Common risk difference test and interval estimation of risk difference for stratified bilateral correlated data

Journal: *Statistical Methods in Medical Research*

Manuscript ID: Draft

Manuscript Type: Original Article

Keywords: Common risk difference test, Interval estimation, Bilateral correlated data, Intra-class correlation coefficients, Likelihood ratio method, Wald-type method, Score method

Abstract: Bilateral correlated data are often encountered in ophthalmologic (or otolaryngologic) studies, in which each unit contributes information for paired organs to the studies, and the measurements from such paired organs are generally highly correlated. Various statistical methods have been developed to tackle intra-class correlation on bilateral correlated data analysis. In practice, it is important to adjust the effect of confounder on statistical inference, since either ignoring the intra-class correlation or confounding effect may lead to biased inference. In this article, we propose three test procedures for testing common risk difference for stratified bilateral correlated data in the basis of equal correlation model assumption. Five interval estimation of common difference of two proportions are derived. The performance of proposed test procedures and interval estimation is examined through Monte Carlo simulation. The simulation results show that the score test statistics outperforms other statistics in the sense that it produces robust type I error with high power. Score confidence interval with respect to score test statistics performs satisfactorily in terms of good coverage rate with reasonable interval width. One example from an otolaryngologic study is given to illustrate our methodologies.

Common risk difference test and interval estimation of risk difference for stratified bilateral correlated data

Xi Shen[1], Chang-Xing Ma[1], and Guo-Liang Tian[2]

[1]Department of Biostatistics, University at Buffalo, Buffalo, NY 14214, USA

[2]Department of Mathematics, Southern University of Science and Technology, Shenzhen 518055, Guangdong Province, P. R. China

Abstract

Bilateral correlated data are often encountered in ophthalmologic (or otolaryngologic) studies, in which each unit contributes information for paired organs to the studies, and the measurements from such paired organs are generally highly correlated. Various statistical methods have been developed to tackle intra-class correlation on bilateral

Corresponding author: Guo-Liang Tian, tiangl@sustc.edu.cn

correlated data analysis. In practice, it is important to adjust the effect of confounder on statistical inference, since either ignoring the intra-class correlation or confounding effect may lead to biased inference. In this article, we propose three test procedures for testing common risk difference for stratified bilateral correlated data in the basis of equal correlation model assumption. Five interval estimation of common difference of two proportions are derived. The performance of proposed test procedures and interval estimation is examined through Monte Carlo simulation. The simulation results show that the score test statistics outperforms other statistics in the sense that it produces robust type I error with high power. Score confidence interval with respect to score test statistics performs satisfactorily in terms of good coverage rate with reasonable interval width. One example from an otolaryngologic study is given to illustrate our methodologies.

Keywords: Common risk difference test, Interval estimation, Bilateral correlated data, strata, Intra-class correlation coefficients, Likelihood ratio method, Wald-type method, Score method.

1 Introduction

Paired correlated data are often collected from each study participant in many medical group comparative studies. For instance, in an ophthalmologic study, researchers are interested in comparison of two treatments. Participants are randomly administrated into one of two treatment groups. It is of great research interest to decide if the two treatments are clinically equivalent. The efficacy of treatment is evaluated by comparing the number of cured eyes at the end of treatment period in two treatment groups. The possible outcome can be summarized in a contingency table (the recorded outcome can be bilateral cured, unilateral cured and none cured). It is noteworthy that the measurements of both eyes from each participant are likely to be correlated. Under this assumption, various test procedures for assessing equality of proportions and confidence interval construction approaches for paired body part have been studied. Rosner [1] proposed a "constant R model" based on dependency assuming that the probability of a response at one side given a response at other side is proportional to the prevalence rate of corresponding group for ophthalmolohic data. This model was shown to empirically perform well, and Tang et al. [2], Ma et al. [3], Shan and Ma [4], and Liu et al. [5] have discussed its corresponding asymptotic and exact testing methods. However, the drawback of Rosner's model was reported by Dallal [6] that it could give a poor fit if the characteristic was almost certain to occur bilaterally with widely varying group-specific prevalence. Donner [7] further suggested an alternative model which assumes that all the treatment groups share a common intra-class correlation coefficient ("ρ model"). Thompson [8] evaluated this model by simulation and confirmed this model is robust for paired data, and various asymptotic and exact testing methods have been investigated by Tang et al. [9], Pei et al. [10], and Ma and Liu [11]. In addition, Confidence interval estimation for risk difference of proportions based on aforementioned two models has received considerable attention in statistical literature. For instance, Tang et al. [12] and Pei et al. [13] investigated asymptotic confidence interval construction in the basis of two pre-specified models for the difference of proportions between two groups. Additionally, Yang et al. [14] construct asymptotic confidence intervals for many-to-one comparisons of proportion differences with multiplicity adjustment.

However, another important feature for consideration in practice is confounding effect. Ignoring the confounding effect could yield incorrect inference. Stratified data analysis therefore arise. With the aforementioned models in hand, computational methods for testing or constructing confidence intervals on stratified data analysis for bilateral binary data have evolved dramatically these years. Pei et al. [15] proposed homogeneity test of proportion ratios for stratified bilateral data based on Donner's model. Tang and Qiu [16] applied Rosner's model on common difference test of two proportions, in which they specified that common difference is zero. Moreover, Shen and Ma [17] proposed three alternative maximum likelihood estimates based testing procedures for testing homogeneity of difference of two proportions for stratified correlated bilateral data under a common intra-cluster correlation assumption. Particularly, if we obtain the result of failing to reject the null hypothesis that the difference of two proportions are equal among strata, furthermore, the problem of interest may shift to test whether that equivalent value are equal to a specific value. Therefore, in this article, we develop several procedures for testing equality of difference of two proportions in a stratified bilateral sample design under a common intra-cluster correlation model with the condition that the MLEs are derived from the restriction of equal common difference, and construct asymptotic confidence intervals for that common difference.

The rest part of this article is organized as follows. In Section 2, we briefly delineate the data structure. Then the maximum likelihood estimates, three different test procedures and confidence interval estimators are introduced in Section 3. In Section 4, simulation study is conducted to investigate the performance of the three tests and five confidence intervals. A real example from otolaryngologic study is used to illustrate our proposed methods in section 5. Some concluding remark and future works are discussed in Section 6.

2 Data Structure

Suppose our purpose is to test if two treatments of some eye disease are clinical equivalent among different age strata in a medical comparative study. The data structure of interest appears as in table 1. A total of N_j patients are randomly allocated into one of two treatment groups for each age stratum. Let m_{lij} represent the number of patients having l (l=0,1,2) eyes with improvement responses in the i^{th} (i=1,2) group from the j^{th} (j=1,...,J) stratum, and $m_{\cdot ij} = \sum_{l=0}^{2} m_{lij}$ be the total number of patients in i^{th} group from j^{th} stratum. Let $Z_{hijk} = 1$ denote the improvement of the h^{th} (h=1,2) eye of k^{th} (k=1,2..., $m_{\cdot ij}$) patient in i^{th} group from j^{th} stratum, and 0 otherwise.

We assume that the probability of improvement at one eye for patients in the i^{th} group from j^{th} stratum is $Pr(Z_{hijk} = 1) = \pi_{ij}$ $(0 \leq \pi_{ij} \leq 1$, $h = 1, 2$, $i = 1, 2)$. Under the "ρ model" assumption (Donner [7]), let constant $\rho_{ij}(-1 \leq \rho_{ij} \leq 1)$ denote a measure of within subject correlation coefficients. We can straightforwardly show that the improvement probabilities for none, one, or both eyes in the i^{th} group form the j^{th} stratum are $(1 - \pi_{ij})(1 - \pi_{ij} + \rho_{ij}\pi_{ij})$, $2\pi_{ij}(1 - \rho_{ij})(1 - \pi_{ij})$, and $\pi_{ij}^2 + \rho_{ij}\pi_{ij}(1 - \pi_{ij})$. Note that we assume intra-cluster correlation coefficients from two groups are equal within each stratum, whereas they are different among strata. In what follows, we replace ρ_{ij} with ρ_j.

Table 1: Data Structure for the jth stratum in a bilateral design ($j = 1, 2, \dots, J$).

	Groups (i)		
Number of Responses (l)	1	2	Total
0	m_{01j}	m_{02j}	S_{0j}
1	m_{11j}	m_{12j}	S_{1j}
2	m_{21j}	m_{22j}	S_{2j}
Total	$m_{\cdot 1j}$	$m_{\cdot 2j}$	N_j

3 Proposed Methods

3.1 Testing Methods

We would like to test if the risk differences between two groups among all strata are equal to a common d_0:

H_0: $d_1 = d_2 = \cdots = d_J \triangleq d = d_0$, versus H_a: $d_1 = d_2 = \cdots = d_J \triangleq d \neq d_0$, where $d_j = \pi_{2j} - \pi_{1j}$.

The log-likelihood of a given table from each stratum and the overal log-likelihood can be expressed as

$$l_j(\pi_{1j}, \pi_{2j}, \rho_j; \mathbf{m}_j) = \sum_{i=1}^{2} [m_{0ij} \log((1 - \pi_{ij})(\rho_j \pi_{ij} - \pi_{ij} + 1))$$
$$+ m_{1ij} \log(2\pi_{ij}(1 - \rho_j)(1 - \pi_{ij})) + m_{2ij} \log(\pi_{ij}^2 + \rho_j \pi_{ij}(1 - \pi_{ij}))],$$

and

$$l = \sum_{j=1}^{J} l_j,$$

where $\mathbf{m}_j = (m_{01j}, m_{11j}, m_{21j}, m_{02j}, m_{12j}, m_{22j})$.

Global maximum likelihood estimates

Naively, we first derive related maximum likelihood estimates from a global setup. Setting the differentiation of l_j with respect to π_{ij}'s and ρ_j's equal to zero yields the MLEs of the parameters,

$$\begin{aligned}
\frac{\partial l}{\partial \pi_{ij}} &= \frac{(2\,\pi_{ij}-1)\;m_{1ij}}{\pi_{ij}\;(\pi_{ij}-1)} + \frac{m_{2ij}\;(\rho_j + 2\,\pi_{ij} - 2\,\rho_j\,\pi_{ij})}{\pi_{ij}\;(\rho_j + \pi_{ij} - \rho_j\,\pi_{ij})} \\
&\quad - \frac{m_{0i}\;(\rho_j + 2\,\pi_{ij} - 2\,\rho_j\,\pi_{ij} - 2)}{(\pi_{ij}-1)\;(\rho_j\,\pi_{ij} - \pi_{ij} + 1)}, i = 1, 2 \\
\frac{\partial l}{\partial \rho_j} &= \sum_{i=1}^{2} \left(\frac{m_{1ij}}{(\rho_j - 1)} - \frac{(\pi_{ij}-1)\;m_{2ij}}{(\rho_j + \pi_{ij} - \rho_j\,\pi_{ij})} + \frac{\pi_{ij}\,m_{0ij}}{(\rho_j\,\pi_{ij} - \pi_{ij} + 1)} \right),
\end{aligned}$$

denoting $\tilde{\pi}_{ij}$ and $\tilde{\rho}_j$ as the MLEs of π_{ij}'s and ρ_j, respectively. There is no closed form solution for above equations. We therefore consider implementing some iterative methods. Classical techniques, such as Newton-Raphson method and Fisher scoring method are usually recommended in these cases. However, the present problem involves high dimensional equations which may pose computational challenges. Therefore, these MLEs can be computed by repeating the following steps derived by Ma and Liu [11] and Shen and Ma [17].

We simplify the first equation as third order polynomial, then obtain the MLE of π_{ij} by solving the real root of it.

$$(4\rho_j - 2{\rho_j}^2 - 2)m_{ij}\pi_{ij}^3 + [3\rho_j^2 m_{ij} - \rho_j(5m_{0ij} + 6m_{1ij} + 7m_{2ij}) + 2m_{0ij} + 3m_{1ij} + 4m_{2ij}]\pi_{ij}^2$$

$$+[(4\rho_j - \rho_j^2)m_{ij} - 2\rho_j m_{0ij} - m_{1ij} - 2m_{2ij}]\pi_{ij} - \rho_j(m_{1ij} + m_{2ij}) = 0.$$

Then ρ_j can be updated by Fisher scoring method. The $(t+1)$th update for ρ_j is

$$\rho_j^{(t+1)} = \rho_j^{(t)} - \left(\frac{\partial^2 l}{\partial \rho_j^2}(\pi_{1j}^{(t)}, \pi_{2j}^{(t)}; \rho_j^{(t)})\right)^{-1} \frac{\partial l}{\partial \rho_j}(\pi_{1j}^{(t)}, \pi_{2j}^{(t)}; \rho_j^{(t)}),$$

where $j = 1, 2, \ldots, J$. The $(t+1)^{th}$ update of π_{ij} can be assessed by the solution of 3rd order polynomial with replacing ρ_j by $\rho_j^{(t+1)}$. Repeat the above steps until convergence. Formula of $\frac{\partial^2 l}{\partial \rho_j^2}$ will be given in Appendix.

Unconstrained maximum likelihood estimates

We now derive the unconstrained maximum likelihood estimates. Based on alternative hypothesis, π_{2j} can be expressed as $\pi_{1j}+d$, where $d \neq d_0$. Thus, the parameters here involve ρ_j, π_{1j}, and a common given d. Differentiation of l_j with respect to ρ_j, π_{1j}'s and d equal to zero yields the maximum likelihood estimates of the parameters $\hat{\rho}_j$, $\hat{\pi}_{1j}$ and $\hat{d}$.

These estimating equations cannot be solved to obtain close-form solutions. Similarly, We consider implementing the Fisher scoring or Newton-Raphson method. The major criticism of Fisher scoring or Newton Raphson method in high dimensional cases is its computational difficulty. Here, we follow Shen and Ma [17] and repeat their two-step approach. Firstly, we update common d by using Newton-Raphson method. Secondly, we apply Fisher scoring method to estimate π_{1j} and ρ_j with a given d from each stratum. The iteration procedure can be presented as follows.

1. The initial values of d and π_{1j} are set as $d^{(0)} = \frac{1}{J}\sum_{j=1}^{J}\tilde{d}_j$, $\pi_{1j}^{(0)} = \frac{1}{J}\sum_{j=1}^{J}\tilde{\pi}_{1j}$, $\rho_j^{(0)} = \frac{1}{J}\sum_{j=1}^{J}\tilde{\rho}_j$, where $\tilde{d}_j$, $\tilde{\pi}_{1j}$ and $\tilde{\rho}_j$ are MLEs from global setups.

2. Update
$$d^{(t+1)} = d^{(t)} - \frac{1}{I_1^{(t)}} \times V^{(t)},$$
where $V^{(t)} = \sum_{j=1}^{J}\frac{\partial l_j}{\partial d}({\pi_{1j}}^{(t)}, {\rho_j}^{(t)}, d^{(t)})$ and $I_1^{(t)} = \sum_{j=1}^{J}\frac{\partial^2 l_j}{\partial d^2}({\pi_{1j}}^{(t)}, {\rho_j}^{(t)}, d^{(t)})$. See Appendix for details.

3. Update
$$\begin{bmatrix} {\pi_{1j}}^{(t+1)} \\ {\rho_j}^{(t+1)} \end{bmatrix} = \begin{bmatrix} {\pi_{1j}}^{(t)} \\ {\rho_j}^{(t)} \end{bmatrix} + I_2^{-1}({\pi_{1j}}^{(t)}, {\rho_j}^{(t)}, d^{(t)}) \begin{bmatrix} \frac{\partial l}{\partial \pi_{1j}}({\pi_{1j}}^{(t)}, {\rho_j}^{(t)}, d^{(t)}) \\ \frac{\partial l}{\partial \rho_j}({\pi_{1j}}^{(t)}, {\rho_j}^{(t)}, d^{(t)}) \end{bmatrix}, j = 1, 2, \ldots, J,$$
where I_2 is the information matrix for π_{1j} and ρ_j.

The formula for I_2 and corresponding differential equations with respect to d will be given in Appendix.

4. Repeating the iteration process 2-3 until convergence.

After above iteration procedures, we assess MLEs under alternative hypothesis denoted by $(\hat{\pi}_{11}, \ldots, \hat{\pi}_{1J}; \hat{\rho}_1, \ldots, \hat{\rho}_J; \hat{d})$.

Constrained maximum likelihood estimates

Next, we investigate constrained maximum likelihood estimates. Under null hypothesis H_0: $d_1 = d_2 = \cdots = d_J \triangleq d = d_0$, π_{2j} can be expressed as $\pi_{1j} + d_0$, where d_0 is a given value from H_0. The maximum likelihood estimates of the parameter here only involve ρ_j and π_{1j}. Therefore, we can simply utilize the iterative step 3 from solving unconstrained MLEs with a given d_0. Denote $(\hat{\pi}_{11H_0}, \ldots, \hat{\pi}_{1JH_0}; \hat{\rho}_{1H_0}, \ldots, \hat{\rho}_{JH_0})$ as the constrained maximum likelihood estimates of nuisance parameter $(\pi_{11}, \ldots, \pi_{1J}; \rho_1, \ldots, \rho_J)$. With all MLEs derived, we consider following test procedures and confidence interval estimation approaches.

3.1.1 Likelihood ratio test (T_L)

The Likelihood ratio test statistic is given by

$$T_L = 2[l(\hat{\pi}_{11}, \hat{\pi}_{21}, \ldots, \hat{\pi}_{1J}, \hat{\pi}_{2J}; \hat{\rho}_1, \ldots, \hat{\rho_J}) - l(\hat{\pi}_{11H_0}, \hat{\pi}_{11H_0} + d_0 \ldots, \hat{\pi}_{1JH_0}, \hat{\pi}_{1JH_0} + d_0; \hat{\rho}_{1H_0}, \ldots, \hat{\rho}_{JH_0})],$$

which under the null hypothesis is asymptotically distributed as a chi-square distribution with 1 degree of freedom.

3.1.2 Wald-type test (T_W)

First, we rewrite the hypothesis as H_0: $d_1 = d_2 = \cdots = d_J \triangleq d = d_0$, versus H_a: $d_1 = d_2 = \cdots = d_J \triangleq d \neq d_0$, where $d_j = \pi_{2j} - \pi_{1j}$. Let $\beta = (d, \pi_{11}, \rho_1, \ldots, \pi_{1J}, \rho_J)$, the corresponding unconstrained MLE is $\hat{\beta} = (\hat{d}, \hat{\pi}_{11}, \hat{\rho}_1, \ldots, \hat{\pi}_{1J}, \hat{\rho}_J)$. Then, the MLE of d is $\hat{d} = K \times \hat{\beta}^T$, where K is a row vector that $K = (1, 0, \ldots 0)_{(2J+1)\times 1}$.

Wald-type test statistic has the form

$$T_W = \frac{(\hat{d} - d_0)^2}{\mathrm{Var}(\hat{d})} = \frac{(\hat{d} - d_0)^2}{K\mathrm{Var}(\hat{\beta}^T)K^T},$$

According to asymptotic normality property of MLEs, one can show that $\mathrm{Var}(\hat{\beta}^T) = \hat{I_n}^{-1}$, where I_n^{-1} is the inverse of the information matrix for β^T which will be derived in Appendix, and $\hat{I_n}^{-1}$ is the

MLE of ${I_n}^{-1}$. Therefore, we can rewrite the Wald-type statistic as

$$T_W = \frac{(\hat{d} - d_0)^2}{\hat{I}_n^{-1}(1,1)}$$

Here $I_n^{-1}(1,1)$ stands for the $(1,1)^{th}$ element of I_n^{-1}.

T_W is asymptotically distributed as a chi-square distribution with 1 degree of freedom.

3.1.3 Score test (T_{SC})

The score test statistic T_{SC} utilizes the MLEs under H_0. The score is a row vector: $\boldsymbol{U}(d, \boldsymbol{\pi}, \boldsymbol{\rho}) = \left(\frac{\partial l}{\partial d}, \frac{\partial l}{\partial \pi_{11}}, \frac{\partial l}{\partial \rho_1}, \ldots \frac{\partial l}{\partial \pi_{1J}}, \frac{\partial l}{\partial \rho_J}\right)$, where $\boldsymbol{\pi} = (\pi_{11}, \pi_{12}, \ldots, \pi_{1J})$ and $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots \rho_J)$.

Then T_{SC} for testing the equality of proportion difference is expressed as

$$T_{SC} = \boldsymbol{U}\boldsymbol{I}^{-1}\boldsymbol{U}^T|_{H_0},$$

where I is the information matrix for $\beta^T = (d, \pi_{11}, \rho_1, \ldots, \pi_{1J}, \rho_J)^T$. Here, d is the parameter of interest, π_{1j} and ρ_j are nuisance parameters. Therefore, the score function is $\boldsymbol{U} = (\frac{\partial l}{\partial d}, 0, 0, \ldots 0)|_{d=d_0}$. The test statistics can be simplified as

$$T_{SC} = (\sum_{j=1}^{J} \frac{\partial l_j}{\partial d})^2 I_n^{-1}(1,1),$$

where $I_n^{-1}(1,1)$ represents the $(1,1)^{th}$ element of I_n^{-1}, and formula of $\frac{\partial l_j}{\partial d}$ will be given in Appendix.

T_{SC} is asymptotically distributed as a chi-square distribution with 1 degree of freedom.

3.2 Confidence Interval Estimation

3.2.1 Global Wald-type CI and alternative Wald-type CI (GW, AW)

Recall that we derive MLE of $\beta = (\pi_{11}, \pi_{21}, \rho_1, \dots, \pi_{1J}, \pi_{2J}, \rho_J)$ from both global setup and alternative hypothesis, and denoted by $\tilde{\beta} = (\tilde{\pi}_{11}, \tilde{\pi}_{21}, \tilde{\rho}_1, \dots, \tilde{\pi}_{1J}, \tilde{\pi}_{2J}, \tilde{\rho}_J)$, and $\hat{\beta} = (\hat{\pi}_{11}, \hat{\pi}_{21}, \hat{\rho}_1, \dots \hat{\pi}_{1J}, \hat{\pi}_{2J}, \hat{\rho}_J)$ respectively, where $\tilde{\pi}_{2j} = \tilde{\pi}_{1j} + \tilde{d}$, and $\hat{\pi}_{2j} = \hat{\pi}_{1j} + \hat{d}$, $j = 1, \cdots J$.

Intuitively, we consider that there exists a weight w_j assigned to each estimate of d_j from different stratum. The choice of weights is not trivial. Here, we provide two examples. (1) Uniformly weighted: $w_j = \frac{1}{J}$. (2) Sample size weighted: $w_j = \frac{N_j}{N}$. where $j = 1, \dots J$, $\sum_{j=1}^{J} w_j = 1$, and N_j is the sample size from jth statum, N is the total number of outcomes.

We apply the algorithm to constuct confidence interval for d_0 by a row vector $\boldsymbol{W}$ and a constant matrix $\boldsymbol{K}$, where $\boldsymbol{W} = (w_1, w_2, \ldots w_J)$, and

$$\boldsymbol{K} = \begin{pmatrix} -1 & 1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & -1 & 1 & 0 & \cdots & \cdots & \cdots & \cdots \\ \vdots & & & \ddots & \ddots & \ddots & & & & \vdots \\ \vdots & & & & \ddots & \ddots & \ddots & & & \vdots \\ \vdots & & & & & \ddots & \ddots & \ddots & & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & -1 & 1 & 0 \end{pmatrix}_{J \times 3J}$$

Thus the MLEs of d from both setups can be obtained by a simple linear transformation:

$$\tilde{d} = \sum_{j=1}^{J} w_j \tilde{d}_j = \boldsymbol{C} \tilde{\beta}^T,$$

and

$$\hat{d} = \sum_{j=1}^{J} w_j \hat{d}_j = \boldsymbol{C} \hat{\beta}^T,$$

where $\boldsymbol{C} = \boldsymbol{W} \cdot \boldsymbol{K} = (-w_1, w_1, 0, -w_2, w_2, 0, \cdots, -w_j, w_j, 0)_{1 \times 3J}$.

It is easy to show that $\frac{(\tilde{d} - d_0)}{\sqrt{\mathrm{Var}(\tilde{d})}}$ and $\frac{(\hat{d} - d_0)}{\sqrt{\mathrm{Var}(\hat{d})}}$ are asymptotically distributed as the standard normal distribution as the sample size is large. In addition, according to the asymptotic normality property of MLE, we can express the variance of the d in terms of $\boldsymbol{C}$ and the information matrix of β^T, that is $\text{Var}(\boldsymbol{C}\beta^T) = \boldsymbol{C}I^{-1}\boldsymbol{C^T}$, where I is the information matrix of β^T.

Therefore, the $100(1-\alpha)\%$ confidence interval of d_0 based on above two setups are respectively, given by

$$[\max(-1, \tilde{d} - Z_{1-\alpha/2}\sqrt{\boldsymbol{C}\tilde{I}^{-1}\boldsymbol{C^T}}), \min(1, \tilde{d} + Z_{1-\alpha/2}\sqrt{\boldsymbol{C}\tilde{I}^{-1}\boldsymbol{C^T}})],$$

and

$$[\max(-1, \hat{d} - Z_{1-\alpha/2}\sqrt{\boldsymbol{C}\hat{I}^{-1}\boldsymbol{C^T}}), \min(1, \hat{d} + Z_{1-\alpha/2}\sqrt{\boldsymbol{C}\hat{I}^{-1}\boldsymbol{C^T}})],$$

where $Z_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution.

3.2.2 Complete Wald-type CI (W)

As aforementioned Wald-type test in section 3.1.2, $\frac{(\hat{d}-d_0)}{\sqrt{\text{Var}(\hat{d})}}$ is asymptotically distributed as the standard normal distribution, where $\text{Var}(\hat{d}) = \hat{I}_n^{-1}(1,1)$, $I_n^{-1}(1,1)$ is the $(1,1)^{th}$ element of the inverse of information matrix under H_a.

Therefore, the $100(1-\alpha)\%$ confidence interval for $d_0 \in [-1, 1]$ is defined as

$$[\max(-1, \hat{d} - Z_{1-\alpha/2}\sqrt{\hat{I_n}^{-1}(1,1)}), \min(1, \hat{d} + Z_{1-\alpha/2}\sqrt{\hat{I_n}^{-1}(1,1)})].$$

where $Z_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution.

3.2.3 Profile likelihood CI (PL)

With the pre-specified common test in section 3.1.1, we intuitively propose an approach to assess the confidence interval estimation from χ^2 distribution by inverting the likelihood ratio test of H_0: $d_1 = d_2 = \cdots = d_J \triangleq d = d_0$, versus H_a: $d_1 = d_2 = \cdots = d_J \triangleq d \neq d_0$, where $d_j = \pi_{2j} - \pi_{1j}$. Since the likelihood ratio test statistic follows χ^2 distribution with 1 degree of freedom under the null hypothesis, the $100(1-\alpha)\%$ confidence interval satisfies

$$2(l(\hat{d}_0, \hat{\pi_{1j}}, \hat{\rho_j}) - l(d_0, \hat{\pi}_{1jH_0}, \hat{\rho}_{jH_0})) \leq \chi^2_{1,1-\alpha},$$

where $\chi^2_{1,1-\alpha}$ is the upper $1-\alpha$ critical point of the Chi-square distribution with 1 degree of freedom.

Bisection method can be used to obtain the limits of above inequation(Yang, Liu, Liu, and Ma [14]). To assess the upper limit, the iteration procedure can be performed as follow.

1. Start with initial values $d^{(0)} = \hat{d}$, stepsize=0.1, flag=1, where $\hat{d}$ is unconstained MLE for d.

2. Update $\hat{d}^{(t+1)} = \hat{d}^{(t)}$+stepsize×flag, and compute constrained

MLE for $(\pi_{11}, \ldots, \pi_{1J}; \rho_1, \ldots, \rho_J)^{(t+1)}$. Then the log-likelihood can be calculated according to the constrained MLEs and the data, denoted by $\hat{l}^{(t+1)}$.

3. Evaluate the aforementioned requirement of confidence interval. If the condition of $2\times$ flag $\times(l(\hat{d}, \hat{\pi}_{11}, \ldots, \hat{\pi}_{1J}; \hat{\rho}_1, \ldots, \hat{\rho}_J)) - \hat{l}^{(t+1)}) \leq$ flag$\times\chi^2_{1,1-\alpha}$ is satisfied, return to step 2. Otherwise, we change the direction to search the bound. That is, set flag $= -$ flag, step size $= 0.1\times$ step size, then return to step 2.

4. Repeating the iteration process 2-3 until convergence.

Similarly, we can use the same iteration procedure with initial value $d^{(0)} = \hat{d}$, stepsize=0.1, flag=1 to assess the lower limit of confidence interval.

3.2.4 Score CI (SC)

Since the score test statistic follows χ^2 distribution with 1 degree of freedom under the null hypothesis, one can assess the $100(1-\alpha)\%$ confidence interval by including all $-1 \leq d_0 \leq 1$ which satisfies

$$T_{SC} \leq \chi^2_{1,1-\alpha},$$

where T_{SC} is the test statistics given in section 3.1.3, and$\chi^2_{1,1-\alpha}$ is the $(1-\alpha)$ quantile of the Chi-square distribution with 1 degree of freedom.

Similarly, bisection method is used to search the lower limit and upper limit.

4 Simulation Study

4.1 Common risk difference test

We now investigate the performance of the proposed statistics for testing the equality of risk difference.

We first evaluate the behavior of the type I error under different parameter settings, where we have $m = m_{\cdot 11} = m_{\cdot 21} = \cdots = m_{\cdot 1J} = m_{\cdot 2J}$=25, 50 or 100 in J=2, 4 or 8 strata respectively. The parameter setups are presented in Table 2, and we consider three values for common differences across strata under $H_0 : d_0 = 0$, 0.1 or 0.2, with various sets of parameters under different sample sizes. For each setup, 10,000 samples are randomly generated under null hypothesis and empirical type I error rates are computed by dividing the number of times of null hypothesis rejected by 10,000. All tests are conducted at 5% significance level.

Following Tang et al. [18], at 0.05 nomial level, we define a test is liberal if empirical type I error is greater than 0.06, conservative if the type I error is less than 0.04, and otherwise robust. The results (Tables 3-5) show that score test and likelihood ratio test are robust

Table 2: Parameter setups for computing empirical type I error and power.

	Cases	Number of strata J=2	J=4	J=8
ρ	I	$(0.2, 0.4)$	$(0.2, 0.4, 0.2, 0.4)$	$(0.2, 0.4, 0.2, 0.4, 0.2, 0.4, 0.2, 0.4)$
	II	$(0.3, 0.3)$	$(0.3, 0.3, 0.3, 0.3)$	$(0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3)$
	III	$(0.3, 0.5)$	$(0.3, 0.5, 0.3, 0.5)$	$(0.3, 0.5, 0.3, 0.5, 0.3, 0.5, 0.3, 0.5)$
	IV	$(0.6, 0.6)$	$(0.6, 0.6, 0.6, 0.6)$	$(0.6, 0.6, 0.6, 0.6, 0.6, 0.6, 0.6, 0.6)$
π_1	a	$(0.2, 0.4)$	$(0.2, 0.4, 0.2, 0.4)$	$(0.2, 0.4, 0.2, 0.4, 0.2, 0.4, 0.2, 0.4)$
	b	$(0.3, 0.3)$	$(0.3, 0.3, 0.3, 0.3)$	$(0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3)$
	c	$(0.4, 0.4)$.	$(0.4, 0.4, 0.4, 0.4)$	$(0.4, 0.4, 0.4, 0.4, 0.4, 0.4, 0.4, 0.4)$

in terms of satisfactory type I error for all scenarios. Wald-type test mostly works well at larger sample size ($m = 50$ or 100), but becomes inflated at smaller sample scenario ($m = 25$) and lower strata scenario ($J = 2$). Moreover, a set of boxplots (Figure 1) are displayed to show the distribution for empirical type I error for all the tests when we have balanced data for J=2, 4 or 8 respectively. We can observe that score test behaves satisfactorily, in the sense that its type I error is close to pre-determined nominal significant level $\alpha = 0.05$ for any configuration. Likelihood ratio test is inflated, while Wald-type test is even worse. However, as the sample size increases, both Likelihood ratio test and Wald-type test perform better.

Next, we investigate the performance of power for proposed test statistics under different parameters settings. To be specific, we consider the same sample size and parameter setups as we did for computing empirical type I error. Table 6 to Table 8 report empirical power associated with three proposed tests for various configurations. Since powers produce by three tests under different d_0 performs similarly, results from one case ($d_0 = 0.1$) are presented. We can observe that, under same parameter settings, the powers of different statistics are very close. Wald-type test tends to produce larger power than other two tests. Powers produced by all three tests increase when the difference between the true d(note by d_a) and d_0 increases. Powers increase when the number of strata J goes larger. Overall, score test procedure is highly recommended, since it is satisfactory on type I error control and has good performance on power.

4.2 Confidence interval estimation

We now investigate the performance of proposed confidence interval estimators of difference proposed in section 3 with one existing method from balanced to unbalanced designs in terms of empirical coverage probabilities (ECPs) and mean interval widths (MIWs). The ECP is defined as the proportion of events that d_0 falls within the constructed CI, and the MIW is calculated by dividing the sum of all widths by 10,000. Following Yang et al. [14], confidence interval can be constructed with pooling data, which the objective of interest is only the treatment group variable. We only present the result of the marginal confidence interval from score method(MSC). In addition, we construct Global Wald-type CIs with uniformly weighted adjustment and sample size weighted respectively, namely GW1 and GW2 and construct

Alternative Wald-type CIs with uniformly weighted adjustment and sample size weighted adjustment respectively, namely AW1 and AW2. The parameter setup is given in Table 9. Under each configuration, 10,000 Monte Carlo samples are generated, and 95% confidence interval is constructed for each replicate. Results are presented in Tables 10-12. Accordingly, we display a set of boxplots to investigate the distribution of ECPs and MIWs(Figure 2). Generally, CIs based on strata assumption outperform CIs based on Marginal model since the ECPs of those are closer than pre-determined confidence level. Among those CIs considering strata assumption, score CIs behave satisfactory, since the ECPs is closest to pre-determined confidence level, and MIWs are reasonable short. It is hence recommended. Likelihood ratio statistics produces CIs with shorter MIWs, but it yields deflated ECPs. Wald-type statistics(without weighted correction) can hardly well control its ECP, but produces the shortest MIW. The CIs based on Global Wald statistics with weighted correction (GW1 and GW2) and Alternative Wald statistics with weighted correction(AW1 and AW2) appear to perform poorly, especially when the number of strata is large(J=4 or J=8). Therefore, CIs produced from score statistics is strongly recommended in practice.

Table 9: Parameter setups for computing interval estimation.

		Number of strata		
	Cases	J=2	J=4	J=8
ρ	A	$(0.2, 0.3)$	$(0.2, 0.3, 0.2, 0.3)$	$(0.2, 0.3, 0.2, 0.3, 0.2, 0.3, 0.2, 0.3)$
	B	$(0.6, 0.6)$	$(0.6, 0.6, 0.6, 0.6)$	$(0.6, 0.6, 0.6, 0.6, 0.6, 0.6, 0.6, 0.6)$
π_1	a	$(0.3, 0.5)$	$(0.3, 0.5, 0.3, 0.5)$	$(0.3, 0.5, 0.3, 0.5, 0.3, 0.5, 0.3, 0.5$
	b	$(0.4, 0.4)$.	$(0.4, 0.4, 0.4, 0.4)$	$(0.4, 0.4, 0.4, 0.4, 0.4, 0.4, 0.4, 0.4)$

5 A real work example

We include a real work example in this section to further evaluate the performance of aforementioned approaches. Mandel (1982) et al [19] reported a data set from a double-blind randomized clinical trial to compare cefaclor and amoxicillin for the treatment of otitis media with effusion (OME) in children with bilateral tympanocentesis. Children with OME were randomized into two groups, and children in each group received a 14-day course with one of two antibiotics (amoxicillin or cefaclor). After the treatment, the number of cured ears for each child was recorded. We first classify the children as three age groups, and then discuss whether the cured rates between the amoxicillin or cefaclor among age are clinical equivalent, assuming that the differences of that cured rates are not influenced by the effect of age. We summarize the observed data in Table 13.

Table 13: Number of children whose ears has improvement cross different strata. (Group 1: Cefaclor; Group 2: Amoxicillin)

Age Groups	Age <2yrs		Age 2-5yrs		Age ≥6yrs	
Number of Responses	1	2	1	2	1	2
0	8	11	6	3	0	1
1	2	2	6	1	1	0
2	8	2	10	5	3	6
Total	18	15	22	9	4	7

Base on the data given above, parameter of estimates, test statistics values and corresponding p-values are reported in Table 14 and Table 15. We notice that all p-values are greater than the nominal level $\alpha = 0.05$, then it implies that there are no significant differences between two groups among strata for all proposed statistics.

Table 14: MLEs of parameters based on observed data.

	Global MLEs			Unconstrained MLEs			Constrained MLEs		
Age Groups	$\tilde{\rho}$	$\tilde{\pi}_1$	$\tilde{d}$	$\hat{\rho}$	$\hat{\pi}_1$	$\hat{d}$	$\hat{\rho}_{H_0}$	$\hat{\pi}_{1H_0}$	$\hat{d}_{H_0}$
Age <2yrs	0.7112	0.5000	-0.2904	0.7282	0.4017	-0.0945	0.7381	0.3636	0
Age 2-5yrs	0.5307	0.5881	0.0324	0.5330	0.6205		0.5308	0.5968	
Age ≥6yrs	0.6153	0.8341	0.0499	0.6332	0.8982		0.6140	0.8636	

Table 15: Statistic values and p-values of different test statistics with three strata.

	T_L	T_W	T_{SC}
Statistic	0.8845	0.9372	0.8537
p-value	0.3470	0.3330	0.3555

Next, we investigate the confidence interval estimator based on $d_0 = 0$. The results are presented in Table 16, and all the 8 CIs lead to the same result that there is no evidence to reject the null hypothesis at 5% nominal level.

Table 16: 95% CI of $d_0 = 0$ based on observed data.

	CI
W	$[-0.2859, 0.0969]$
PL	$[-0.2906, 0.1015]$
SC	$[-0.3954, 0.1018]$
GW1	$[-0.2622, 0.1234]$
GW2	$[-0.2909, 0.0947]$
AW1	$[-0.2885, 0.0994]$
AW2	$[-0.2885, 0.0994]$
MSC	$[-0.3138, 0.1016]$

6 Conclusions

In this article, we first consider test for common risk difference of two proportions on stratified bilateral correlated data. Three MLE based test procedures (likelihood ratio test, Wald-type test, score test) are investigated. Classical approaches, such as Fisher scoring and Newton Raphson method are usually criticized for computational difficulty in high dimensional cases. We derive two-step approaches for iteration process to obtain the unconstrained and constrained MLEs, which is very efficient. Then, we propose five confidence intervals of common difference of two proportions on stratified bilateral correlated data, which include two weight adjusted approaches (Global Wald-type confidence interval and Alternative Wald-type confidence interval) and three test based approaches.

Simulation study shows that (i) statistics derived from score test behaves satisfactorily in the sense that it has robust type I error, and reasonable power regardless of number of strata, sample size or parameter configurations. Wald-type test and Likelihood ratio test yield inflated type I error when sample size is relatively small. (ii) confidence interval estimation derived from score test statistics perform well in the sense that its ECPs is closest to pre-determined confidence level and MIWs is relatively short. As we expected, interval based on marginal model performs worse, since ignorance of the strata effect may lead to incorrect inference. For these reasons, we highly recommend the score test statistics in practical use for stratified bilateral-sample designs.

In this article, we consider the scenario in which we treat strata as nominal categories. In clinical trials, one interesting research goal is to test if there is a trend among the strata. Some information between strata may be ignored when there exists ordinal classifications relationship. We can further consider developing either asymptotic or exact trend test as future work.

Appendix A Information matrix and formula derivation

The second order differential equations from jth stratum respect to $\pi_{ij}, i=1,2$ and ρ_j yield

$$
\begin{aligned}
\frac{\partial^2 l}{\partial \pi_{ij}^2} &= -\frac{(2\,\pi_{ij}^2 - 2\,\pi_{ij} + 1)m_{1ij}}{\pi_{ij}^2\,(\pi_{ij}-1)^2} - \frac{(2\,\rho_j{}^2\,\pi_{ij}^2 - 2\,\rho_j{}^2\,\pi_{ij} + \rho_j{}^2 - 4\,\rho_j\,\pi_{ij}^2 + 2\,\rho_j\,\pi_{ij} + 2\,\pi_{ij}^2)m_{2ij}}{\pi_{ij}^2\,(\rho_j + \pi_{ij} - \rho_j\,\pi_{ij})^2} \\
&\quad - \frac{(2\,\rho_j{}^2\,\pi_{ij}^2 - 2\,\rho_j{}^2\,\pi_{ij} + \rho_j{}^2 - 4\,\rho_j\,\pi_{ij}^2 + 6\,\rho_j\,\pi_{ij} - 2\,\rho_j + 2\,\pi_{ij}^2 - 4\,\pi_{ij} + 2)m_{0ij}}{(\pi_{ij}-1)^2\,(\rho_j\,\pi_{ij} - \pi_{ij} + 1)^2}, \\
\frac{\partial^2 l}{\partial \pi_{ij}\partial \rho_j} &= \frac{m_{0ij}}{(\rho_j\,\pi_{ij} - \pi_{ij} + 1)^2} - \frac{m_{2ij}}{(\rho_j\,\pi_{ij} - \pi_{ij} - \rho_j)^2}, \\
& \qquad i = 1,2 \\
\frac{\partial^2 l}{\partial \pi_{ij}\partial \pi_{kj}} &= 0, i \neq k, \\
\frac{\partial^2 l}{\partial \rho_j^2} &= -\sum_{i=1}^{2}\left[\frac{m_{1ij}}{(\rho_j - 1)^2} + \frac{\pi_{ij}^2\,m_{0ij}}{(\rho_j\,\pi_{ij} - \pi_{ij} + 1)^2} + \frac{(\pi_{ij}-1)^2\,m_{2ij}}{(\rho_j + \pi_{ij} - \rho_j\,\pi_{ij})^2}\right].
\end{aligned}
$$

Then from jth stratum,we have information matrix express as,

$$
I_j(\pi_{ij}, \rho_j) = \begin{bmatrix} I_{11(j)} & 0 & I_{13(j)} \\ 0 & I_{22(j)} & I_{23(j)} \\ I_{13(j)} & I_{23(j)} & I_{33(j)} \end{bmatrix},
$$

where

$$I_{ii(j)} = E\left(-\frac{\partial^2 l}{\partial \pi_{ij}^2}\right) = \frac{m_{\cdot ij}\ \left(-4\,\rho_j{}^2\,\pi_{ij}^2 + 4\,\rho_j{}^2\,\pi_{ij} - \rho_j{}^2 + 6\,\rho_j\,\pi_{ij}^2 - 6\,\rho_j\,\pi_{ij} + 2\,\rho_j - 2\,\pi_{ij}^2 + 2\,\pi_{ij}\right)}{\pi_{ij}\ (1-\pi_{ij})\ (\rho_j + \pi_{ij} - \rho_j\,\pi_{ij})\ (\rho_j\,\pi_{ij} - \pi_{ij} + 1)}$$

$$i = 1, 2$$

$$I_{i3(j)} = E\left(-\frac{\partial^2 l}{\partial \pi_{ij}\partial \rho_j}\right) = \frac{m_{\cdot ij}\rho_j\ (2\,\pi_{ij} - 1)}{(\rho_j + \pi_{ij} - \rho_j\,\pi_{ij})\ (\rho_j\,\pi_{ij} - \pi_{ij} + 1)},$$

$$i = 1, 2$$

$$I_{33(j)} = E\left(-\frac{\partial^2 l}{\partial \rho_j^2}\right) = \sum_{i=1}^{2} \frac{m_{\cdot ij}\,\pi_{ij}\ (\rho_j + 1)\ (1-\pi_{ij})}{(1-\rho_j)\ (\rho_j + \pi_{ij} - \rho_j\,\pi_{ij})\ (\rho_j\,\pi_{ij} - \pi_{ij} + 1)}.$$

Therefore, information matrix for J strata has the form that

$$I = \begin{bmatrix} I_1(\pi_{i1}, \rho_1) & & & \\ & I_2(\pi_{i2}, \rho_2) & & \\ & & \ddots & \\ & & & I_J(\pi_{iJ}, \rho_J) \end{bmatrix}_{3J \times 3J},$$

The inverse of the information matrix is

$$I^{-1} = \begin{bmatrix} I_1(\pi_{i1}, \rho_1)^{-1} & & & \\ & I_2(\pi_{i2}, \rho_2)^{-1} & & \\ & & \ddots & \\ & & & I_J(\pi_{iJ}, \rho_J)^{-1} \end{bmatrix}_{3J \times 3J},$$

where $I_j^{-1}(\pi_{ij}, \rho_j) = \frac{1}{k(j)} \times z(j)$,

$$z(j) = \begin{bmatrix} {I_{23(j)}}^2 - I_{22(j)}\, I_{33(j)} & -I_{13(j)}\, I_{23(j)} & I_{22(j)}\, I_{13(j)} \\ -I_{13(j)}\, I_{23(j)} & {I_{13(j)}}^2 - I_{11(j)}\, I_{33(j)} & I_{11(j)}\, I_{23(j)} \\ I_{22(j)}\, I_{13(j)} & I_{11(j)}\, I_{23(j)} & -I_{11(j)}\, I_{22(j)} \end{bmatrix},$$

$$k(j) = I_{22(j)}\, {I_{13(j)}}^2 + I_{11(j)}\, {I_{23(j)}}^2 - I_{11(j)}\, I_{22(j)}\, I_{33(j)}.$$

$i = 1, 2,\ j = 1, \ldots, J.$

Let π_{2j} denote $\pi_{1j} + d,\ j = 1, 2, \ldots J$. The first order and second order differential equations from jth stratum respect to d are

$$\frac{\partial l_j}{\partial d} = \frac{m_{02j}(2\pi_{2j}\rho_j - \rho_j - 2\pi_{2j} + 2)}{(\pi_{2j}\rho_j - \pi_{2j} + 1)(\pi_{2j} - 1)} + \frac{m_{12j}(2\pi_{2j} - 1)}{\pi_{2j}(\pi_{2j} - 1)} + \frac{m_{22j}(2\pi_{2j}\rho_j - 2\pi_{2j} - \rho_j)}{\pi_{2j}^2\rho_j - \pi_{2j}\rho_j - \pi_{2j}^2}$$

$$\frac{\partial^2 l_j}{\partial d^2} = -\frac{m_{02j}(2\pi_{2j}^2\rho_j^2 - 4\pi_{2j}^2\rho_j + 2\pi_{2j}^2 - 2\pi_{2j}\rho_j^2 + 6\pi_{2j}\rho_j - 4\pi_{2j} + \rho_j^2 - 2\rho_j + 2)}{(\pi_{2j}\rho_j - \pi_{2j} + 1)^2(\pi_{2j} - 1)^2}$$
$$+ \frac{m_{12j}(-2\pi_{2j}^2 + 2\pi_{2j} - 1)}{\pi_{2j}^2(\pi_{2j} - 1)^2} - \frac{m_{22j}(2\pi_{2j}^2\rho_j^2 - 2\pi_{2j}^2\rho_j + 2\pi_{2j}^2 - 2\pi_{2j}\rho_j^2 + 2\pi_{2j}\rho j + \rho_j^2)}{\pi_{2j}^2(\pi_{2j}\rho_j - \pi_{2j} - \rho_j)^2}$$

Moreover, with a given d information matrix I_2 for π_{1j} and ρ_j is

$$I(\pi_{1j}, \rho_j, d) = \begin{bmatrix} I_{11(j)} & I_{12(j)} \\ I_{12(j)} & I_{22(j)} \end{bmatrix}.$$

Thus, the inverse of the information matrix can be expressed as

$$I^{-1}(\pi_{1j}, \rho_j, d) = \frac{1}{I_{11(j)} \times I_{22(j)} - I_{12(j)}^2} \begin{bmatrix} -I_{22(j)} & I_{12(j)} \\ I_{12(j)} & -I_{11(j)} \end{bmatrix},$$

where

$$\begin{aligned} I_{11(j)} &= E\left(-\frac{\partial^2 l}{\partial \pi_{1j}^2}\right) = \sum_{i=1}^{2} \frac{m_{\cdot ij}\ (-4\,\pi_{ij}^2\,{\rho_j}^2 + 6\,\pi_{ij}^2\,\rho_j - 2\,\pi_{ij}^2 + 4\,\pi_{ij}\,{\rho_j}^2 - 6\,\pi_{ij}\,\rho_j + 2\,\pi_{ij} - {\rho_j}^2 + 2}{\pi_{ij}\ (1-\pi_{ij})\ (\pi_{ij}\,\rho_j - \pi_{ij} + 1)\ (\pi_{ij} + \rho_j - \pi_{ij}\,\rho_j)} \\ I_{12(j)} &= E\left(-\frac{\partial^2 l}{\partial \pi_{1j} \partial \rho_j}\right) = \sum_{i=1}^{2} \frac{m_{\cdot ij}\,\rho_j\ (2\,\pi_{ij} - 1)}{(\pi_{ij}\,\rho_j - \pi_{ij} + 1)\ (\pi_{ij} + \rho_j - \pi_{ij}\,\rho_j)}, \\ I_{22(j)} &= E\left(-\frac{\partial^2 l}{\partial \rho_j^2}\right) = \sum_{i=1}^{2} m_{\cdot ij} \left(\frac{2\,\pi_{ij}\ (\pi_{ij} - 1)}{\rho_j - 1} - \frac{\pi_{ij}^2\ (\pi_{ij} - 1)}{\pi_{ij}\,\rho_j - \pi_{ij} + 1} + \frac{\pi_{ij}\,(\pi_{ij} - 1)^2}{\pi_{ij} + \rho_j - \pi_{ij}\,\rho_j}\right), \end{aligned}$$

References

[1] B. Rosner. Statistical methods in ophthalmology:an adjustment for the intraclass correlation between eyes. *Biometrics*, 38:105–114, March 1982.

[2] M.-L Tang, N.-S Tang, and B. Rosner. Statistical inference for correlated data in ophthalmologic studies. *Statistics in Medicine*, 25(16):2771–2783, 2006.

[3] C.-X Ma, G. Shan, and S. Liu. Homogeneity test for binary correlated data. *PLoS ONE*, 10(4):e0124337, 2015.

[4] G. Shan and C.-X Ma. Exact methods for testing the equality of proportions for clustered data from otolaryngologic studies. *Statistics in biopharmaceutical research*, 6(1):115–122, 2014.

[5] X. Liu, G. Shan, L. Tian, and C.-X Ma. Exact methods for testing homogeneity of proportions for correlated multiple groups paired binary data. *Communications in Statistics - Simulation and Computation*, 2016.

[6] G. E. Dallal. Paired bernoulli trials. *Biometrics*, 44:253–257, March 1988.

[7] A. Donner. Statistical methods in opthalmology: an adjusted chi-square approach. *Biometrics*, 45(2):605–611, June 1989.

[8] J. R. Thompson. The chi-square test for data collected on eyes. *British Journal of Ophthalmology*, 77:115–117, 1993.

[9] N.-S Tang, S.-F Qiu, M.-L Tang, and Y.-B Pei. Asymptotic confidence interval construction for proportion difference in medical studies with bilateral data. *Statistical Methods in Medical Research*, 20(3):233–259, June 2011.

[10] Y.-B Pei, M.-L Tang, and J.-H Guo. Testing the equality of two proportions for combined unilateral and bilateral data. *Communications in Statistics - Simulation and Computation*, 37:1515–1529, 2008.

[11] C.-X Ma and S. Liu. Testing equality of proportions for correlated binary data in ophthalmologic studies. *Journal of Biopharmaceutical Statistics*, pages 1–9, 2016.

[12] N.-S Tang, S.-F Qiu, M.-L Tang, and Y.-B Pei. Asymptotic confidence interval construction for proportion difference in medical studies with bilateral data. *Statistical Methods in Medical Research*, 20(3):233–259, 2011.

[13] Y.-B Pei, M.-L Tang, W.-K Wong, and J.-H Guo. Confidence intervals for correlated proportion differences from paired data in a two-arm randomised clinical trial. *Statistical Methods in Medical Research*, 21(2):167–187, 2012.

[14] Z.-Y Yang, X. Liu, S. Liu, and C.-X Ma. Simultaneous confidence interval construction for many-to-one comparisons of proportion differences based on correlated paired data. *Unpublished manuscript*, 2017.

[15] Y.-B Pei, G.-L Tian, and M.-L Tang. Testing homogeneity of proportion ratios for stratified correlated bilateral data in two-arm randomized clinical trials. *Statistics in Medicine*, 33:4370–4386, 2014.

[16] N.-S Tang and S.-F Qiu. Homogeneity test,sample size determination and interval construction of difference of two proportions in

stratified bilateral-sample designs. *Journal of Statisticsl Planning and Inference*, 142(5):1242–1251, 2012.

[17] X. Shen and C.-X Ma. Testing homogeneity of difference of two proportions for stratified correlated paired binary data. *Journal of Applied Statistics*, pages 1–16, 2017.

[18] N.-S Tang, M.-L Tang, and S.-F Qiu. Testing the equality of proportions for correlated otolaryngologic data. *Computational Statistics and Data Analysis*, 52(7):2719–3729, 2008.

[19] E. M. Mandel et al. Duration of effusion after antibiotic treatment for acute otitis media: comparison of cefaclor and amoxicillin. *Pediatric Infectious Disease*, 1(5):310–316, 1982.

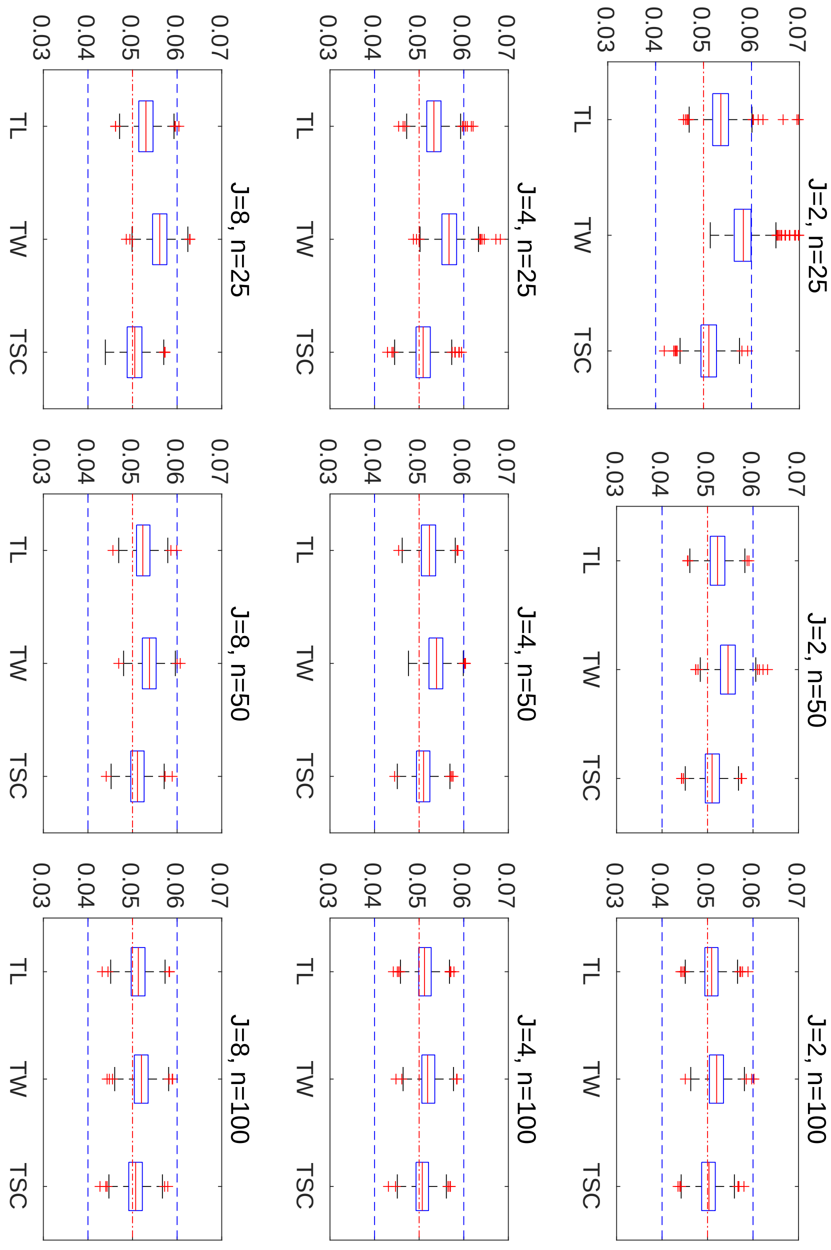

Figure 1: Box-plots of empirical sizes

Table 3: Simulation results of the empirical sizes for 2 strata.

			$m=25$			$m=50$			$m=100$		
d	ρ	π_1	T_L	T_W	T_{SC}	T_L	T_W	T_{SC}	T_L	T_W	T_{SC}
0	I	a	4.58	5.05	4.31	5.04	5.29	4.92	5.06	5.23	4.94
		b	5.21	5.58	4.94	5.15	5.36	5.03	5.22	5.34	5.17
		c	5.64	**6.09**	5.41	5.26	5.48	5.17	5.53	5.62	5.50
	II	a	5.02	5.47	4.78	5.21	5.41	5.07	5.29	5.38	5.23
		b	5.44	5.87	5.11	5.33	5.60	5.22	5.28	5.37	5.18
		c	5.38	5.77	5.15	5.21	5.36	5.09	5.27	5.37	5.23
	III	a	4.78	5.28	4.60	5.20	5.40	5.04	4.95	5.00	4.86
		b	5.34	5.68	5.02	5.23	5.44	5.09	5.08	5.19	4.99
		c	5.68	**6.03**	5.44	5.27	5.45	5.14	4.88	5.02	4.84
	IV	a	5.39	5.96	5.20	5.17	5.37	5.02	5.26	5.34	5.15
		b	5.23	5.64	4.96	5.08	5.30	4.97	5.38	5.53	5.28
		c	4.89	5.26	4.63	5.24	5.47	5.09	4.95	5.00	4.93
0.1	I	a	5.08	5.49	4.85	4.89	5.15	4.79	5.11	5.19	5.03
		b	4.98	5.41	4.75	5.26	5.41	5.09	5.01	5.20	4.94
		c	5.84	**6.32**	5.52	5.07	5.28	5.01	4.74	4.79	4.71
	II	a	5.03	5.48	4.71	5.02	5.24	4.90	5.19	5.27	5.09
		b	5.48	5.91	5.23	5.19	5.38	5.08	5.32	5.42	5.23
		c	5.27	5.70	5.02	5.54	5.77	5.38	5.42	5.55	5.38
	III	a	4.96	5.45	4.73	5.56	5.74	5.39	5.31	5.42	5.24
		b	5.31	5.71	5.10	4.81	5.05	4.71	5.18	5.28	5.10
		c	5.40	6.00	5.13	4.99	5.19	4.94	5.92	**6.03**	5.90
	IV	a	5.16	5.47	4.86	5.49	5.67	5.28	5.22	5.31	5.14
		b	5.28	5.73	4.99	5.11	5.36	5.01	5.15	5.24	5.09
		c	5.52	5.92	5.31	4.91	5.09	4.85	5.21	5.38	5.18
0.2	I	a	5.49	5.99	5.20	5.19	5.39	5.09	5.27	5.40	5.19
		b	5.41	5.94	5.09	5.42	5.63	5.31	5.11	5.14	5.06
		c	5.37	5.85	5.05	5.00	5.15	4.88	5.23	5.33	5.18
	II	a	5.06	5.52	4.76	5.11	5.34	4.96	5.47	5.56	5.40
		b	4.74	5.05	4.47	5.23	5.40	5.07	5.00	5.15	4.98
		c	5.56	5.97	5.32	5.41	5.62	5.23	5.00	5.05	4.95
	III	a	5.16	5.68	5.02	5.26	5.56	5.22	4.90	5.00	4.86
		b	5.26	5.72	5.06	4.78	5.07	4.73	4.97	5.05	4.91
		c	5.16	5.59	5.02	4.89	5.24	4.85	5.08	5.21	5.05
	IV	a	5.28	5.58	5.02	5.37	5.70	5.21	5.53	5.65	5.50
		b	5.51	5.97	5.20	5.36	5.59	5.31	5.22	5.29	5.19
		c	4.81	5.47	4.78	4.78	4.90	4.75	4.61	4.76	4.56

Table 4: Simulation results of the empirical sizes for 4 strata.

			$m = 25$			$m = 50$			$m = 100$		
d	ρ	π_1	T_L	T_W	T_{SC}	T_L	T_W	T_{SC}	T_L	T_W	T_{SC}
0	I	a	4.61	4.89	4.22	5.20	5.39	5.04	4.83	4.90	4.74
		b	5.10	5.42	4.85	4.90	5.13	4.75	5.34	5.48	5.30
		c	5.63	5.90	5.40	5.03	5.18	4.87	5.01	5.12	4.96
	II	a	5.22	5.41	4.86	5.34	5.54	5.24	4.91	5.02	4.85
		b	5.09	5.30	4.75	5.28	5.45	5.13	4.94	4.99	4.89
		c	5.56	5.86	5.36	5.23	5.36	5.15	4.92	4.96	4.87
	III	a	4.69	5.03	4.50	4.64	4.80	4.46	4.98	5.08	4.86
		b	5.50	5.89	5.22	4.90	5.10	4.74	5.03	5.12	5.00
		c	5.46	5.81	5.30	5.49	5.60	5.41	4.87	4.92	4.84
	IV	a	5.17	5.59	4.94	5.21	5.42	5.00	5.08	5.19	5.00
		b	5.29	5.63	5.01	5.10	5.24	4.99	5.04	5.10	5.00
		c	5.13	5.44	4.93	5.12	5.26	5.02	5.19	5.27	5.13
0.1	I	a	4.92	5.35	4.70	5.37	5.57	5.16	5.20	5.26	5.11
		b	5.74	**6.06**	5.47	5.28	5.48	5.22	5.40	5.50	5.31
		c	5.24	5.50	5.10	5.15	5.22	5.02	4.96	5.02	4.93
	II	a	5.10	5.51	4.86	5.06	5.29	4.90	4.79	4.90	4.69
		b	5.36	5.67	5.15	5.36	5.55	5.20	5.30	5.37	5.26
		c	5.52	5.74	5.34	5.18	5.40	5.07	5.25	5.35	5.20
	III	a	5.06	5.40	4.77	5.37	5.53	5.21	5.04	5.17	5.00
		b	5.20	5.63	4.96	4.88	5.06	4.73	5.16	5.30	5.12
		c	5.34	5.61	5.20	5.27	5.42	5.19	4.95	5.03	4.92
	IV	a	4.89	5.29	4.66	5.46	5.61	5.33	5.25	5.34	5.17
		b	5.19	5.53	5.00	4.89	5.10	4.75	5.23	5.30	5.18
		c	5.29	5.55	5.11	5.35	5.46	5.28	5.32	5.39	5.26
0.2	I	a	4.99	5.31	4.80	4.96	5.23	4.83	5.40	5.48	5.32
		b	5.15	5.34	4.88	4.82	4.86	4.73	4.88	4.92	4.92
		c	5.21	5.50	4.91	5.43	5.58	5.35	5.66	5.63	5.62
	II	a	4.88	5.20	4.63	5.23	5.44	5.14	4.97	4.98	4.96
		b	5.36	5.64	5.06	5.53	5.78	5.45	5.24	5.25	5.23
		c	5.33	5.65	5.15	5.33	5.47	5.19	5.17	5.23	5.13
	III	a	5.00	5.40	4.82	5.13	5.32	5.05	5.31	5.38	5.26
		b	5.50	5.81	5.29	4.90	5.07	4.79	5.21	5.26	5.16
		c	5.30	5.71	5.13	5.09	5.34	5.04	5.44	5.51	5.41
	IV	a	5.14	5.50	5.07	5.17	5.32	5.11	5.62	5.69	5.61
		b	5.38	5.72	5.23	5.57	5.72	5.48	5.28	5.38	5.27
		c	5.13	5.48	5.02	5.15	5.31	5.07	5.60	5.54	5.59

Table 5: Simulation results of the empirical sizes for 8 strata.

			$m = 25$			$m = 50$			$m = 100$		
d	ρ	π_1	T_L	T_W	T_{SC}	T_L	T_W	T_{SC}	T_L	T_W	T_{SC}
0	I	a	4.57	4.93	4.20	5.10	5.34	4.96	5.04	5.14	4.99
		b	5.01	5.34	4.80	5.65	5.81	5.55	5.14	5.20	5.06
		c	5.01	5.27	4.77	5.27	5.40	5.13	5.40	5.45	5.35
	II	a	4.84	5.14	4.44	5.02	5.24	4.81	5.42	5.56	5.34
		b	4.93	5.24	4.61	5.52	5.68	5.27	4.88	4.94	4.82
		c	5.29	5.52	5.02	4.88	5.00	4.75	5.36	5.43	5.31
	III	a	5.27	5.56	4.91	5.22	5.38	5.06	5.35	5.44	5.25
		b	5.12	5.40	4.86	5.38	5.58	5.26	5.36	5.41	5.26
		c	5.36	5.56	5.19	5.19	5.37	5.03	4.80	4.83	4.75
	IV	a	4.77	5.25	4.48	5.24	5.50	5.09	5.08	5.19	5.00
		b	5.52	5.84	5.23	5.08	5.19	4.93	5.61	5.68	5.51
		c	5.39	5.56	5.31	4.94	5.06	4.90	5.04	5.09	5.04
0.1	I	a	4.83	5.24	4.56	5.11	5.22	4.92	5.21	5.25	5.12
		b	5.09	5.33	4.85	5.47	5.58	5.36	5.46	5.52	5.42
		c	5.14	5.43	4.95	5.50	5.58	5.41	5.10	5.13	5.05
	II	a	5.18	5.43	4.92	5.31	5.46	5.13	4.75	4.81	4.71
		b	5.48	5.74	5.27	4.88	5.01	4.71	5.22	5.26	5.14
		c	5.28	5.49	5.10	5.31	5.39	5.17	4.73	4.76	4.69
	III	a	5.14	5.49	4.91	5.41	5.54	5.24	5.20	5.27	5.13
		b	5.34	5.63	5.08	5.20	5.28	5.10	5.21	5.25	5.13
		c	5.23	5.43	5.07	5.08	5.14	4.99	4.90	4.94	4.89
	IV	a	5.26	5.62	5.00	4.75	4.87	4.63	5.00	5.06	4.95
		b	5.25	5.53	5.01	5.30	5.43	5.22	4.98	5.03	4.96
		c	5.52	5.75	5.34	5.13	5.20	5.10	5.42	5.43	5.41
0.2	I	a	5.47	5.75	5.28	5.31	5.44	5.20	5.07	5.14	5.00
		b	5.99	**6.02**	5.46	5.15	5.22	5.01	4.88	4.98	4.85
		c	5.63	5.78	5.33	5.29	5.46	5.15	5.05	5.18	4.99
	II	a	5.12	5.38	4.81	4.98	5.07	4.90	5.58	5.63	5.50
		b	5.48	5.52	5.12	4.97	5.10	4.87	4.96	5.03	4.91
		c	5.45	5.64	5.22	5.09	5.17	5.00	5.25	5.22	5.16
	III	a	5.55	5.81	5.29	4.92	5.09	4.83	5.17	5.22	5.10
		b	5.40	5.55	5.16	5.31	5.32	5.24	4.94	4.93	4.91
		c	5.46	5.59	5.27	5.35	5.46	5.23	5.24	5.29	5.23
	IV	a	5.12	5.39	4.87	5.19	5.30	5.12	5.31	5.40	5.28
		b	5.12	5.36	5.03	5.16	5.22	5.06	5.04	5.05	4.96
		c	5.07	5.22	4.94	5.04	5.14	4.98	5.03	5.01	4.93

Table 6: Part of simulation results of the empirical powers for 2 strata (where $H_0 : d_0 = 0.1,\ H_A : d_a = 0.05, 0.15$ or 0.25).

			$m = 25$			$m = 50$			$m = 100$		
d_a	ρ	π_1	T_L	T_W	T_{SC}	T_L	T_W	T_{SC}	T_L	T_W	T_{SC}
0.05	I	a	0.109	0.113	0.105	0.172	0.175	0.170	0.292	0.293	0.292
		b	0.106	0.109	0.103	0.159	0.161	0.158	0.267	0.267	0.267
		c	0.101	0.106	0.100	0.155	0.156	0.154	0.247	0.246	0.247
	II	a	0.109	0.114	0.105	0.174	0.177	0.172	0.286	0.286	0.284
		b	0.109	0.114	0.106	0.163	0.165	0.161	0.281	0.283	0.281
		c	0.106	0.111	0.102	0.169	0.170	0.167	0.269	0.269	0.269
	III	a	0.111	0.114	0.108	0.148	0.150	0.147	0.252	0.252	0.252
		b	0.108	0.112	0.106	0.147	0.148	0.147	0.241	0.240	0.242
		c	0.099	0.102	0.097	0.137	0.138	0.137	0.223	0.222	0.223
	IV	a	0.099	0.104	0.095	0.148	0.151	0.147	0.244	0.246	0.243
		b	0.099	0.102	0.096	0.138	0.140	0.138	0.231	0.231	0.231
		c	0.094	0.097	0.093	0.129	0.129	0.129	0.212	0.211	0.213
0.15	I	a	0.098	0.106	0.092	0.162	0.170	0.158	0.271	0.275	0.267
		b	0.111	0.120	0.105	0.155	0.162	0.150	0.264	0.271	0.261
		c	0.101	0.111	0.095	0.146	0.153	0.142	0.246	0.252	0.243
	II	a	0.102	0.110	0.096	0.160	0.166	0.156	0.275	0.281	0.272
		b	0.102	0.110	0.097	0.153	0.160	0.149	0.254	0.260	0.251
		c	0.101	0.111	0.096	0.153	0.162	0.149	0.255	0.262	0.252
	III	a	0.100	0.110	0.095	0.151	0.157	0.147	0.239	0.244	0.234
		b	0.100	0.109	0.096	0.147	0.156	0.144	0.244	0.249	0.241
		c	0.098	0.108	0.094	0.147	0.154	0.143	0.225	0.232	0.222
	IV	a	0.097	0.108	0.092	0.141	0.147	0.137	0.225	0.229	0.223
		b	0.092	0.101	0.087	0.136	0.142	0.133	0.223	0.230	0.220
		c	0.091	0.101	0.086	0.129	0.136	0.125	0.213	0.220	0.211
0.25	I	a	0.502	0.523	0.490	0.792	0.801	0.788	0.979	0.980	0.978
		b	0.492	0.514	0.481	0.780	0.788	0.773	0.970	0.971	0.970
		c	0.483	0.507	0.471	0.762	0.773	0.757	0.969	0.969	0.967
	II	a	0.513	0.534	0.500	0.799	0.808	0.794	0.978	0.979	0.977
		b	0.490	0.512	0.480	0.773	0.783	0.767	0.971	0.972	0.970
		c	0.482	0.504	0.471	0.775	0.787	0.769	0.971	0.972	0.970
	III	a	0.468	0.490	0.458	0.746	0.757	0.742	0.960	0.962	0.959
		b	0.481	0.501	0.469	0.758	0.770	0.753	0.967	0.968	0.966
		c	0.449	0.471	0.438	0.737	0.747	0.731	0.954	0.956	0.953
	IV	a	0.435	0.452	0.425	0.714	0.725	0.709	0.946	0.949	0.945
		b	0.409	0.430	0.399	0.686	0.698	0.680	0.930	0.933	0.929
		c	0.402	0.424	0.391	0.675	0.688	0.669	0.922	0.925	0.921

Table 7: Part of simulation results of the empirical powers for 4 strata (where $H_0 : d_0 = 0.1,\ H_A : d_a = 0.05, 0.15$ or 0.25).

			$m = 25$			$m = 50$			$m = 100$		
d_a	ρ	π_1	T_L	T_W	T_{SC}	T_L	T_W	T_{SC}	T_L	T_W	T_{SC}
0.05	I	a	0.176	0.182	0.170	0.292	0.297	0.288	0.512	0.515	0.510
		b	0.159	0.162	0.156	0.272	0.274	0.270	0.473	0.473	0.471
		c	0.151	0.155	0.148	0.241	0.241	0.239	0.436	0.435	0.436
	II	a	0.180	0.186	0.173	0.284	0.289	0.281	0.504	0.505	0.503
		b	0.171	0.178	0.165	0.277	0.280	0.274	0.482	0.483	0.481
		c	0.163	0.166	0.159	0.275	0.276	0.272	0.470	0.470	0.469
	III	a	0.154	0.158	0.151	0.254	0.256	0.253	0.452	0.452	0.452
		b	0.152	0.155	0.151	0.245	0.246	0.244	0.430	0.429	0.430
		c	0.146	0.149	0.145	0.228	0.228	0.228	0.400	0.399	0.401
	IV	a	0.155	0.161	0.150	0.258	0.261	0.255	0.434	0.437	0.432
		b	0.143	0.146	0.141	0.229	0.231	0.228	0.411	0.411	0.411
		c	0.129	0.131	0.129	0.209	0.208	0.209	0.368	0.367	0.368
0.15	I	a	0.152	0.161	0.145	0.272	0.279	0.268	0.479	0.484	0.475
		b	0.160	0.168	0.154	0.265	0.271	0.260	0.453	0.457	0.450
		c	0.154	0.163	0.148	0.252	0.259	0.247	0.429	0.434	0.426
	II	a	0.158	0.165	0.150	0.276	0.282	0.269	0.471	0.477	0.468
		b	0.142	0.151	0.135	0.255	0.261	0.250	0.447	0.452	0.444
		c	0.160	0.167	0.153	0.258	0.267	0.254	0.452	0.457	0.449
	III	a	0.142	0.152	0.136	0.245	0.253	0.241	0.428	0.432	0.424
		b	0.145	0.154	0.141	0.245	0.252	0.241	0.427	0.431	0.424
		c	0.145	0.155	0.140	0.228	0.235	0.224	0.401	0.406	0.398
	IV	a	0.137	0.146	0.133	0.231	0.239	0.226	0.399	0.404	0.396
		b	0.137	0.145	0.132	0.219	0.226	0.216	0.379	0.385	0.377
		c	0.129	0.138	0.124	0.212	0.219	0.209	0.357	0.362	0.354
0.25	I	a	0.794	0.805	0.787	0.977	0.979	0.976	1.000	1.000	1.000
		b	0.774	0.787	0.768	0.970	0.971	0.968	1.000	1.000	1.000
		c	0.777	0.787	0.770	0.970	0.971	0.969	1.000	1.000	1.000
	II	a	0.799	0.810	0.792	0.975	0.977	0.975	1.000	1.000	1.000
		b	0.777	0.787	0.770	0.969	0.971	0.967	1.000	1.000	1.000
		c	0.773	0.784	0.766	0.971	0.972	0.970	1.000	1.000	1.000
	III	a	0.742	0.755	0.735	0.961	0.963	0.960	1.000	1.000	1.000
		b	0.767	0.776	0.759	0.966	0.968	0.964	1.000	1.000	1.000
		c	0.745	0.755	0.738	0.954	0.956	0.953	0.999	0.999	0.999
	IV	a	0.707	0.721	0.699	0.945	0.947	0.944	0.999	0.999	0.999
		b	0.693	0.706	0.685	0.936	0.938	0.935	0.998	0.998	0.998
		c	0.680	0.695	0.673	0.932	0.934	0.930	0.998	0.998	0.998

Table 8: Part of simulation results of the empirical powers for 8 strata (where $H_0: d_0 = 0.1,\ H_A: d_a = 0.05, 0.15$ or 0.25).

			$m = 25$			$m = 50$			$m = 100$		
d_a	ρ	π_1	T_L	T_W	T_{SC}	T_L	T_W	T_{SC}	T_L	T_W	T_{SC}
0.05	I	a	0.308	0.315	0.300	0.519	0.523	0.514	0.804	0.806	0.802
		b	0.287	0.293	0.280	0.484	0.487	0.481	0.770	0.771	0.770
		c	0.252	0.256	0.249	0.440	0.440	0.438	0.716	0.716	0.716
	II	a	0.297	0.306	0.291	0.504	0.509	0.500	0.792	0.794	0.791
		b	0.291	0.300	0.283	0.484	0.488	0.479	0.773	0.774	0.771
		c	0.286	0.291	0.279	0.490	0.492	0.488	0.764	0.764	0.763
	III	a	0.268	0.272	0.262	0.449	0.451	0.447	0.728	0.729	0.727
		b	0.248	0.251	0.245	0.435	0.436	0.433	0.713	0.713	0.713
		c	0.237	0.239	0.234	0.418	0.419	0.417	0.676	0.676	0.677
	IV	a	0.249	0.256	0.243	0.436	0.440	0.433	0.700	0.702	0.699
		b	0.228	0.233	0.223	0.402	0.403	0.399	0.670	0.670	0.669
		c	0.209	0.210	0.208	0.363	0.363	0.363	0.621	0.620	0.621
0.15	I	a	0.256	0.266	0.248	0.476	0.483	0.471	0.771	0.774	0.770
		b	0.259	0.269	0.251	0.456	0.464	0.450	0.730	0.734	0.728
		c	0.243	0.252	0.236	0.435	0.442	0.430	0.709	0.712	0.707
	II	a	0.263	0.272	0.255	0.477	0.483	0.470	0.768	0.771	0.766
		b	0.254	0.264	0.245	0.451	0.457	0.446	0.734	0.738	0.732
		c	0.258	0.267	0.250	0.450	0.456	0.444	0.736	0.740	0.733
	III	a	0.241	0.248	0.234	0.434	0.441	0.429	0.714	0.718	0.712
		b	0.251	0.262	0.243	0.435	0.443	0.431	0.705	0.708	0.703
		c	0.231	0.240	0.226	0.401	0.408	0.397	0.669	0.673	0.666
	IV	a	0.232	0.240	0.226	0.406	0.413	0.402	0.678	0.680	0.676
		b	0.228	0.235	0.221	0.381	0.386	0.377	0.649	0.653	0.646
		c	0.214	0.223	0.211	0.355	0.361	0.352	0.609	0.613	0.607
0.25	I	a	0.973	0.975	0.972	1.000	1.000	1.000	1.000	1.000	1.000
		b	0.971	0.973	0.970	1.000	1.000	1.000	1.000	1.000	1.000
		c	0.964	0.966	0.961	1.000	1.000	1.000	1.000	1.000	1.000
	II	a	0.971	0.973	0.969	1.000	1.000	1.000	1.000	1.000	1.000
		b	0.968	0.970	0.967	1.000	1.000	1.000	1.000	1.000	1.000
		c	0.969	0.971	0.968	1.000	1.000	1.000	1.000	1.000	1.000
	III	a	0.961	0.963	0.960	1.000	1.000	1.000	1.000	1.000	1.000
		b	0.964	0.966	0.963	1.000	1.000	1.000	1.000	1.000	1.000
		c	0.950	0.953	0.949	0.999	0.999	0.999	1.000	1.000	1.000
	IV	a	0.946	0.949	0.944	0.999	0.999	0.999	1.000	1.000	1.000
		b	0.932	0.936	0.929	0.999	0.999	0.999	1.000	1.000	1.000
		c	0.927	0.930	0.925	0.998	0.998	0.998	1.000	1.000	1.000

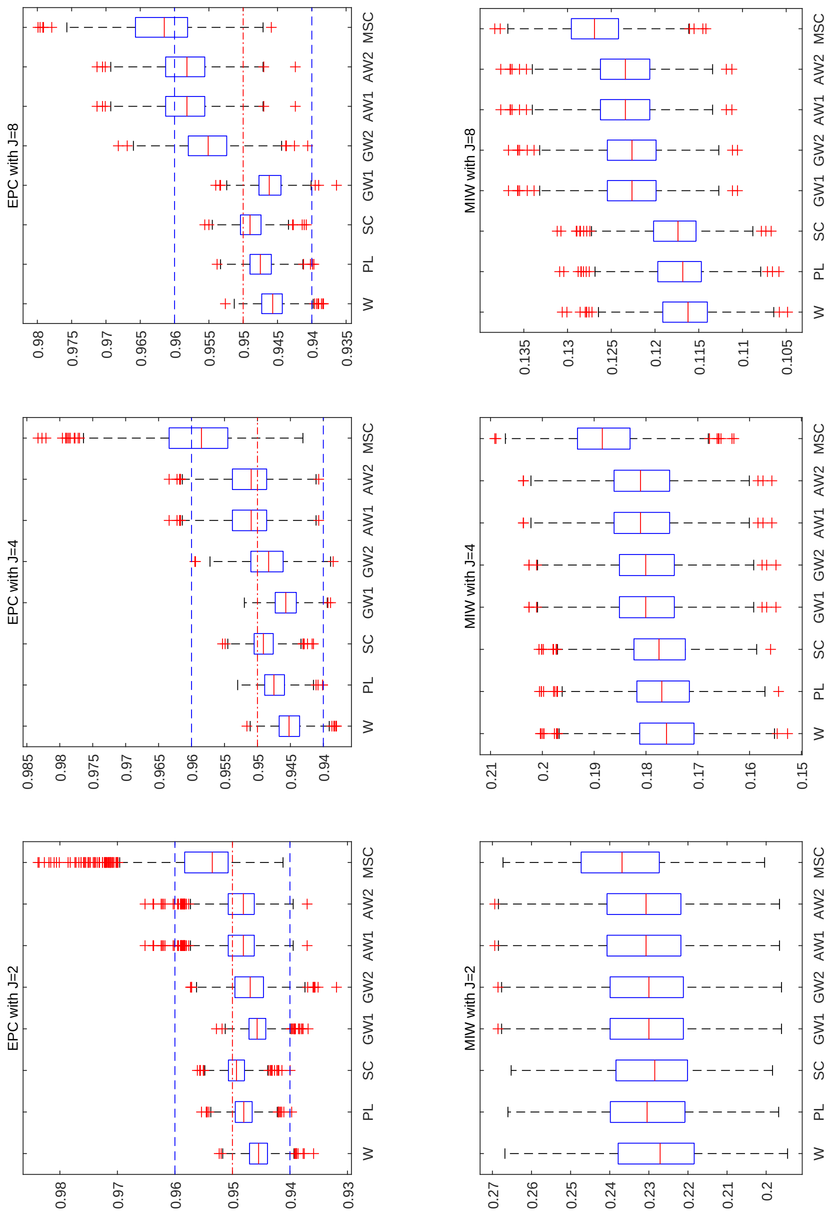

Figure 2: Box-plots of Empirical Coverage Probabilities and Mean Interval Widths

Table 10: Comparison of eight interval estimation methods for 2 strata.

m	d_0	ρ	π_1	Empirical Coverage Probability×100 (ECP×100)								Mean Interval Width (MIW)							
				W	PL	SC	$GW1$	$GW2$	$AW1$	$AW2$	MSC	W	PL	SC	$GW1$	$GW2$	$AW1$	$AW2$	M
I	0	A	a	95.0	95.3	95.5	94.4	94.4	96.3	96.3	96.3	0.253	0.257	0.259	0.270	0.270	0.272	0.272	0.2
			b	94.7	95.0	95.3	94.7	94.7	95.3	95.3	96.6	0.273	0.275	0.275	0.280	0.280	0.281	0.281	0.3
		B	a	94.9	95.1	95.3	94.8	94.8	96.0	96.0	96.6	0.263	0.268	0.270	0.276	0.276	0.278	0.278	0.2
			b	94.1	94.4	94.6	93.9	93.9	94.4	94.4	96.0	0.286	0.287	0.287	0.288	0.288	0.289	0.289	0.3
	0.1	A	a	94.6	94.9	95.2	94.4	94.4	95.6	95.6	96.5	0.264	0.266	0.267	0.275	0.275	0.276	0.276	0.2
			b	94.1	94.5	94.8	94.3	94.3	94.5	94.5	96.3	0.275	0.277	0.277	0.278	0.278	0.279	0.279	0.2
		B	a	94.4	94.9	95.1	94.3	94.3	95.1	95.1	96.2	0.275	0.277	0.278	0.282	0.282	0.283	0.283	0.3
			b	94.3	94.7	94.9	94.3	94.3	94.4	94.4	96.1	0.287	0.288	0.288	0.288	0.288	0.289	0.289	0.3
	0.2	A	a	94.7	95.1	95.3	94.7	94.7	95.3	95.3	96.6	0.267	0.270	0.270	0.273	0.273	0.275	0.275	0.2
			b	94.4	94.7	94.9	94.5	94.5	94.6	94.6	96.4	0.270	0.274	0.273	0.270	0.270	0.271	0.271	0.2
		B	a	94.6	95.0	95.2	94.5	94.5	95.0	95.0	96.5	0.279	0.281	0.281	0.282	0.282	0.283	0.283	0.3
			b	94.8	95.2	95.3	94.8	94.8	94.9	94.9	96.7	0.280	0.284	0.283	0.280	0.280	0.282	0.282	0.3
II	0	A	a	94.8	94.8	95.0	94.8	94.8	96.2	96.2	96.4	0.196	0.198	0.199	0.210	0.210	0.211	0.211	0.2
			b	94.3	94.5	94.5	94.5	94.5	94.9	94.9	96.1	0.213	0.214	0.214	0.218	0.218	0.219	0.219	0.2
		B	a	94.4	94.5	94.7	94.5	94.5	95.6	95.6	96.2	0.205	0.207	0.208	0.215	0.215	0.216	0.216	0.2
			b	94.9	95.1	95.2	94.7	94.7	95.2	95.2	96.4	0.223	0.223	0.223	0.225	0.225	0.225	0.225	0.2
	0.1	A	a	94.2	94.4	94.6	94.2	94.2	95.3	95.3	96.1	0.205	0.206	0.207	0.214	0.214	0.215	0.215	0.2
			b	94.6	94.7	94.9	94.6	94.6	94.8	94.8	96.5	0.215	0.216	0.216	0.217	0.217	0.217	0.217	0.2
		B	a	94.1	94.4	94.5	94.3	94.3	94.7	94.7	96.1	0.215	0.216	0.216	0.220	0.220	0.221	0.221	0.2
			b	94.7	94.9	95.0	94.8	94.8	94.9	94.9	96.3	0.224	0.225	0.224	0.224	0.224	0.225	0.225	0.2
	0.2	A	a	94.5	94.9	95.0	94.6	94.6	95.2	95.2	96.5	0.208	0.210	0.209	0.213	0.213	0.214	0.214	0.2
			b	94.6	94.8	95.0	94.7	94.7	94.7	94.7	96.6	0.210	0.214	0.212	0.211	0.211	0.211	0.211	0.2
		B	a	94.8	95.0	95.1	94.7	94.7	95.1	95.1	96.3	0.218	0.219	0.219	0.220	0.220	0.220	0.220	0.2
			b	94.5	94.6	94.7	94.4	94.4	94.5	94.5	96.1	0.219	0.224	0.220	0.219	0.219	0.219	0.219	0.2
III	0	A	a	94.8	95.0	95.2	94.6	93.8	95.7	95.7	96.4	0.211	0.214	0.215	0.219	0.219	0.219	0.219	0.2
			b	94.7	94.9	95.1	94.8	94.6	94.8	94.8	96.4	0.227	0.228	0.228	0.228	0.228	0.229	0.229	0.2
		B	a	94.6	94.8	95.0	94.6	93.7	95.2	95.2	96.3	0.219	0.222	0.223	0.224	0.224	0.225	0.225	0.2
			b	94.7	94.9	95.0	94.7	94.4	94.8	94.8	96.3	0.236	0.237	0.237	0.236	0.236	0.237	0.237	0.2
	0.1	A	a	94.7	95.0	95.1	94.6	94.0	95.1	95.1	96.3	0.220	0.221	0.222	0.224	0.224	0.224	0.224	0.2
			b	94.5	94.8	95.0	94.6	94.5	94.6	94.6	96.2	0.228	0.229	0.229	0.228	0.228	0.229	0.229	0.2
		B	a	94.7	94.9	95.0	94.7	94.2	95.0	95.0	96.2	0.229	0.230	0.231	0.231	0.231	0.231	0.231	0.2
			b	94.3	94.5	94.6	94.3	94.3	94.4	94.4	96.1	0.237	0.237	0.237	0.237	0.237	0.237	0.237	0.2
	0.2	A	a	94.7	94.9	95.0	94.7	94.2	94.9	94.9	96.3	0.222	0.224	0.224	0.223	0.223	0.224	0.224	0.2
			b	94.7	94.9	95.1	94.7	94.8	94.8	94.8	96.7	0.222	0.227	0.224	0.222	0.222	0.223	0.223	0.2
		B	a	94.9	95.2	95.4	94.9	94.8	95.0	95.0	96.6	0.231	0.233	0.232	0.231	0.231	0.232	0.232	0.2
			b	94.6	94.8	95.0	94.5	94.8	94.8	94.8	96.5	0.230	0.234	0.232	0.232	0.232	0.232	0.232	0.2

Sample size m: Case$I$:m= $(30, 30, 30, 30)$; CaseII: m= $(50, 50, 50, 50)$; CaseIII: m= $(40, 40, 50, 50)$.

Table 11: Comparison of eight interval estimation methods for 4 strata.

				Empirical Coverage Probability×100 (ECP×100)								Mean Interval Width (MIW)							
m	d_0	ρ	π_1	W	PL	SC	$GW1$	$GW2$	$AW1$	$AW2$	MSC	W	PL	SC	$GW1$	$GW2$	$AW1$	$AW2$	M
I	0	A	a	95.1	95.3	95.5	94.8	94.8	96.6	96.6	96.5	0.178	0.180	0.182	0.191	0.191	0.192	0.192	0.2
			b	94.4	94.7	94.9	94.6	94.6	95.1	95.1	96.3	0.193	0.195	0.196	0.198	0.198	0.199	0.199	0.2
		B	a	94.3	94.6	95.0	94.5	94.5	95.6	95.6	96.3	0.186	0.188	0.190	0.195	0.195	0.197	0.197	0.2
			b	94.3	94.5	94.7	94.6	94.6	94.6	94.6	96.3	0.202	0.203	0.204	0.204	0.204	0.205	0.205	0.2
	0.1	A	a	94.5	94.8	94.9	94.5	94.5	95.4	95.4	96.2	0.186	0.188	0.189	0.194	0.194	0.196	0.196	0.2
			b	94.8	95.0	95.2	94.7	94.7	95.1	95.1	96.6	0.195	0.196	0.197	0.197	0.197	0.198	0.198	0.2
		B	a	94.4	94.7	95.0	94.7	94.7	95.3	95.3	96.5	0.195	0.196	0.198	0.200	0.200	0.201	0.201	0.2
			b	94.4	94.6	94.8	94.5	94.5	94.5	94.5	96.4	0.203	0.204	0.205	0.203	0.203	0.205	0.205	0.2
	0.2	A	a	94.6	94.9	95.1	94.4	94.4	95.2	95.2	96.4	0.189	0.191	0.192	0.194	0.194	0.195	0.195	0.2
			b	94.5	94.8	94.9	94.4	94.4	94.7	94.7	96.5	0.191	0.192	0.193	0.191	0.191	0.192	0.192	0.2
		B	a	94.9	95.2	95.4	94.9	94.9	95.3	95.3	96.6	0.198	0.199	0.200	0.199	0.199	0.201	0.201	0.2
			b	94.5	94.8	95.0	94.7	94.7	94.7	94.7	96.5	0.198	0.201	0.201	0.198	0.198	0.200	0.200	0.2
II	0	A	a	94.9	95.1	95.3	94.9	94.9	96.4	96.4	96.4	0.139	0.140	0.141	0.149	0.149	0.149	0.149	0.1
			b	94.8	94.9	95.0	94.8	94.8	95.4	95.4	96.4	0.151	0.151	0.152	0.154	0.154	0.155	0.155	0.1
		B	a	94.4	94.6	94.8	94.6	94.6	95.6	95.6	96.1	0.145	0.146	0.147	0.152	0.152	0.153	0.153	0.1
			b	94.3	94.5	94.6	94.6	94.6	94.6	94.6	96.1	0.158	0.158	0.158	0.159	0.159	0.160	0.160	0.1
	0.1	A	a	94.6	94.9	95.1	94.9	94.9	95.8	95.8	96.3	0.145	0.146	0.147	0.152	0.152	0.152	0.152	0.1
			b	94.4	94.5	94.7	94.6	94.6	94.7	94.7	96.2	0.152	0.152	0.153	0.153	0.153	0.154	0.154	0.1
		B	a	94.8	95.0	95.1	94.8	94.8	95.5	95.5	96.2	0.152	0.153	0.153	0.156	0.156	0.156	0.156	0.1
			b	94.5	94.6	94.7	94.4	94.4	94.5	94.5	95.9	0.158	0.159	0.159	0.159	0.159	0.159	0.159	0.1
	0.2	A	a	94.6	94.8	94.9	94.6	94.6	95.3	95.3	96.4	0.147	0.148	0.148	0.151	0.151	0.151	0.151	0.1
			b	94.6	94.7	94.9	94.5	94.5	94.7	94.7	96.2	0.149	0.149	0.150	0.149	0.149	0.150	0.150	0.1
		B	a	94.7	94.8	95.0	95.0	95.0	95.1	95.1	96.7	0.154	0.154	0.155	0.155	0.155	0.156	0.156	0.1
			b	94.6	94.7	94.8	94.7	94.7	94.6	94.6	96.4	0.155	0.155	0.156	0.155	0.155	0.155	0.155	0.1
III	0	A	a	94.6	94.8	95.0	94.7	94.4	96.1	96.1	96.3	0.162	0.163	0.165	0.171	0.171	0.172	0.172	0.1
			b	94.5	94.7	94.9	94.6	94.2	95.0	95.0	96.1	0.175	0.176	0.176	0.178	0.178	0.179	0.179	0.1
		B	a	94.5	94.9	95.1	94.6	94.5	95.7	95.7	96.4	0.168	0.170	0.172	0.175	0.175	0.176	0.176	0.1
			b	94.3	94.6	94.8	94.6	94.7	94.9	94.9	96.4	0.182	0.183	0.183	0.184	0.184	0.185	0.185	0.1
	0.1	A	a	94.5	94.8	95.0	94.5	94.2	95.4	95.4	96.2	0.169	0.170	0.171	0.175	0.175	0.176	0.176	0.1
			b	94.6	94.8	95.0	94.6	94.7	95.0	95.0	96.4	0.176	0.176	0.177	0.178	0.178	0.178	0.178	0.1
		B	a	94.4	94.7	94.9	94.5	94.6	95.1	95.1	96.3	0.176	0.177	0.178	0.180	0.180	0.181	0.181	0.1
			b	94.5	94.8	95.0	94.6	94.7	94.8	94.8	96.3	0.183	0.183	0.184	0.184	0.184	0.185	0.185	0.1
	0.2	A	a	94.5	94.7	94.9	94.6	94.3	95.1	95.1	96.4	0.171	0.172	0.173	0.174	0.174	0.175	0.175	0.1
			b	94.3	94.6	94.8	94.5	94.8	94.8	94.8	96.4	0.172	0.172	0.173	0.173	0.173	0.174	0.174	0.1
		B	a	94.6	94.8	95.0	94.8	94.8	95.1	95.1	96.5	0.178	0.179	0.179	0.180	0.180	0.181	0.181	0.1
			b	94.4	94.6	94.8	94.5	94.9	94.8	94.8	96.2	0.178	0.179	0.179	0.180	0.180	0.181	0.181	0.1

sample size m: Case$I$:$m=$ $(30, 30, 30, 30, 30, 30, 30, 30)$; Case$II$: $m=$ $(50, 50, 50, 50, 50, 50, 50, 50)$; Case$III$: $m=$ $(30, 30, 35, 35, 40, 40, 45, 45)$

Table 12: Comparison of eight interval estimation methods for 8 strata.

				Empirical Coverage Probability×100 (ECP×100)								Mean Interval Width (MIW)							
m	d_0	ρ	π_1	W	PL	SC	$GW1$	$GW2$	$AW1$	$AW2$	MSC	W	PL	SC	$GW1$	$GW2$	$AW1$	$AW2$	M
I	0	A	a	95.2	95.4	95.7	94.8	94.8	96.7	96.7	96.5	0.126	0.128	0.129	0.135	0.135	0.136	0.136	0.1
			b	94.7	94.9	95.2	94.7	94.7	95.5	95.5	96.3	0.137	0.138	0.139	0.140	0.140	0.141	0.141	0.1
		B	a	94.6	94.9	95.2	94.7	94.7	96.0	96.0	96.6	0.131	0.133	0.134	0.138	0.138	0.139	0.139	0.1
			b	94.0	94.2	94.5	94.3	94.3	94.4	94.4	95.9	0.143	0.144	0.145	0.144	0.144	0.145	0.145	0.1
	0.1	A	a	94.7	94.9	95.1	94.7	94.7	95.8	95.8	96.4	0.132	0.133	0.134	0.138	0.138	0.139	0.139	0.1
			b	94.6	94.8	95.0	94.8	94.9	95.0	95.0	96.7	0.138	0.139	0.140	0.139	0.139	0.140	0.140	0.1
		B	a	94.3	94.6	94.9	94.7	94.7	95.1	95.1	96.3	0.138	0.139	0.140	0.141	0.141	0.142	0.142	0.1
			b	94.4	94.6	94.8	94.4	94.4	94.6	94.6	96.2	0.144	0.144	0.145	0.144	0.144	0.145	0.145	0.1
	0.2	A	a	94.9	95.2	95.4	94.7	94.7	95.6	95.6	96.6	0.134	0.135	0.136	0.137	0.137	0.138	0.138	0.1
			b	94.2	94.4	94.6	94.3	94.3	94.4	94.4	96.2	0.135	0.136	0.137	0.135	0.135	0.136	0.136	0.1
		B	a	94.2	94.4	94.6	94.4	94.4	94.7	94.7	96.2	0.140	0.141	0.141	0.141	0.141	0.142	0.142	0.1
			b	94.1	94.3	94.6	94.3	94.3	94.3	94.3	96.3	0.140	0.141	0.142	0.140	0.140	0.141	0.141	0.1
II	0	A	a	94.8	94.9	95.0	94.7	94.7	96.4	96.4	96.1	0.098	0.099	0.099	0.105	0.105	0.106	0.106	0.1
			b	94.6	94.7	94.8	94.7	94.7	95.3	95.3	96.2	0.107	0.107	0.108	0.109	0.109	0.110	0.110	0.1
		B	a	94.7	94.8	95.1	95.1	95.1	95.8	95.8	96.7	0.102	0.103	0.104	0.108	0.108	0.108	0.108	0.1
			b	94.3	94.5	94.7	94.7	94.7	94.8	94.8	96.1	0.111	0.112	0.112	0.112	0.112	0.113	0.113	0.1
	0.1	A	a	94.5	94.7	94.8	94.6	94.6	95.7	95.7	96.2	0.103	0.103	0.104	0.107	0.107	0.108	0.108	0.1
			b	94.8	95.0	95.1	94.6	94.6	95.1	95.1	96.4	0.107	0.108	0.108	0.108	0.108	0.109	0.109	0.1
		B	a	94.7	94.9	95.0	94.8	94.8	95.4	95.4	96.3	0.107	0.108	0.108	0.110	0.110	0.111	0.111	0.1
			b	94.9	95.0	95.1	94.8	94.8	95.0	95.0	96.4	0.112	0.112	0.113	0.112	0.112	0.113	0.113	0.1
	0.2	A	a	95.0	95.2	95.3	94.9	94.9	95.6	95.6	96.5	0.104	0.105	0.105	0.107	0.107	0.107	0.107	0.1
			b	94.3	94.4	94.6	94.3	94.3	94.4	94.4	96.3	0.105	0.106	0.106	0.105	0.105	0.106	0.106	0.1
		B	a	94.5	94.7	94.8	94.6	94.6	94.9	94.9	96.0	0.109	0.109	0.110	0.110	0.110	0.110	0.110	0.1
			b	94.8	94.9	95.0	94.8	94.8	94.8	94.8	96.4	0.109	0.110	0.110	0.109	0.109	0.110	0.110	0.1
III	0	A	a	94.7	95.0	95.1	95.0	95.2	96.7	96.7	96.3	0.107	0.108	0.109	0.117	0.117	0.118	0.118	0.1
			b	94.9	95.0	95.2	94.7	95.3	96.1	96.1	96.6	0.116	0.117	0.117	0.122	0.122	0.123	0.123	0.1
		B	a	94.4	94.7	95.0	94.8	95.0	96.1	96.1	96.3	0.112	0.113	0.114	0.120	0.120	0.121	0.121	0.1
			b	94.8	94.9	95.1	94.9	95.5	95.8	95.8	96.5	0.121	0.122	0.122	0.126	0.126	0.127	0.127	0.1
	0.1	A	a	94.5	94.7	94.9	94.9	95.0	96.0	96.0	96.2	0.112	0.113	0.114	0.120	0.120	0.120	0.120	0.1
			b	94.8	95.1	95.2	95.0	95.5	95.6	95.6	96.4	0.117	0.117	0.118	0.121	0.121	0.122	0.122	0.1
		B	a	94.5	94.6	94.8	94.8	95.0	95.8	95.8	96.1	0.117	0.118	0.118	0.123	0.123	0.124	0.124	0.1
			b	94.8	94.9	95.1	94.7	95.7	95.8	95.8	96.5	0.122	0.122	0.122	0.126	0.126	0.127	0.127	0.1
	0.2	A	a	94.3	94.5	94.5	94.6	95.3	95.5	95.5	96.4	0.114	0.114	0.115	0.119	0.119	0.120	0.120	0.1
			b	95.0	95.2	95.3	94.9	95.8	95.9	95.9	96.7	0.114	0.115	0.115	0.118	0.118	0.119	0.119	0.1
		B	a	94.9	95.1	95.3	95.0	95.6	96.1	96.1	96.8	0.118	0.119	0.119	0.123	0.123	0.124	0.124	0.1
			b	94.5	94.7	94.9	94.4	95.4	95.5	95.5	96.3	0.118	0.120	0.120	0.123	0.123	0.124	0.124	0.1

sample size m: Case$I$:$m=$ $(30,30,30,30,30,30,30,30,30,30,30,30,30,30,30,30)$; Case$II$: $m=$ $(50,50,50,50,50,50,50,50,50,50,50,50,50,50,50,50)$; Case$III$: $m=$ $(25,25,30,30,35,35,40,40,45,45,50,50,55,55,60,60)$