 0.5$, $\pi_{11}^{(1)} = 0.25$, $\pi_{00}^{(2)} = 0.4$, $\pi_{11}^{(2)} = 0.35$, half-width $\omega = 0.3$, and (**a**) $r = 0.9$, $\pi_{10}^{(2)} = 0.1$; (**b**) $r = 0.9$, $\pi_{10}^{(2)} = 0.05$; (**c**) $r = 1$, $\pi_{10}^{(2)} = 0.1$; (**d**) $r = 1$, $\pi_{10}^{(2)} = 0.05$.

The simulation results indicate that (i) the required sample size increases with the increase in ϕ and decreases with the increase in r; (ii) based on the estimated sample sizes, the ECPs of all CIs are very close to the pre-specified confidence level, and the half-interval widths are also well controlled.



(**a**)

(**b**)

(**c**)

(**d**)

Figure 5. Plot of sample sizes, ECPs(%) and ECWs of CIs against r with $\pi_{00}^{(1)} = 0.5$, $\pi_{11}^{(1)} = 0.2$, $\pi_{00}^{(2)} = 0.4$, $\pi_{11}^{(2)} = 0.3$, half-width $\omega = 0.15$, and (**a**) $\phi = 0.2$, $\pi_{10}^{(2)} = 0.15$; (**b**) $\phi = 0.3$, $\pi_{10}^{(2)} = 0.15$; (**c**) $\phi = 0.2$, $\pi_{10}^{(2)} = 0.1$; (**d**) $\phi = 0.3$, $\pi_{10}^{(2)} = 0.1$.

7. Real Example

7.1. Example of Two New Devices Delivering Salbutamol

To demonstrate the practicality and effectiveness of the proposed methods, we first consider an example of the AB/BA crossover test conducted by 3M Riker to compare the

applicability of two new inhalation devices (A and B) in patients using standard inhalation devices to deliver salbutamol (Ezzet and Whitehead [13]). The response of the patient is either 'yes' or 'no', and neglecting the missing results of very few patients; the frequency of known reactions in patients is summarized in Table 7.

Table 7. Frequency of patient responses in a cross-over study of salbutamol inhalation devices A and B.

AB Sequence				
		Period 2		
		1	0	Total
Period 1	1	26	41	67
	0	15	57	72
	Total	41	98	139
BA Sequence				
		Period 2		
		1	0	Total
Period 1	1	38	16	54
	0	32	54	86
	Total	70	70	140

'1' represents 'Yes' response; '0' represents 'No' response.

Assume that there is no significant difference between the preferences for device B and A, i.e., we are interested in the hypothesis testing: $H_0 : \phi = 1$ versus $H_1 : \phi \neq 1$. Based on the observed data, the MLE of the interest parameter ϕ is $\hat{\phi} = 0.1829$, and MLEs of $\pi_{01}^{(1)}$, $\pi_{10}^{(1)}$, $\pi_{01}^{(2)}$ and $\pi_{10}^{(2)}$ are given by $\hat{\pi}_{01}^{(1)} = 0.1079$, $\hat{\pi}_{10}^{(1)} = 0.2950$, $\hat{\pi}_{01}^{(2)} = 0.2286$ and $\hat{\pi}_{10}^{(2)} = 0.1143$ according to Euqation (3). The CMLEs of $\pi_{01}^{(1)}$, $\pi_{10}^{(1)}$, $\pi_{01}^{(2)}$ and $\pi_{10}^{(2)}$ are given by $\tilde{\pi}_{01}^{(1)} = 0.1821$, $\tilde{\pi}_{10}^{(1)} = 0.2208$, $\tilde{\pi}_{01}^{(2)} = 0.1549$ and $\tilde{\pi}_{10}^{(2)} = 0.1879$ according to Equation (7). The p-values of the asymptotic test procedures are based on T_{w1}. T_{w2}, T_l and T_{sc} are less than 0.001. The corresponding 95% confidence intervals for CI_{w1}, CI_{w2}, CI_l and CI_{sc} are [0.0788, 0.4248], [0.0710, 0.4041], [0.0767, 0.4163] and [0.0792, 0.4222], respectively. Therefore, we would reject the null hypothesis, i.e., there is a significant difference in patient preference rates between devices A and B, this conclusion is consistent with Li et al. [18]. Let $\phi = 0.2$, $\pi_{00}^{(1)} = 0.4100$, $\pi_{11}^{(1)} = 0.1871$, $\pi_{00}^{(2)} = 0.3857$, $\pi_{11}^{(2)} = 0.2714$, $\pi_{10}^{(2)} = 0.1143$ and $r = 1$, we consider the sample size determination, and the desired sample sizes based on CI_{w1}, CI_{w2}, CI_l and CI_{sc} for controlling the interval width within $2\omega = 0.3$ are $N_{w1} = 441$, $N_{w2} = 423$, $N_l = 429$, $N_{sc} = 438$, respectively. The corresponding empirical coverage probabilities are 95.36%, 94.80%, 95.08%, and 95.10%, respectively.

7.2. Example of Relieving Heartburn

Koch et al. [33] investigated an AB/BA crossover trial for comparing the efficacy of active drugs and placebo in relieving heartburn after two symptomatic meals (two meals corresponding to two cycles) from two centers. At each center, 30 patients participated in this trial, of which 15 were randomly assigned to the (A:P) sequence group (treated with active medication for heartburn in the first meal and placebo in the second meal), and the other 15 were assigned to the (P:A) sequence group (treated with placebo in the first meal and active medication in the second meal). The allocation of sequence groups adopts a double-blind method. The interval between each patient's two periods (two meals) in the crossover design is several days, which is considered long enough to rule out any residual

effects of treatment. The frequency data on whether or not the patient experienced relief within 15 min of the first dose of treatment in Center 2 are given in Table 8.

Table 8. Data on the first and second phase first dose response in a cross-design clinical trial for the relief of heartburn in center 2.

Sequence	Period I	Period II		Total
		R	NR	
A:P	R	0	7	7
	NR	1	7	8
	Total	1	14	15
P:A	R	0	3	3
	NR	10	2	12
	Total	10	5	15

R = Relief, NR = No Relief. A:P = active treatment for Period 1 and placebo for Period 2. P:A = placebo treatment for Period 1 and active for Period 2.

Suppose that there is no significant difference in the efficacy of active drugs and placebo, i.e., we consider the hypothesis testing $H_0 : \phi = 1$. With the data from Table 8, we have $\hat{\phi} = 0.0430$, the MLEs of $\pi_{01}^{(1)}$, $\pi_{10}^{(1)}$, $\pi_{01}^{(2)}$ and $\pi_{10}^{(2)}$ are given by $\hat{\pi}_{01}^{(1)} = 0.0667$, $\hat{\pi}_{10}^{(1)} = 0.4667$, $\hat{\pi}_{01}^{(2)} = 0.6667$ and $\hat{\pi}_{10}^{(2)} = 0.2000$ according to Euqation (3), and the CMLEs of $\pi_{01}^{(1)}$, $\pi_{10}^{(1)}$, $\pi_{01}^{(2)}$ and $\pi_{10}^{(2)}$ are given by $\tilde{\pi}_{01}^{(1)} = 0.2794$, $\tilde{\pi}_{10}^{(1)} = 0.2540$, $\tilde{\pi}_{01}^{(2)} = 0.4540$ and $\tilde{\pi}_{10}^{(2)} = 0.4127$ according to Equation (7). Since the sample size is very small, we consider both the asymptotic and approximate unconditional test procedures for the hypothesis testing $H_0 : \phi = 1$; the asymptotic test p-values based on test statistics T_{w1}. T_{w2}, T_l and T_{sc} are 0.0127, 0.0015, 0.0047, 0.0063, and the corresponding approximate unconditional test p-values are 0.0048, 0.0077, 0.0040, 0.0036, respectively.

Obviously, the p-values of test procedures based on test statistics T_{w1}, T_{w2}, T_l and T_{sc} are less than 0.05, so these p-values strongly support a significant difference in heartburn relief between active treatment and placebo. Moreover, 95% CIs, i.e., CI_{w1}, CI_{w2}, CI_l and CI_{sc} are [0.0079, 0.5609], [0.0, 0.3747], [0.0054, 0.4597] and [0.0094, 0.5018], respectively. Let $\phi = 0.1$, $\pi_{00}^{(1)} = 0.4412$, $\pi_{11}^{(1)} = 0.0294$, $\pi_{00}^{(2)} = 0.1471$, $\pi_{11}^{(2)} = 0.0294$, $\pi_{10}^{(2)} = 0.2059$ and $r = 1$, with the half interval width $\omega = 0.05$, the desired sample sizes are $N_{w1} = 605$, $N_{w2} = 587$, $N_l = 596$, $N_{sc} = 602$, respectively. The corresponding empirical coverage probabilities are 95.12%, 94.60%, 94.74%, and 95.26%, respectively.

8. Conclusions and Discussion

The equivalence test of odds ratio (OR) in an AB/BA crossover study with binary outcomes is considered in this article, two test procedures, including the asymptotical and approximate unconditional test procedures based on four test statistics, i.e., two Wald type test statistics, likelihood ratio test statistic and score test statistic are proposed to test the equivalence hypothesis. Four confidence intervals for the OR and the corresponding sample sizes, which can control the width of a confidence interval with a pre-specified confidence level, are developed. Simulation results indicate that the asymptotical test procedures based on four test statistics perform well when the sample size is not small, and the approximate test procedures can produce close to the nominal level even if the sample size is very small (e.g., $n_1 = n_2 = 10$ or $n_1 = n_2 = 15$). Confidence intervals derived from the score test statistic, likelihood ratio statistic, and Wald-type test statistics generally perform satisfactorily in terms of coverage. Take into account that in small sample sizes, the Wald CI based on T_{w2} performs poorly under certain parameter settings. In general, sample size estimation methods can be recommended for practical applications in the sense

that the empirical coverage probabilities are close to the pre-specified confidence level under the estimated sample sizes.

Under the assumption that there is no carryover effect or stage effect in an AB/BA crossover design, we considered the equivalence test, confidence interval and sample size determination based on OR for the treatment effects of two treatments. Python 5.4.1 codes that implement the proposed methodologies and calculations are available from the second author by request. However, the carryover effect and/or period effect may exist in some trials with binary outcomes. Therefore, statistical inference for the treatment effects in the crossover trials with carryover effect and/or period effect will be considered in the future. In this article, we provide some estimation methods for sample size determination which can control the width of a confidence interval, but we do not take the cost into account for the sample size estimation, it also may be the interest topic in the future research.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/axioms14080582/s1.

Author Contributions: Formal analysis, X.-Q.Y.; Funding acquisition, S.-F.Q.; Methodology, S.-F.Q., X.-Q.Y. and W.-Y.P.; Project administration, W.-Y.P.; Supervision, S.-F.Q.; Writing—original draft, X.-Q.Y.; Writing—review and editing, S.-F.Q. and W.-Y.P. All authors have read and agreed to the published version of the manuscript.

Funding: The work of Dr. Qiu was sponsored by Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJZD-K202201101) ,the Natural Science Foundation of Chongqing, China (Grant No. CSTB2024NSCQ-LZX0136), and the National Natural Science Foundation of China (Grant No. 11871124).

Data Availability Statement: No new data were generated during this study. All real examples are publicly available and cited appropriately.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Derivation of the CMLE $\tilde{\pi}_{10}^{(2)}$ of $\pi_{10}^{(2)}$ Given ϕ

The CMLE $\tilde{\pi}_{10}^{(2)}$ of $\pi_{10}^{(2)}$ is the root of the following quadratic equation

$$A(\pi_{10}^{(2)})^2 + B\pi_{10}^{(2)} + C = 0. \tag{A1}$$

where $A = (\phi-1)(n_{01}^{(2)} + n_{10}^{(2)})$, $B = m_2(n_{10}^{(2)} - n_{01}^{(1)}) - m_2(n_{01}^{(2)} + n_{10}^{(1)} + 2n_{10}^{(2)})$, $C = \phi(n_{10}^{(1)} + n_{10}^{(2)})m_2^2$ with $m_1 = (n_{01}^{(1)} + n_{10}^{(1)})/n_1$, $m_2 = (n_{01}^{(2)} + n_{10}^{(2)})/n_2$. Therefore, we have

$$\begin{aligned}
B^2 - 4AC &= m_2^2[(n_{10}^{(2)} - n_{01}^{(1)}) - \phi(n_{01}^{(2)} + n_{10}^{(1)} + n_{10}^{(2)} + n_{10}^{(2)})]^2 - 4\phi(\phi-1)(n_{01}^{(2)} + n_{10}^{(2)})(n_{10}^{(1)} + n_{10}^{(2)})m_2^2 \\
&= m_2^2\{[(n_{10}^{(2)} - n_{01}^{(1)})^2 + \phi^2(n_{01}^{(2)} + n_{10}^{(1)} + n_{10}^{(2)} + n_{10}^{(2)})^2 - 2(n_{10}^{(2)} - n_{01}^{(1)})(n_{01}^{(2)} + n_{10}^{(1)} + n_{10}^{(2)} + n_{10}^{(2)})\phi] \\
&\quad - 4(n_{01}^{(2)} + n_{10}^{(2)})(n_{10}^{(1)} + n_{10}^{(2)})\phi^2 + 4(n_{01}^{(2)} + n_{10}^{(2)})(n_{10}^{(1)} + n_{10}^{(2)})\phi\} \\
&= m_2^2\{\phi^2(n_{01}^{(2)} - n_{10}^{(1)})^2 + 2\phi[(n_{10}^{(1)} + n_{10}^{(2)})(n_{01}^{(2)} + n_{01}^{(1)}) + (n_{01}^{(2)} \\
&\quad + n_{10}^{(2)})(n_{10}^{(1)} + n_{01}^{(1)})] + (n_{10}^{(2)} - n_{01}^{(1)})^2\}
\end{aligned}$$

Obviously, $B^2 - 4AC$ is aways not less than 0, which indicates that the quadratic equation about $\pi_{10}^{(2)}$, i.e., $A(\pi_{10}^{(2)})^2 + B\pi_{10}^{(2)} + C = 0$ must have roots. Moreover, it is easily seen that $4AC = \phi(\phi-1)(n_{01}^{(2)} + n_{10}^{(2)})(n_{10}^{(1)} + n_{10}^{(2)})m_2^2$, so $|B| > \sqrt{B^2 - 4AC}$ if $\phi > 1$, and $|B| < \sqrt{B^2 - 4AC}$ if $\phi < 1$.

Case 1: $\phi > 1$

In this case, $A > 0$, $B < 0$ and $-B \mp \sqrt{B^2 - 4AC} > 0$, the quadratic Equation (A1) has two positive roots:

$$(-B - \sqrt{B^2 - 4AC})/(2A) \text{ and } (-B + \sqrt{B^2 - 4AC})/(2A).$$

Due to $0 \leq \pi_{00}^{(2)}, \pi_{01}^{(2)}, \pi_{10}^{(2)}, \pi_{11}^{(2)} \leq 1$ and $\pi_{00}^{(2)} + \pi_{01}^{(2)} + \pi_{10}^{(2)} + \pi_{11}^{(2)} = 1$, the CMLE $\tilde{\pi}_{10}^{(2)}$ of $\pi_{10}^{(2)}$ should be the smaller positive root, i.e., $\tilde{\pi}_{10}^{(2)} = (-B - \sqrt{B^2 - 4AC})/(2A)$.

Case 2: $\phi < 1$

In this case, since $A < 0$, then $(-B - \sqrt{B^2 - 4AC}) < 0$ and $(-B + \sqrt{B^2 - 4AC}) > 0$. The quadratic Equation (A1) has a positive root $(-B - \sqrt{B^2 - 4AC})/(2A)$ and a negative root $(-B + \sqrt{B^2 - 4AC})/(2A))$). Due to $\pi_{10}^{(2)} > 0$, then the CMLE $\tilde{\pi}_{10}^{(2)}$ of $\pi_{10}^{(2)}$ should be $\tilde{\pi}_{10}^{(2)} = (-B + \sqrt{B^2 - 4AC})/(2A)$.

To sum up, when $A \neq 0$, the CMLE $\tilde{\pi}_{10}^{(2)}$ of $\pi_{10}^{(2)}$ is given by $\tilde{\pi}_{10}^{(2)} = (-B - \sqrt{B^2 - 4AC})/(2A)$. When $A = 0$, Equation (A1) becomes $B\pi_{10}^{(2)} + C = 0$, it has a single root $\tilde{\pi}_{10}^{(2)} = -C/B$.

Appendix B. Derivation of the Asymptotical Distribution of the Test Statistic $T_{w1}(T_{w2})$

Let $\boldsymbol{\pi}^{(1)} = (\pi_{01}^{(1)}, \pi_{10}^{(1)})'$, $\boldsymbol{\pi}^{(2)} = (\pi_{01}^{(2)}, \pi_{10}^{(2)})'$, the corresponding MLEs of $\boldsymbol{\pi}^{(1)}$ and $\boldsymbol{\pi}^{(2)}$ are $\hat{\boldsymbol{\pi}}^{(1)}$ and $\hat{\boldsymbol{\pi}}^{(2)}$, respectively. Under the regularity conditions $\pi_{00}^{(1)}, \pi_{10}^{(1)}, \pi_{01}^{(1)}, \pi_{11}^{(1)}, \pi_{00}^{(2)}, \pi_{10}^{(2)}, \pi_{01}^{(2)}, \pi_{11}^{(2)} > 0$, the MLEs $\hat{\boldsymbol{\pi}}^{(1)}$ and $\hat{\boldsymbol{\pi}}^{(2)}$ of the multinomial distribution parameter vectors $\boldsymbol{\pi}^{(1)}$ and $\boldsymbol{\pi}^{(2)}$ are asymptotically distributed as the normal distributions when $\min\{n_1, n_2\} \to \infty$, that is

$$\sqrt{n_i}(\hat{\boldsymbol{\pi}}^{(i)} - \boldsymbol{\pi}^{(i)}) \xrightarrow{d} N_2(\mathbf{0}, \Sigma_\pi^{(i)}), \Sigma_\pi^{(i)} = diag(\boldsymbol{\pi}^{(i)}) - \boldsymbol{\pi}^{(i)}\boldsymbol{\pi}^{(i)'}, \ i = 1, 2.$$

Let $\boldsymbol{\beta}^{(1)} = (\log \pi_{01}^{(1)}, \log \pi_{10}^{(1)})'$, $\boldsymbol{\beta}^{(2)} = (\log \pi_{01}^{(2)}, \log \pi_{10}^{(2)})'$. According to the delta method, we have

$$\sqrt{n_i}(\hat{\boldsymbol{\beta}}^{(i)} - \boldsymbol{\beta}^{(i)}) \xrightarrow{d} N_2(\mathbf{0}, \Sigma_\beta^{(i)}), \Sigma_\beta^{(i)} = \begin{pmatrix} \frac{1-\pi_{01}^{(i)}}{\pi_{01}^{(i)}} & -1 \\ -1 & \frac{1-\pi_{10}^{(i)}}{\pi_{10}^{(i)}} \end{pmatrix}, \ i = 1, 2.$$

Therefore, we have

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \begin{pmatrix} \hat{\boldsymbol{\beta}}^{(1)} \\ \hat{\boldsymbol{\beta}}^{(2)} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\beta}^{(1)} \\ \boldsymbol{\beta}^{(2)} \end{pmatrix} \xrightarrow{d} N_4(\mathbf{0}_{4\times 1}, \Sigma), \ \ \Sigma = \begin{pmatrix} \frac{1}{n_1}\Sigma_\beta^{(1)} & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \frac{1}{n_2}\Sigma_\beta^{(2)} \end{pmatrix}.$$

Thus,

$$c'\hat{\boldsymbol{\beta}}c - c'\boldsymbol{\beta}c \xrightarrow{d} N(0, c'\Sigma c)$$

Since $\hat{\Sigma}_\pi^{(i)} = diag(\hat{\boldsymbol{\pi}}^{(i)}) - \hat{\boldsymbol{\pi}}^{(i)}(\hat{\boldsymbol{\pi}}^{(i)})'$ is the consistent estimator of $\Sigma_\pi^{(i)}$ for $i = 1, 2$, and $g(x) = \log(x)$ is a continuously differentiable function, then the estimator of Σ, which is obtained by replacing the parameters (i.e., $\pi_{01}^{(1)}$, $\pi_{10}^{(1)}$, $\pi_{01}^{(2)}$, $\pi_{10}^{(2)}$) with their MLEs or their constrained MLEs under $H_0 : \phi = 1$, is the consistent estimator of Σ. Therefore, under the null hypothesis $H_0 : c'\boldsymbol{\beta} = 0$, we have

$$Z_1 = (c'\hat{\boldsymbol{\beta}}c)/\sqrt{c'\hat{\Sigma}c} \xrightarrow{d} N(0, 1), \ \ Z_2 = (c'\hat{\boldsymbol{\beta}}c)/\sqrt{c'\hat{\Sigma}c} \xrightarrow{d} N(0, 1).$$

When $\min\{n_1, n_2\} \to \infty$. Then $T_{w1} = Z_1^2$ and $T_{w2} = Z_2^2$ are asymptotically distributed as the χ^2 distribution with one degree of freedom.

Appendix C. Derivation of Score Test Statistic

Differentiating $l_2(\phi, \pi_{00}^{(1)}, \pi_{11}^{(1)}, \pi_{00}^{(2)}, \pi_{11}^{(2)}, \pi_{10}^{(2)})$ respect to ϕ, $\pi_{00}^{(1)}$, $\pi_{11}^{(1)}$, $\pi_{00}^{(2)}$, $\pi_{11}^{(2)}$, and $\pi_{10}^{(2)}$ yields

$$\frac{\partial l_2}{\partial \phi} = \frac{n_{01}^{(1)}}{\phi} - \frac{(M_2 - \pi_{10}^{(2)})(n_{01}^{(1)} + n_{10}^{(1)})}{\pi_{10}^{(2)} + \phi(M_2} - \pi_{10}^{(2)}),$$

$$\frac{\partial l_2}{\partial \pi_{10}^{(2)}} = -\frac{n_{01}^{(1)} + n_{01}^{(2)}}{M_2 - \pi_{10}^{(2)}} - \frac{(1-\phi)(n_{01}^{(1)} + n_{10}^{(1)})}{\pi_{10}^{(2)} + \phi(M_2 - \pi_{10}^{(2)})} + \frac{n_{10}^{(1)} + n_{10}^{(2)}}{\pi_{10}^{(2)}},$$

$$\frac{\partial l_2}{\partial \pi_{00}^{(1)}} = \frac{n_{00}^{(1)}}{\pi_{00}^{(1)}} - \frac{n_{01}^{(1)} + n_{10}^{(1)}}{M_1},$$

$$\frac{\partial l_2}{\partial \pi_{11}^{(1)}} = \frac{n_{11}^{(1)}}{\pi_{11}^{(1)}} - \frac{n_{01}^{(1)} + n_{10}^{(1)}}{M_1},$$

$$\frac{\partial l_2}{\partial \pi_{00}^{(2)}} = \frac{n_{00}^{(2)}}{\pi_{00}^{(2)}} - \frac{n_{01}^{(1)} + n_{01}^{(2)}}{M_2 - \pi_{10}^{(2)}} + \frac{\phi(n_{10}^{(1)} + n_{01}^{(1)})}{\pi_{10}^{(2)} + \phi(M_2 - \pi_{10}^{(2)})}, \text{and } \frac{\partial l_2}{\partial \pi_{11}^{(2)}} = \frac{n_{11}^{(2)}}{\pi_{11}^{(2)}} - \frac{n_{01}^{(1)} + n_{01}^{(2)}}{M_2 - \pi_{10}^{(2)}} + \frac{\phi(n_{10}^{(1)} + n_{01}^{(1)})}{\pi_{10}^{(2)} + \phi(M_2 - \pi_{10}^{(2)})}.$$

Differentiating $\frac{\partial l_2}{\partial \phi}$, $\frac{\partial l_2}{\partial \pi_{10}^{(2)}}$, $\frac{\partial l_2}{\partial \pi_{00}^{(1)}}$, $\frac{\partial l_2}{\partial \pi_{11}^{(1)}}$, $\frac{\partial l_2}{\partial \pi_{00}^{(2)}}$, $\frac{\partial l_2}{\partial \pi_{11}^{(2)}}$ with respect to ϕ, $\pi_{00}^{(1)}$, $\pi_{11}^{(1)}$, $\pi_{00}^{(2)}$, $\pi_{11}^{(2)}$, and $\pi_{10}^{(2)}$, respectively, and yields

$$\frac{\partial^2 l_2}{\partial \phi \partial \phi} = -\frac{n_{01}^{(1)}}{\phi^2} + \frac{(M_2 - \pi_{10}^{(2)})^2 (n_{01}^{(1)} + n_{10}^{(1)})}{(\pi_{10}^{(2)} + \phi(M_2 - \pi_{10}^{(2)}))^2},$$

$$\frac{\partial^2 l_2}{\partial \phi \partial \pi_{10}^{(2)}} = \frac{\partial^2 l_2}{\partial \pi_{10}^{(2)} \partial \phi} = \frac{M_2 (n_{01}^{(1)} + n_{10}^{(1)})}{(\pi_{10}^{(2)} + \phi(M_2 - \pi_{10}^{(2)}))^2},$$

$$\frac{\partial^2 l_2}{\partial \phi \partial \pi_{00}^{(2)}} = \frac{\partial^2 l_2}{\partial \pi_{00}^{(2)} \partial \phi} = \frac{\partial^2 l_2}{\partial \phi \partial \pi_{11}^{(2)}} = \frac{\partial^2 l_2}{\partial \pi_{11}^{(2)} \partial \phi} = \frac{(n_{01}^{(1)} + n_{10}^{(1)}) \pi_{10}^{(2)}}{(\pi_{10}^{(2)} + \phi(M_2 - \pi_{10}^{(2)}))^2},$$

$$\frac{\partial^2 l_2}{\partial \pi_{10}^{(2)} \partial \pi_{10}^{(2)}} = -\frac{(n_{01}^{(1)} + n_{01}^{(2)})}{(M_2 - \pi_{10}^{(2)})^2} + \frac{(1-\phi)^2 (n_{01}^{(1)} + n_{10}^{(2)})}{(\pi_{10}^{(2)} + \phi(M_2 - \pi_{10}^{(2)}))^2} - \frac{n_{10}^{(1)} + n_{10}^{(2)}}{(\pi_{10}^{(2)})^2},$$

$$\frac{\partial^2 l_2}{\partial \pi_{10}^{(2)} \partial \pi_{00}^{(2)}} = \frac{\partial^2 l_2}{\partial \pi_{00}^{(2)} \partial \pi_{10}^{(2)}} = \frac{\partial^2 l_2}{\partial \pi_{10}^{(2)} \partial \pi_{11}^{(2)}} = \frac{\partial^2 l_2}{\partial \pi_{11}^{(2)} \partial \pi_{10}^{(2)}} = -\frac{n_{01}^{(2)} + n_{01}^{(1)}}{(M_2 - \pi_{10}^{(2)})^2} - \frac{(1-\phi)\phi(n_{01}^{(1)} + n_{10}^{(1)})}{(\pi_{10}^{(2)} + \phi(M_2 - \pi_{10}^{(2)}))^2},$$

$$\frac{\partial^2 l_2}{\partial \phi \partial \pi_{11}^{(1)}} = \frac{\partial^2 l_2}{\partial \phi \partial \pi_{00}^{(1)}} = \frac{\partial^2 l_2}{\partial \pi_{10}^{(2)} \partial \pi_{11}^{(1)}} = \frac{\partial^2 l_2}{\partial \pi_{10}^{(2)} \partial \pi_{00}^{(1)}} = 0,$$

$$\frac{\partial^2 l_2}{\partial \pi_{00}^{(1)} \partial \pi_{00}^{(1)}} = -\frac{n_{00}^{(1)}}{(\pi_{00}^{(1)})^2} - \frac{n_{01}^{(1)} + n_{10}^{(1)}}{{M_1}^2},$$

$$\frac{\partial^2 l_2}{\partial \pi_{00}^{(1)} \partial \pi_{11}^{(1)}} = \frac{\partial^2 l_2}{\partial \pi_{11}^{(1)} \partial \pi_{00}^{(1)}} = -\frac{n_{01}^{(1)} + n_{10}^{(1)}}{{M_1}^2},$$

$$\frac{\partial^2 l_2}{\partial \pi_{00}^{(1)} \partial \phi} = \frac{\partial^2 l_2}{\partial \pi_{00}^{(1)} \partial \pi_{10}^{(2)}} = \frac{\partial^2 l_2}{\partial \pi_{00}^{(1)} \partial \pi_{00}^{(2)}} = \frac{\partial^2 l_2}{\partial \pi_{00}^{(1)} \partial \pi_{11}^{(2)}} = 0,$$

$$\frac{\partial^2 l_2}{\partial \pi_{11}^{(1)} \partial \pi_{11}^{(1)}} = -\frac{n_{11}^{(1)}}{(\pi_{11}^{(1)})^2} - \frac{n_{01}^{(1)} + n_{10}^{(1)}}{{M_1}^2},$$

$$\frac{\partial^2 l_2}{\partial \pi_{11}^{(1)} \partial \phi} = \frac{\partial^2 l_2}{\partial \pi_{11}^{(1)} \partial \pi_{10}^{(2)}} = \frac{\partial^2 l_2}{\partial \pi_{11}^{(1)} \partial \pi_{11}^{(2)}} = \frac{\partial^2 l_2}{\partial \pi_{11}^{(1)} \partial \pi_{00}^{(2)}} = 0,$$

$$\frac{\partial^2 l_2}{\partial \pi_{00}^{(2)} \partial \pi_{00}^{(2)}} = -\frac{n_{00}^{(2)}}{(\pi_{00}^{(2)})^2} - \frac{n_{01}^{(1)} + n_{01}^{(2)}}{(M_2 - \pi_{10}^{(2)})^2} + \frac{\phi^2 (n_{10}^{(1)} + n_{01}^{(1)})}{(\pi_{10}^{(2)} + \phi(M_2 - \pi_{10}^{(2)}))^2},$$

$$\frac{\partial^2 l_2}{\partial \pi_{00}^{(2)} \partial \pi_{11}^{(2)}} = \frac{\partial^2 l_2}{\partial \pi_{11}^{(2)} \partial \pi_{00}^{(2)}} = -\frac{n_{01}^{(1)} + n_{01}^{(2)}}{(M_2 - \pi_{10}^{(2)})} + \frac{\phi^2 (n_{01}^{(1)} + n_{10}^{(1)})}{(\pi_{10}^{(2)} + \phi(M_2 - \pi_{10}^{(2)}))^2},$$

$$\frac{\partial^2 l_2}{\partial \pi_{11}^{(2)} \partial \pi_{11}^{(2)}} = -\frac{n_{11}^{(2)}}{(\pi_{11}^{(2)})^2} - \frac{n_{01}^{(1)} + n_{01}^{(2)}}{(M_2 - \pi_{10}^{(2)})^2} + \frac{\phi^2 (n_{10}^{(1)} + n_{01}^{(1)})}{(\pi_{10}^{(2)} + \phi(M_2 - \pi_{10}^{(2)}))^2},$$

$$\frac{\partial^2 l_2}{\partial \pi_{00}^{(2)} \partial \pi_{00}^{(1)}} = \frac{\partial^2 l_2}{\partial \pi_{00}^{(2)} \partial \pi_{11}^{(1)}} = \frac{\partial^2 l_2}{\partial \pi_{11}^{(2)} \partial \pi_{00}^{(1)}} = \frac{\partial^2 l_2}{\partial \pi_{11}^{(2)} \partial \pi_{11}^{(1)}} = 0.$$

Let $Q_1 = \pi_{10}^{(2)} + \phi(M_2 - \pi_{10}^{(2)}) = \pi_{10}^{(2)} + \phi\pi_{01}^{(2)}$; $Q_2 = n_1(\pi_{01}^{(1)} + \pi_{10}^{(1)})$; $Q_3 = (n_1\pi_{01}^{(1)} + n_2\pi_{01}^{(2)})/(\pi_{01}^{(2)})^2$. Next, we have

$$I_{\phi\phi} = E(-\frac{\partial^2 l_2}{\partial \phi \partial \phi}) = \frac{n_1 \pi_{01}^{(1)}}{\phi^2} - \frac{Q_2 (\pi_{01}^{(2)})^2}{Q_1^2},$$

$$
\begin{aligned}
&I_{\phi\pi_{10}^{(2)}} = I_{\pi_{10}^{(2)}\phi} = E(-\frac{\partial^2 l_2}{\partial\phi\partial\pi_{10}^{(2)}}) = -\frac{Q_2M_2}{Q_1^2},\\
&I_{\phi\pi_{00}^{(2)}} = I_{\pi_{00}^{(2)}\phi} = E(-\frac{\partial^2 l_2}{\partial\pi_{00}^{(2)}\partial\phi}) = -\frac{Q_2\pi_{10}^{(2)}}{Q_1^2} = I_{\phi\pi_{11}^{(2)}} = I_{\pi_{11}^{(2)}\phi},\\
&I_{\pi_{10}^{(2)}\pi_{10}^{(2)}} = E(-\frac{\partial^2 l_2}{\partial\pi_{10}^{(2)}\partial\pi_{10}^{(2)}}) = \frac{n_1\pi_{10}^{(1)}+n_2\pi_{10}^{(2)}}{(\pi_{10}^{(2)})^2} - \frac{(1-\phi)^2Q_2}{Q_1^2} + Q_3,\\
&I_{\pi_{10}^{(2)}\pi_{00}^{(2)}} = I_{\pi_{00}^{(2)}\pi_{10}^{(2)}} = E(-\frac{\partial^2 l_2}{\partial\pi_{10}^{(2)}\partial\pi_{00}^{(2)}}) = Q_3 + \frac{\phi(1-\phi)Q_2}{Q_1^2} = I_{\pi_{10}^{(2)}\pi_{11}^{(2)}} = I_{\pi_{11}^{(2)}\pi_{10}^{(2)}},\\
&I_{\pi_{00}^{(1)}\pi_{00}^{(1)}} = E(-\frac{\partial^2 l_2}{\partial\pi_{00}^{(1)}\partial\pi_{00}^{(1)}}) = \frac{n_1}{\pi_{00}^{(1)}} + \frac{Q_2}{M_1^2},\\
&I_{\pi_{00}^{(1)}\pi_{11}^{(1)}} = I_{\pi_{11}^{(1)}\pi_{00}^{(1)}} = E(-\frac{\partial^2 l_2}{\partial\pi_{00}^{(1)}\partial\pi_{11}^{(1)}}) = \frac{Q_2}{M_1^2},\\
&I_{\pi_{11}^{(1)}\pi_{11}^{(1)}} = E(-\frac{\partial^2 l_2}{\partial\pi_{11}^{(1)}\partial\pi_{11}^{(1)}}) = \frac{n_1}{\pi_{11}^{(1)}} + \frac{Q_2}{M_1^2},\\
&I_{\pi_{00}^{(2)}\pi_{00}^{(2)}} = E(-\frac{\partial^2 l_2}{\partial\pi_{00}^{(2)}\partial\pi_{00}^{(2)}}) = \frac{n_2}{\pi_{00}^{(2)}} + Q_3 - \frac{Q_2\phi^2}{Q_1^2},\\
&I_{\pi_{00}^{(2)}\pi_{11}^{(2)}} = I_{\pi_{11}^{(2)}\pi_{00}^{(2)}} = E(-\frac{\partial^2 l_2}{\partial\pi_{00}^{(2)}\partial\pi_{11}^{(2)}}) = Q_3 - \frac{Q_2\phi^2}{Q_1^2},\\
&I_{\pi_{11}^{(2)}\pi_{11}^{(2)}} = E(-\frac{\partial^2 l_2}{\partial\pi_{11}^{(2)}\partial\pi_{11}^{(2)}}) = \frac{n_2}{\pi_{11}^{(2)}} + Q_3 - \frac{Q_2\phi^2}{Q_1^2},\\
&I_{\phi\pi_{00}^{(1)}} = I_{\phi\pi_{11}^{(1)}} = I_{\pi_{00}^{(1)}\phi} = I_{\pi_{11}^{(1)}\phi} = I_{\pi_{10}^{(2)}\pi_{00}^{(1)}} = I_{\pi_{10}^{(2)}\pi_{11}^{(1)}} = I_{\pi_{00}^{(1)}\pi_{10}^{(2)}} = I_{\pi_{00}^{(1)}\pi_{00}^{(2)}} = I_{\pi_{00}^{(1)}\pi_{11}^{(2)}},\\
&I_{\pi_{11}^{(1)}\pi_{10}^{(2)}} = I_{\pi_{11}^{(1)}\pi_{00}^{(2)}} = I_{\pi_{11}^{(2)}\pi_{11}^{(2)}} = I_{\pi_{00}^{(2)}\pi_{00}^{(1)}} = I_{\pi_{00}^{(2)}\pi_{11}^{(1)}} = I_{\pi_{11}^{(2)}\pi_{00}^{(1)}} = I_{\pi_{11}^{(2)}\pi_{11}^{(1)}} = 0.
\end{aligned}
$$

Thus, the Fisher information matrix is given by

$$
I(\phi,\pi_{10}^{(2)},\pi_{00}^{(1)},\pi_{11}^{(1)},\pi_{00}^{(2)},\pi_{11}^{(2)}) = \begin{pmatrix}
I_{\phi\phi} & I_{\phi\pi_{10}^{(2)}} & I_{\phi\pi_{00}^{(1)}} & I_{\phi\pi_{11}^{(1)}} & I_{\phi\pi_{00}^{(2)}} & I_{\phi\pi_{11}^{(2)}}\\
I_{\pi_{10}^{(2)}\phi} & I_{\pi_{10}^{(2)}\pi_{10}^{(2)}} & I_{\pi_{10}^{(2)}\pi_{00}^{(1)}} & I_{\pi_{10}^{(2)}\pi_{11}^{(1)}} & I_{\pi_{10}^{(2)}\pi_{00}^{(2)}} & I_{\pi_{10}^{(2)}\pi_{11}^{(2)}}\\
I_{\pi_{00}^{(1)}\phi} & I_{\pi_{00}^{(1)}\pi_{10}^{(2)}} & I_{\pi_{00}^{(1)}\pi_{00}^{(1)}} & I_{\pi_{00}^{(1)}\pi_{11}^{(1)}} & I_{\pi_{00}^{(1)}\pi_{00}^{(2)}} & I_{\pi_{00}^{(1)}\pi_{11}^{(2)}}\\
I_{\pi_{11}^{(1)}\phi} & I_{\pi_{11}^{(1)}\pi_{10}^{(2)}} & I_{\pi_{11}^{(1)}\pi_{00}^{(1)}} & I_{\pi_{11}^{(1)}\pi_{11}^{(1)}} & I_{\pi_{11}^{(1)}\pi_{00}^{(2)}} & I_{\pi_{11}^{(1)}\pi_{00}^{(2)}}\\
I_{\pi_{00}^{(2)}\phi} & I_{\pi_{00}^{(2)}\pi_{10}^{(2)}} & I_{\pi_{00}^{(2)}\pi_{00}^{(1)}} & I_{\pi_{00}^{(2)}\pi_{11}^{(1)}} & I_{\pi_{00}^{(2)}\pi_{00}^{(2)}} & I_{\pi_{00}^{(2)}\pi_{11}^{(2)}}\\
I_{\pi_{11}^{(2)}\phi} & I_{\pi_{11}^{(2)}\pi_{10}^{(2)}} & I_{\pi_{11}^{(2)}\pi_{00}^{(1)}} & I_{\pi_{11}^{(2)}\pi_{11}^{(1)}} & I_{\pi_{11}^{(2)}\pi_{00}^{(2)}} & I_{\pi_{11}^{(2)}\pi_{11}^{(2)}}
\end{pmatrix}
$$

Under $H_0 : \phi = 1$, we have the Fisher information matrix

$$
I_0(\phi,\pi_{10}^{(2)},\pi_{00}^{(1)},\pi_{11}^{(1)},\pi_{00}^{(2)},\pi_{11}^{(2)}) = \begin{pmatrix}
I^0_{\phi\phi} & I^0_{\phi\pi_{10}^{(2)}} & 0 & 0 & I^0_{\phi\pi_{00}^{(2)}} & I^0_{\phi\pi_{11}^{(2)}}\\
I^0_{\pi_{10}^{(2)}\phi} & I^0_{\pi_{10}^{(2)}\pi_{10}^{(2)}} & 0 & 0 & I^0_{\pi_{10}^{(2)}\pi_{00}^{(2)}} & I^0_{\pi_{10}^{(2)}\pi_{11}^{(2)}}\\
0 & 0 & I^0_{\pi_{00}^{(1)}\pi_{00}^{(1)}} & I^0_{\pi_{00}^{(1)}\pi_{11}^{(1)}} & 0 & 0\\
0 & 0 & I^0_{\pi_{11}^{(1)}\pi_{00}^{(1)}} & I^0_{\pi_{11}^{(1)}\pi_{11}^{(1)}} & 0 & 0\\
I^0_{\pi_{00}^{(2)}\phi} & I^0_{\pi_{00}^{(2)}\pi_{10}^{(2)}} & 0 & 0 & I^0_{\pi_{00}^{(2)}\pi_{00}^{(2)}} & I^0_{\pi_{00}^{(2)}\pi_{11}^{(2)}}\\
I^0_{\pi_{11}^{(2)}\phi} & I^0_{\pi_{11}^{(2)}\pi_{10}^{(2)}} & 0 & 0 & I^0_{\pi_{11}^{(2)}\pi_{00}^{(2)}} & I^0_{\pi_{11}^{(2)}\pi_{11}^{(2)}}
\end{pmatrix}
$$

and its inverse matrix is given by

$$
I_0^{-1}(\phi,\pi_{10}^{(2)},\pi_{00}^{(1)},\pi_{11}^{(1)},\pi_{00}^{(2)},\pi_{11}^{(2)}) = \begin{pmatrix}
I^{11} & I^{12} & I^{13} & I^{14} & I^{15} & I^{16}\\
I^{21} & I^{22} & I^{23} & I^{24} & I^{25} & I^{26}\\
I^{31} & I^{32} & I^{33} & I^{34} & I^{35} & I^{36}\\
I^{41} & I^{42} & I^{43} & I^{44} & I^{45} & I^{46}\\
I^{51} & I^{52} & I^{53} & I^{54} & I^{55} & I^{56}\\
I^{61} & I^{62} & I^{63} & I^{64} & I^{65} & I^{66}
\end{pmatrix}
$$

where

$$I^{11} = \left[I^0_{\phi\phi} - \begin{pmatrix} I^0_{\phi\pi^{(2)}_{10}} & 0 & 0 & I^0_{\phi\pi^{(2)}_{00}} & I^0_{\phi\pi^{(2)}_{11}} \end{pmatrix} \begin{pmatrix} I^0_{\pi^{(2)}_{10}\pi^{(2)}_{10}} & 0 & 0 & I^0_{\pi^{(2)}_{10}\pi^{(2)}_{00}} & I^0_{\pi^{(2)}_{10}\pi^{(2)}_{11}} \\ 0 & I^0_{\pi^{(1)}_{00}\pi^{(1)}_{00}} & I^0_{\pi^{(1)}_{00}\pi^{(1)}_{11}} & 0 & 0 \\ 0 & I^0_{\pi^{(1)}_{11}\pi^{(1)}_{00}} & I^0_{\pi^{(1)}_{11}\pi^{(1)}_{11}} & 0 & 0 \\ I^0_{\pi^{(2)}_{00}\pi^{(2)}_{10}} & 0 & 0 & I^0_{\pi^{(2)}_{00}\pi^{(2)}_{00}} & I^0_{\pi^{(2)}_{00}\pi^{(2)}_{11}} \\ I^0_{\pi^{(2)}_{11}\pi^{(2)}_{10}} & 0 & 0 & I^0_{\pi^{(2)}_{11}\pi^{(2)}_{00}} & I^0_{\pi^{(2)}_{11}\pi^{(2)}_{11}} \end{pmatrix}^{-1} \begin{pmatrix} I^0_{\pi^{(2)}_{10}\phi} \\ 0 \\ 0 \\ I^0_{\pi^{(2)}_{00}\phi} \\ I^0_{\pi^{(2)}_{11}\phi} \end{pmatrix} \right]^{-1}$$

References

1. Hills, M.; Armitage, P. The two-period cross-over clinical trial. *Br. J. Pharmacol.* **1979**, *8*, 7. [CrossRef] [PubMed]
2. Fleiss, J.L. *The Design and Analysis of Clinical Experiments*; John Wiley & Sons: New York, NY, USA, 1986.
3. Senn, S.J. *Cross-Over Trials in Clinical Research*, 2nd ed.; John Wiley & Sons, Ltd.: Chichester, UK, 2002.
4. Sever, P.S.; Poulter, N.R.; Bulpitt, C.J. Double-blind crossover versus parallel groups in hypertension. *Am. Heart J.* **1989**, *117*, 735–739. [CrossRef] [PubMed]
5. Ménard, J.; Serrurier, D.; Bautier, P.; Plouin, P.-F.; Alexandre, J.-M.; Corvol, P. Crossover design to test antihypertensive drugs with self-recorded blood pressure. *Hypertension* **1988**, *117*, 153–159. [CrossRef] [PubMed]
6. Grenet, G.; Blanc, C.; Bardel, C.; Francillard, I.; Combret, S.; Pivot, X.; Roy, P. Comparison of crossover and parallel-group designs for the identification of a binary predictive biomarker of the treatment effect. *Basic Clin. Physiol. Pharmacol.* **2020**, *126*, 59–64. [CrossRef]
7. Jones, B.; Kenward, M.G. *Design and Analysis of Cross-Over Trials*; Chapman and Hall: London, UK, 1989.
8. Senn, S. Cross-over trials in Statistics in Medicine: The first '25' years. *Stat. Med.* **2006**, *25*, 3430–3442. [CrossRef]
9. Mills, E.J.; Chan, A.W.; Wu, P.; Vail, A.; Guyatt, G.H.; Altman, D.G. Design, Analysis, and Presentation of Crossover Trials. *Trials* **2009**, *10*, 27 . [CrossRef]
10. Fava, G.M.; Patel, H.I. A Survey of Crossover Designs Used in Industry. Unpublished manuscript, 1986.
11. Jones, B.; Kenward, M.G. *Design and Analysis of Cross-Over Trials*, 3rd ed.; Chapman & Hall/CRC, Taylor & Francis: Boca Raton, FL, USA, 2014.
12. Kershner, R.P.; Federer, W.T. Two-treatment crossover designs for estimating a variety of effects. *J. Am. Stat. Assoc.* **1981**, *76*, 612–619. [CrossRef]
13. Ezzet, F.; Whitehead, J. A random effects model for binary data from crossover clinical trials. *J. R. Stat. Soc. C-Appl.* **1992**, *41*, 117–126. [CrossRef]
14. Becker, M.P.; Balagtas, C.C. Marginal modeling of binary cross-over data. *Biometrics* **1993**, *49*, 997–1009. [CrossRef]
15. Jaki, T.; Pallmann, P. Estimation in AB/BA crossover trials with application to bioequivalence studies with incomplete and complete data designs. *Stat. Med.* **2013**, *32*, 5469–5483. [CrossRef]
16. Lui, K.J.; Chang, K.C. Hypothesis testing and estimation in ordinal data under a simple crossover design. *J. Biopharm. Stat.* **2012**, *22*, 1137–1147. [CrossRef]
17. Lui, K.J. *Crossover Designs: Testing, Estimation, and Sample Size*; John Wiley & Sons, Ltd.: Chichester, UK, 2016.
18. Li, X.; Li, H.; Jin, M.; Goldberg, J.D. Likelihood ratio and score tests to test the non-inferiority (or equivalence) of the odds ratio in a crossover study with binary outcomes. *Stat. Med.* **2016**, *35*, 3471–3481. [CrossRef]
19. Lui, K.J. Estimation of the treatment effect under an incomplete block crossover design in binary data-a conditional likelihood approach. *Stat. Methods Med. Res.* **2017**, *26*, 2197–2209. [CrossRef]
20. Lui, K.J. Testing equality of treatments under an incomplete block crossover design with ordinal responses. *Int. J. Biostat.* **2017**, *13*, 20160069 . [CrossRef]
21. Lui, K.J.; Chang, K.C. Exact tests in binary data under an incomplete block crossover design. *Stat. Methods Med. Res.* **2018**, *27*, 579–592. [CrossRef] [PubMed]
22. Zhu, L.; Lui, K.J. Notes on misspecifying the random effects distribution regarding analysis under the AB/BA crossover trial in dichotomous data-a Monte Carlo evaluation. *Commun. Stat. Simul. Comput.* **2020**, *49*, 419–435. [CrossRef]
23. Rao, C.R. *Linear Statistical Inference and Its Applications*, 2nd ed.; Wiley: New York, NY, USA, 1985.
24. Tang, N.S.; Tang, M.L.; Qiu, S.F. Testing the equality of proportions for correlated otolaryngologic data. *Comput. Stat. Data Anal.* **2008**, *52*, 3719–3729. [CrossRef]
25. Alhija, F.N.A.; Levy, A. Effect size reporting practices in published articles. *Educ. Psychol. Meas.* **2009**, *69*, 245–265. [CrossRef]
26. Odgaard, E.C.; Fowler, R.L. Confidence intervals for effect sizes: Compliance and clinical significance in the Journal of Consulting and clinical Psychology. *J. Consult. Clin. Psychol.* **2010**, *78*, 287–297. [CrossRef]

27. Sun, S.; Pan, W.; Wang, L.L. A comprehensive review of effect size reporting and interpreting practices in academic journals in education and psychology. *J. Educ. Psychol.* **2010**, *102*, 989–1004. [CrossRef]
28. Dunst, C.J.; Hamby, D.W. Guide for calculating and interpreting effect sizes and confidence intervals in intellectual and developmental disability research studies. *J. Intellect. Dev. Disabil.* **2012**, *37*, 89–99. [CrossRef]
29. Fritz, C.O.; Morris, P.E.; Richler, J.J. Effect size estimates: Current use, calculations, and interpretation. *J. Exp. Psychol. Gen.* **2012**, *141*, 2–18. [CrossRef] [PubMed]
30. American Psychological Association. *Publication Manual of the American Psychological Association*, 6th ed.; American Psychological Association: Washington, DC, USA, 2009.
31. Traub, J.F. *Iterative Methods for the Solution of Equations*; American Mathematical Society: Providence, RI, USA, 1982.
32. Lui, K.J.; Chang, K.C. Exact Sample-Size Determination in Testing Non-Inferiority under a Simple Crossover Trial. *Pharm. Stat.* **2012**, *11*, 129–134. [CrossRef] [PubMed]
33. Koch, G.G.; Gitomer, S.L.; Skalland, L.; Stokes, M.E. Some non-parametric and categorical data analysis for a change-over design study and discussion of apparent carry-over effects. *Stat. Med.* **1983**, *2*, 397–412. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.