axioms

MDPI

Article

Comparative Analysis of Exact Methods for Testing Equivalence of Prevalences in Bilateral and Unilateral Combined Data with and without Assumptions of Correlation

Shuyi Liang and Changxing Ma *

Department of Biostatistics, University at Buffalo, Buffalo, NY 14214, USA
* Correspondence: cxma@buffalo.edu

Abstract: In clinical studies focusing on paired body parts, diseases can manifest on either both sides (bilateral) or just one side (unilateral) of the organs. Consequently, the data in these studies may consist of records from both bilateral and unilateral cases. There are two different methods of analyzing the data. One of the methods is assuming that the pair of measurements from the same subject are independent, while the other considers the correlation between paired organs. In terms of the homogeneity test of proportions, asymptotic methods have been proposed given the moderate size of data. This article extends the existing work by proposing exact methods to deal with the scenarios when the sample size is small and asymptotic methods perform poorly. The impact of the correlation assumption is also explored. Among the proposed methods, calculating *p*-values by replacing unknown parameters with estimated values while accounting for the correlation is recommended based on its satisfactory type I error controls and statistical powers. The proposed methods are applied to three real examples for illustration.

Keywords: equal correlation coefficients model; homogeneity test; unilateral and bilateral data; exact method

MSC: 62F03

check for updates

Citation: Liang, S.; Ma, C. Comparative Analysis of Exact Methods for Testing Equivalence of Prevalences in Bilateral and Unilateral Combined Data with and without Assumptions of Correlation. *Axioms* **2024**, *13*, 430. https://doi.org/10.3390/axioms13070430

Academic Editors: Xiaonan Zhu and Jong-Min Kim

Received: 21 May 2024
Revised: 18 June 2024
Accepted: 23 June 2024
Published: 26 June 2024

Copyright: © 2024 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

Data collected from paired organs of human bodies can be bilateral or unilateral depending on the availability of each site or the disease status. Whether to assume or ignore the correlation between two sites of an organ remains an interesting discussion. The majority of articles published in the *British Journal of Ophthalmology* (BJO) in 1995 and 2017 assumed independence of the sites according to Zhang and Ying [1]. For the homogeneity tests of group-specific response rates given the outcome is binary, classical methods such as Pearson's chi-squared test [2] can be applied. However, Rosner [3], Donner, and Banting [4]) criticized this approach by pointing out that the omission of the correlation will lead to biased inferences. However, there are not many studies that demonstrate such a bias with numerical presentations.

To avoid the possible bias, Donner [5] assumed that the number of successes from each pair of organs follows a binomial distribution with probability parameter P, where P varies among subjects and has an expectation of p. The variance of P is proportional to $p(1-p)$; that is, $Var(P) = \rho p(1-p)$. This concept can be extended easily by assuming the intra-person correlation is ρ, where ρ ranges from -1 to 1. The correlation is assumed to be constant for all individuals and is referred to as the equal correlation coefficients model. In the context of bilateral data, Ma and Liu [6] derived three asymptotic methods, the likelihood ratio test, the score test, and the Wald-type test, to assess the homogeneity of prevalences of multiple groups. Mou and Li [7] compared multiple algorithms for estimating parameters and investigated asymptotic statistics for the homogeneity test

of many-to-one risk differences. Liu et al. [8] considered four exact approaches, the E approach, the M approach, the E + M approach, and the C approach, as alternatives to the methods proposed by Ma and Liu [6] when the sample size cannot ensure a good asymptotic approximation. In general, exact approaches calculate the probability of a tail area, which is defined as a set of cases that are more extreme than the observed data against the null hypothesis. The E approach, initially proposed by Liddell [9], replaces nuisance parameters with their corresponding maximum likelihood estimate. Basu [10] introduced the M approach, which involves determining the p-value by maximizing the likelihood of observing the tail area across the entire parameter space. By combining the E and the M methods, Lloyd [11] first determined the tail area by regarding the p-value from the E approach as a statistic and then applied the M method to calculate the final p-value. The previous three methods assume the group totals are fixed, while the C method in Liu's work [8] fixes both margins of a table. In the context of bilateral and unilateral combined data, Ma and Wang [12] derived three asymptotic approaches for testing the equality of proportions using the equal correlation coefficients model. It is worth noting that these methods still require a large sample size, and there is a need to extend Liu's work to combined data.

This article proposes five exact approaches given bilateral and unilateral combined data with limited samples for testing the equivalence of proportions of multiple groups. Section 2 presents details of Donner's equal correlation coefficients model. In Section 3, we introduce the individual site model, which omits the correlation between sites. Classical methods based on the individual site model are described in Section 4, and we introduce five exact approaches in Section 5 using Donner's model. In Section 4, we provide numerical studies that compare the two models and various methods with regard to their ability to control Type I errors (TIEs) and their statistical powers. The application of the proposed methods through two real-world examples is illustrated in Section 6. The final conclusions and discussion are given in Section 7.

2. Equal Correlation Coefficients Model

Define $Z_{ijk} = 1$ as the event of the jth subject in the ith group having a response at the kth site and π_i as $Prob(Z_{ijk} = 1) = Prob(Z_{ij(3-k)} = 1)$, where $i = 1, 2, ..., g$ and $k = 1, 2$. The equal correlation coefficients model proposed by Donner [5] assumes the correlation coefficient $Corr(Z_{ijk}, Z_{ij(3-k)})$ between measurements from two sites to be a constant ρ. In the bilateral cohort where subjects contribute one measurement at each site of the paired organs, let m_{ir} stand for the number of subjects with r response(s) in the ith group, where $r = 0, 1, 2$. In the unilateral cohort where subjects provide one measurement on only one site of the paired organs, n_{ir^*} denotes the number of subjects with r^* response(s) in the ith group, where $r^* = 0, 1$. The data layout on the subject level can be found in Table 1.

Table 1. Data layout on subject level.

	Group (i)		**1**	**2**	**...**	**g**	**Total**
Bilateral	Response (r)	0	m_{10}	m_{20}	...	m_{g0}	M_0
		1	m_{11}	m_{21}		m_{g1}	M_1
		2	m_{12}	m_{22}		m_{g2}	M_2
	Total		m_1	m_2	...	m_g	M^+
Unilateral	Response (r^*)	0	n_{10}	n_{20}	...	n_{g0}	N_0
		1	n_{11}	n_{21}		n_{g1}	N_1
	Total		n_1	n_2	...	n_g	N^+

Given the nature of clinical study designs, $(m_1, m_2, ..., m_g, n_1, n_2, ..., n_g)$ are fixed, and the plausible distribution assumptions are as follows:

$$\begin{cases} n_{i0} \sim \text{binomial distribution Bin}(n_i, P_{ui0}); \\ n_{i1} \sim \text{binomial distribution Bin}(n_i, P_{ui1}); \\ (m_{i0}, m_{i1}, m_{i2}) \sim \text{multinomial distribution } Multi(m_i; P_{bi0}, P_{bi1}, P_{bi2}). \end{cases}$$

In observational studies, the sample size of each group can be random, which is beyond the scope of this article. It can be easily shown that $P_{bi0} = (1-\pi_i)(1-\pi_i+\rho\pi_i)$, $P_{bi1} = 2(1-\rho)(\pi_i - \pi_i^2)$, $P_{bi2} = \rho\pi_i + (1-\rho)\pi_i^2$,$P_{ui0} = 1-\pi_i$, and $P_{ui1} = \pi_i$. The null and the alternative hypotheses of interest are as follows:

$$H_0 : \pi_1 = \pi_2 = ...\pi_g = \pi \text{ versus } H_1 : \text{not all of them are equal.}$$

3. Individual Site Model without Considering Correlation

Many studies encounter challenges due to the limited sample size of participants. One of the common strategies that researchers consider is conducting analyses on the site level of each subject to extract as much information as possible. To our knowledge, the impact of ignoring the potential correlation of the two sites of a subject remains unknown. In order to explore this effect, data from the equal correlation coefficients model are transformed to new data on the site level. Define x_{ie} as the number of sites with e response, where $e = 0, 1$ and $i = 1, 2, ..., g$. The new data structure is presented in Table 2. The relationship between the two types of data is presented as follows:

$$\begin{cases} x_{i0} = 2m_{i0} + m_{i1} + n_{i0}, \\ x_{i1} = 2m_{i2} + m_{i1} + n_{i1}, \\ x_i = 2m_i + n_i. \end{cases}$$

The random variables x_{i0} and x_{i1} are assumed to follow binomial distributions $\text{Bin}(x_i, 1-\pi_i)$ and $\text{Bin}(x_i, \pi_i)$, respectively. The null and the alternative hypotheses are the same as stated in Section 2.

Table 2. Data layout on site level.

Group (i)		**1**	**2**	**...**	**g**	**Total**
Response (e)	0	x_{10}	x_{20}	...	x_{g0}	T_0
	1	x_{11}	x_{21}	...	x_{g1}	T_1
Total		x_1	x_2	...	x_g	T

4. Methods for Individual Site Model

4.1. The Pearson Chi-Squared Test

Define the notation X^* as an observed table $(x_{10}, x_{11}, x_{20}, x_{21}, ..., x_{g0}, x_{g1})$ from Table 2. Given the ignorance of correlations between the sites, the individual site model becomes an unordered $g \times 2$ table where the Pearson chi-squared test statistic is given by Equation (1).

$$T_{Pearson}(X^*) = \sum_{i=1}^{g} \left[\frac{(T\, x_{i1} - T_1\, x_i)^2}{T\, T_1\, x_i} + \frac{(T\, x_{i0} - T\, x_i + T_1\, x_i)^2}{T\, x_i\, (T - T_1)} \right] \quad (1)$$

The p-value is defined as $P_{Pearson}(X^*) = Prob(\mathcal{X}^2_{g-1} \geq T_{Pearson}(X^*))$ since the test statistic $T_{Pearson}(X^*)$ asymptotically follows a chi-squared distribution with $g-1$ degrees of freedom when the sample size is sufficient (Fagerland et al. [13]). The null hypothesis is rejected if the calculated p-value is less than 0.05 given the significance level of 0.05.

4.2. The Fisher–Freeman–Halton Exact Test

The asymptotic behavior described above does not hold when the numbers in Table 2 are small. To overcome this limitation, an alternative test called the Fisher–Freeman–Halton (FFH) exact test can be employed (Fagerland et al. [13]). The point probability, $Prob(X)$, conditions both the row and column totals and is used as the test statistic. In the null hypothesis, the probability distribution of Table 2 follows the multiple hypergeometric distribution. The Fisher–Freeman–Halton exact test treats tables with a smaller point probability than $Prob(X^*)$ as a piece of evidence against H_0, where $Prob(X^*)$ is the probability of observing X^*. Let $\Omega_{FFH}(X^*)$ represent the collection of all tables sharing the same row and column margins as the observed table X^*. Therefore, the exact p-value is defined by Equation (2), where $I()$ is the indicator function.

$$P_{FFH}(X^*) = \sum_{X \in \Omega_{FFH}(X^*)} I(Prob(X) \leq Prob(X^*))Prob(X) \tag{2}$$

4.3. The Mid-P Test

The Fisher–Freeman–Halton exact test is often criticized for being unnecessarily conservative due to discreteness in certain scenarios. One of the adjustments to $P_{FFH}(X^*)$ is called the mid-P test. This test downsizes the weight of tables with the same point probability as the observed table X^* (Fagerland et al. [13]). In other words, the adjusted p-value can be defined by Equation (3).

$$\begin{aligned} P_{mid-P}(X^*) = & \sum_{X \in \Omega_{FFH}(X^*)} I(Prob(X) < Prob(X^*))Prob(X) \\ & + \frac{1}{2} \sum_{X \in \Omega_{FFH}(X^*)} I(Prob(X) = Prob(X^*))Prob(X) \end{aligned} \tag{3}$$

5. Methods for Equal Correlation Coefficients Model

Given observed data $M^* = (m_{10}, m_{11}, m_{12}, n_{10}, n_{11}, ..., m_{g0}, m_{g1}, m_{g2}, n_{g0}, n_{g1})$, the likelihood function is given by Equation (4), and the log-likelihood function can be written as a function of $\pi_1, \pi_2, ..., \pi_g$, and ρ as shown in Equation (5).

$$\begin{aligned} \mathcal{L}(M^*; \pi_1, \pi_2, ..., \pi_g, \rho) = \prod_{i=1}^{g} \Bigg[& \frac{m_i!}{m_{i0}! m_{i1}! m_{i2}!} ((1-\pi_i)(1-\pi_i+\rho\pi_i))^{m_{i0}} \\ & \times (2(1-\rho)(\pi_i - \pi_i^2))^{m_{i1}} (\rho\pi_i + (1-\rho)\pi_i^2)^{m_{i2}} \\ & \times \frac{n_i!}{n_{i0}! n_{i1}!} (1-\pi_i)^{n_{i0}} \pi_i^{n_{i1}} \Bigg] \end{aligned} \tag{4}$$

$$\begin{aligned} l(\pi_1, \pi_2, ..., \pi_g, \rho) = \sum_{i=1}^{g} \Big[& m_{i0} \log((1-\pi_i)(1-\pi_i+\rho\pi_i)) + m_{i1} \log(2(1-\rho)(\pi_i - \pi_i^2)) \\ & + m_{i2} \log(\rho\pi_i + (1-\rho)\pi_i^2) \\ & + n_{i0} \log(1-\pi_i) + n_{i1} \log(\pi_i) \Big] \end{aligned} \tag{5}$$

Let $\mathcal{L}(M^*; \pi, \rho)$ represent $\mathcal{L}(M^*; \pi_1 = \pi_2 = ... = \pi_g = \pi, \rho)$ for ease of notation. With H_0, the maximum likelihood estimates (MLEs) can be obtained by setting the partial derivatives with respect to π_i and ρ to zero and solving Equations (6) and (7).

$$\frac{\partial l}{\partial \rho} = \frac{M_1}{\rho - 1} - \frac{M_2\,(\pi - 1)}{\pi + \rho - \pi\,\rho} + \frac{M_0\,\pi}{\pi\,\rho - \pi + 1} = 0 \tag{6}$$

$$\frac{\partial l}{\partial \pi} = \frac{N_1}{\pi} + \frac{N_0}{\pi - 1} - \frac{M_2\,(\rho - 2\,\pi\,(\rho - 1))}{\pi^2\,(\rho - 1) - \pi\,\rho} - \frac{M_1\,(2\,\pi - 1)}{\pi - \pi^2} + \frac{M_0\,((\pi - 1)\,(\rho - 1) - \pi + \pi\,\rho + 1)}{(\pi - 1)\,(\pi\,\rho - \pi + 1)} = 0 \tag{7}$$

Let $\hat{\pi}$ and $\hat{\rho}$ denote the constrained MLEs for H_0. The closed-form solutions to the above equations can be found in Ma and Wang's work [12].

5.1. Score Test

Ma and Wang [12] examined the likelihood ratio test, the score test, and the Wald-type test and concluded that the score test outperformed the other two approaches based on simulation studies using various combinations of parameter settings and sample sizes. Therefore, we choose the score test to explore the asymptotic behaviors given smaller sample sizes compared to Ma and Wang's settings. This approach is referred to as the A method. The score test statistic is given by Equation (8), where $\mathbf{U} = (U_1, U_2, ..., U_g, U_{g+1}) = (\frac{\partial l}{\partial \pi_1}, \frac{\partial l}{\partial \pi_2}, ..., \frac{\partial l}{\partial \pi_g}, \frac{\partial l}{\partial \rho})$, and $\mathbf{I}(\pi_1, \pi_2, ..., \pi_g, \rho)$ is the Fisher information assuming the group-specific proportion π_i's.

$$T_{SC}(M^*) = \mathbf{U}\mathbf{I}(\pi_1, \pi_2, ..., \pi_g, \rho)^{-1}\mathbf{U}^T | \pi_1 = \pi_2 = ... = \pi_g = \hat{\pi}, \rho = \hat{\rho} \tag{8}$$

We kindly recommend that readers refer to Ma and Wang's work for a simplified form of $T_{SC}(M^*)$. Asymptotically, the score test statistic follows a chi-square distribution with $g - 1$ degrees of freedom given the null hypothesis is true according to Rao [14]. The corresponding p-value can be calculated by Equation (9).

$$P_A(M^*) = Prob(\mathcal{X}^2_{g-1} > T_{SC}(M^*)) \tag{9}$$

Similar to the Pearson chi-squared statistic, the approximation to $\mathcal{X}^2_{g-1}$ does not hold if the sample size is insufficient, leading to uncontrolled type I error rates. To alleviate this difficulty, four exact methods are proposed by fixing $(m_1, m_2, ..., m_g, n_1, n_2, ..., n_g)$ and are described in detail in Sections 5.2–5.5. Define the collection of all tables with the same row margin as M^* in Table 1 as $\Omega_1(M^*)$. In Section 5.6, we propose another exact method by fixing both the row and column totals.

5.2. E Method

There are several methods to calculate p-values based on the summation of probabilities of tables that are more extreme than the observed table M^* when H_0 is true. One of the approaches treats the values of constrained MLEs $\hat{\pi}$ and $\hat{\rho}$ as the true parameter values, and it is referred to as the E method because these estimates are used as substitutes for unknown parameters in calculations. As a result, the p-value based on the E method is defined by Equation (10).

$$P_E(M^*) = \sum_{M \in \Omega_1(M^*)} I(T_{SC}(M) \geq T_{SC}(M^*))\mathcal{L}(M | \pi = \hat{\pi}, \rho = \hat{\rho}) \tag{10}$$

5.3. M Method

Instead of using the estimated parameters directly, an alternative approach is to maximize the above summation of probabilities in the whole parameter space. This technique is denoted as the M method. The exact p-value can be calculated by Equation (11).

$$P_M(M^*) = \sup_{\pi \in (0,1), \rho \in (-1,1)} \left[\sum_{M \in \Omega_1(M^*)} I(T_{SC}(M) \geq T_{SC}(M^*))\mathcal{L}(M; \pi, \rho) \right] \tag{11}$$

5.4. E + M Method

The M method discussed above calculates p-values based on the score test statistic. Notice that the exact p-value, $P_E(M^*)$, can be regarded as a statistic as well and replace the role of the score statistic in the M method. The fusion of the two approaches is referred to as the E + M method. The exact p-value is defined by Equation (12).

$$P_{E+M}(M^*) = \sup_{\pi\in(0,1),\rho\in(-1,1)} \left[\sum_{M\in\Omega_1(M^*)} I(P_E(M) \le P_E(M^*))\mathcal{L}(M;\pi,\rho) \right] \tag{12}$$

5.5. CI Method

Unlike the M method, a smaller parameter space consisting of the $100(1-\beta)\%$ confidence intervals (CIs) of π and ρ can be used when maximizing the summation of probabilities, and this is called the CI method. The confidence intervals of π and ρ are denoted as CI_π and CI_ρ, respectively. We kindly ask the readers to refer to Berger [15] and Silvapulle's [16] work for details. In essence, the rationale is that the M method can be conservative since the generated p-values may be close to one due to the nature of the supremum. The CI method shrinks the parameter space, resulting in smaller p-values than the M approach. The corresponding p-value is defined by Equation (13).

$$P_{CI}(M^*) = \sup_{\pi\in CI_\pi,\rho\in CI_\rho} \left[\sum_{M\in\Omega_1(M^*)} I(T_{SC}(M) \ge T_{SC}(M^*))\mathcal{L}(M;\pi,\rho) \right] + 3\beta \tag{13}$$

A definition of the p-value is considered valid if $Prob(p\text{-value} \le \alpha|H_0) \le \alpha$ holds according to Vexler [17]). A simple proof that $P_{CI}(M^*)$ is a valid p-value is given in the Appendix A.

The confidence intervals of π and ρ are derived using another score test statistic (T^*_{SC}) with the null hypothesis $H_0 : \pi_1 = \pi_2 = ... = \pi_g = \pi$. The conditional MLE of π given known ρ can be obtained by solving Equation (2), and the conditional MLE of ρ given π can be obtained by solving Equation (1). The score test statistic T^*_{SC} can be expressed by Equation (14),

$$T^*_{SC}(M^*) = \mathbf{U}_C\mathbf{I}_C(\pi,\rho)^{-1}\mathbf{U}_C^T \tag{14}$$

where

$$\mathbf{U}_C = (\frac{\partial l}{\partial \pi}, \frac{\partial l}{\partial \rho}),\ \mathbf{I}_C(\pi,\rho) = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix},$$

$$\begin{aligned} I_{11} =& E(-\frac{\partial^2 l}{\partial \pi^2}) = \frac{N}{\pi} - 4\,M\,(\rho-1) - \frac{N}{\pi-1} + 2\,M\,(2\,\rho-2) + \frac{M\,(\rho-2\,\pi\,(\rho-1))^2}{\pi\,\rho-\pi^2\,(\rho-1)} \\ & - \frac{M\,(\pi\rho-\pi+(\pi-1)(\rho-1)+1)}{\pi-1} - \frac{2\,M\,(2\,\pi-1)^2\,(\rho-1)}{\pi-\pi^2} \\ & - \frac{M\,(\rho-1)\,(\pi\rho-\pi+(\pi-1)(\rho-1)+1)}{\pi\,\rho-\pi+1}, \\ I_{12} =& I_{21} = E(-\frac{\partial^2 l}{\partial \pi \partial \rho}) = \frac{M\,\rho\,(2\,\pi-1)}{(\pi\,\rho-\pi+1)\,(\rho+\pi-\pi\,\rho)}, \\ \text{and } I_{22} =& E(-\frac{\partial^2 l}{\partial \rho^2}) = \frac{\pi\,M\,(\pi-1)\,(\rho+1)}{(\rho-1)\,(\pi\,\rho-\pi+1)\,(\rho+\pi-\pi\,\rho)}. \end{aligned}$$

The following iterative procedure outlines the details of finding the upper limit of the CI of π:

(1) Set $\hat{\pi}$ as the starting point of π, where $\hat{\pi}$ is the constrained MLE of π for H_0. Initialize flag = 1 and stepsize = $min\{0.01, (1-\hat{\pi})/10\}$ so that the updated upper bound does not prematurely exceed 1;
(2) Update $\pi^{(t+1)} = \pi^{(t)}$ + flag $\times$ stepsize and calculate the conditional MLE $\tilde{\rho}$ given $\pi^{(t+1)}$. Then, the score test statistic is given by $(T^*_{SC}(M^*))^{(t+1)} = \mathbf{U}_C\mathbf{I}_C(\pi,\rho)^{-1}\mathbf{U}_C^T|\pi = \pi^{(t+1)}, \rho = \tilde{\rho}$;
(3) If $(T^*_{SC}(M^*))^{(t+1)} > \mathcal{X}^2_{1,1-\beta}$, where $\mathcal{X}^2_{1,1-\beta}$ is the $1-\beta$ quantile of the chi-square distribution with one degree of freedom, turn to the opposite searching direction by letting flag = -1 and reduce the stepsize by multiplying it by $1/3.1416$, then return to step (2). Otherwise, keep flag = 1 and return to step (2);
(4) Repeat steps (2) and (3) until the stepsize is sufficiently small (e.g., 10^{-4}).

The lower limit of the CI of π can be determined by letting flag = -1 and stepsize = $min\{0.01, \hat{\pi}/10\}$ in step (1) and then multiplying it by -1 if $(T^*_{SC}(M^*))^{(t+1)} > \mathcal{X}^2_{1,1-\beta}$ or multiplying it by 1 if $(T^*_{SC}(M^*))^{(t+1)} \leq \mathcal{X}^2_{1,1-\beta}$.

The following procedures can be used to determine the upper limit of the CI of ρ:

(1) Set $\hat{\rho}$ as the starting point of ρ, where $\hat{\rho}$ is the constrained MLE of ρ for H_0. Initialize flag = 1 and stepsize = $min\{0.01, (1-\hat{\rho})/10\}$;
(2) Update $\rho^{(t+1)} = \rho^{(t)}$ + flag $\times$ stepsize and calculate the conditional MLE $\tilde{\pi}$ given $\rho^{(t+1)}$. Then, the score test statistic is given by $(T^*_{SC}(M^*))^{(t+1)} = \mathbf{U}_C\mathbf{I}_C(\pi,\rho)^{-1}\mathbf{U}_C^T|\rho = \rho^{(t+1)}, \pi = \tilde{\pi}$;
(3) If $(T^*_{SC}(M^*))^{(t+1)} > \mathcal{X}^2_{1,1-\beta}$, turn to the opposite searching direction by letting flag = -1 and reduce the stepsize by multiplying it by $1/3.1416$, then return to step (2). Otherwise, keep flag = 1 and return to step (2);
(4) Repeat steps (2) and (3) until the stepsize is sufficiently small (e.g., 10^{-4}).

The lower bound of the CI of ρ can be found by initializing flag = -1 and stepsize = $min\{0.01, (1+\hat{\rho})/10\}$ in step (1) and multiply it by -1 if $(T^*_{SC}(M^*))^{(t+1)} > \mathcal{X}^2_{1,1-\beta}$ or multiply it by 1 if $(T^*_{SC}(M^*))^{(t+1)} \leq \mathcal{X}^2_{1,1-\beta}$.

5.6. C Method

The conditional method, denoted as the C method, assumes both the column and row totals are fixed in Table 1, and it does not involve nuisance parameters in determining p-values. Let $(m_1^*, n_1^*, m_2^*, n_2^*, ..., m_g^*, n_g^*, S_0^*, S_1^*, S_2^*, N_0^*, N_1^*)$ represent the margins of the observed table M^* and define $\Omega_C(M^*)$ as the collection of all tables having the same margins as M^*. The exact p-value is determined by Equation (15).

$$P_C(M^*) = \sum_{M \in \Omega_C(M^*)} \left(I(T_{SC}(M) \geq T_{SC}(M^*)) \times \frac{\prod_{i=1}^{g} \frac{m_i!}{m_{i0}!m_{i1}!m_{i2}!}}{\frac{M!}{S_0!S_1!S_2!}} \times \frac{\prod_{i=1}^{g} \frac{n_i!}{n_{i0}!n_{i1}!}}{\frac{N!}{N_0!N_1!}} \right) \quad (15)$$

6. Numerical Study

The performance of the approaches discussed in Sections 4 and 5 was evaluated with respect to type I error controls and statistical powers. To simulate practical scenarios characterized by small sample sizes, the total sample size $M+N$ was set to around 20 and 10. We further assumed a balanced design with $m_1 = m_2 = ... = m_g$ and $n_1 = n_2 = ... = n_g$. It is worth noting that unbalanced studies can be explored in a similar manner. The nominal level $\alpha = 5\%$ was used throughout the numerical study.

For the equal correlation coefficients model, all tables having a given margin $(m_1, m_2, ..., m_g, n_1, ..., n_g)$ were enumerated, and the corresponding p-values based on the six methods discussed can be calculated. The type I errors and powers of each proposed method were determined by the summation of probabilities of tables with a p-value < 0.05 for H_0 and H_a, respectively. We set the parameter $\beta = 0.0001$ in the CI method. Figures 1–3 demonstrate the contour plots of type I error rates as a function of π and ρ when the total sample size is

around 10. The contour plots of type I error rates given a sample size of approximately 20 are shown in Figures A1 and A2. The blank areas in the plots represent combinations of π and ρ for which at least one of the parameters (P_{bi0}, P_{bi1}, P_{bi2}, P_{ui0}, P_{ui1}) falls outside the interval [0, 1]. If the type I error of a test is greater than 6%, the test is considered liberal. And if the type I error of a test is less than 4%, the test is considered conservative according to Tang et al. [18].

Regardless of the sample size and the group number, the asymptotic method, the M method, the E + M method, the C method, and the CI method typically yield conservative type I errors since a large area of combinations of π and ρ producing TIEs ≤ 0.04 is observed. The CI approach and the M approach generate very similar results due to their similar constructions of p-values. The CI approach defines the p-values within the space consisting of two confidence intervals, and the M approach considers the whole parameter space. When the total sample size is around 10, the C method produces extremely conservative outcomes with all parameter settings generating TIEs that are less than 3% due to discreteness. Although the E method generates a small proportion of area in the contour plots where the type I errors exceed 6%, it maintains a larger portion of type I error rates within the range of 4% and 6% compared to the other five methods.

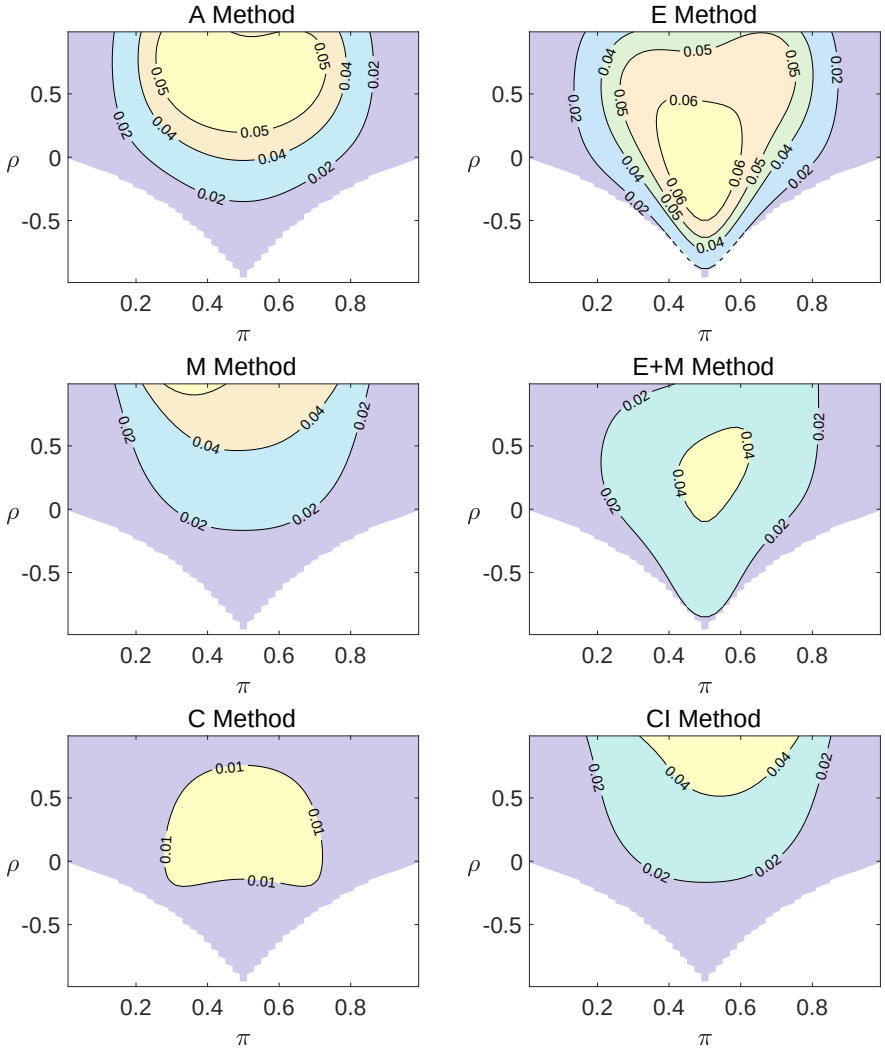


Figure 1. Contour plots of type I errors for the equal correlation coefficients model ($g = 2$, $m_i = 3$, $n_i = 2$, $i = 1, 2$).

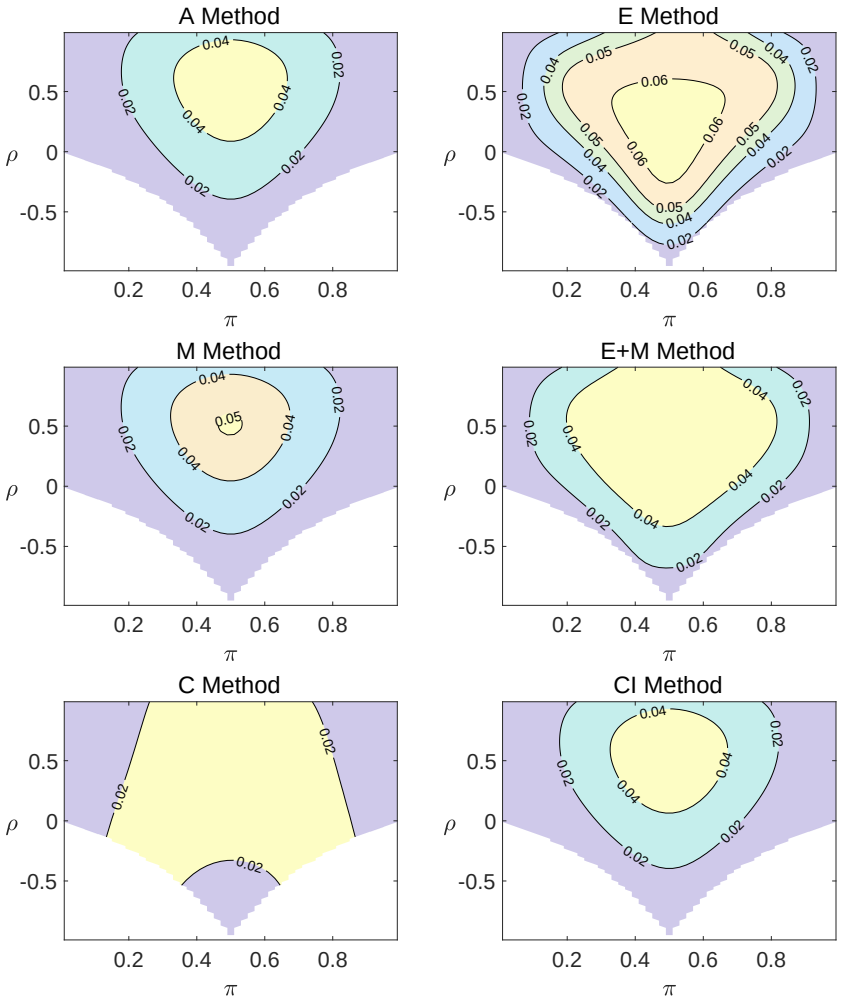


Figure 2. Contour plots of type I errors for the equal correlation coefficients model ($g = 3$, $m_i = 3$, $n_i = 1$, $i = 1, 2, 3$).

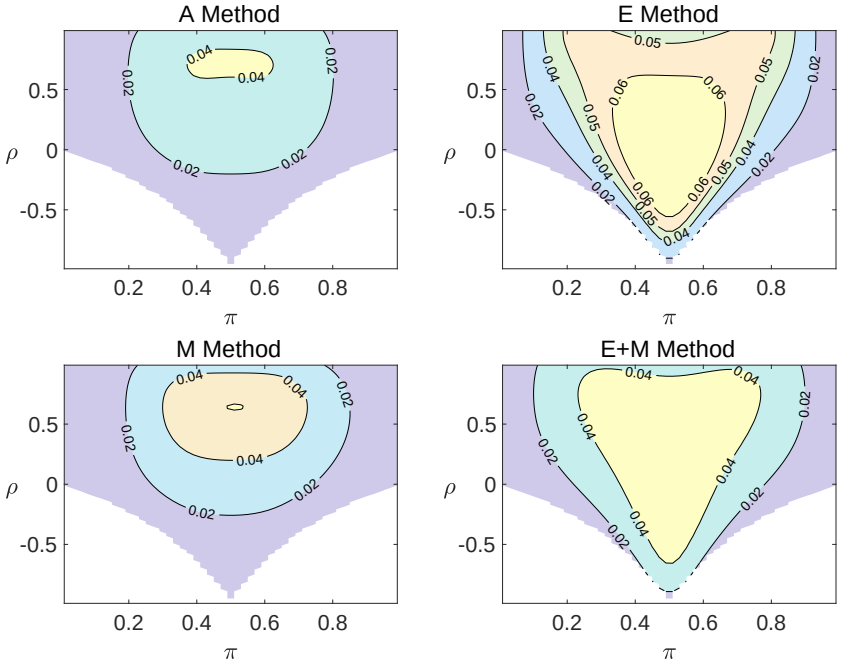


Figure 3. *Cont.*

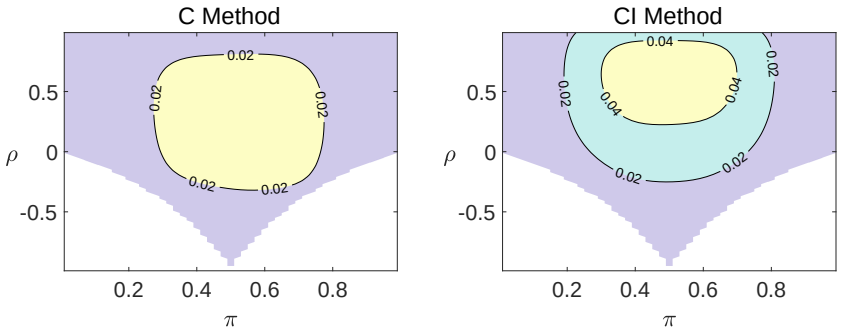


Figure 3. Contour plots of type I errors for the equal correlation coefficients model ($g = 4$, $m_i = 2$, $n_i = 1$, $i = 1, 2, 3, 4$).

For each combination of (g, m_i, n_i) from Figures 1–3 and from Figures A1 and A2, a parallel study was conducted using the individual site model. For example, tables with $g = 2, m_i = 5$, and $n_i = 5$ can be transformed to $(x_{10}, x_{11}, x_{20}, x_{21})$, where $x_1 = x_2 = 2 \times 5 + 5 = 15$. The performance of type I error controls of the Pearson chi-squared test, the FFH test, and the mid-P test were investigated, and the contour plots are exhibited in Figures 4–6, A3 and A4.

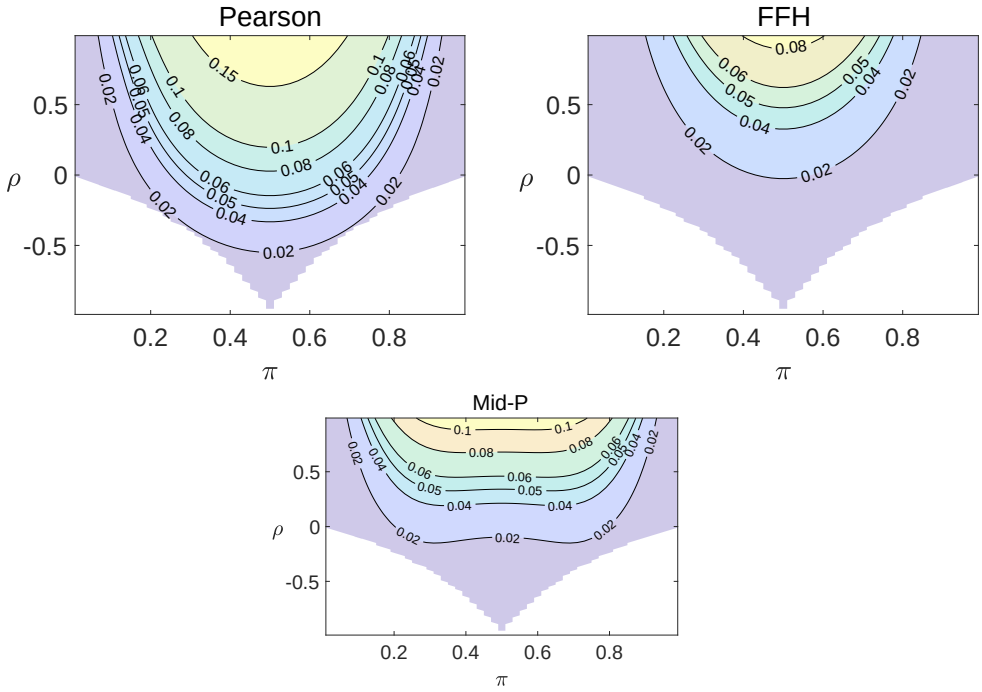


Figure 4. Contour plots of type I errors for the individual site model ($g = 2$, $m_i = 3$, $n_i = 2$, $i = 1, 2$).

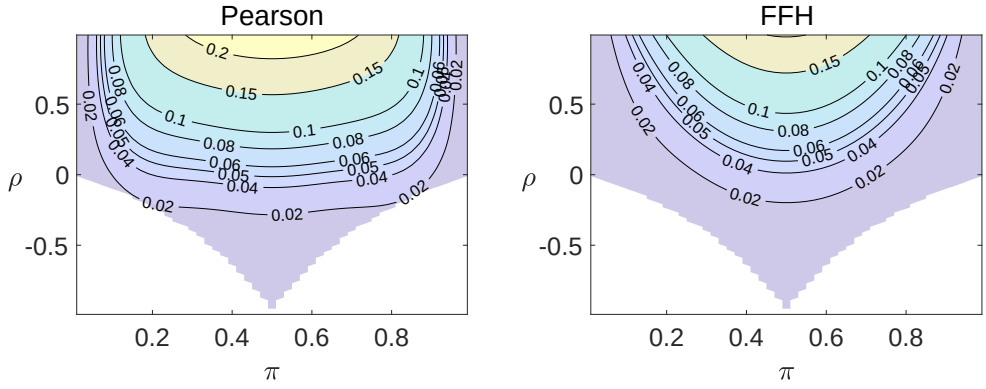


Figure 5. *Cont.*

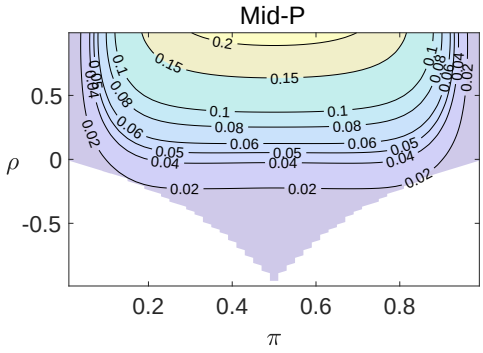


Figure 5. Contour plots of type I errors for the individual site model ($g = 3, m_i = 3, n_i = 1, i = 1, 2, 3$).

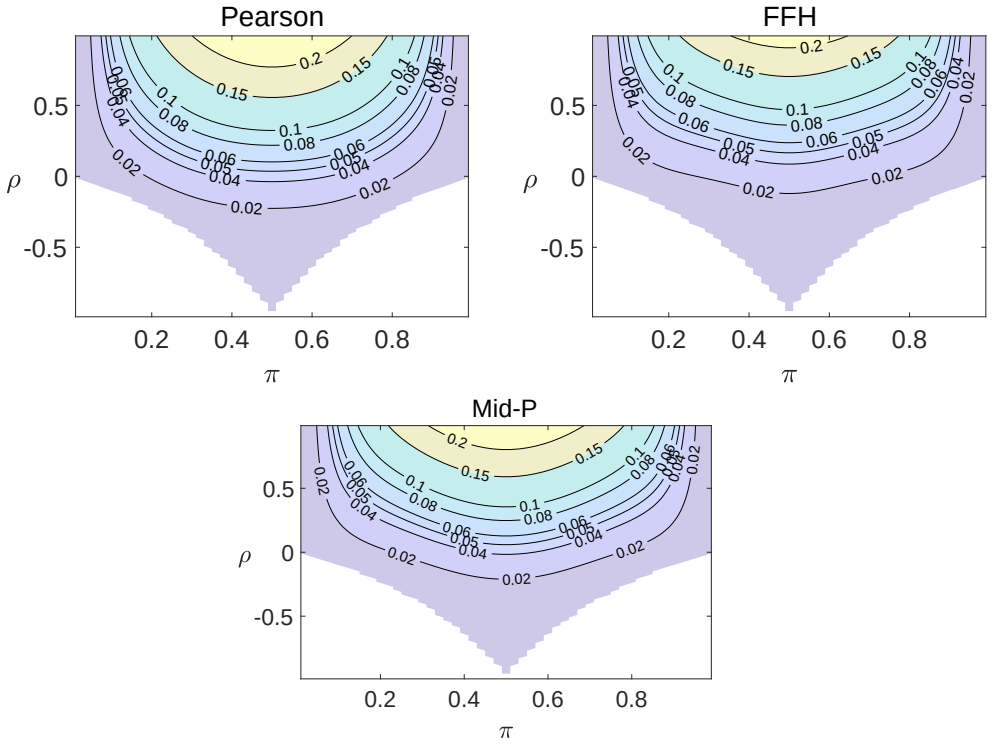


Figure 6. Contour plots of type I errors for the individual site model ($g = 4, m_i = 2, n_i = 1, i = 1, 2, 3, 4$).

Using the individual site model, the Pearson, the FFH, and the mid-P test generate inflated type I errors when the underlying correlation coefficient ρ is close to 1, while the FFH approach outputs conservative TIEs when $g = 2, m_i = 5, n_i = 5$. If the two sites of a subject have a weak correlation ($\rho \approx 0$), type I errors from the three methods can be controlled within the satisfied region. Hence, the equal correlation coefficients model is recommended in practice given the unsatisfied type I error controls when the correlation is ignored.

The power performance is investigated in the following part. The exact power is determined by varying the value of π_g while fixing $\pi_1, ..., \pi_{g-1}$, and ρ. Based on Figures 7–10 and A5–A8, the E method outperforms the other five methods and have the highest powers in all scenarios. The power of the C approach is generally the lowest, which is expected given the low type I error rates observed in the numerical study above. Again, the CI and the M method generate very similar powers due to the intrinsic similarities. Note that if the absolute difference between π_g and the value of $\pi_1 = ...\pi_{g-1}$ decrease to 0, the powers of all methods become close to 0.05 as indicated by the horizontal dash lines. This behavior is anticipated because the smaller the difference, the more likely the tables will support the null hypothesis. The statistical test powers increase as the sample size increases from 10 to 20 no matter what method is employed.

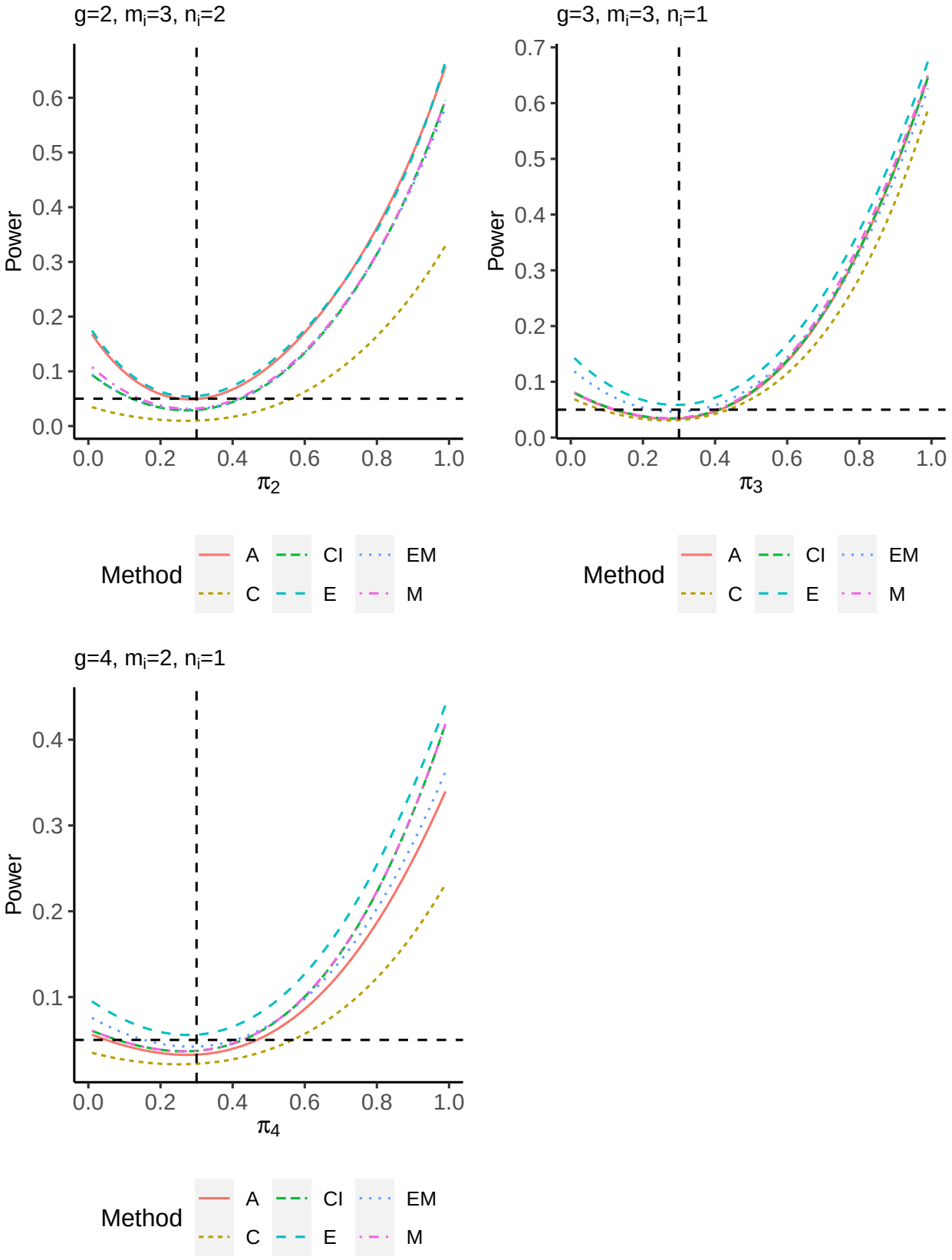


Figure 7. Power plots for $\pi_1 = ... = \pi_{g-1} = 0.3$, and $\rho = 0.4$.

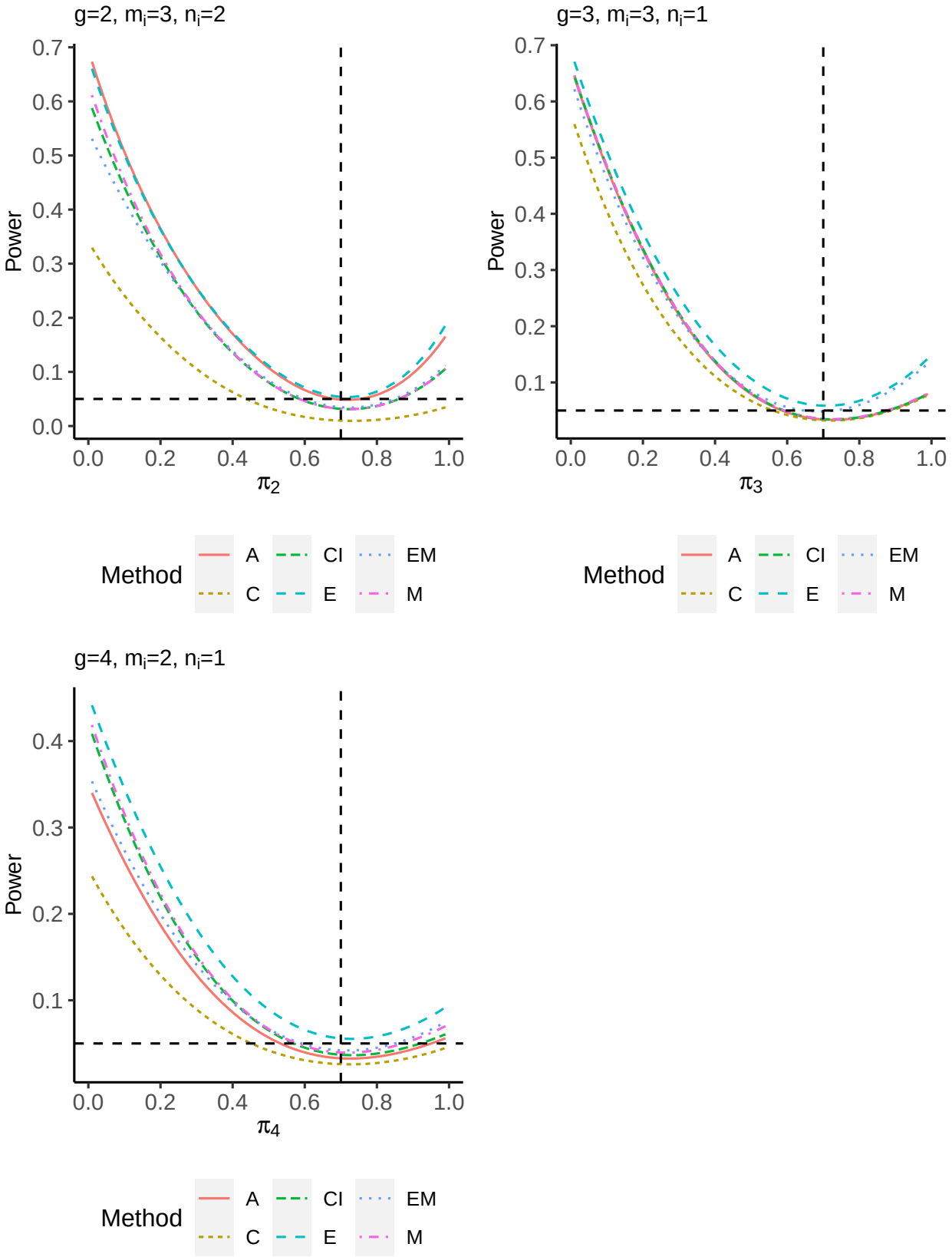


Figure 8. Power plots for $\pi_1 = ... = \pi_{g-1} = 0.7$, and $\rho = 0.4$.

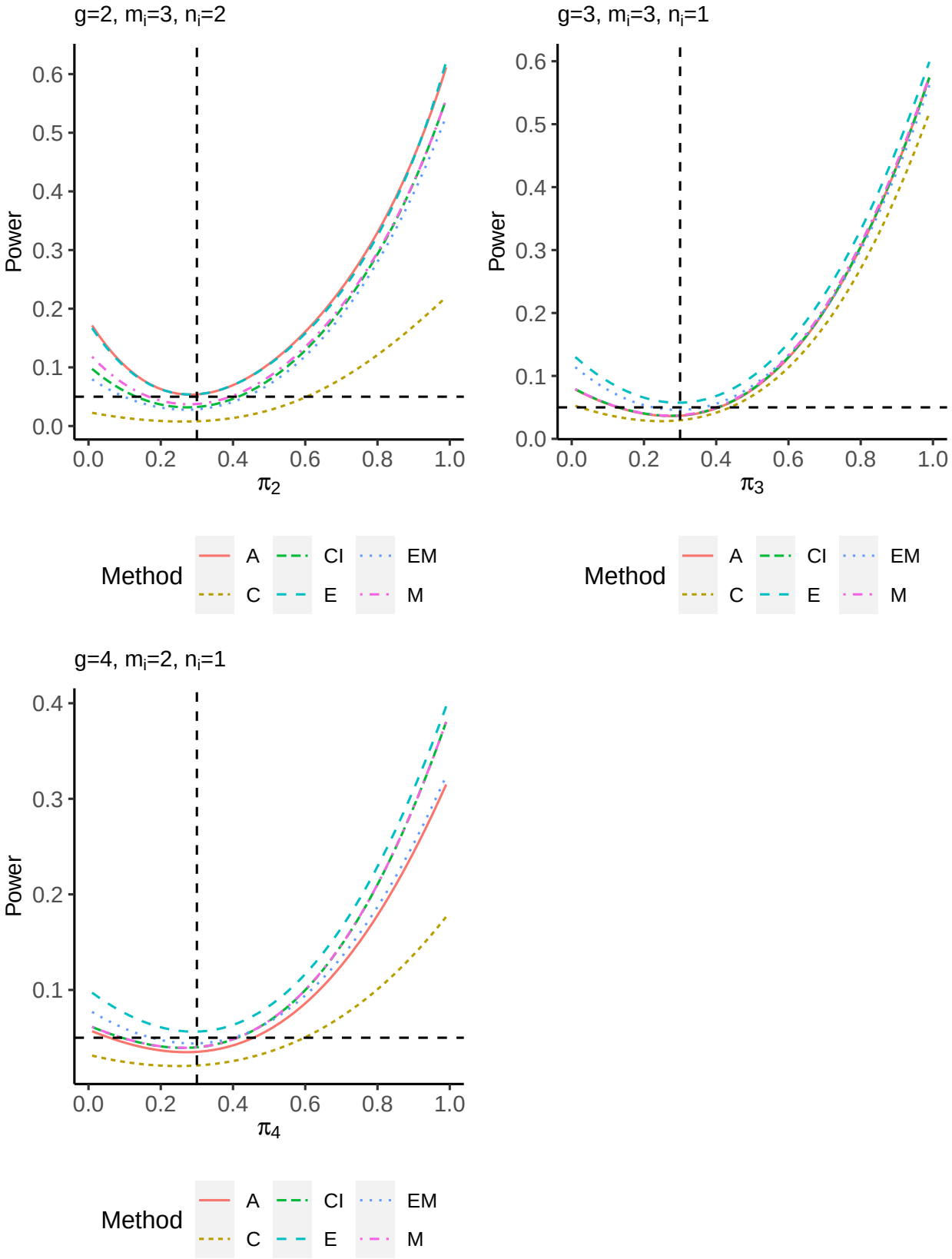


Figure 9. Power plots for $\pi_1 = ... = \pi_{g-1} = 0.3$, and $\rho = 0.6$.

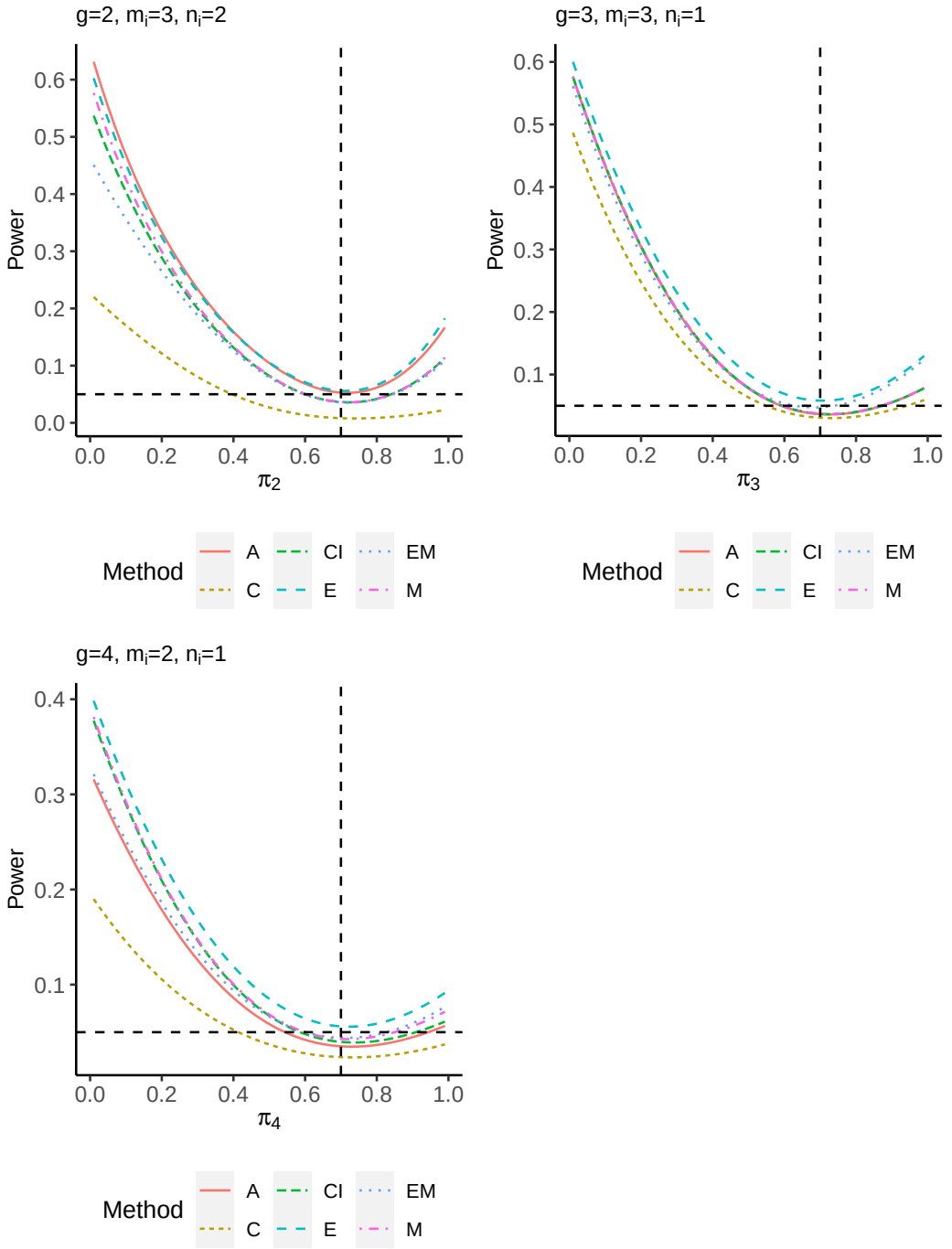


Figure 10. Power plots for $\pi_1 = ... = \pi_{g-1} = 0.7$, and $\rho = 0.6$.

7. Real Examples

Three real-world examples are described in detail to demonstrate the application of the proposed methods. The first example is a double-blind randomized clinical trial studying acute otitis media with effusion (OME) as described by Mandel [19]. The total number of children recruited in the study was 214, and there were two hundred and ninety-three ears registered in the study among these children. Every participant was categorized into either the bilateral disease or unilateral disease group at the study entry. All subjects underwent either unilateral or bilateral tympanocentesis and were subsequently randomized into one of two treatment groups, cefaclor or amoxicillin, for 14 days. After the course of treatment, the disease status of each ear was recorded, and 11 children were dropped from the study

since they met the exclusion criteria. Table 3 displays the distribution of a subgroup of children aged ≥ 6 years. To test the homogeneity of cure rates of the two treatments, the six methods discussed in Section 5 were applied, and the corresponding p-values can be found in Table 4. All p-values are greater than 0.05, suggesting a lack of evidence to reject the null hypothesis at $\alpha = 5\%$.

Table 3. The distribution of children aged ≥ 6 years at 14 days by treatment, number of cured ears, and disease status at study entry.

Treatment	Bilateral at Entry			Unilateral at Entry	
	No. of Cured Ears			No. of Cured Ears	
	0	1	2	0	1
Cefaclor	0	1	3	8	11
Amoxicillin	1	0	6	7	11

Table 4. p-values using different approaches for example 1.

Approach	p-Value
A	0.6629
E	0.6712
M	0.7218
E + M	0.8608
C	0.7667
CI	0.7197

The second example consists of 60 subjects receiving Orthokeratology (Ortho-k) to treat nearsightedness. This is an observational study conducted at the First Affiliated Hospital of Xiamen University in 2023 (Liang et al. [20]). Ortho-k is a vision correction procedure that uses specially designed contact lenses to temporarily reshape the cornea and improve vision. The study involved various brands that can be further classified into two categories based on the type of lens designs. Vision shaping treatment (VST) is one of the two designs constructing the edge of the lens that touches the edge of the cornea, while the other design, called corneal refractive therapy (CRT), constructs the edge of the lens so that it does not touch the cornea (Lu et al. [21]). Patients can opt to wear the lenses on either one or both eyes, depending on their individual requirements. An eye is said to have a response to the treatment if the axial length growth < 0.3 mm according to Rose et al. [22]. A subgroup of female subjects were included in the demonstration to test whether there was a difference between the response rates of the two designs. There were 3 subjects who received unilateral Ortho-k and 26 subjects who underwent bilateral Ortho-k, as indicated in Table 5. The p-values from Table 6 are all less than the nominal level of 5%, suggesting a rejection of the null hypothesis.

Table 5. The number of subjects by design and number of eyes demonstrating response after treatment.

Design	Bilateral			Unilateral	
	No. of Eyes with Response			No. of Eyes with Response	
	0	1	2	0	1
VST	9	3	7	2	1
CRT	7	0	0	0	0

Table 6. *p*-values using different approaches for example 2.

Approach	*p*-Value
A	0.0188
E	0.0171
M	0.0220
E + M	0.0249
C	0.0265
CI	0.0219

The data of the third example were collected from patients with retinitis pigmentosa referred from the outpatient facilities of the Massachusetts Eye and Ear Infirmary or from private ophthalmologists. Details of the study can be found in Berson, Rosner, and Simonoff [23]. The patients were classified into four groups based on their genetic types, which are autosomal dominant RP (DOM), autosomal recessive RP (AR), sex-linked RP (SL), and isolate RP (ISO). The type ISO was dropped from the analysis since the sample size of this group is large and asymptotic methods are preferred. The distribution of patients can be found in Table 7. This example demonstrates the capability of the proposed methods in handling bilateral data. The calculation time of the E + M method is excessively long; therefore, the corresponding *p*-value is omitted. All the *p*-values of the proposed methods displayed in Table 8 are less than 0.05, indicating a rejection of the null hypothesis, and the prevalences of the three groups are not the same.

Table 7. The number of patients by genetic type and number of affected eyes.

Design	Bilateral			Unilateral	
	No. of Eyes with Response			No. of Eyes with Response	
	0	1	2	0	1
DOM	15	6	7	0	0
AR	7	5	9	0	0
SL	3	2	14	0	0

Table 8. *p*-values using different approaches for example 3.

Approach	*p*-Value
A	0.0048
E	0.0042
M	0.0044
E + M	NA
C	0.0045
CI	0.0047

8. Discussion

The individual site model assumes the measurements from sites of individuals are independent and does not consider the potential correlation between two sites of the same subject. Given prior knowledge of the nonexistence of the correlation, an investigator may choose to proceed with the individual site model and perform the Pearson chi-squared test or exact tests without doubt. However, it is often the case that there is no proven fact of such an assumption. As the correlation moves further from zero, the asymptotic and exact methods lose the type I error controls based on the numerical studies in Section 6. One may argue that the individual site model may carry out the hypothesis test using an appropriate method with acceptable TIEs falling in the interval of $[0.04, 0.06]$. It is important to note that the equal correlation coefficients model also produces satisfactory TIE controls with a suitable approach. Therefore, the model taking into account the correlation takes precedence if the data collection involves paired sites of the same subject.

Given the priority of the equal correlation coefficients model, testing the homogeneity of group-specific proportions using the score test can be the first candidate method that re-

searchers consider. Nevertheless, the asymptotic behavior of the score test is not guaranteed if the sample size is small. As a consequence, five exact tests are proposed, and their performance in terms of TIE controls and statistical powers was examined in different scenarios. When the sample size is relatively small as seen in Figure 1–3, the E method generally performs well in controlling TIEs within 4% to 6%, while other methods are conservative with larger proportions in the contour plots below 4%. This superiority of the E method can be found in Figure A1 and A2 where the sample size is relatively large. Regarding statistical powers, the E method is superior or similar to other methods for all sample sizes and parameter settings. For example, the E method has similar powers to the asymptotic method when $g = 2$ and differentiates from other methods when $g = 3$ or 4 in Figure 7. Overall, the E method is recommended as it controls type I error in a satisfactory region and has higher powers than the other methods. The M, E + M, and CI methods generally produce comparable TIEs and powers, whereas the C method becomes the most conservative specifically when the total sample size is approximately 10. Future works will include validation of the equal correlation coefficient assumption before performing the homogeneity test.

9. Conclusions

There are two aims of this article: first, to investigate and compare the two models with and without the assumption of correlation coefficients, and second, to propose exact methods to address the lack of type I error controls caused by poor approximations of the score test when the sample size is small. The numerical study indicates the relative appropriateness of the equal correlation model when analyzing paired binary data and highlights the superiority of the E method for exact tests of homogeneity of proportions.

Author Contributions: All authors contribute equally to the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

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

Appendix A

Let π_0 and ρ_0 be the true parameters with the null hypothesis H_0 and note that $Prob(T_{SC}(M) \geq T_{SC}(M^*)|\pi_0, \rho_0)$ is uniformly distributed on [0, 1] with H_0. Following the definition of $P_{CI}(M^*)$, it is clear that

$$\begin{aligned}
Prob(P_{CI}(M^*) \leq \alpha|H_0) =& Prob(P_{CI}(M^*) \leq \alpha, \pi_0 \in CI_\pi, \rho_0 \in CI_\rho|H_0) \\
&+ Prob(P_{CI}(M^*) \leq \alpha, \pi_0 \in CI_\pi, \rho_0 \notin CI_\rho|H_0) \\
&+ Prob(P_{CI}(M^*) \leq \alpha, \pi_0 \notin CI_\pi, \rho_0 \in CI_\rho|H_0) \\
&+ Prob(P_{CI}(M^*) \leq \alpha, \pi_0 \notin CI_\pi, \rho_0 \notin CI_\rho|H_0) \\
\leq& Prob(P_{CI}(M^*) \leq \alpha, \pi_0 \in CI_\pi, \rho_0 \in CI_\rho|H_0) + Prob(\rho_0 \notin CI_\rho|H_0) \\
&+ Prob(\pi_0 \notin CI_\pi|H_0) + Prob(\pi_0 \notin CI_\pi|H_0) \\
=& Prob\left[\sup_{\pi \in CI_\pi, \rho \in CI_\rho}\left[\sum_{M \in \Omega_1(M^*)} I(T_{SC}(M) \geq T_{SC}(M^*))\mathcal{L}(M;\pi,\rho)\right] + 3\beta \leq \alpha,\right. \\
&\left.\pi_0 \in CI_\pi, \rho_0 \in CI_\rho|H_0\right] \\
&+ Prob(\rho_0 \notin CI_\rho|H_0) + 2 \times Prob(\pi_0 \notin CI_\pi|H_0) \\
\leq& Prob(Prob(T_{SC}(M) \geq T_{SC}(M^*)|\pi_0, \rho_0) + 3\beta \leq \alpha, \pi_0 \in CI_\pi, \rho_0 \in CI_\rho|H_0) \\
&+ Prob(\rho_0 \notin CI_\rho|H_0) + 2 \times Prob(\pi_0 \notin CI_\pi|H_0) \\
\leq& Prob(Prob(T_{SC}(M) \geq T_{SC}(M^*)|\pi_0, \rho_0) \leq \alpha - 3\beta|H_0) + 3\beta \\
=& \alpha - 3\beta + 3\beta \\
=& \alpha,
\end{aligned}$$

Appendix B

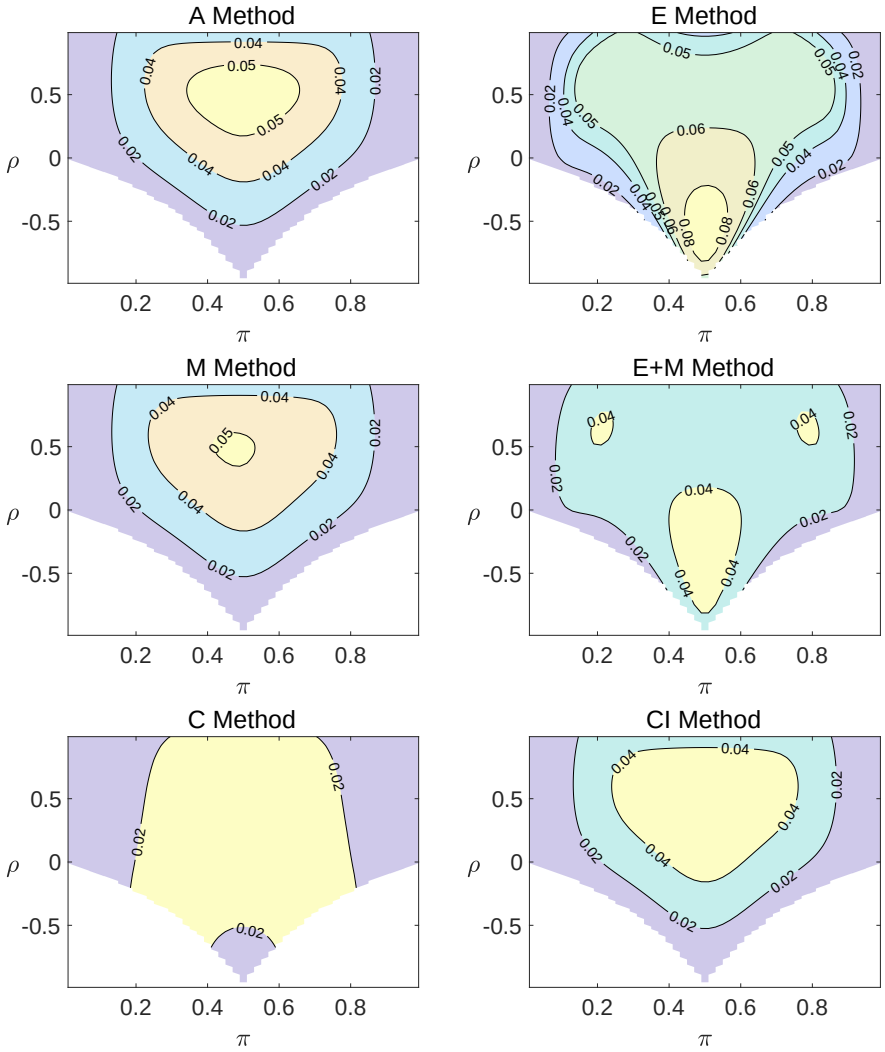


Figure A1. Contour plots of type I errors for the equal correlation coefficients model ($g = 2$, $m_i = 5$, $n_i = 5$, $i = 1, 2$).

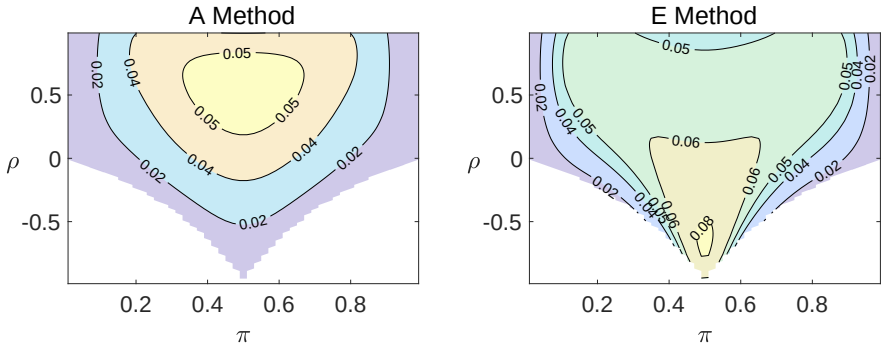


Figure A2. *Cont.*

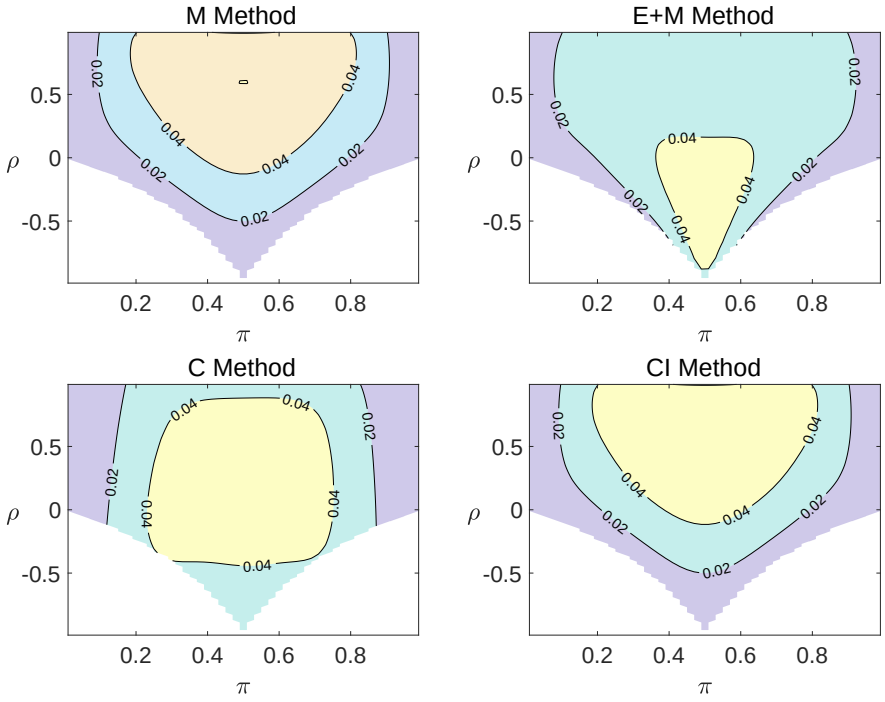


Figure A2. Contour plots of type I errors for the equal correlation coefficients model ($g = 3$, $m_i = 4$, $n_i = 3$, $i = 1, 2, 3$).

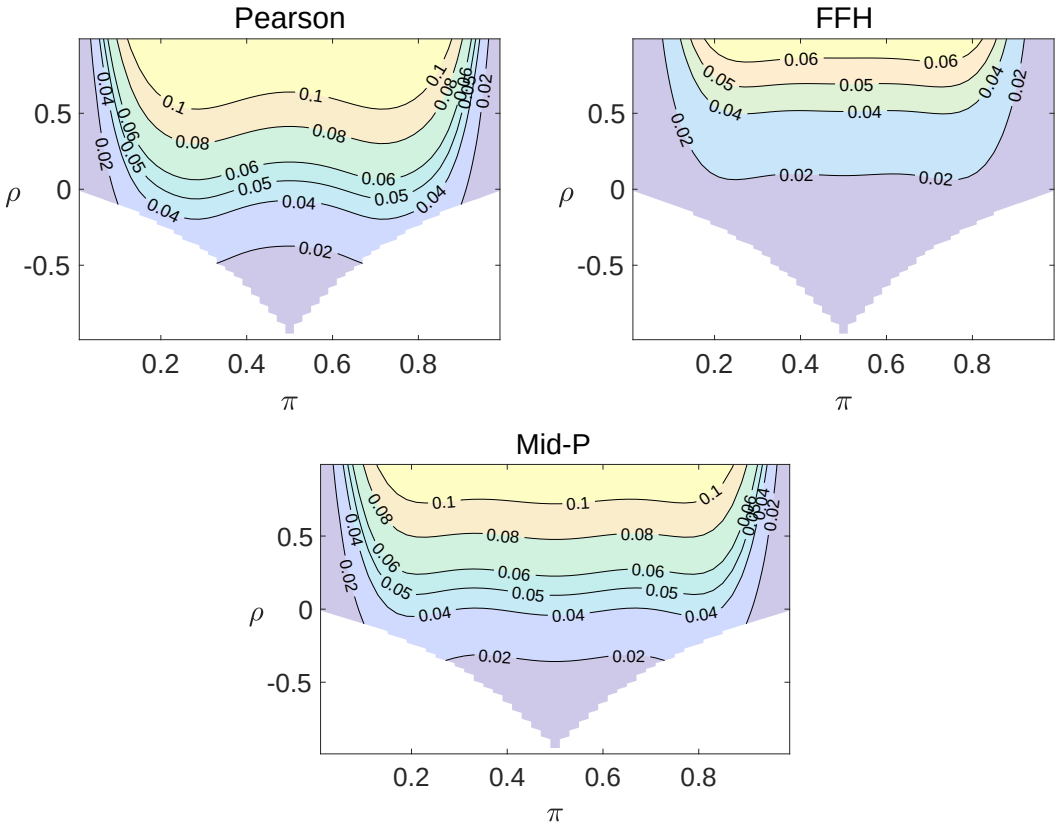


Figure A3. Contour plots of type I errors for the individual site model ($g = 2$, $m_i = 5$, $n_i = 5$, $i = 1, 2$).

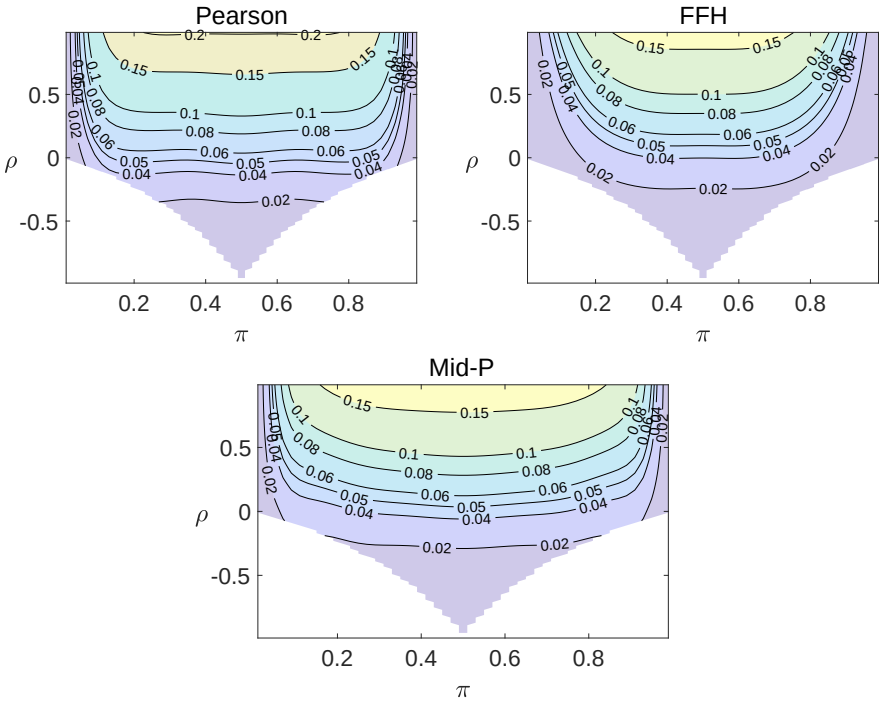


Figure A4. Contour plots of type I errors for the individual site model ($g = 3$, $m_i = 4$, $n_i = 3$, $i = 1, 2, 3$).

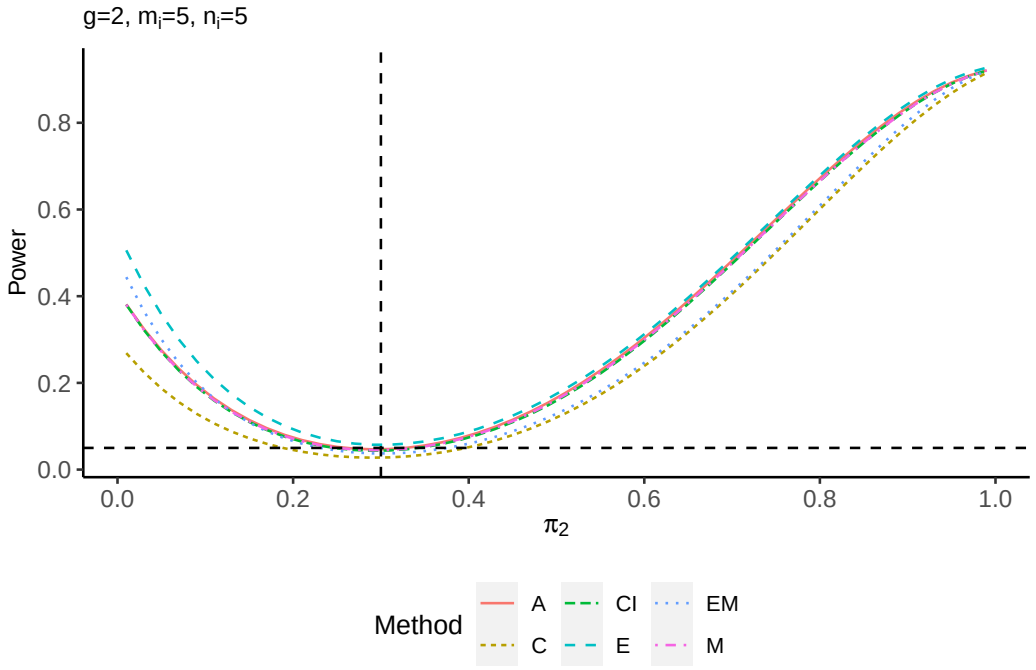


Figure A5. *Cont.*

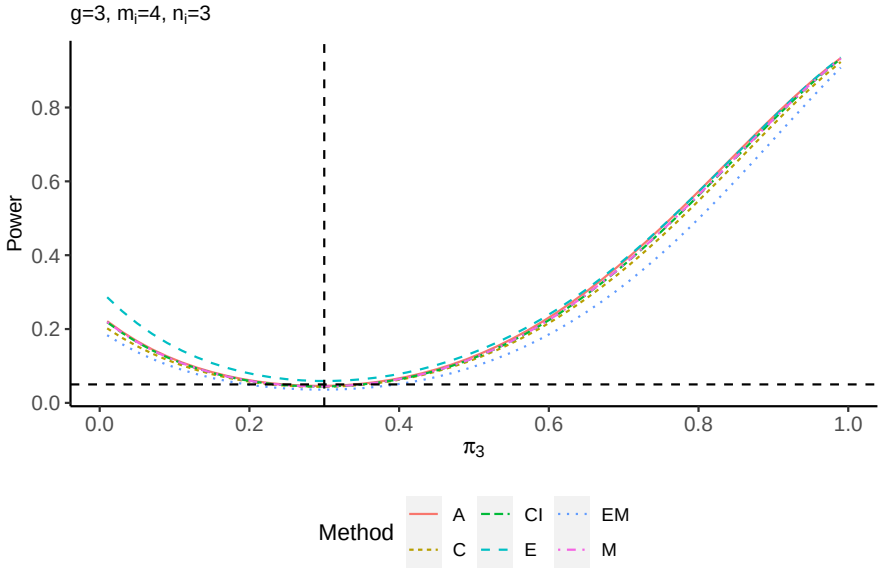


Figure A5. Power plots for $\pi_1 = ... = \pi_{g-1} = 0.3$, and $\rho = 0.4$.

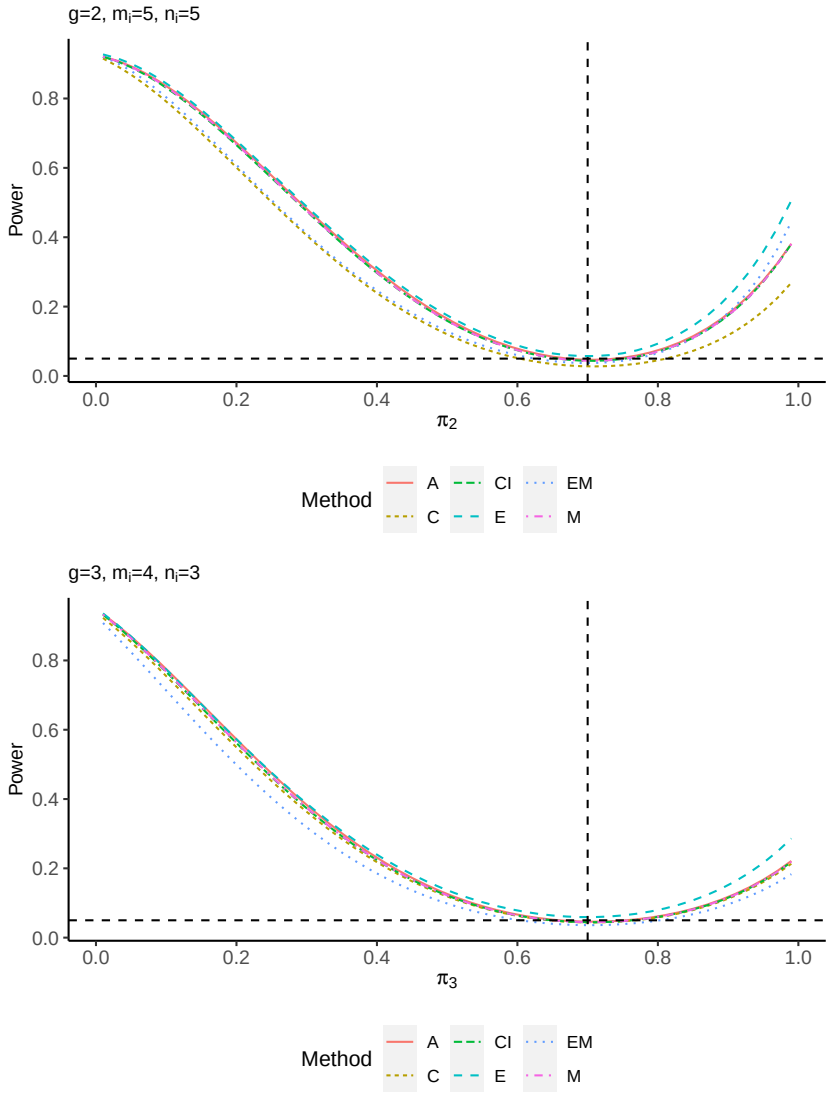


Figure A6. Power plots for $\pi_1 = ... = \pi_{g-1} = 0.7$, and $\rho = 0.4$.

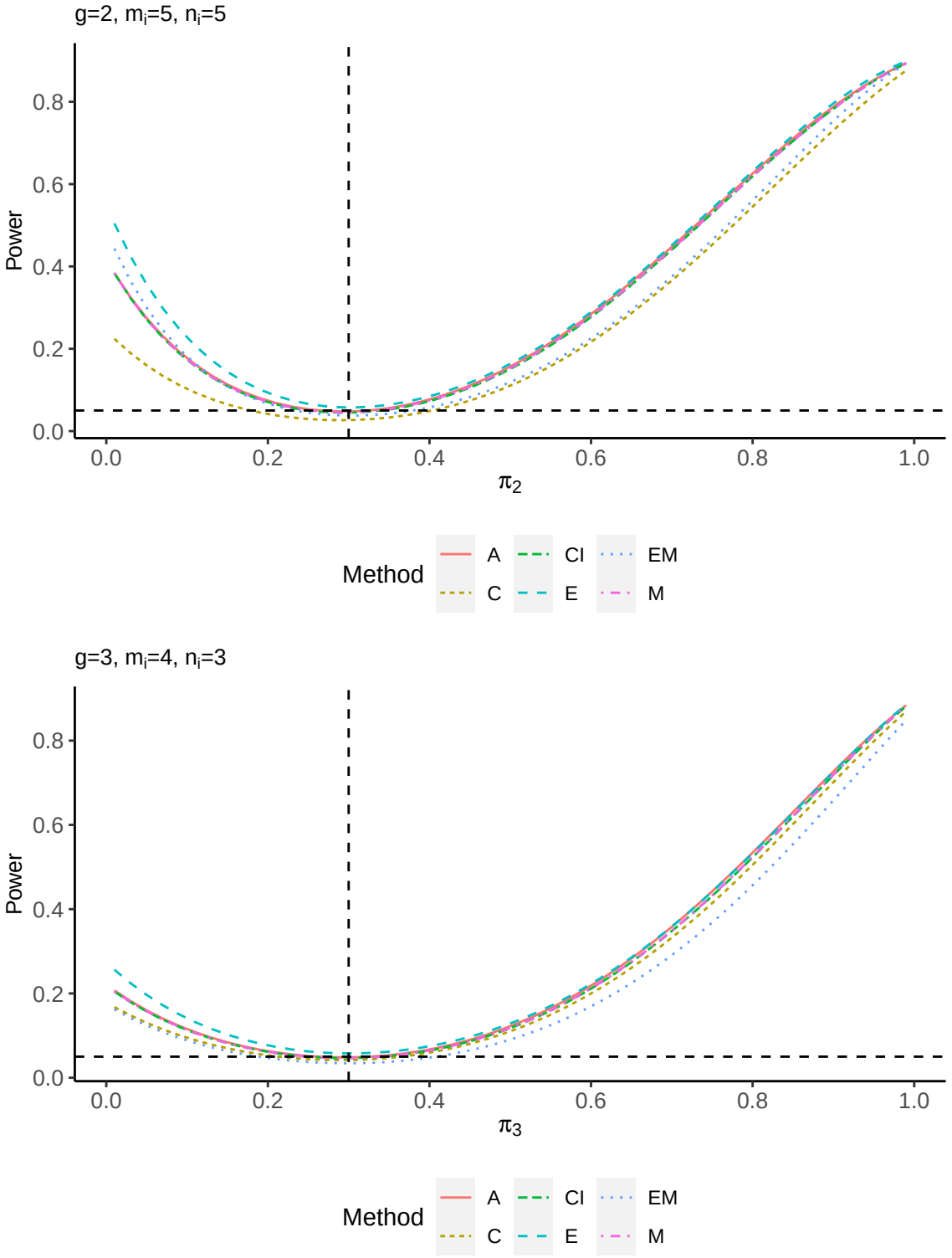


Figure A7. Power plots for $\pi_1 = ... = \pi_{g-1} = 0.3$, and $\rho = 0.6$.

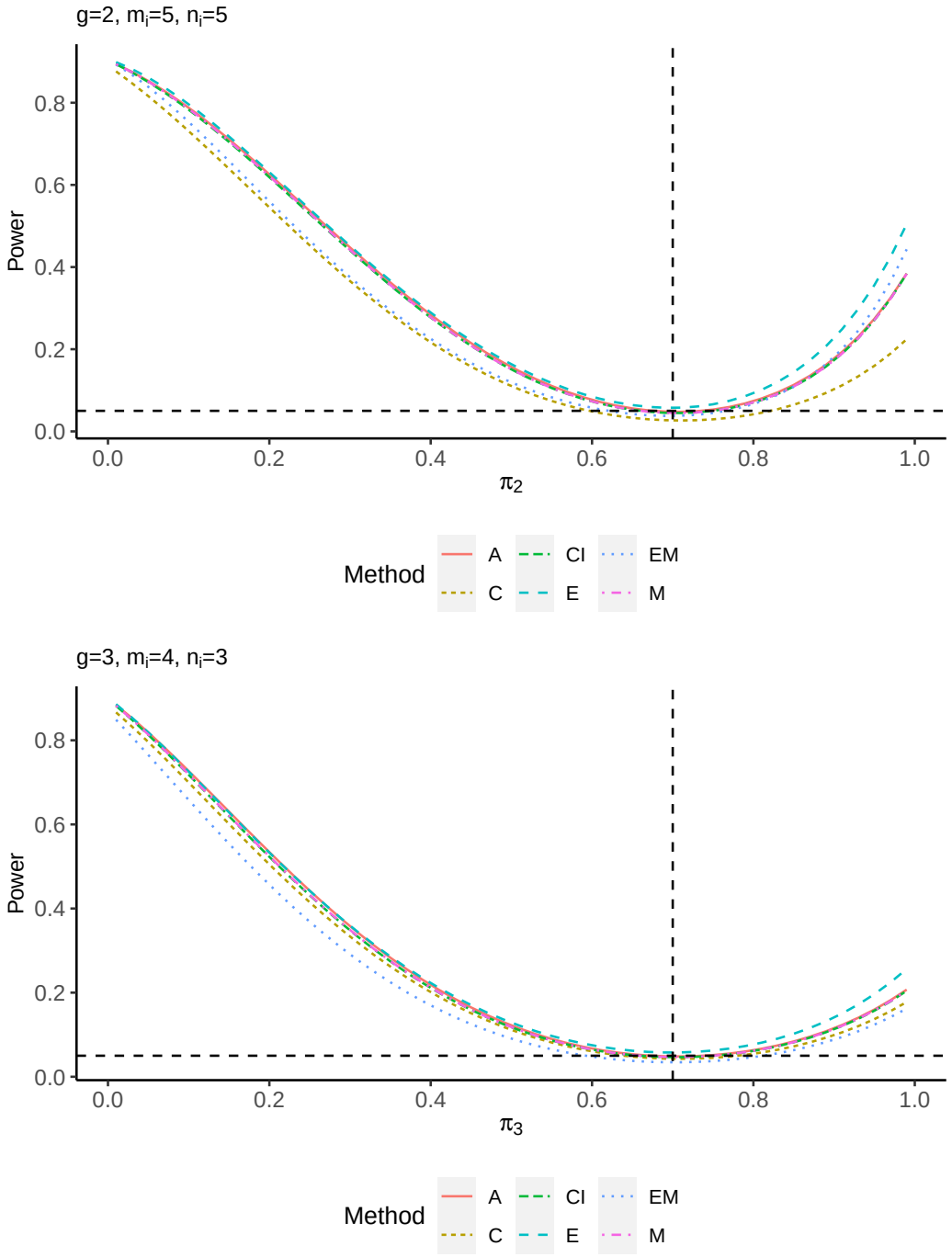


Figure A8. Power plots for $\pi_1 = ... = \pi_{g-1} = 0.7$, and $\rho = 0.6$.

References

1. Zhang, H.G.; Ying, G.s. Statistical approaches in published ophthalmic clinical science papers: A comparison to statistical practice two decades ago. *Br. J. Ophthalmol.* **2018**, *102*, 1188–1191. [CrossRef] [PubMed]
2. Pearson, K. X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *London, Edinburgh, Dublin Philos. Mag. J. Sci.* **1900**, *50*, 157–175. [CrossRef]
3. Rosner, B. Statistical methods in ophthalmology: An adjustment for the intraclass correlation between eyes. *Biometrics* **1982**, *38*, 105–114. [CrossRef] [PubMed]
4. Donner, A.; Banting, D. Analysis of site-specific data in dental studies. *J. Dent. Res.* **1988**, *67*, 1392–1395. [CrossRef] [PubMed]
5. Donner, A. Statistical methods in ophthalmology: An adjusted chi-square approach. *Biometrics* **1989**, *45*, 605–611. [CrossRef] [PubMed]

6. Ma, C.X.; Liu, S. Testing equality of proportions for correlated binary data in ophthalmologic studies. *J. Biopharm. Stat.* **2017**, *27*, 611–619. [CrossRef] [PubMed]
7. Mou, K.; Li, Z. Homogeneity Test of Many-to-One Risk Differences for Correlated Binary Data under Optimal Algorithms. *Complexity* **2021**, *2021*, 6685951. [CrossRef]
8. Liu, X.; Yang, Z.; Liu, S.; Ma, C.X. Exact methods of testing the homogeneity of prevalences for correlated binary data. *J. Stat. Comput. Simul.* **2017**, *87*, 3021–3039. [CrossRef]
9. Liddell, D. Practical Tests of 2Times2 Contingency Tables. *J. R. Stat. Soc. Ser. D (Stat.)* **1976**, *25*, 295–304.
10. Basu, D. On the Elimination of Nuisance Parameters. *J. Am. Stat. Assoc.* **1977**, *72*, 279–290. [CrossRef]
11. Lloyd, C.J. Exact *P*-Values Discret. Model. Obtained Estim. Maximization. *Aust. N. Z. J. Stat.* **2008**, *50*, 329–345. [CrossRef]
12. Ma, C.X.; Wang, H. Testing the equality of proportions for combined unilateral and bilateral data under equal intraclass correlation model. *Stat. Biopharm. Res.* **2022**, *15*, 608–617. [CrossRef]
13. Fagerland, M.; Lydersen, S.; Laake, P. *Statistical Analysis of Contingency Tables*; CRC Press: Boca Raton, FL, USA, 2017.
14. Rao, C.R. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Math. Proc. Camb. Philos. Soc.* **1948**, *44*, 50–57.
15. Berger, R.L.; Boos, D.D. P values maximized over a confidence set for the nuisance parameter. *J. Am. Stat. Assoc.* **1994**, *89*, 1012–1016. [CrossRef]
16. Silvapulle, M.J. A test in the presence of nuisance parameters. *J. Am. Stat. Assoc.* **1996**, *91*, 1690–1693. [CrossRef]
17. Vexler, A. Valid *P*-Values Expect. *p*-Values Revisited. *Ann. Inst. Stat. Math.* **2021**, *73*, 227–248. [CrossRef]
18. 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]
19. Mandel, E.M.; Bluestone, C.D.; Rockette, H.E.; BLATTER, M.M.; Reisinger, K.S.; Wucher, F.P.; Harper, J. Duration of effusion after antibiotic treatment for acute otitis media: Comparison of cefaclor and amoxicillin. *Pediatr. Infect. Dis. J.* **1982**, *1*, 310–316. [CrossRef]
20. Liang, S.; Fang, K.T.; Huang, X.W.; Xin, Y.; Ma, C. Homogeneity Tests and Interval Estimations of Risk Differences for Stratified Bilateral and Unilateral Correlated Data. *arXiv* **2023**, arXiv:2304.00162.
21. Lu, W.; Ning, R.; Diao, K.; Ding, Y.; Chen, R.; Zhou, L.; Lian, Y.; McAlinden, C.; Sanders, F.W.; Xia, F.; et al. Comparison of two main orthokeratology lens designs in efficacy and safety for myopia control. *Front. Med.* **2022**, *9*, 798314. [CrossRef]
22. Rose, L.V.; Schulz, A.M.; Graham, S.L. Use baseline axial length measurements in myopic patients to predict the control of myopia with and without atropine 0.01%. *PLoS ONE* **2021**, *16*, e0254061. [CrossRef] [PubMed]
23. Berson, E.L.; Rosner, B.; Simonoff, E. Risk factors for genetic typing and detection in retinitis pigmentosa. *Am. J. Ophthalmol.* **1980**, *89*, 763–775. [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.