



Article

Asymptotic Sample Size for Common Test of Relative Risk Ratios in Stratified Bilateral Data

Keyi Mou ¹, Zhiming Li ^{1,*}  and Changxing Ma ² 

¹ College of Mathematics and System Sciences, Xinjiang University, Urumqi 830000, China; 107556521095@stu.xju.edu.cn

² Department of Biostatistics, University at Buffalo, Buffalo, NY 14214, USA; cxma@buffalo.edu

* Correspondence: zml@xju.edu.cn

Abstract: In medical clinical studies, various tests usually relate to the sample size. This paper proposes several methods to calculate sample sizes for a common test of relative risk ratios in stratified bilateral data. Under the prespecified significant level and power, we derive some explicit formulae and an algorithm of the sample size. The sample sizes of the stratified intra-class model are obtained from the likelihood ratio, score, and Wald-type tests. Under pooled data, we calculate sample size based on the Wald-type test and its log-transformation form. Numerical simulations show that the proposed sample sizes have empirical power close to the prespecified value for given significance levels. The sample sizes from the iterative method are more stable and effective.

Keywords: paired data; stratified design; relative risk ratio; sample size

MSC: 62F25



Citation: Mou, K.; Li, Z.; Ma, C. Asymptotic Sample Size for Common Test of Relative Risk Ratios in Stratified Bilateral Data. *Mathematics* **2023**, *11*, 4198. <https://doi.org/10.3390/math11194198>

Academic Editor: Stelios Psarakis

Received: 2 September 2023

Revised: 29 September 2023

Accepted: 5 October 2023

Published: 8 October 2023



Copyright: © 2023 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 medical clinical studies, we often encounter observations from patients' paired organs or parts (e.g., eyes, ears, and arms). The observations are often binary, that is, response or no response. When each patient in the study contributes such paired measurements, the overall outcome can be bilateral, unilateral, or no response(s). We call this type of data bilateral data. A correlation always exists between paired parts' responses. Several intra-class correlation models have been proposed for analyzing such paired data under a probability model [1–3]. Rosner [1] introduced a constant R model with the assumption that the conditional probability of a response at one side of the paired body parts given response at the other side was R times the unconditional probability. Donner [3] proposed a model by assuming that the correlation coefficient was a fixed constant ρ in each group. Based on the three models mentioned above, there have been many achievements, such as asymptotic tests [4–8], exact tests [9–11], and confidence intervals [12,13]. However, comparatively little work has been done on the sample size problem of bilateral data.

In practice, sample size is one of the essential factors in designing clinical trials. Up to now, a considerable body of literature has grown up relating to binary data. For example, Tang et al. [8] proposed a power-controlled sample size, which can guarantee a desired power and control type I error rates (TIEs) at the fixed significance level. Qiu et al. [14] studied sample sizes for common tests of disease prevalence. Further, Qiu et al. [15] used an iterative algorithm to calculate the asymptotic sample sizes of common tests for binomial proportions. Sun et al. [16] extended the iterative algorithm to the homogeneity test in stratified unilateral and bilateral data. For more detailed results of sample sizes, we refer to [17–19]. Through the above analysis, there are mainly two approaches: sample size formula and algorithms. The former is practical for some explicit test statistics, while the latter often handles complex cases. However, there is little research to evaluate these methods' performance.

Under Donner’s model, Zhuang et al. [20] proposed several statistics for testing common relative risk ratios across J strata. In this paper, we extend the hypotheses to the general case. Thus, their results are a special case of our results. What is more important, our work focuses on calculating the sample sizes of various test statistics, while Zhuang et al. [20] did not take it into account. The novelty and contribution of our work are shown by three main aspects as follows: (i) the common hypotheses mentioned above are extended to the general case $H_0 : \delta = \delta_0$ versus $H_a : \delta \neq \delta_0$, where $\delta_0 \in (0, \infty)$. (ii) Under the dependent and independent models, the approximate equations and algorithms of sample size are derived at prespecified power and significance levels. (iii) Numerical simulations are conducted to compare these methods.

The rest of the work is organized as follows. Section 2 reviews data structure and establishes Donner’s model. Several test statistics are derived for a common test of relative risk ratios. Based on these statistics, sample size calculations are proposed in Section 3. In Section 4, simulations show the performance of eight asymptotic sample sizes in terms of empirical power and TIEs. We also compare estimated values with actual sample sizes. A real example is provided to illustrate the proposed methods in Section 5. Conclusions are given in Section 6.

2. Donner’s Model and Common Test

Let M be the total number of patients for a stratified bilateral design. There are M_j ($j = 1, 2, \dots, J$) patients in the j th stratum. Thus, $M = \sum_{j=1}^J M_j$. Let m_{lij} be the number of patients with l responses ($l = 0, 1, 2$) in the i th group ($i = 1, 2$) of the j th stratum ($j = 1, 2, \dots, J$), and $N_{ij} = \sum_{l=0}^2 m_{lij}$ be a fixed total number of patients in the i th group of the j th stratum. Let p_{lij} ($l = 0, 1, 2, i = 1, 2, j = 1, \dots, J$) be the probabilities of no, unilateral, and bilateral response(s) in the i th group of the j th stratum, where $p_{0ij} + p_{1ij} + p_{2ij} = 1$ for any fixed i and j . The data structure of the j th stratum is shown in Table 1.

Table 1. Data structure in the j th stratum.

Number of Responses (l)	Group (i)		Total
	1	2	
0	m_{01j} (p_{01j})	m_{02j} (p_{02j})	S_{0j}
1	m_{11j} (p_{11j})	m_{12j} (p_{12j})	S_{1j}
2	m_{21j} (p_{21j})	m_{22j} (p_{22j})	S_{2j}
Total	N_{1j}	N_{2j}	M_j

Denote $\mathbf{m}_{ij} = (m_{0ij}, m_{1ij}, m_{2ij})^T$ and $\mathbf{p}_{ij} = (p_{0ij}, p_{1ij}, p_{2ij})^T$ ($i = 1, 2, j = 1, \dots, J$). For the j th stratum, \mathbf{m}_{ij} ($i = 1, 2$) follows a trinomial distribution. Thus, its probability density is expressed as follows:

$$f(\mathbf{m}_{ij} | \mathbf{p}_{ij}) = \frac{N_{ij}!}{m_{0ij}!m_{1ij}!m_{2ij}!} p_{0ij}^{m_{0ij}} p_{1ij}^{m_{1ij}} p_{2ij}^{m_{2ij}}.$$

Let Z_{hijk} be an indicator of the k th organ’s response ($k = 1, 2$) for the h th patient ($h = 1, 2, \dots, N_{ij}$) in the i th group from the j th stratum. If there is a response, $Z_{hijk} = 1$, and 0 otherwise. Under Donner’s model, $\Pr(Z_{hijk} = 1) = \pi_{ij}$ ($0 \leq \pi_{ij} \leq 1$), and $\text{Corr}(Z_{hijk}, Z_{hij(3-k)}) = \rho_j$ ($0 \leq \rho_j \leq 1$). The correlation coefficient ρ_j ($j = 1, 2, \dots, J$) is constant in each stratum. The organ responses are affected by other factors such as group (e.g., treatments) and stratum (e.g., age, sex) factors. Through a simple calculation, the probabilities p_{lij} can be obtained by $p_{0ij} = \rho_j(1 - \pi_{ij}) + (1 - \rho_j)(1 - \pi_{ij})^2$, $p_{1ij} = 2\pi_{ij}(1 - \rho_j)(1 - \pi_{ij})$, $p_{2ij} = \rho_j\pi_{ij} + (1 - \rho_j)\pi_{ij}^2$ for $i = 1, 2, j = 1, \dots, J$.

Let $\delta_j = \pi_{2j} / \pi_{1j}$ ($j = 1, \dots, J$) be the relative risk ratio between two groups in the j th stratum. Denote $\boldsymbol{\pi}_i = (\pi_{i1}, \dots, \pi_{ij})^T$ for $i = 1, 2$, and $\boldsymbol{\rho} = (\rho_1, \dots, \rho_J)^T$. Suppose that the

relative risk ratios from all strata are equal, that is, $\delta_1 = \delta_2 = \dots = \delta_J \triangleq \delta$. The effects of strata include the correlation coefficients ρ_j , probabilities $p_{lij} (l = 0, 1, 2)$, sample sizes $M_j (j = 1, 2, \dots, J)$, and ratios N_{2j}/N_{1j} between two groups. Based on the observed data $\mathbf{m}_{ij} (i = 1, 2, j = 1, 2, \dots, J)$, the log-likelihood function can be written as

$$\ell(\boldsymbol{\pi}_1, \delta, \boldsymbol{\rho}) = \log \left(\prod_{j=1}^J \prod_{i=1}^2 f(\mathbf{m}_{ij} | \mathbf{p}_{ij}) \right) = \sum_{j=1}^J l_j(\pi_{1j}, \delta, \rho_j) + \log C, \tag{1}$$

where $C = \prod_{j=1}^J \prod_{i=1}^2 \frac{N_{ij}!}{m_{0ij}!m_{1ij}!m_{2ij}!}$ and

$$\begin{aligned} l_j(\pi_{1j}, \delta, \rho_j) = & m_{01j} \log[\rho_j(1 - \pi_{1j}) + (1 - \rho_j)(1 - \pi_{1j})^2] \\ & + m_{11j} \log[2\pi_{1j}(1 - \rho_j)(1 - \pi_{1j})] + m_{21j} \log[\rho_j\pi_{1j} + (1 - \rho_j)\pi_{1j}^2] \\ & + m_{02j} \log[\rho_j(1 - \pi_{1j}\delta) + (1 - \rho_j)(1 - \pi_{1j}\delta)^2] \\ & + m_{12j} \log[2\pi_{1j}\delta(1 - \rho_j)(1 - \pi_{1j}\delta)] + m_{22j} \log[\rho_j\pi_{1j}\delta + (1 - \rho_j)(\pi_{1j}\delta)^2]. \end{aligned}$$

Under the assumption that relative risk ratios from all strata are equal, we propose the hypotheses of the common test as

$$H_0 : \delta = \delta_0 \text{ versus } H_a : \delta \neq \delta_0, \tag{2}$$

where $\delta_0 \in (0, \infty)$. Under H_0 , that is $\pi_{2j} = \pi_{1j}\delta_0 (j = 1, 2, \dots, J)$, the log-likelihood function can be rewritten as

$$\ell_0(\boldsymbol{\pi}_1, \delta_0, \boldsymbol{\rho}) = \sum_{j=1}^J l_{0j}(\pi_{1j}, \delta_0, \rho_j) + \log C, \tag{3}$$

where l_{0j} is written in Appendix A. Next, we introduce several asymptotic statistics to test the hypotheses (2) and solve the sample size calculations for the prespecified power and significance level based on these statistics.

Let $\hat{\boldsymbol{\pi}}_1 = (\hat{\pi}_{11}, \hat{\pi}_{12}, \dots, \hat{\pi}_{1J})^T, \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\rho}} = (\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_J)^T$ be the global maximum likelihood estimations (MLEs) of $\boldsymbol{\pi}_1, \delta, \boldsymbol{\rho}$, and $\tilde{\boldsymbol{\pi}}_1 = (\tilde{\pi}_{11}, \tilde{\pi}_{12}, \dots, \tilde{\pi}_{1J})^T, \tilde{\boldsymbol{\rho}} = (\tilde{\rho}_1, \tilde{\rho}_2, \dots, \tilde{\rho}_J)^T$ be the constrained MLEs of $\boldsymbol{\pi}_1, \boldsymbol{\rho}$ under H_0 . The global MLEs are the solutions of the following equations

$$\frac{\partial \ell}{\partial \pi_{1j}} = 0, \quad \frac{\partial \ell}{\partial \delta} = 0, \quad \frac{\partial \ell}{\partial \rho_j} = 0, \quad j = 1, 2, \dots, J. \tag{4}$$

The constrained MLEs are the solutions of the following equations

$$\frac{\partial \ell_0}{\partial \pi_{1j}} = 0, \quad \frac{\partial \ell_0}{\partial \rho_j} = 0, \quad j = 1, 2, \dots, J. \tag{5}$$

However, there are no closed-form solutions of Equations (4) and (5). Following Zhuang et al. [20], the global and constrained MLEs under $\delta = \delta_0$ can be obtained through the Fisher scoring algorithm.

2.1. Likelihood Ratio Test

Based on the log-likelihood functions (1) and (3), the likelihood ratio test is defined as

$$T_L = 2[\ell(\hat{\boldsymbol{\pi}}_1, \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\rho}}) - \ell_0(\tilde{\boldsymbol{\pi}}_1, \delta_0, \tilde{\boldsymbol{\rho}})] = 2 \sum_{j=1}^J [l_j(\hat{\pi}_{1j}, \hat{\delta}, \hat{\rho}_j) - l_{0j}(\tilde{\pi}_{1j}, \delta_0, \tilde{\rho}_j)].$$

2.2. Score Test

Denote a score function $\mathbf{U} = (\sum_{j=1}^J \frac{\partial l_j}{\partial \delta}, 0, \dots, 0)$. A score test can be expressed by

$$T_{SC} = \mathbf{U} \mathbf{I}^{-1} \mathbf{U}^T \Big|_{\delta=\delta_0, \rho=\tilde{\rho}, \pi_1=\tilde{\pi}_1'}$$

where I is a Fisher information matrix with respect to the parameters $\delta, \rho = (\rho_1, \rho_2, \dots, \rho_J)^T, \pi_1 = (\pi_{11}, \pi_{12}, \dots, \pi_{1J})^T$. Denote

$$\begin{aligned}
 A_j &= \frac{\pi_{1j}(2(1 - \rho_j)(1 - \delta\pi_{1j}) + \rho_j)}{\rho_j(1 - \delta\pi_{1j}) + (1 - \rho_j)(1 - \delta\pi_{1j})^2}, \\
 B_j &= \frac{(1 - 2\delta\pi_{1j})}{\delta(1 - \delta\pi_{1j})}, \\
 C_j &= \frac{(2\delta\pi_{1j}(1 - \rho_j) + \rho_j)}{\delta\rho_j + (1 - \rho_j)\delta^2\pi_{1j}}.
 \end{aligned}$$

Further, it can be simplified as

$$T_{SC} = \left(\sum_{j=1}^J e_j \right)^2 I^{-1}(1, 1) \Big|_{\delta=\delta_0, \rho=\hat{\rho}, \pi_1=\hat{\pi}_1},$$

where $e_j = m_{02j}A_j + m_{12j}B_j + m_{22j}C_j$, and $I^{-1}(1, 1)$ is the first diagonal element of the inverse matrix of the Fisher information matrix I . See Appendix B for more details.

2.3. Wald-Type Test

Denote $\beta = (\delta, \rho^T, \pi_1^T)^T$, and $C = (1, 0, \dots, 0)_{1 \times (1+2J)}$. Thus, $H_0 : \delta = \delta_0$ is equivalent to $C\beta^T = \delta_0$. We can construct a Wald-type test statistic as

$$T_W = (\beta C^T - \delta_0)(CI^{-1}C^T)^{-1}(C\beta^T - \delta_0) \Big|_{\delta=\hat{\delta}, \rho=\hat{\rho}, \pi_1=\hat{\pi}_1},$$

where $\hat{\pi}_1 = (\hat{\pi}_{11}, \hat{\pi}_{12}, \dots, \hat{\pi}_{1J})^T, \hat{\delta}$ and $\hat{\rho} = (\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_J)^T$ are global MLEs under H_a . Through calculation, it can be simplified as

$$T_W = (\hat{\delta} - \delta_0)^2 I^{-1}(1, 1) \Big|_{\delta=\hat{\delta}, \rho=\hat{\rho}, \pi_1=\hat{\pi}_1}.$$

2.4. Pooled MLE-Based Wald-Type Test

Suppose that relative risk ratios from all strata are equal regardless of the stratification effects. Thus, we can pool the data from J strata together (Zhuang et al. [20]) and summarize all observations into a 3×2 table. The pooled data structure is shown in Table 2. Let M_{li} ($l = 0, 1, 2; i = 1, 2$) be the total number of subjects with l response(s) in the i th group, and $M_{l+} = \sum_{i=1}^2 M_{li}$ be the number of subjects with l responses, where $l = 0, 1, 2$. M_{+i} ($i = 1, 2$) is the total sample size in the i th group.

Table 2. Pooled data structure.

Number of Responses (l)	Group (i)		Total
	1	2	
0	$M_{01} = \sum_{j=1}^J m_{01j}$	$M_{02} = \sum_{j=1}^J m_{02j}$	M_{0+}
1	$M_{11} = \sum_{j=1}^J m_{11j}$	$M_{12} = \sum_{j=1}^J m_{12j}$	M_{1+}
2	$M_{21} = \sum_{j=1}^J m_{21j}$	$M_{22} = \sum_{j=1}^J m_{22j}$	M_{2+}
Total	M_{+1}	M_{+2}	M

For the pooled data, let π_i ($i = 1, 2$) be the response rate of the i th group. Denote the relative risk ratio $\delta = \pi_2/\pi_1$. If we ignore the relevance between the paired bilateral responses, the MLEs of π_i for $i = 1, 2$ and δ can be derived as

$$\hat{\pi}_i = \frac{M_{1i} + 2M_{2i}}{2M_{+i}}, \quad \hat{\delta} = \frac{\hat{\pi}_2}{\hat{\pi}_1} = \frac{(M_{12} + 2M_{22})M_{+1}}{(M_{11} + 2M_{21})M_{+2}}.$$

Based on the assumption that the observations from the paired organs are independent, the estimate of $Var(\bar{\pi}_i)$ is given in Tang et al. [9] as

$$V_i = \frac{4M_{0i}M_{2i} + M_{1i}(M_{0i} + M_{2i})}{4M_{+i}^3},$$

for $i = 1, 2$. By the delta method, we have $E(\bar{\delta}) \approx \delta$ and $Var(\bar{\delta}) \approx [\delta^2 Var(\bar{\pi}_1) + Var(\bar{\pi}_2)]/\bar{\pi}_1^2$. The estimate of $Var(\bar{\delta})$ is directly obtained by $\bar{\sigma}^2 = [\bar{\delta}^2 V_1 + V_2]/\bar{\pi}_1^2$. A Wald-type test can be given by

$$T_{LS} = \frac{(\bar{\delta} - \delta_0)^2}{\bar{\sigma}^2} = \frac{(\bar{\delta} - \delta_0)^2}{(\bar{\delta}^2 V_1 + V_2)/\bar{\pi}_1^2}.$$

2.5. Pooled MLE-Based Log-Transformation Test

According to the delta method, $\log(\bar{\delta})$ has the mean $\log(\delta)$ and the variance $Var[\log(\bar{\delta})] \approx Var(\bar{\delta})/\delta^2$. In this case, an estimate of $Var[\log(\bar{\delta})]$ is $\bar{\sigma}^2 = \bar{\sigma}^2/\bar{\delta}^2 = (V_1 + V_2/\bar{\delta}^2)/\bar{\pi}_1^2$. The χ^2 test statistic based on pooled MLE-based log-transformation is given by

$$T_{log} = \frac{(\log(\bar{\delta}) - \log(\delta_0))^2}{\bar{\sigma}^2} = \frac{(\log(\bar{\delta}) - \log(\delta_0))^2}{(V_1 + V_2/\bar{\delta}^2)/\bar{\pi}_1^2}.$$

Under H_0 , the test statistics T_i ($i = L, SC, W, LS, log$) are asymptotically distributed as a chi-square distribution with one degree of freedom. Thus, H_0 should be rejected if the value of the test statistic is larger than $\chi_{1,1-\alpha}^2$ at the significant level α , where $\chi_{1,1-\alpha}^2$ is the 100(1 - α) percentile of the chi-square distribution with one degree of freedom.

3. Sample Size Determination

Determining an appropriate sample size for a statistical test is essential in any clinical trial design. This section proposes the approximate formulae and iterative algorithm to determine sample sizes based on the above test statistics. However, the approximate formula method is not entirely applicable to all statistics, but only some of them. For convenience, denote $k_j = N_{1j}/M_{+1}$, and $t_j = N_{2j}/N_{1j}$ ($j = 1, 2, \dots, J$).

3.1. Asymptotic Sample Size

Based on the statistics above T_{SC}, T_{LS} , and T_{log} , we derive the asymptotic sample size under the desired power level. Under the alternative hypothesis, T_i ($i = SC, LS, log$) asymptotically follows a non-central chi-square distribution with one degree of freedom. The non-central parameter, denoted by τ_i , can be obtained by solving the equation: $\chi_{1,1-\beta}^2(\tau_i) = \chi_{1,\alpha}^2$, where $\chi_{1,1-\beta}^2(\tau_i)$ is the 100(1 - β) percentile of the non-central χ^2 distribution. When the observed frequencies are replaced by their expected frequencies (i.e., replace m_{lij} with $M_{+i}p_{lij}$ for $l = 0, 1, 2, i = 1, 2$ and $j = 1, 2, \dots, J$), denote the solutions from the Equation (5) for π_{1j}, ρ_j as $\bar{\pi}_{1j}^*, \bar{\rho}_j^*$, respectively. The Fisher scoring algorithm can be used to solve $\bar{\pi}_{1j}^*, \bar{\rho}_j^*$ for $j = 1, 2, \dots, J$.

Theorem 1. Given the prespecified power $1 - \beta$ and significance level α , the asymptotic sample size for the score test T_{SC} is expressed by

$$M_{SC}^a \approx \sum_{j=1}^J \frac{(\tau_{SC} + 1)/W(\delta_0, \bar{\rho}_j^*, \bar{\pi}_j^*) - \sum_{j=1}^J k_j t_j P_j}{(\sum_{j=1}^J k_j t_j E_j)^2} \times (k_j + k_j t_j),$$

where $W(\delta, \rho_j, \pi_j) = I^{-1}(1, 1)M_{+1}$, and

$$P_j = A_j^2 p_{02j}(1 - p_{02j}) + B_j^2 p_{12j}(1 - p_{12j}) + C_j^2 p_{22j}(1 - p_{22j}) - 2A_j B_j p_{02j} p_{12j} - 2A_j C_j p_{02j} p_{22j} - 2B_j C_j p_{12j} p_{22j} \Big|_{\delta=\delta_0, \pi_{1j}=\bar{\pi}_{1j}^*, \rho_j=\bar{\rho}_j^*}$$

$$E_j = p_{02j}A_j + p_{12j}B_j + p_{22j}C_j \Big|_{\delta=\delta_0, \pi_{1j}=\tilde{\pi}_{1j}^*, \rho_j=\tilde{\rho}_j^*}$$

Proof. Given a significance level α , the power of the test T_{SC} can be expressed as

$$Pr(T_{SC} \geq \chi_{1,\alpha}^2 \mid H_a) = Pr(\chi_1^2(\tau_{SC}) \geq \chi_{1,\alpha}^2).$$

Thus, it follows that

$$\tau_{SC} + 1 = E(T_{SC} \mid H_a). \tag{6}$$

We first substitute M_{1j} and M_{2j} with $k_j M_{+1}$ and $k_j t_j M_{+1}$ into the Fisher information matrix I and obtain the symmetrical matrix $I = M_{+1} I_{sym}$. Letting $W(\delta, \rho_j, \pi_j)$ be the first main-diagonal element of the inverse of the matrix I_{sym} , we have $I^{-1}(1, 1) = W(\delta, \rho_j, \pi_j) / M_{+1}$. The conditional expectation can be written as

$$E(T_{SC} \mid H_a) = E\left(\left(\sum_{j=1}^J e_j\right)^2 I^{-1}(1, 1) \mid H_a\right) \approx W(\delta_0, \tilde{\rho}_j^*, \tilde{\pi}_j^*) / M_{+1} E\left(\left(\sum_{j=1}^J e_j\right)^2 \mid H_a\right),$$

where

$$E\left(\left(\sum_{j=1}^J e_j\right)^2 \mid H_a\right) = \text{Var}\left(\sum_{j=1}^J e_j \mid H_a\right) + \left(E\left(\sum_{j=1}^J e_j \mid H_a\right)\right)^2.$$

Since $e_j = m_{02j}A_j + m_{12j}B_j + m_{22j}C_j$, the variance of $\sum_{j=1}^J e_j$ is derived through the properties of the trinomial distribution of $m_{l2j} (l = 0, 1, 2, j = 1, 2, \dots, J)$. Thus,

$$\begin{aligned} \text{Var}\left(\sum_{j=1}^J e_j\right) &= \text{Var}\left(\sum_{j=1}^J m_{02j}A_j + m_{12j}B_j + m_{22j}C_j\right) = \sum_{j=1}^J \text{Var}(m_{02j}A_j + m_{12j}B_j + m_{22j}C_j) \\ &= \sum_{j=1}^J \left\{ A_j^2 \text{Var}(m_{02j}) + B_j^2 \text{Var}(m_{12j}) + C_j^2 \text{Var}(m_{22j}) + 2A_j B_j \text{Cov}(m_{02j}, m_{12j}) \right. \\ &\quad \left. + 2A_j C_j \text{Cov}(m_{02j}, m_{22j}) + 2B_j C_j \text{Cov}(m_{12j}, m_{22j}) \right\}, \end{aligned}$$

where $\text{Var}(m_{02j}) = N_{2j} p_{02j} (1 - p_{02j})$, $\text{Var}(m_{12j}) = N_{2j} p_{12j} (1 - p_{12j})$, $\text{Var}(m_{22j}) = N_{2j} p_{22j} (1 - p_{22j})$, $\text{Cov}(m_{02j}, m_{12j}) = -N_{2j} p_{02j} p_{12j}$, $\text{Cov}(m_{02j}, m_{22j}) = -N_{2j} p_{02j} p_{22j}$ and $\text{Cov}(m_{12j}, m_{22j}) = -N_{2j} p_{12j} p_{22j}$. Substitute the MLEs $\tilde{\pi}_{1j}, \tilde{\rho}_j$ with $\tilde{\pi}_{1j}^*, \tilde{\rho}_j^*$ for $j = 1, 2, \dots, J$. The approximate conditional expectation of e_j under H_a can be given as $E(e_j \mid H_a) \approx M_{+1} k_j t_j E_j$. The approximate conditional variance of e_j under H_a is $\text{Var}(e_j \mid H_a) \approx M_{+1} k_j t_j P_j$. According to the Equation (6), we have

$$\tau_{SC} + 1 \approx W(\delta_0, \tilde{\rho}_j^*, \tilde{\pi}_j^*) \left(\sum_{j=1}^J k_j t_j P_j + M_{+1} \left(\sum_{j=1}^J k_j t_j E_j \right)^2 \right).$$

The total sample size of the first group can be solved as

$$M_{+1} \approx \frac{(\tau_{SC} + 1) / W(\delta_0, \tilde{\rho}_j^*, \tilde{\pi}_j^*) - \sum_{j=1}^J k_j t_j P_j}{\left(\sum_{j=1}^J k_j t_j E_j \right)^2}.$$

Thus, the asymptotic sample size of score statistic can be obtained as

$$M_{SC}^a \approx \sum_{j=1}^J \frac{(\tau_{SC} + 1) / W(\delta_0, \tilde{\rho}_j^*, \tilde{\pi}_j^*) - \sum_{j=1}^J k_j t_j P_j}{\left(\sum_{j=1}^J k_j t_j E_j \right)^2} \times (k_j + k_j t_j).$$

The proof is complete. \square

For the Wald-type statistic under the pooled model, the sample size formula can be given according to the non-central chi-square distribution under H_a . Denote $t = M_{+2}/M_{+1}$ for the pooled data.

Theorem 2. *Given the prespecified power $1 - \beta$ and significance level α , the asymptotic sample size for the test T_{LS} is expressed by*

$$M_{LS}^a \approx (1 + t) \frac{(\tau_{SL} + 1) \frac{\delta^2 W_1 + W_2/t}{(\bar{\pi}_1)^2} - \frac{\delta^2 W_1 + W_2/t}{\bar{\pi}_1^2}}{(\delta - \delta_0)^2},$$

where $W_i = p_{0i}p_{2i} + p_{1i}(p_{0i} + p_{2i})/4$ for $i = 1, 2$.

Proof. The test statistic T_{LS} is asymptotically distributed as a non-central χ^2 distribution with one degree of freedom and non-central parameter τ_{LS} under the alternative hypothesis H_a . The non-central parameter τ_{LS} can be obtained by solving the equation: $\chi^2_{1,1-\beta}(\tau_{LS}) = \chi^2_{1,\alpha}$. Thus, it follows that $\tau_{LS} + 1 = E(T_{LS} | H_a)$. Substitute M_{11} and M_{12} with $M_{+1}p_{11}$ and $kM_{+1}p_{12}$ ($l=0,1,2$) for $\bar{\sigma}^2 = [\delta^2 V_1 + V_2]/\bar{\pi}_1^2$. Hence,

$$\begin{aligned} E(T_{LS} | H_a) &= E\left(\frac{(\bar{\delta} - \delta_0)^2}{(\bar{\delta}^2 V_1 + V_2)/\bar{\pi}_1^2} | H_a\right) \\ &\approx E((\bar{\delta} - \delta_0)^2 | H_a) \frac{\bar{\pi}_1^2}{\bar{\delta}^2 W_1/M_{+1} + W_2/(tM_{+1})}. \end{aligned}$$

Based on the conditional expectation $E((\bar{\delta} - \delta_0)^2 | H_a) \approx Var(\bar{\delta} | H_a) + (E(\bar{\delta} - \delta_0 | H_a))^2$, and the conditional variance $Var(\bar{\delta} | H_a) \approx [\delta^2 W_1/M_{+1} + W_2/(tM_{+1})]/\bar{\pi}_1^2$, we have

$$\tau_{SL} + 1 \approx \left(\frac{\delta^2 W_1 + W_2/t}{\bar{\pi}_1^2} + (\delta - \delta_0)^2 M_{+1} \right) \frac{(\bar{\pi}_1)^2}{\bar{\delta}^2 W_1 + W_2/t}.$$

We can obtain the sample size of the first group by

$$M_{+1} \approx \frac{(\tau_{SL} + 1) \frac{\delta^2 W_1 + W_2/t}{(\bar{\pi}_1)^2} - \frac{\delta^2 W_1 + W_2/t}{\bar{\pi}_1^2}}{(\delta - \delta_0)^2}.$$

Then, the sample size for the statistic T_{LS} is given as

$$M_{LS}^a \approx (1 + t) \frac{(\tau_{LS} + 1) \frac{\delta^2 W_1 + W_2/t}{(\bar{\pi}_1)^2} - \frac{\delta^2 W_1 + W_2/t}{\bar{\pi}_1^2}}{(\delta - \delta_0)^2}.$$

The proof is complete. \square

Next, we present the corresponding sample size formula of the statistic T_{log} for the prespecified power at a fixed significance level.

Theorem 3. *Given the prespecified power $1 - \beta$ and significance level α , the asymptotic sample size for the test T_{log} is expressed by*

$$M_{log}^a \approx (1 + t) \frac{(\tau_{log} + 1) \frac{W_1 + W_2/(\delta^2 t)}{(\bar{\pi}_1)^2} - \frac{W_1 + W_2/(\delta^2 t)}{\bar{\pi}_1^2}}{(\log \delta - \log \delta_0)^2},$$

where $W_i = p_{0i}p_{2i} + p_{1i}(p_{0i} + p_{2i})/4$ for $i = 1, 2$.

Proof. The test statistic T_{log} is asymptotically distributed as a non-central χ^2 distribution with one degree of freedom and non-central parameter τ_{log} under the alternative hypothesis H_a . Thus, it follows that

$$\tau_{log} + 1 = E(T_{log} | H_a).$$

Substitute M_{l1} and M_{l2} with $M_{+1}p_{l1}$ and $kM_{+1}p_{l2}$ ($l = 0, 1, 2$) in $\tilde{\sigma}^2 = (V_1 + V_2/\bar{\delta}^2)/\bar{\pi}_1^2$. The conditional expectation of T_{log} can be obtained

$$\begin{aligned} E(T_{log} | H_a) &= E\left(\frac{(\log(\bar{\delta}) - \log(\delta_0))^2}{(V_1 + V_2/\bar{\delta}^2)/\bar{\pi}_1^2} \mid H_a\right) \\ &= E((\log \bar{\delta} - \log \delta_0)^2 \mid H_a) \frac{\bar{\pi}_1^2}{W_1/M_{+1} + W_2/(tM_{+1}\bar{\delta}^2)}, \end{aligned}$$

where $E((\log \bar{\delta} - \log \delta_0)^2 \mid H_a) \approx Var(\log \bar{\delta} \mid H_a) + E((\log \bar{\delta} - \log \delta_0)^2 \mid H_a)$. The value of $Var(\log \bar{\delta} \mid H_a)$ can be approximated by $\frac{W_1/M_{+1} + W_2/(tM_{+1}\bar{\delta}^2)}{\bar{\pi}_1^2}$. Then, we have

$$\begin{aligned} \tau_{log} + 1 &\approx \left\{ (\log \delta - \log \delta_0)^2 + \frac{W_1/M_{+1} + W_2/(tM_{+1}\bar{\delta}^2)}{\bar{\pi}_1^2} \right\} \frac{\bar{\pi}_1^2}{W_1/M_{+1} + W_2/(tM_{+1}\bar{\delta}^2)} \\ &= \left\{ (\log \delta - \log \delta_0)^2 M_{+1} + \frac{W_1 + W_2/(t\bar{\delta}^2)}{\bar{\pi}_1^2} \right\} \frac{\bar{\pi}_1^2}{W_1 + W_2/(t\bar{\delta}^2)}. \end{aligned}$$

We can obtain the sample size of the first group by

$$M_{+1} \approx \frac{(\tau_{log} + 1) \frac{W_1 + W_2/(\bar{\delta}^2 t)}{(\bar{\pi}_1)^2} - \frac{W_1 + W_2/(\bar{\delta}^2 t)}{\bar{\pi}_1^2}}{(\log \delta - \log \delta_0)^2}.$$

The sample size of T_{log} statistic is

$$M_{log}^a \approx \frac{(\tau_{log} + 1) \frac{W_1 + W_2/(\bar{\delta}^2 t)}{(\bar{\pi}_1)^2} - \frac{W_1 + W_2/(\bar{\delta}^2 t)}{\bar{\pi}_1^2}}{(\log \delta - \log \delta_0)^2} \times (1 + t).$$

The result follows. □

Theorems 1–3 provide the sample size formulae according to the asymptotic distribution of the statistics T_{SC} , T_{LS} , and T_{log} under the hypothesis $H_a : \delta \neq \delta_0$. However, it is not easy to obtain explicit expressions in practical applications. The iterative method can be considered to solve the problem.

3.2. The Iterative Method

The power $1 - \beta$ of a common test is the probability that the null hypothesis is rejected when the alternative hypothesis is true. Under the alternative hypothesis H_a , the asymptotic power of statistic T_i ($i = L, SC, W, LS$ or log) is given by

$$Pr(T_i \geq \chi_{1,1-\alpha}^2 \mid H_a) = 1 - \beta, \tag{7}$$

where significance level α is fixed. We can usually derive sample size by solving Equation (2) at a given significance level α and power of $1 - \beta$. However, we cannot know the exact distribution under H_a of some test statistics based on MLEs from the iterative algorithm (e.g., T_L and T_W). In this case, we propose an iterative algorithm for calculating the asymptotic sample size for stratified correlated data. The method can compare empirical power with the given power of $1 - \beta$ for searching sample size. The detailed process of the algorithm is shown as follows:

- (i) Given $k_j, t_j, \delta, \pi_{1j}$ and $\rho_j, j = 1, 2, \dots, J$. The initial values of sample size $\tilde{M}^{(0)} = 0$, the step size $d = 1000$ and flag $f = 1$.
- (ii) The $(t + 1)$ th update of \tilde{M} is $\tilde{M}^{(t+1)} = \tilde{M}^{(t)} + d \times f$. The 10,000 replicates $m_{ij} = (m_{0ij}, m_{1ij}, m_{2ij})^T$ are randomly generated under H_a , where m_{ij} ($i = 1$ or 2) follows a trinomial distribution

$$f(\mathbf{m}_{ij} \mid \mathbf{p}_{ij}) = \frac{N_{ij}!}{m_{0ij}!m_{1ij}!m_{2ij}!} p_{0ij}^{m_{0ij}} p_{1ij}^{m_{1ij}} p_{2ij}^{m_{2ij}}.$$

- (iii) Calculate empirical power based on random samples generated in step (ii) at a given significance level α . The empirical power can be computed by dividing the number of times rejecting H_0 by 10,000. The empirical power is denoted as $p^*(\tilde{M}^{(t+1)})$.
- (iv) Compare $p^*(\tilde{M}^{(t+1)})$ with given power $1 - \beta$. If $f \times p^*(\tilde{M}^{(t+1)}) < f \times (1 - \beta)$, return to step (ii). Otherwise, $d = 0.1 \times d$, $f = -f$ and return to step (ii).
- (v) Repeat the steps (ii)–(iv) until $p^*(\tilde{M}^{(t+1)})$ closes to $1 - \beta$ before d becomes a decimal.

From the test statistics T_i , the sample sizes from power approximation method are denoted as M_i^b , where i is L, SC, W, LS , or log . The MATLAB code of the iterative algorithm is available in the Supplementary Materials.

4. Simulation for Asymptotic Power and Sample Size

In this section, numerical simulations are conducted to evaluate the performance of asymptotic sample sizes from our proposed methods.

4.1. Asymptotic Sample Size, Power and TIE

Take $J = 3$. Under the null hypothesis $H_0 : \delta = 1$, we calculate the sample size $M_{SC}^a, M_{LS}^a, M_{log}^a, M_L^b, M_{SC}^b, M_W^b, M_{LS}^b$, and M_{log}^b for given power $1 - \beta = 0.9$. Considering that the total numbers of strata may be the same or different, we let $k = (1/3, 1/3, 1/3)$ and $(0.5, 0.3, 0.2)$. The correlation coefficient $\rho_j = 0.3$ (low), or 0.5 (moderate), and the response rates $\pi_{1j} = 0.4$ (less than 50%), or 0.6 (bigger than 50%). The scenarios of parameter settings in the simulations are listed in Table 3.

Table 3. The scenarios of parameter settings in the simulations.

Scenario	π_{1j}	ρ_j	k
<i>i</i>	0.4	0.5	(1/3, 1/3, 1/3)
<i>ii</i>	0.4	0.5	(0.5, 0.3, 0.2)
<i>iii</i>	0.4	0.3	(1/3, 1/3, 1/3)
<i>iv</i>	0.4	0.3	(0.5, 0.3, 0.2)
<i>v</i>	0.6	0.5	(1/3, 1/3, 1/3)
<i>vi</i>	0.6	0.5	(0.5, 0.3, 0.2)
<i>vii</i>	0.6	0.3	(1/3, 1/3, 1/3)
<i>viii</i>	0.6	0.3	(0.5, 0.3, 0.2)

Scenarios *i–viii* represent various parameter combinations of π_{1j} , ρ_j and k .

Figures 1–3 show the estimated sample sizes, empirical powers and TIEs of $M_{SC}^a, M_{LS}^a, M_{log}^a, M_L^b, M_{SC}^b, M_W^b, M_{LS}^b$ and M_{log}^b for the given parameters listed in Table 3. We observe that there are obvious differences among M_{SC}^a, M_{LS}^a , and M_{log}^a even though they are all solved by fitting asymptotic χ^2 distributions. Asymptotic sample sizes based on the statistics T_L, T_{SC} , and T_{LS} are close under the iterative method. For the five sample sizes from the iterative algorithm, the values of M_W^b are always lower, and those of M_{log}^b tend to be bigger. The sample sizes of $\delta_a = 0.7$ are smaller than that of $\delta_a = 0.8$. When other parameters are fixed, we can observe two phenomena: (i) the sample sizes become bigger for $\rho_j = 0.5$ than $\rho_j = 0.3$; (ii) the sample sizes tend to decrease when π_{1j} increases from 0.4 to 0.6. The sample sizes increase as δ_a is closer to 1. That is to say, more samples are required when $\delta = \delta_a$ is closer to that under $H_0 : \delta = \delta_0$.

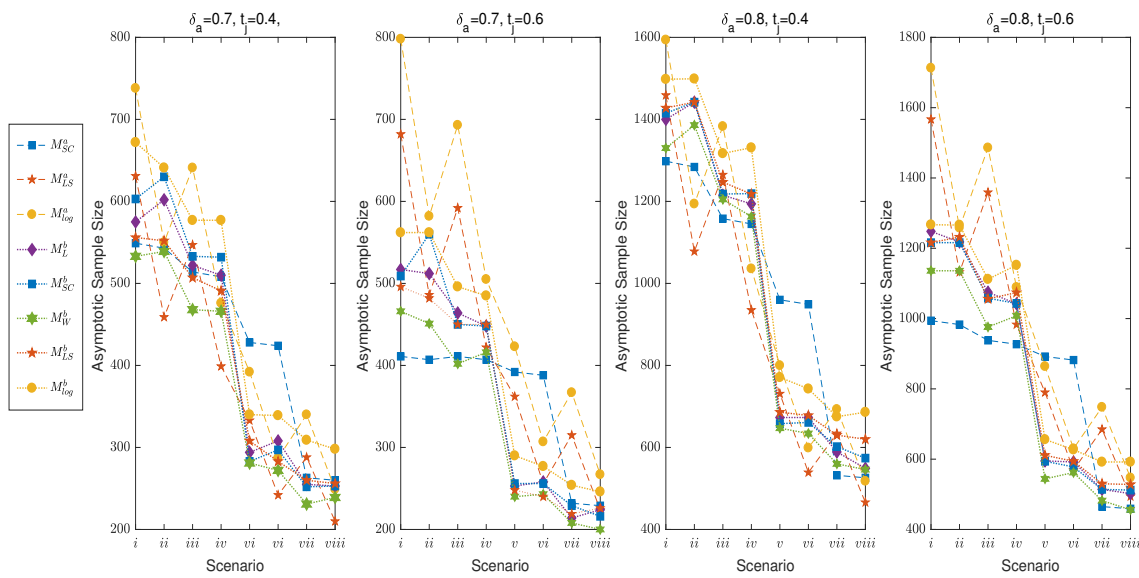


Figure 1. Asymptotic sample size for $1 - \beta = 0.9$.

We can calculate the empirical powers and TIEs based on the estimated sample sizes. For every sample size, 10,000 replicates $m_{ij} = (m_{0ij}, m_{1ij}, m_{2ij})^T$ are randomly generated under H_a . Further, the empirical power can be computed by dividing the number of times rejecting H_0 by 10,000 when the significance level is 0.05. Empirical TIE is calculated through 10,000 replicates randomly generated under $\delta = \delta_0$. The results are listed in Figures 2 and 3. The empirical powers are mostly close to the desired powers of 0.9. Compared with the sample size formula, the iterative algorithm often produces the sample size with empirical power closer to 0.9. In terms of empirical power, there are no significant trend changes for various parameter settings of $\delta_a, k_j, t_j, \pi_{1j}$ and $\rho_j (j = 1, 2, \dots, J)$.

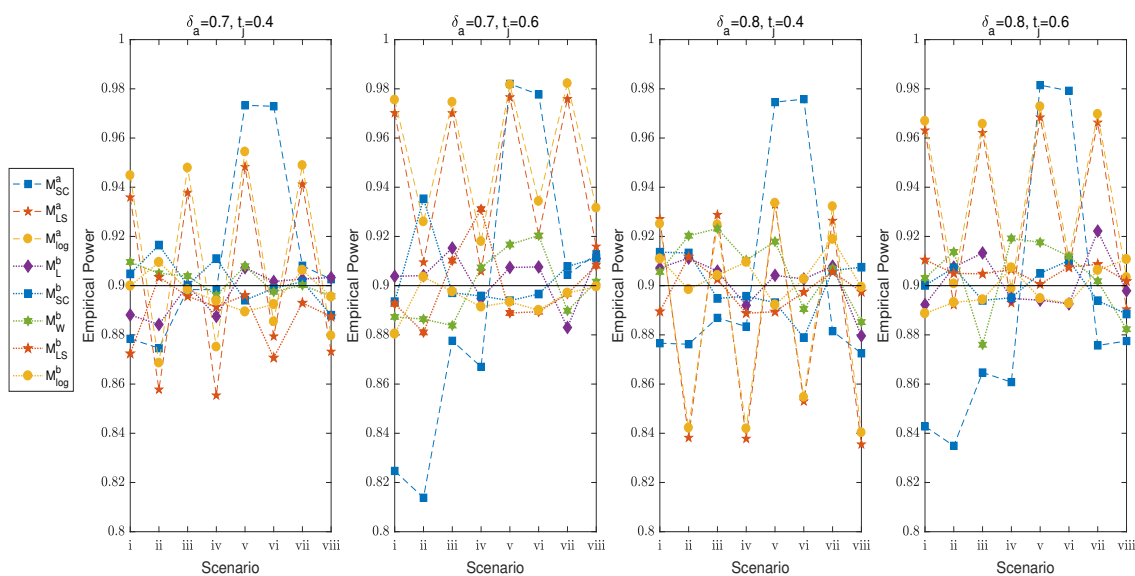


Figure 2. Empirical power (%) based on estimated sample sizes for $1 - \beta = 0.9$.

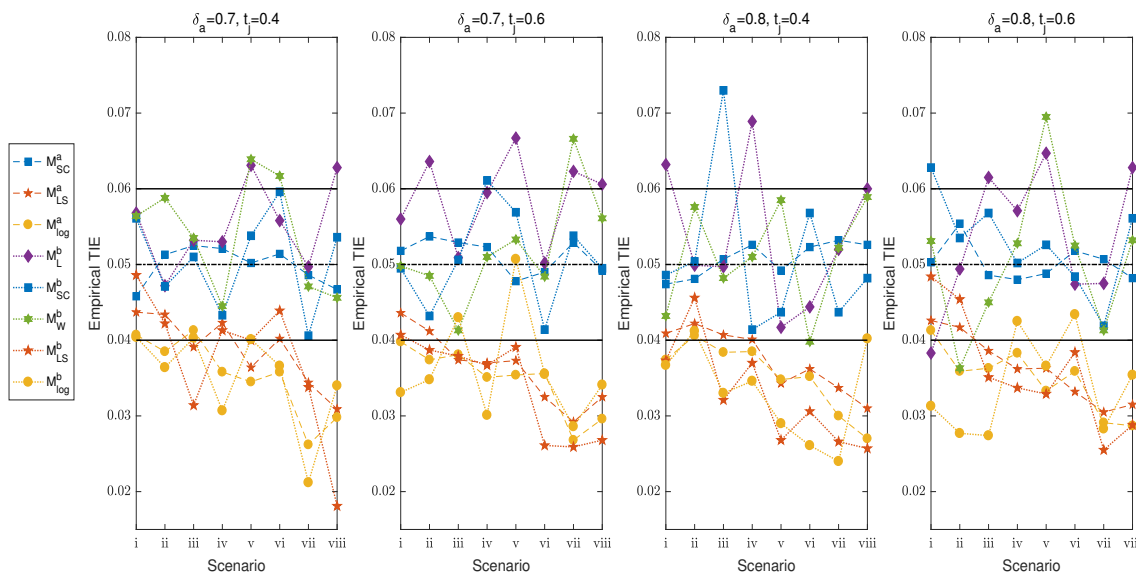


Figure 3. Empirical TIEs (%) based on estimated sample sizes for $1 - \beta = 0.9$.

According to Tang et al. [8], a test is robust if its empirical TIE is between 0.04 and 0.06 when the significance level is 0.05, liberal if it is greater than 0.06, and conservative if it is less than 0.04. As shown in Figure 3, the sample size M_{SC}^a has better performance in empirical TIEs. For the sample size M_{SC}^a , the empirical TIEs are located on $[0.04, 0.06]$ under all considered settings. For some cases, the empirical TIEs based on statistics T_{LS} and T_{log} are lower than 0.04, which means that T_{LS} and T_{log} have conservative TIEs based on sample sizes from the approximate formula or the iterative algorithm. Sample sizes M_L^b , M_{SC}^b , and M_W^b sometimes have liberal TIEs bigger than 0.06.

4.2. Accuracy

This section evaluates the effectiveness of sample size determinations. Let $J = 3$, $\delta_0 = 1$, and the sample size $M_{+1} = 100, 200, 300$. Other parameters k_i, π_{ij}, ρ_j are randomly generated under $H_a : \delta = \delta_a (\neq \delta_0)$, satisfying $0 < k_i < 1, 0 \leq \pi_{ij} \leq 1, -1 \leq \rho_j \leq 1$ and $0 \leq \delta_a \leq 1$. Denote $\mathbf{k} = (k_1, k_2, k_3), t_1 = t_2 = t_3, \boldsymbol{\pi}_1 = (\pi_{11}, \pi_{12}, \pi_{13})^T, \boldsymbol{\rho} = (\rho_1, \rho_2, \rho_3)^T$. We choose 1000 parameter settings for every sample size. There are 10,000 random replications under every parameter setting. Empirical powers are given at a significance level of 0.05. Further, we calculate the corresponding asymptotic sample sizes for the obtained empirical powers at a significance level $\alpha = 0.05$. The boxplots in Figure 4 show all asymptotic sample sizes.

We note that the estimated sample sizes ($M_L^b, M_{SC}^b, M_W^b, M_{LS}^b$, and M_{log}^b) are close to the true sample size for $M_{+1} = 100, 200, 300$, which performs better than the sample size formula method (M_{SC}^a, M_{LS}^a and M_{log}^a). In terms of the sample size formula, the sample size from the score statistic is more accurate than the statistics T_{LS} and T_{log} . The estimated sample sizes M_{LS}^a and M_{log}^a are always bigger than the true values. Thus, the stratified correlated model performs better for effectively estimating sample size through formula derivation.

Above all, all power-controlled methods can provide optimal estimates because the empirical powers are usually close to the prespecified one. The sample sizes from the iterative algorithm have more satisfactory power. Thus, the iterative method is recommended in practice.

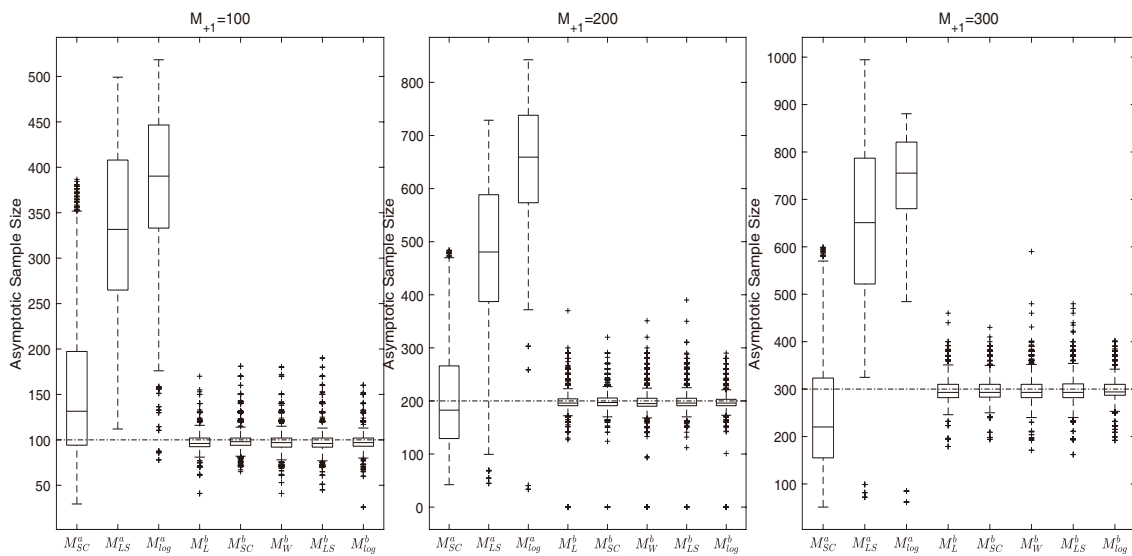


Figure 4. Sample sizes under various methods.

4.3. The Effect of Parameters

Through the iterative algorithm, we further study the effect of parameters on sample sizes. Given a prespecified power of 0.9 and a significance level of 0.05, the hypotheses $H_0 : \delta_0 = 1$ and $H_1 : \delta_a = 0.8$. To study the effect of strata, we take $J = 2, 4, 6$. The influence of π_{1j} , ρ_j and t_j can be observed according to these cases: (i) $\pi_{1j} = 0.3(0.1)0.7$ for $\rho_j = 0.7$ and $t_j = 1$; (ii) $\rho_j = 0.1(0.2)0.9$ for $\pi_{1j} = 0.5$ and $t_j = 1$; and (iii) $t_j = 0.2(0.2)1$ for $\pi_{1j} = 0.5$ and $\rho_j = 0.7$. Under various parameter settings, Figure 5 shows that the asymptotic sample sizes become smaller when π_{1j} ($j = 1, 2, \dots, J$) increases. The five sample sizes increase with bigger ρ_j ($j = 1, 2, \dots, J$). As t_j ($j = 1, 2, \dots, J$) approaches 1, the values of asymptotic sample sizes will be smaller. This reflects that the balanced design (i.e., $N_{1j} = N_{2j}$ for $j = 1, 2, \dots, J$) of two groups needs a smaller sample size for the prespecified power. When the number of strata changes, the sample sizes have similar values for $J = 2, 4, 6$. Strata number has little effect on sample sizes.

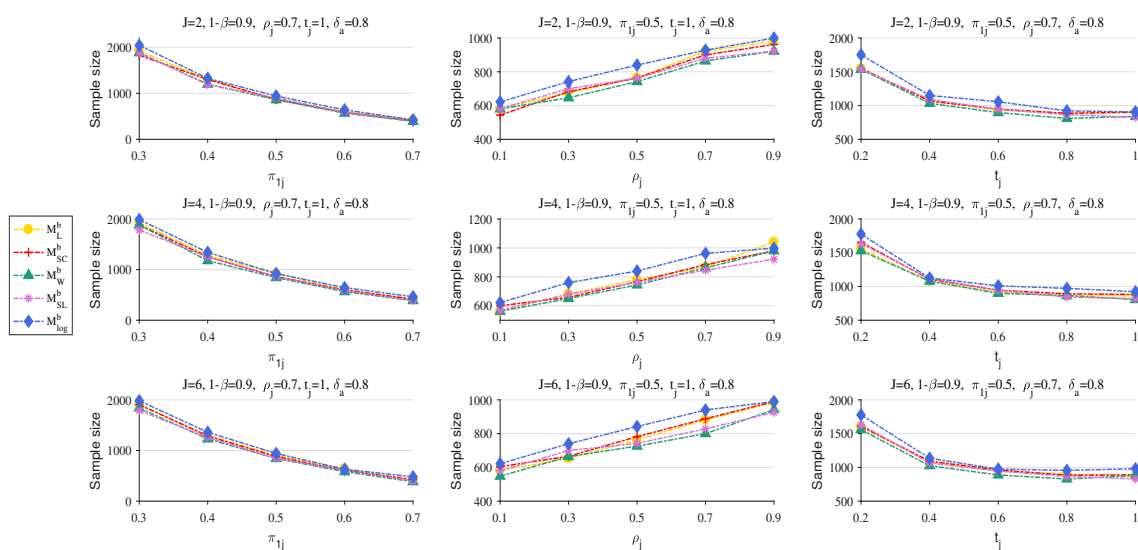


Figure 5. The effect of the parameter settings for sample size ($\delta_0 = 1, k_j = 1/J$).

5. A Real Example

Mandel et al. [21] conducted a double-blinded clinical trial to illustrate the proposed sample size considerations. In this trial, children with otitis media with effusion (OME) were stratified by age and randomly assigned to the cefaclor or amoxicillin treatment groups. There are three strata (<2, 2–5 and >5 yr) divided according to age as shown in Table 4.

Table 4. Frequency of OME-free ears after treatment.

Number of OME-Free Ears	<2 yr		2–5 yr		>5 yr		Total
	Cefaclor	Amoxicillin	Cefaclor	Amoxicillin	Cefaclor	Amoxicillin	
0	8	11	6	3	0	1	29
1	2	2	6	1	1	0	12
2	8	2	10	5	3	6	34
Total	18	15	22	9	4	7	75

From Table 4, it is obvious that $k_1 = (18 + 15)/75 = 33/75$, $k_2 = (22 + 9)/75 = 31/75$, $k_3 = (4 + 7)/75 = 11/75$. Let $t_j = M_{+2}/M_{+1} = (18 + 22 + 4)/(15 + 9 + 7) = 1.4$ for $j = 1, 2, 3$. The Fisher scoring algorithm can derive the global MLEs under $H_a : \delta_1 = \delta_2 = \delta_3 \triangleq \delta$ (see in Table 5). The asymptotic sample sizes are obtained based on the global MLEs of parameters ρ_j, π_{1j} and δ ($j = 1, 2, 3$).

Table 5. Global MLEs in the real example.

Age	Stratum (j)	$\hat{\pi}_{1j}$	$\hat{\rho}_j$	$\hat{\delta}_j$
<2 yr	1	0.377	0.736	0.937
2–5 yr	2	0.606	0.532	0.937
>5 yr	3	0.885	0.624	0.937

Table 6 shows the asymptotic sample sizes for the prespecified power 0.8, 0.9, and 0.95 under $\delta_0 = 0.5, 0.6$. The sample sizes based on T_{LS} have bigger values than other sample size formulae or the iterative methods. The sample sizes from the iterative algorithm are smaller than those from the sample size formula. For $\delta_0 = 0.5$, the sample size 75 in the real example is close to the values of M_{SC}^b and M_L^b even when the power is 0.95. However, only M_{SC}^b and M_L^b for power 0.8 are close to the real sample size for $\delta_0 = 0.6$. That is to say, the sample size 75 can basically guarantee the power of 0.95 for the test statistics T_L and T_{SC} under $H_0 : \delta = 0.5$. Moreover, when the values of δ under H_a and H_0 become closer, we need the larger sample size to keep the power at the fixed significance level. Then, we conduct the common test under the null hypothesis $H_0 : \delta = 0.5, 0.6$. The values and p -values of the test statistics are recorded in Table 7. For $\delta_0 = 0.5$, the p -values of five statistics $T_i(L, SC, W, LS, log)$ are 0.0029, 0.0084, 0.0040, 0.0384, and 0.0311, respectively. Obviously, the p -values are lower than 0.05. Thus, we have full evidence to reject $H_0 : \delta = 0.5$ at the significance level 0.05. Under $H_0 : \delta = 0.6$, there are some p -values smaller than 0.05 but others bigger than 0.05. Among the estimated sample sizes, only M_L^b and M_{SC}^b have close values to real sample sizes given power 0.8. Since p -values of T_L and T_{SC} are 0.0373 and 0.0490, the hypothesis $H_0 : \delta = 0.6$ is rejected at the significance level 0.05.

Table 6. Asymptotic sample sizes for the real example.

δ_0	Power	M_{SC}^a	M_{LS}^a	M_{log}^a	M_L^b	M_{SC}^b	M_W^b	M_{LS}^b	M_{log}^b
0.5	0.80	70	127	67	43	43	53	86	53
	0.90	81	153	82	53	62	72	100	77
	0.95	88	170	91	70	74	86	122	94
0.6	0.80	129	213	134	77	79	91	151	120
	0.90	151	259	163	100	110	122	182	154
	0.95	164	285	180	132	132	146	218	192

Table 7. Values and p -values of the test statistics for the real example.

δ_0	Result	T_L	T_{SC}	T_W	T_{LS}	T_{log}
0.5	Value	8.8475	6.9551	8.2666	4.2853	4.6490
	p -value	0.0029	0.0084	0.0040	0.0384	0.0311
0.6	Value	4.3363	3.8767	4.9158	1.6514	1.7826
	p -value	0.0373	0.0490	0.0266	0.1988	0.1818

6. Conclusions

This paper mainly investigated the asymptotic sample size for the common test of relative risk ratio in stratified bilateral data. Under Donner’s model, we obtained the sample size formula and proposed the iterative method based on the likelihood ratio statistic T_L , score statistic T_{SC} , and Wald-type statistic T_W . We also applied the sample size methods to the pooled MLE-based Wald-type statistic T_{LS} and the log-transformation T_{log} .

Numerical simulations are designed to evaluate the performance of estimated sample sizes. According to the simulations, the main conclusions lie in five aspects. (i) The estimated sample sizes have empirical power closer to the prespecified power level. (ii) The sample sizes $M_L^b, M_{SC}^b, M_W^b, M_{LS}^b$ and M_{log}^b , which are obtained by the iterative algorithm, outperform other asymptotic sample sizes because of accurate values and satisfactory power. (iii) The sample size M_{SC}^a has more robust TIEs than M_{LS}^a and M_{log}^a from the sample size formula. (iv) The sample sizes are sensitive to π_{1j} and ρ_j ($j = 1, 2, \dots, J$) when other parameters are fixed. (v) The sample sizes under unbalanced designs of two groups are greater than those under balanced designs for given power and significance levels.

This paper mainly focuses on the power-controlled sample size of the common test of relative risk ratios for stratified correlated bilateral data. However, one may want to calculate the necessary sample size such that the width of the $(1 - \alpha)\%$ confidence interval (CI) does not exceed a prespecified quantity. This type of sample size is called CI-width controlled sample size determination [8]. In the future, we will study the CI-width controlled sample size for the relative risk ratio in stratified paired data. Through numerical simulation, we compared the performance of the proposed methods in terms of empirical TIEs and powers. An example was used to illustrate our proposed methods. However, we have not analyzed the theoretical results of the efficiency of these methods. These problems will be considered in our future works.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/math11194198/s1>, the MATLAB code of the iterative algorithm.

Author Contributions: Conceptualization, investigation, project administration, writing—original draft, K.M. and Z.L.; Writing—review and editing, supervision, C.M. All authors have read and agreed to the published version of the manuscript.

Funding: The work was supported by 2022 Innovation Project for Excellent Doctoral Postgraduates of Xinjiang University (XJU2022BS027), the National Natural Science Foundation of China (12061070) and the Natural Science Foundation of Xinjiang Uygur Autonomous Region (2021D01E13).

Data Availability Statement: The real example data is from the article: Mandel, E.; Bluestone, C.; Rockette, H.; Blatter, M.; Reisinger, K.; Wucher, F.; 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.

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

Appendix A. Derivation for the Log-Likelihood Function under H_0

Under H_0 , the log-likelihood function can be rewritten as

$$\ell_0(\boldsymbol{\pi}_1, \delta_0, \boldsymbol{\rho}) = \sum_{j=1}^J l_{0j}(\pi_{1j}, \delta_0, \rho_j) + \log C,$$

where

$$\begin{aligned} l_{0j}(\pi_{1j}, \delta_0, \rho_j) = & m_{01j} \log[\rho_j(1 - \pi_{1j}) + (1 - \rho_j)(1 - \pi_{1j})^2] + m_{11j} \log[2\pi_{1j}(1 - \rho_j)(1 - \pi_{1j})] \\ & + m_{21j} \log[\rho_j\pi_{1j} + (1 - \rho_j)\pi_{1j}^2] + m_{02j} \log[\rho_j(1 - \pi_{1j}\delta_0) + (1 - \rho_j)(1 - \pi_{1j}\delta_0)^2] \\ & + m_{12j} \log[2\pi_{1j}\delta_0(1 - \rho_j)(1 - \pi_{1j}\delta_0)] + m_{22j} \log[\rho_j\pi_{1j}\delta_0 + (1 - \rho_j)(\pi_{1j}\delta_0)^2]. \end{aligned}$$

Appendix B. Derivation for Score Statistic

The information matrix I for score test statistic is written as

$$I = \begin{bmatrix} I_{11} & I_{12} & I_{13} \\ I_{12}^T & I_{22} & I_{23} \\ I_{13}^T & I_{23}^T & I_{33} \end{bmatrix},$$

where

$$\begin{aligned} I_{11} &= E\left(-\sum_{j=1}^J \frac{\partial^2 l_j}{\partial \delta^2}\right) \\ &= \sum_{j=1}^J N_{2j} \pi_{1j}^2 \left\{ \frac{(1 + 2\delta^2 \pi_{1j}^2)(1 - \rho_j)^2 - 2\delta \pi_{1j}(1 - \rho_j)(2 - \rho_j) + 1}{\rho_j(1 - \delta \pi_{1j}) + (1 - \rho_j)(1 - \delta \pi_{1j})^2} \right. \\ &\quad \left. + \frac{2(1 - \rho_j)(2\delta^2 \pi_{1j}^2 - 2\delta \pi_{1j} + 1)}{\delta \pi_{1j}(1 - \delta \pi_{1j})} + \frac{2\delta^2 \pi_{1j}^2(1 - \rho_j)^2 + 2\delta \pi_{1j} \rho_j(1 - \rho_j) + \rho_j^2}{\rho_j \delta \pi_{1j} + (1 - \rho_j)(\delta \pi_{1j})^2} \right\} \\ &= \sum_{j=1}^J i_{11}^{(j)}, \\ I_{12} &= \left[E\left(-\frac{\partial^2 l_1}{\partial \delta \partial \rho_1}\right), E\left(-\frac{\partial^2 l_2}{\partial \delta \partial \rho_2}\right), \dots, E\left(-\frac{\partial^2 l_J}{\partial \delta \partial \rho_J}\right) \right] \\ &= \left\{ N_{21} \pi_{11} \left(-\frac{(1 - \delta \pi_{11})^2}{\rho_1(1 - \delta \pi_{11}) + (1 - \rho_1)(1 - \delta \pi_{11})^2} + \frac{(\delta \pi_{11})^2}{\rho_1 \delta \pi_{11} + (1 - \rho_1)(\delta \pi_{11})^2} \right), \right. \\ &\quad N_{22} \pi_{12} \left(-\frac{(1 - \delta \pi_{12})^2}{\rho_2(1 - \delta \pi_{12}) + (1 - \rho_2)(1 - \delta \pi_{12})^2} + \frac{(\delta \pi_{12})^2}{\rho_2 \delta \pi_{12} + (1 - \rho_2)(\delta \pi_{12})^2} \right), \\ &\quad \dots \\ &\quad \left. N_{2J} \pi_{1J} \left(-\frac{(1 - \delta \pi_{1J})^2}{\rho_J(1 - \delta \pi_{1J}) + (1 - \rho_J)(1 - \delta \pi_{1J})^2} + \frac{(\delta \pi_{1J})^2}{\rho_J \delta \pi_{1J} + (1 - \rho_J)(\delta \pi_{1J})^2} \right) \right\} \\ &= [i_{12}^{(1)}, i_{12}^{(2)}, \dots, i_{12}^{(J)}], \\ I_{13} &= \left[E\left(-\frac{\partial^2 l_1}{\partial \pi_{11} \partial \delta}\right), E\left(-\frac{\partial^2 l_2}{\partial \pi_{12} \partial \delta}\right), \dots, E\left(-\frac{\partial^2 l_J}{\partial \pi_{1J} \partial \delta}\right) \right] \\ &= \left[E\left(-\frac{1}{\pi_{11}} \frac{\partial l_1}{\partial \delta} - \frac{\delta}{\pi_{11}} \frac{\partial^2 l_1}{\partial \delta^2}\right), E\left(-\frac{1}{\pi_{12}} \frac{\partial l_2}{\partial \delta} - \frac{\delta}{\pi_{12}} \frac{\partial^2 l_2}{\partial \delta^2}\right), \right. \\ &\quad \left. \dots, E\left(-\frac{1}{\pi_{1J}} \frac{\partial l_J}{\partial \delta} - \frac{\delta}{\pi_{1J}} \frac{\partial^2 l_J}{\partial \delta^2}\right) \right] = \left[\frac{\delta}{\pi_{11}} i_{11}^{(1)}, \frac{\delta}{\pi_{12}} i_{11}^{(2)}, \dots, \frac{\delta}{\pi_{1J}} i_{11}^{(J)} \right], \end{aligned}$$

$$\begin{aligned}
 I_{22} &= \text{diag} \left\{ E \left(- \frac{\partial^2 l_j}{\partial \rho_j^2} \right) \right\} \\
 &= \text{diag} \left\{ N_{1j} \left(\frac{\pi_{1j}^2 (1 - \pi_{1j})^2}{\rho_j (1 - \pi_{1j}) + (1 - \rho_j) (1 - \pi_{1j})^2} + \frac{2\pi_{1j} (1 - \pi_{1j})}{1 - \rho_j} + \frac{(1 - \pi_{1j})^2 \pi_{1j}^2}{\rho_j \pi_{1j} + (1 - \rho_j) \pi_{1j}^2} \right) \right. \\
 &\quad \left. + N_{2j} \left(\frac{\delta^2 \pi_{1j}^2 (1 - \delta \pi_{1j})^2}{\rho_j (1 - \delta \pi_{1j}) + (1 - \rho_j) (1 - \delta \pi_{1j})^2} + \frac{2\pi_{1j} (1 - \pi_{1j})}{1 - \rho_j} + \frac{(1 - \delta \pi_{1j})^2 (\delta \pi_{1j})^2}{\rho_j \delta \pi_{1j} + (1 - \rho_j) \delta \pi_{1j}^2} \right) \right\}, \\
 I_{23} &= \text{diag} \left\{ E \left(- \frac{\partial^2 l_j}{\partial \pi_{1j} \partial \rho_j} \right) \right\} = \text{diag} \left\{ N_{1j} \left(- \frac{(1 - \pi_{1j})^2}{\rho_j (1 - \pi_{1j}) + (1 - \rho_j) (1 - \pi_{1j})^2} \right. \right. \\
 &\quad \left. \left. + \frac{\pi_{1j}^2}{\rho_j \pi_{1j} + (1 - \rho_j) \pi_{1j}^2} \right) + \frac{\delta}{\pi_{1j}} i_{12}^{(j)} \right\}, \\
 I_{33} &= \text{diag} \left\{ E \left(- \frac{\partial^2 l_j}{\partial \pi_{1j}^2} \right) \right\} = \text{diag} \left\{ N_{1j} \left(\frac{(1 + 2\pi_{1j}^2) (1 - \rho_j)^2 - 2\pi_{1j} (1 - \rho_j) (2 - \rho_j) + 1}{\rho_j (1 - \pi_{1j}) + (1 - \rho_j) (1 - \pi_{1j})^2} \right. \right. \\
 &\quad \left. \left. + \frac{2(1 - \rho_j) (2\pi_{1j}^2 - 2\pi_{1j} + 1)}{\pi_{1j} (1 - \pi_{1j})} + \frac{2\pi_{1j}^2 (1 - \rho_j)^2 + 2\pi_{1j} \rho_j (1 - \rho_j) + \rho_j^2}{\pi_{1j} \rho_j + (1 - \rho_j) \pi_{1j}^2} \right) + \frac{\delta^2}{\pi_{1j}^2} i_{11}^{(j)} \right\}.
 \end{aligned}$$

The first diagonal element in the inverse matrix I is

$$I^{-1}(1, 1) = \left(I_{11} - [I_{12}, I_{13}] \begin{bmatrix} I_{22} & I_{23} \\ I_{23} & I_{33} \end{bmatrix}^{-1} \begin{bmatrix} I_{12}^T \\ I_{13}^T \end{bmatrix} \right)^{-1}.$$

Thus, the test statistics T_{SC} can be simplified as

$$T_{SC} = \left(\sum_{j=1}^J e_j \right)^2 I^{-1}(1, 1) \Big|_{\delta=\delta_0, \rho=\tilde{\rho}, \pi_1=\tilde{\pi}_1},$$

where

$$e_j = - \frac{m_{02j} \pi_{1j} (2(1 - \rho_j) (1 - \delta \pi_{1j}) + \rho_j)}{\rho_j (1 - \delta \pi_{1j}) + (1 - \rho_j) (1 - \delta \pi_{1j})^2} + \frac{m_{12j} (1 - 2\delta \pi_{1j})}{\delta (1 - \delta \pi_{1j})} + \frac{m_{22j} (2\delta \pi_{1j} (1 - \rho_j) + \rho_j)}{\delta \rho_j + (1 - \rho_j) \delta^2 \pi_{1j}}.$$

References

1. Rosner, B. Statistical methods in ophthalmology: An adjustment for the correlation between eyes. *Biometrics* **1982**, *38*, 105–114. [\[CrossRef\]](#)
2. Dallal, G. Paired Bernoulli trials. *Biometrics* **1988**, *44*, 253–257. [\[CrossRef\]](#)
3. Donner, A. Statistical methods in ophthalmology: An adjusted chi-square approach. *Biometrics* **1989**, *45*, 605–611. [\[CrossRef\]](#)
4. Tang, N.; Tang, M.; Qiu, S. Testing the equality of proportions for correlated otolaryngologic data. *Comput. Stat. Data Anal.* **2008**, *52*, 3719–3729. [\[CrossRef\]](#)
5. Pei, Y.; Tang, M.; Wong, W.; Tang, N. Testing equality of correlations of two paired binary responses from two treated groups in a randomized trial. *J. Biopharm. Stat.* **2011**, *21*, 511–525. [\[CrossRef\]](#)
6. Mou, K.; Ma, C.; Li, Z. Homogeneity test of relative risk ratios for stratified bilateral data under different algorithms. *J. Appl. Stat.* **2023**, *50*, 1060–1077. [\[CrossRef\]](#)
7. Tang, N.; Qiu, S. Homogeneity test, sample size determination and interval construction of difference of two proportions in stratified bilateral-sample designs. *J. Stat. Plan. Inference* **2012**, *142*, 1242–1251. [\[CrossRef\]](#)
8. Tang, M.; Tang, N.; Carey, V. Sample size determination for 2-step studies with dichotomous response. *J. Stat. Plan. Inference* **2006**, *136*, 1166–1180. [\[CrossRef\]](#)
9. Tang, M.; Tang, N.; Rosner, B. Statistical inference for correlated data in ophthalmologic studies. *Stat. Med.* **2006**, *25*, 2771–2783. [\[CrossRef\]](#)
10. Liu, X.; Shan, G.; Tian, L.; Ma, C. Exact methods for testing homogeneity of proportions for multiple groups of paired binary data. *Commun. Stat.-Simul. C* **2017**, *46*, 6074–6082. [\[CrossRef\]](#)
11. Shan, G. Exact approaches for testing non-inferiority or superiority of two incidence rates. *Stat. Probab. Lett.* **2014**, *85*, 129–134. [\[CrossRef\]](#)
12. Tang, N.; Qiu, S.; Tang, M.; Pei, Y. Asymptotic confidence interval construction for proportion difference in medical studies with bilateral data. *Stat. Methods Med. Res.* **2011**, *20*, 233–259. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Pei, Y.; Tang, M.; Wong, W.; Gao, J. Confidence intervals for correlated proportion differences from paired data in a two-arm randomised clinical trial. *Stat. Methods Med. Res.* **2012**, *21*, 167–187. [\[CrossRef\]](#) [\[PubMed\]](#)

14. Qiu, S.; Poon, W.; Tang, M. Sample size determination for disease prevalence studies with partially validated data. *Stat. Methods Med. Res.* **2012**, *25*, 37–63. [[CrossRef](#)]
15. Qiu, S.; Zeng, X.; Tang, M.; Pei, Y. Test procedure and sample size determination for a proportion study using a doublesampling scheme with two fallible classifiers. *Stat. Methods Med. Res.* **2019**, *28*, 1019–1043. [[CrossRef](#)] [[PubMed](#)]
16. Sun, S.; Li, Z.; Jiang, H. Homogeneity test and sample size of risk difference for stratified unilateral and bilateral data. *Commun. Stat.-Simul. C* **2022**, 1–24. [[CrossRef](#)]
17. Lloyd, C.; Ripamonti, E. comprehensive open-source library for exact required sample size in binary clinical trials. *Contemp. Clin. Trials* **2021**, *107*, 106491. [[CrossRef](#)]
18. Tang, N.; Yu, B. Bayesian sample size determination in a three-arm non-inferiority trial with binary endpoints. *J. Biopharm. Stat.* **2022**, *32*, 768–788. [[CrossRef](#)]
19. Pilz, M. Sample size calculation for one-armed clinical trials with clustered data and binary outcome. *Biom. J.* **2023**, *28*, e2300123. [[CrossRef](#)]
20. Zhuang, T.; Tian, G.; Ma, C. Homogeneity test of ratio of two proportions in stratified bilateral data. *Stat. Biopharm. Res.* **2019**, *11*, 200–209. [[CrossRef](#)]
21. Mandel, E.; Bluestone, C.; Rockette, H.; Blatter, M.; Reisinger, K.; Wucher, F.; 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](#)] [[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.