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

Journal:	*Statistical Methods in Medical Research*
Manuscript ID	SMM-17-0628.R1
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 medical researches such as ophthalmologic (or otolaryngologic) studies, in which each unit contributes information from paired organs to the data analysis, 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 very important to adjust the effect of confounder on statistical inferences, since either ignoring the intra-class correlation or confounding effect may lead to biased results. In this article, we propose three approaches for testing common risk difference for stratified bilateral correlated data under the assumption of equal correlation. Five confidence intervals (CIs) of common difference of two proportions are derived. The performance of the proposed test methods and CI estimations is evaluated by Monte Carlo simulations. The simulation results show that the score test statistic outperforms other statistics in the sense that the former has robust type $\text I$ error rates with high powers. The score CI induced from the score test statistic performs satisfactorily in terms of coverage probabilities with reasonable interval widths. A real data set from an otolaryngologic study is used to illustrate the proposed methodologies.

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

Xi Shen[1], Chang-Xing Ma[1], Kam Chuen Yuen[2], and Guo-Liang Tian[3]

[1]Department of Biostatistics, The State University of New York at Buffalo, Buffalo, NY 14214, USA

[2]Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong, P. R. China

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

Abstract

Bilateral correlated data are often encountered in medical researches such as ophthalmologic (or otolaryngologic) studies, in which each unit

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

contributes information from paired organs to the data analysis, 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 very important to adjust the effect of confounder on statistical inferences, since either ignoring the intra-class correlation or confounding effect may lead to biased results. In this article, we propose three approaches for testing common risk difference for stratified bilateral correlated data under the assumption of equal correlation. Five confidence intervals (CIs) of common difference of two proportions are derived. The performance of the proposed test methods and CI estimations is evaluated by Monte Carlo simulations. The simulation results show that the score test statistic outperforms other statistics in the sense that the former has robust type I error rates with high powers. The score CI induced from the score test statistic performs satisfactorily in terms of coverage probabilities with reasonable interval widths. A real data set from an otolaryngologic study is used to illustrate the proposed methodologies.

Keywords: Bilateral correlated data; Common risk difference test; Intra-class correlation coefficients; Interval estimation; Strata.

1 Introduction

Paired correlated data are often collected from all participants in medical studies of group comparisons. For instance, in an ophthalmologic study, researchers are interested in the comparison of two treatments. Participants are randomly administrated into one of the two treatment groups. It is of great interest to investigate if or not the two treatments are clinically equivalent. The efficacy of treatment is evaluated by comparing the numbers of cured eyes at the end of the trials of the two treatment groups. The possible outcomes can be summarized in a contingency table (the recorded outcome would be bilateral cured, unilateral cured or none cured). It is noteworthy that the measurements of both eyes from each participant are likely to be correlated.

Under this framework, various test methods for assessing the equality of proportions and various confidence interval (CI) construction approaches for parameters of interest have been developed. Rosner [1] proposed a so-called "constant R model" based on dependency by assuming that the probability of a response at one side given a response at the other side is proportional to the prevalence rate of corresponding group for the ophthalmologic data. Tang et al. [2], Ma et al. [3], Shan and Ma [4], and Liu et al. [5] have developed asymptotic and exact testing methods for this model, which was empirically shown to perform well. However, Dallal [6] pointed out one drawback of Rosner's model; i.e., it could lead a poor fit if the characteristic is almost to occur bilaterally with widely varying group-specific prevalence. Later, Donner [7] suggested an alternative model by assuming that all treatment groups share an intra-class correlation coefficient ("ρ model"). Thompson [8] evaluated this "ρ model" by simulation and confirmed that this model is robust for paired data. Furthermore, various asymptotic and exact testing methods have been proposed by Tang et al. [9], Pei et al. [10], and Ma and Liu [11]. In addition, CI estimation for risk difference of two proportions based on aforementioned two models has received considerable attentions in statistical literature. For instance, Tang et al. [12] and Pei et al. [13] investigated asymptotic CI construction in two pre-specified models for the difference of proportions between two groups. Recently, Yang et al. [14] constructed asymptotic CIs for many-to-one comparisons of proportion differences with multiplicity adjustment.

However, one important feature for practical consideration is the stratification factor or confounding effect. Many randomized controlled trials (RCTs) recruit patients to multiple centers or hospitals, rather than to a single center, and we expect that the patients in the same center tend to have correlated outcomes. It is often necessary to account for the center-effect in the data analysis, since ignoring the stratification factor will lead to incorrect assumptions in the study, and will result in invalid inference [15] [16][17]. In addition, in some RCTs with large number of confounders, some confounders, by chance, could appear imbalanced with treatment arms, making it desirable to adjust for the stratification factors in the analysis to obtain valid inferences. For these reasons, it has been emphasized that extra care should be taken in the analysis of the stratified data.

With the aforementioned models in hand, computational methods for testing or constructing CIs on the stratified data analysis for bilateral binary observations have evolved dramatically in recent years. Pei et al. [18] proposed a homogeneity test of proportion ratios for stratified bilateral data based on Donner's model. Tang and Qiu [19] applied Rosner's model on common difference test of two proportions, in which they specified the common difference being zero. Moreover, Shen and Ma [20] introduced three alternative testing procedures based on maximum likelihood estimates (MLEs) for testing homogeneity of difference of two proportions for stratified correlated bilateral data under a common intra-cluster correlation assumption. Particularly, if we fail to reject the null hypothesis that the differences of two proportions are equal among strata, the problem of interest may shift to explore what is the equivalent value. Therefore, in this article, we develop several procedures for testing equality of difference of two proportions in a stratified bilateral 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 CIs for that common difference.

The rest of this article is organized as follows. In Section 2, we briefly delineate the data structure. Then the MLEs, three different test procedures and CI estimators are derived in Section 3. Simulation studies are conducted to investigate the performance of the three tests and five CIs in Section 4. A real example from otolaryngologic study is used to illustrate our proposed methods in Section 5. Some concluding remarks and future works are discussed in Section 6.

2 Data Structure

Suppose that our purpose is to test if or not two treatments of some eye disease are clinically equivalent among different age strata in a medical comparative study. The data structure of interest is shown in Table 1. A total of N_j patients are randomly allocated into one of two treatment groups for the j^{th} age stratum ($j = 1, \ldots, J$). Let m_{lij} represent the number of patients having l ($l = 0, 1, 2$) eyes with improvement response(s) in the i^{th} ($i = 1, 2$) group from the j^{th} stratum, and $m_{\cdot ij} = \sum_{l=0}^{2} m_{lij}$ be the total number of patients in the i^{th} group from the j^{th} stratum. Define $Z_{hijk} = 1$ if there exists an improvement for the h^{th} ($h = 1, 2$) eye of the k^{th} ($k = 1, \ldots, m_{\cdot ij}$) patient in the i^{th} group from the j^{th} stratum, and 0 otherwise.

We assume that the probability of improvement at one eye for the patient in the i^{th} group from the j^{th} stratum is $\Pr(Z_{hijk} = 1) = \pi_{ij}$ $(0 \leq \pi_{ij} \leq 1,\ h = 1, 2,\ i = 1, 2)$. For the "$\rho$ model" (Donner [7]), the constant ρ_{ij} $(-1 \leq \rho_{ij} \leq 1)$ denotes a measure of within-subject correlation. It is easy to show that the improvement probabilities for none, one, and both eyes in the i^{th} group from 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 $\pi_{ij}^2+\rho_{ij}\pi_{ij}(1-\pi_{ij})$, respectively. 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 $j^{th}(j = 1, \ldots, J)$ stratum in a stratified bilateral design

	Group (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 want to test if the risk differences between two groups among all strata are equal to a common d_0; i.e., the considered null hypothesis is H_0: $d_1 = \cdots = d_J \triangleq d = d_0$, versus H_a: $d \neq d_0$, where $d_j = \pi_{2j} - \pi_{1j}$. Let $\mathbf{m}_j = \{m_{01j}, m_{11j}, m_{21j}; m_{02j}, m_{12j}, m_{22j}\}$ denote the observed data for the j^{th} stratum as shown in Table 1. Then, the log-likelihood of parameters of interest based on $\mathbf{m}_j$ is

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

$$+m_{1ij} \log[2\pi_{ij}(1-\rho_j)(1-\pi_{ij})] + m_{2ij} \log[\pi_{ij}^2 + \rho_j \pi_{ij}(1-\pi_{ij})]\} + \text{constant},$$

so that the overall log-likelihood function is

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

(a) Global MLEs

We first derive the MLEs of parameters from a global setup. Setting the partial differentiation of l_j with respect to π_{ij}'s and ρ_j's equal to zero yields the MLEs of the parameters, denoted by $\tilde{\pi}_{ij}$ and $\tilde{\rho}_j$, respectively, where

$$\frac{\partial l}{\partial \pi_{ij}} = \frac{(2\,\pi_{ij}-1)\,m_{1ij}}{\pi_{ij}\,(\pi_{ij}-1)} + \frac{m_{2ij}\,(\rho_j+2\,\pi_{ij}-2\,\rho_j\,\pi_{ij})}{\pi_{ij}\,(\rho_j+\pi_{ij}-\rho_j\,\pi_{ij})} - \frac{m_{0i}\,(\rho_j+2\,\pi_{ij}-2\,\rho_j\,\pi_{ij}-2)}{(\pi_{ij}-1)\,(\rho_j\,\pi_{ij}-\pi_{ij}+1)},\ i=1,2,$$

$$\frac{\partial l}{\partial \rho_j} = \sum_{i=1}^{2}\left[\frac{m_{1ij}}{(\rho_j-1)} - \frac{(\pi_{ij}-1)\,m_{2ij}}{(\rho_j+\pi_{ij}-\rho_j\,\pi_{ij})} + \frac{\pi_{ij}\,m_{0ij}}{(\rho_j\,\pi_{ij}-\pi_{ij}+1)}\right].$$

There are no closed form solutions for $\tilde{\pi}_{ij}$ and $\tilde{\rho}_j$. Therefore, classical techniques such as the Newton–Raphson or the Fisher scoring algorithms are usually recommended in these cases. However, for the current problem with high-dimensional parameters there are computational challenges. Therefore, these MLEs can be computed by repeating the following steps derived by Ma and Liu [11] and Shen and Ma [20]. We can simplify the first equation into a cubic equation,

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

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

and obtain the MLE of π_{ij} by solving the real root of it. Then ρ_j can be updated by the Fisher scoring algorithm. The $(t+1)^{th}$ approximate of ρ_j is

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

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

(b) Unconstrained MLEs

We now consider the unconstrained MLEs. Based on the alternative hypothesis, we can see that π_{2j} can be expressed as $\pi_{1j} + d$, where $d \neq d_0$. Thus, the parameters here only involve ρ_j, π_{1j}, and a common given d. Differentiating l_j with respect to (ρ_j, π_{1j}, d) and setting them equal to zero yield the MLEs of the parameters $\hat{\rho}_j$, $\hat{\pi}_{1j}$ and $\hat{d}$.

Closed-form solutions of $(\hat{\rho}_j, \hat{\pi}_{1j}, \hat{d})$ are not available. Similarly, we can employ the two-step approach of Shen and Ma [20] by updating the common d via the Newton–Raphson algorithm. Then, we apply the Fisher scoring algorithm to estimate π_{1j} and ρ_j with a given d from each stratum. The iteration procedure is described 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 $\tilde{\pi}_{1j}$ and $\tilde{\rho}_j$ are global MLEs, and $\tilde{d}_j \triangleq \tilde{\pi}_{2j} - \tilde{\pi}_{1j}$.

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

See Appendix A.2 for the more details.

3. Update

$$\begin{bmatrix} {\pi_{1j}}^{(t+1)} \\ {\rho_j}^{(t+1)} \end{bmatrix} = \begin{bmatrix} {\pi_{1j}}^{(t)} \\ {\rho_j}^{(t)} \end{bmatrix} + I_2^{-1}({\pi_{1j}}^{(t)}, {\rho_j}^{(t)}, d^{(t)}) \begin{bmatrix} \frac{\partial l({\pi_{1j}}^{(t)}, {\rho_j}^{(t)}, d^{(t)})}{\partial \pi_{1j}} \\ \frac{\partial l({\pi_{1j}}^{(t)}, {\rho_j}^{(t)}, d^{(t)})}{\partial \rho_j} \end{bmatrix}, j = 1, \ldots, J,$$

where I_2 is the Fisher information matrix for π_{1j} and ρ_j. The formula of I_2 and the corresponding differential equations with respect to d are given in Appendix A.2.

4. Repeating Steps 2–3 until convergence.

We denote the MLEs of parameters under the alternative hypothesis by $(\hat{\pi}_{11}, \ldots, \hat{\pi}_{1J}; \hat{\rho}_1, \ldots, \hat{\rho}_J; \hat{d})$.

(c) Constrained MLEs

Finally, we investigate the constrained MLEs. Under the null hypothesis H_0: $d_1 = \cdots = d_J \triangleq d = d_0$, the parameter π_{2j} can be expressed as $\pi_{1j} + d_0$, where d_0 is a known value. The parameters here only involve ρ_j and π_{1j}. Therefore, we can simply utilize the Step 3 in solving unconstrained MLEs with a given d_0. Let $(\hat{\pi}_{11H_0}, \ldots, \hat{\pi}_{1JH_0}; \hat{\rho}_{1H_0}, \ldots, \hat{\rho}_{JH_0})$ be the constrained MLEs of the nuisance parameters $(\pi_{11}, \ldots, \pi_{1J}; \rho_1, \ldots, \rho_J)$. With all MLEs obtained, we consider the following test procedures and CI estimation approaches.

3.1.1 Likelihood ratio test (T_L)

The likelihood ratio test (LRT) 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 asymptotically follows a chi-square distribution with one degree of freedom under the null hypothesis.

3.1.2 Wald-type test (T_W)

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

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

Based on asymptotic normality of the MLEs, one can show that $\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \hat{I}_n^{-1}$, where I_n^{-1} is the inverse of the Fisher information matrix for $\boldsymbol{\beta}$, and $\hat{I}_n^{-1}$ is the MLE of I_n^{-1}. Therefore, we can rewrite the Wald-type statistic as

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

where $I_n^{-1}(1,1)$ stands for the $(1,1)^{th}$ entry of I_n^{-1}. Under the null hypothesis, T_W is asymptotically distributed as a chi-square distribution with one degree of freedom.

3.1.3 Score test (T_{SC})

The score test statistic T_{SC} utilizes the MLEs of parameters 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 $\boldsymbol{\beta} = (d, \pi_{11}, \rho_1, \ldots, \pi_{1J}, \rho_J)^T$. Here, d is the parameter of interest, while 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

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

where $I_n^{-1}(1,1)$ represents the $(1,1)^{th}$ entry of I_n^{-1}, and the formula of $\frac{\partial l_j}{\partial d}$ is given in Appendix. Under the null hypothesis, T_{SC} is asymptotically distributed as a chi-square distribution with one 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 have derived the MLE of $\boldsymbol{\beta} = (\pi_{11}, \pi_{21}, \rho_1, \ldots, \pi_{1J}, \pi_{2J}, \rho_J)^T$ from the global setup and alternative hypothesis, and denoted them by $\tilde{\boldsymbol{\beta}} = (\tilde{\pi}_{11}, \tilde{\pi}_{21}, \tilde{\rho}_1, \ldots, \tilde{\pi}_{1J}, \tilde{\pi}_{2J}, \tilde{\rho}_J)^T$ and $\hat{\boldsymbol{\beta}} = (\hat{\pi}_{11}, \hat{\pi}_{21}, \hat{\rho}_1, \ldots, \hat{\pi}_{1J}, \hat{\pi}_{2J}, \hat{\rho}_J)^T$, respectively, where $\tilde{\pi}_{2j} = \tilde{\pi}_{1j} + \tilde{d}$, and $\hat{\pi}_{2j} = \hat{\pi}_{1j} + \hat{d}$, for $j = 1, \ldots, J$.

Intuitively, we consider that there exist weights $\{w_j\}$ assigned to each stratum satisfying $d = \sum_{j=1}^{J} w_j d_j$ and $\sum_{j=1}^{J} w_j = 1$, where $j = 1, \ldots, J$. The choice of weights is not trivial. Here, we suggest two ways. (1) Uniformly weights: $w_j = \frac{1}{J}$. (2) Sample size weights: $w_j = \frac{N_j}{N}$, where N_j is the sample size of the j^{th} stratum and $N = \sum_{j=1}^{J} N_j$ is the total number of patients.

We apply the algorithm to construct CI for d_0 by a row vector $\boldsymbol{W} = (w_1, \ldots, w_J)$ and a constant matrix,

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

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

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

and

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

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

It is straightforward 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 of the MLE, we can express the variance of the d in terms of $\boldsymbol{C}$ and the information matrix of $\boldsymbol{\beta}$; that is $\mathrm{Var}(\boldsymbol{C}\boldsymbol{\beta}) = \boldsymbol{C}I^{-1}\boldsymbol{C}^{\boldsymbol{T}}$, where I is the information matrix of $\boldsymbol{\beta}$.

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

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

and

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

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, $\frac{(\hat{d}-d_0)}{\sqrt{\mathrm{Var}(\hat{d})}}$ asymptotically follows the standard normal distribution, where $\mathrm{Var}(\hat{d}) = \hat{I}_n^{-1}(1,1)$, $I_n^{-1}(1,1)$ is the $(1,1)^{th}$ element of the inverse of information matrix under H_a. Therefore, the $100(1-\alpha)\%$ CI for $d_0 \in [-1,1]$ is defined as

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

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 CI estimation from χ^2 distribution by inverting the LRT of H_0: $d_1 = \cdots = d_J \triangleq d = d_0$ versus H_a: $d \neq d_0$, where $d_j = \pi_{2j} - \pi_{1j}$. Since the LRT statistic follows a chi-square distribution with one degree of freedom under the null hypothesis, the $100(1-\alpha)\%$ CI 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 $1-\alpha$ quantile of the chi-square distribution with one degree of freedom.

The bisection method can be used to obtain the lower/upper limits of above inequality (Yang, Tian, Liu, and Ma [14]). To assess the upper limit, the iteration procedure can be performed as follows.

1. Start with the initial values $d^{(0)} = \hat{d}$, stepsize=0.1, and flag=1, where $\hat{d}$ is unconstrained MLE of 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 function can be calculated according to the constrained MLEs and the data, denoted by $\hat{l}^{(t+1)}$.

3. Evaluate the aforementioned requirement of CI. 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 (that is, the stepsize is sufficiently small, say, 10^{-5}).

Similarly, we repeat the iteration procedure with flag$=-1$ to assess the lower limit of CI.

3.2.4 Score CI (SC)

Since the score test statistic also follows a chi-square distribution with one degree of freedom under the null hypothesis, one can assess the $100(1-\alpha)\%$ CI 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. Similarly, the bisection method is used to search the lower and upper limits.

4 Simulation Studies

4.1 Common risk difference test

We now investigate the performance of the proposed three statistics for testing the equality of risk differences. We first evaluate the behavior of the type I error rate under various parameter settings, where $m = m_{\cdot 11} = m_{\cdot 21} = \ldots = m_{\cdot 1J} = m_{\cdot 2J}$=25, 50 or 100 in J=2, 4 or 8 strata, respectively. The parameter setups are displayed 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 rejecting the null hypothesis with

10,000. All tests are conducted at 5% significance level.

Table 2: Parameter setups for computing empirical type I error rates and powers

	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	$(0.4, 0.4)$	$(0.4, 0.4, 0.4, 0.4)$	$(0.4, 0.4, 0.4, 0.4, 0.4, 0.4, 0.4, 0.4)$

Following Tang et al. [21], at 0.05 nominal level, we define a test is liberal if the 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 the score test and the LRT are robust in terms of satisfactory type I error for all scenarios. The 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 ($J = 2$). Additionally, a set of boxplots (Figure 1) showed the distribution for the empirical type I error rates for all tests when we have balanced data for J=2, 4 or 8, respectively. We can observe that the score test behaves satisfactorily, in the sense that its type I error rate is close to the pre-determined nominal level 0.05 for any configuration. The LRT is inflated, while the Wald-type test is even worse. However, as the sample size increases, both the LRT and the Wald-type test perform better.

Next, we investigate the performance of power for proposed test statistics under various parameters settings. To be specific, we consider the same sample sizes and parameter setups as we did for computing empirical type I error. Tables 6–8 report empirical power associated with three proposed tests for various configurations. Since powers produced by three tests under different d_0 perform similarly, results from one case ($d_0 = 0.1$) are presented. We can also observe that, under the same parameter settings, the powers of different test statistics are very close. The 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 (denoted by d_a) and d_0 increases. Powers increase when the number of strata J goes larger. Overall, the score test is highly recommended, since it is satisfactory on type I error control and has a good performance on power.

4.2 Confidence interval estimation

In this subsection, we compare the proposed five CI estimators with one existing CI estimator from balanced to unbalanced designs in terms of empirical coverage probability (ECP) and mean interval width (MIW). 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 with 10,000. Following Yang et al. [14], CI can be constructed with pooling data, where the objective of interest is only the treatment group variable. We only present the result of this marginal CI for score method (MSC). In addition, we construct global Wald-type CIs with uniformly weighted adjustment and sample size weighted adjustment, respectively, namely GW1 and GW2, and also 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% CI is constructed for each replicate. Results are shown in Tables 10–12. Accordingly, we display a set of boxplots to investigate the distribution of ECPs and MIWs for unbalanced cases (Figure 2). Generally, CIs based on strata assumption outperform CIs based on marginal model since the ECPs of those are closer than pre-determined CI. Among those CIs considering strata assumption, score CIs behave satisfactorily, since the ECPs are the closest to the pre-determined confidence level, and MIWs are reasonable short. It is hence recommended. Likelihood ratio statistic produces CIs with shorter MIWs, but it yields deflated ECPs. Wald-type statistic (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, CI produced from the score statistic 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 Example

We analyze a real example in this section to further evaluate the performance of aforementioned methods. Mandel et al [22] 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 clinically equivalent. We summarize the observed data in Table 13.

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

Age groups	Age < 2yrs		Age 2–5 yrs		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

Based on the data given above, all MLEs of parameters are reported in Table 14. First, we consider testing homogeneity proposed by Shen and Ma [20]. For the homogeneity test, the null hypothesis is H_0: $d_1 = d_2 = d_3 \triangleq d$ versus H_a: some of d_js are not equal for $j \in \{1, 2, 3\}$, values of the three test statistics are $T_L = 2.83$, $T_W = 2.93$, $T_{SC} = 2.76$ and the corresponding p-values are 0.24, 0.23, 0.25, respectively. We note that all p-values are greater than the nominal level $\alpha = 0.05$, indicating that the differences of cured ears between two groups are not correlated to the age effect. Next, we consider a common test to check whether or not $d = 0$. The corresponding values of common test statistics and their p-values are presented in Table 15. In addition, CI estimators are given in Table 16. These results imply that there are no significant differences between two groups among age strata (i.e., the cured rates between the amoxicillin and cefaclor among age are clinically equivalent).

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 2–5 yrs	0.5307	0.5881	0.0323	0.5330	0.6205	-	0.5308	0.5968	-
Age $\geq$ 6yrs	0.6153	0.8341	0.0499	0.6332	0.8982	-	0.6140	0.8636	-

Table 15: The values of statistics and p-values for three different tests.

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

Table 16: 95% CIs for common risk difference ($\hat{d} = -0.0945$).

	CI
W	$[-0.2859, 0.0969]$
PL	$[-0.2938, 0.1015]$
SC	$[-0.3039, 0.1018]$
GW1	$[-0.2622, 0.1234]$
GW2	$[-0.3005, 0.0863]$
AW1	$[-0.2885, 0.0994]$
AW2	$[-0.2939, 0.1048]$
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 (LRT, Wald-type test and score test) are investigated. Classical algorithms, such as the Fisher scoring and the Newton–Raphson methods are usually criticized for computational difficulty in high-dimensional cases. We derived the two-step approaches for obtaining the unconstrained and constrained MLEs, which are very efficient. Then, we proposed five CIs of common difference of two proportions on stratified bilateral correlated data, which include two weight-adjusted approaches (global Wald-type CI and alternative Wald-type CI) and three test-based approaches.

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

For correlated data with binomial distributions, there are many well-built model-based methods to calculate the MLE iteratively or perform statistical analysis, e.g. “GENMOD” and “GLIMMIX” procedure in SAS. Ying et al. [23] described and demonstrated appropriate linear regression analysis involving both eyes, including mixed effects and marginal models under various covariance structures to account for inter-eye correlation. These methods also offered the flexibility to incorporate covariates in the model. However, those model-based methods fail to provide a closed-form solution for either MLEs or test statistics. The explicit form of solutions improves computational efficiency, especially when we further develop exact test for small sample situation in the future. All proposed methods are asymptotic, and do not perform well at relatively small sample sizes. Thus, exact tests are necessary to overcome inflated type I error rate as future work. To perform exact test, extensive calculations will be required, which make it very difficult using model-based methods.

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 among strata may be ignored when there exists ordinal classification relationship. We can further develop either asymptotic or exact trend test as an interesting future work.

A user-friendly online calculator is available via the link http://www.buffalo.edu/ cxma/CommonRiskDifferenceRhoModelStrafified.htm. Readers can simulate data by user-specified parameters, or input their own data to perform tests or construct CIs proposed in the article.

Appendix A Information matrix and formula derivation

A.1 Information matrix for computing global MLEs

The second order differential equations from the j^{th} stratum with respect to π_{ij} $(i = 1, 2)$ and ρ_j yield

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

Then from the j^{th} stratum, we have,

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

where

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

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

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

Therefore, the information matrix for J strata has the form

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

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

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

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

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

A.2 Information matrix for computing unconstrained and constrained MLEs

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

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

and

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

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

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

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*, 46(8):6074–6082, 2017.

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

[7] A. Donner. Statistical methods in opthalmology: an adjusted chi-square approach. *Biometrics*, 45(2):605–611, 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, 2011.

[10] Y.-B Pei, M.-L Tang, and J.-H Guo. Testing the equality of two proportions for combined unilateral and bilateral data. *Communi-*

cations 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*, 27(4):611–619, 2017.

[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, G.-L Tian, X. 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] B. C. Kahan and T. P. Morris. Assessing potential sources of clustering in individually randomised trials. *BMC Medical Research Methodology*, 13(1):58, 2013.

[16] B. C. Kahan and T. P. Morris. Analysis of multicentre trials with

continuous outcomes: when and how should we account for centre effects? *Statistics in Medicine*, 32(7):1136–1149, 2013.

[17] B. C. Kahan. Accounting for centre-effects in multicentre trials with a binary outcome–when, why, and how? *BMC Medical Research Methodology*, 14(1):20, 2014.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] G.-S Ying, M. G. Maguire, R. Glynn, and B. Rosner. Tutorial on biostatistics: linear regression analysis of continuous correlated eye data. *Ophthalmic Epidemiology*, 24(2):130–140, 2017.

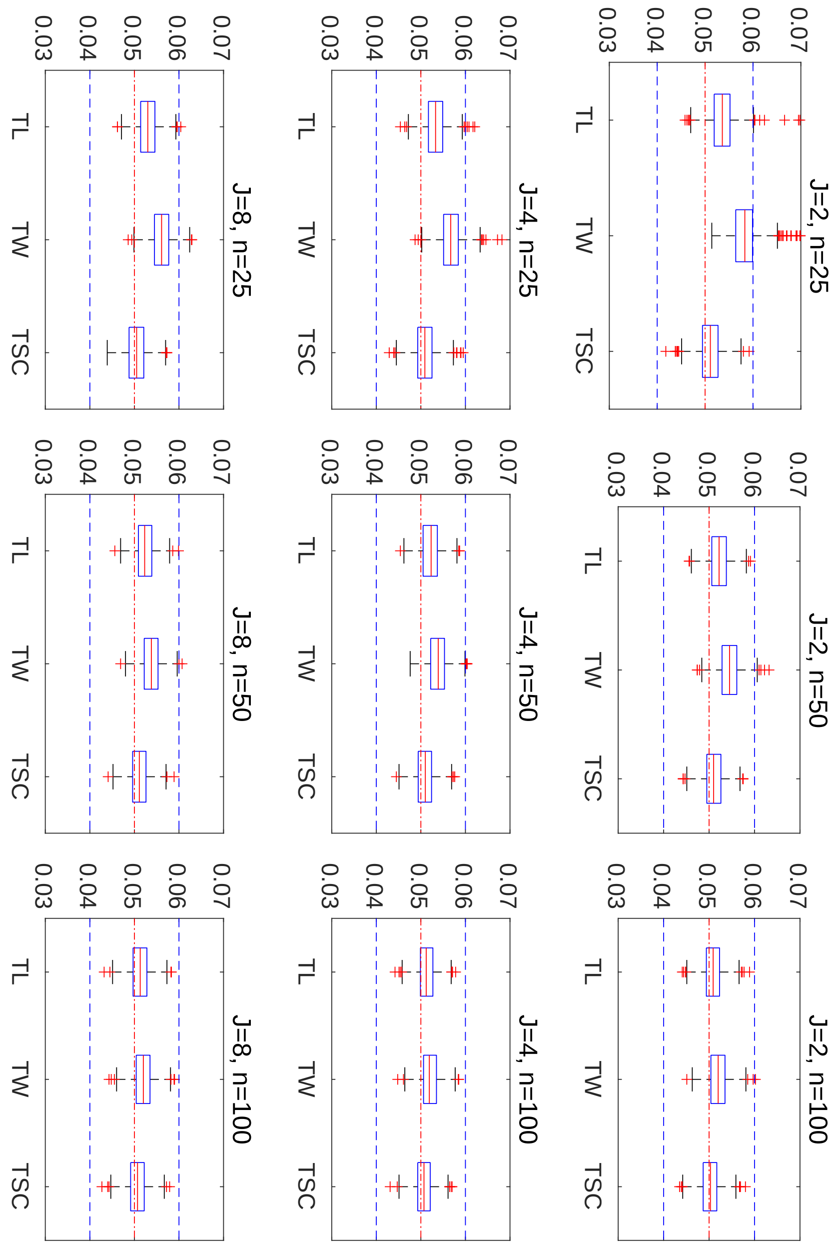

Figure 1: Box plots of empirical sizes.

Table 3: Simulation results of the empirical sizes for two 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 four 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 eight 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 two 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 four 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 eight 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

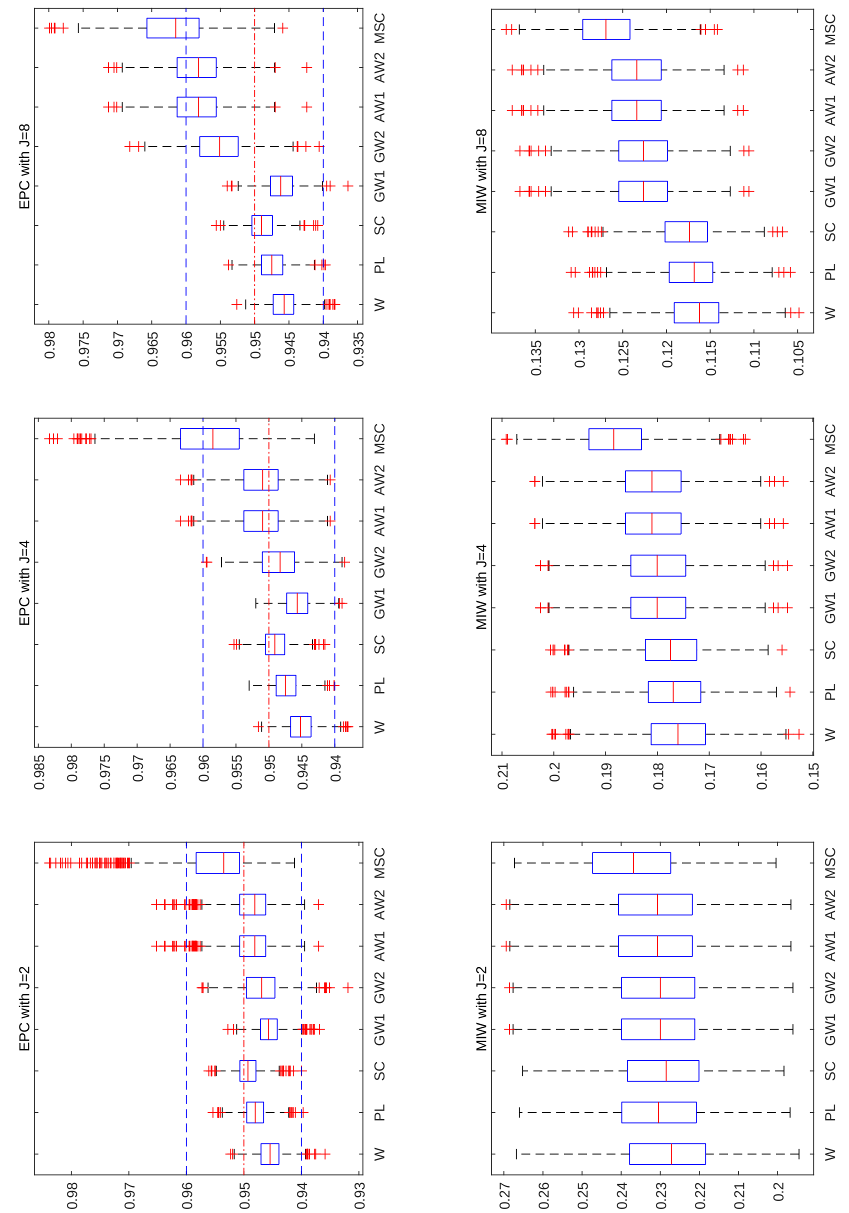

Figure 2: Box plots of empirical coverage probabilities and mean interval widths.

Table 10: Comparisons of eight interval estimation methods for two strata.

Case	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	$GW1$	$GW2$	$AW1$	$AW2$	MSC
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	0.290
			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	0.300
		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	0.297
			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	0.308
	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	0.295
			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	0.299
		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	0.302
			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	0.307
	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	0.294
			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	0.292
		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	0.302
			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	0.301
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	0.226
			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	0.234
		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	0.231
			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	0.240
	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	0.236
			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	0.232
		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	0.236
			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	0.240
	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	0.229
			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	0.228
		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	0.236
			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	0.235
III	0	A	a	94.8	95.0	95.2	95.1	95.0	95.9	96.5	96.4	0.211	0.214	0.215	0.219	0.226	0.219	0.227	0.242
			b	94.7	94.9	95.1	94.8	94.7	95.0	95.5	96.4	0.227	0.228	0.228	0.228	0.232	0.229	0.233	0.248
		B	a	94.6	94.8	95.0	94.8	94.9	95.3	95.8	96.3	0.219	0.222	0.223	0.224	0.231	0.225	0.231	0.247
			b	94.7	94.9	95.0	94.6	94.7	94.6	94.8	96.3	0.236	0.237	0.237	0.236	0.238	0.237	0.239	0.254
	0.1	A	a	94.7	95.0	95.1	94.4	94.5	94.8	95.4	96.3	0.220	0.221	0.222	0.224	0.229	0.224	0.230	0.245
			b	94.5	94.8	95.0	94.5	94.5	94.5	94.8	96.2	0.228	0.229	0.229	0.228	0.230	0.229	0.231	0.247
		B	a	94.7	94.9	95.0	94.7	94.6	94.5	95.0	96.2	0.229	0.230	0.231	0.230	0.235	0.231	0.235	0.251
			b	94.3	94.5	94.6	94.6	94.7	94.7	94.7	96.1	0.237	0.237	0.237	0.237	0.237	0.238	0.238	0.253
	0.2	A	a	94.7	94.9	95.0	94.7	94.6	94.7	95.2	96.3	0.222	0.224	0.224	0.223	0.227	0.224	0.228	0.244
			b	94.7	94.9	95.1	94.5	94.6	94.6	94.6	96.7	0.222	0.227	0.224	0.222	0.222	0.223	0.223	0.240
		B	a	94.9	95.2	95.4	94.5	94.4	94.6	94.8	96.6	0.231	0.233	0.232	0.231	0.233	0.232	0.234	0.250
			b	94.6	94.8	95.0	94.8	95.1	95.3	95.1	96.5	0.230	0.234	0.232	0.232	0.230	0.232	0.231	0.247

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

Table 11: Comparisons of eight interval estimation methods for four strata.

Case	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	$GW1$	$GW2$	$AW1$	$AW2$	MSC
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	0.207
			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	0.214
		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	0.211
			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	0.220
	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	0.210
			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	0.213
		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	0.216
			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	0.219
	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	0.210
			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	0.208
		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	0.215
			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	0.215
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	0.166
			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	0.166
		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	0.164
			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	0.171
	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	0.163
			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	0.166
		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	0.163
			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	0.171
	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	0.163
			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	0.162
		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	0.167
			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	0.167
III	0	A	a	94.6	94.8	95.0	94.5	94.6	96.1	96.3	96.3	0.162	0.163	0.165	0.171	0.173	0.172	0.174	0.187
			b	94.5	94.7	94.9	94.5	94.6	95.1	95.1	96.1	0.175	0.176	0.176	0.178	0.179	0.179	0.180	0.193
		B	a	94.5	94.9	95.1	94.5	94.7	95.6	95.8	96.4	0.168	0.170	0.172	0.175	0.177	0.176	0.178	0.191
			b	94.3	94.6	94.8	94.8	94.7	95.0	94.9	96.4	0.182	0.183	0.183	0.184	0.184	0.185	0.185	0.198
	0.1	A	a	94.5	94.8	95.0	94.6	94.6	95.4	95.6	96.2	0.169	0.170	0.171	0.175	0.176	0.176	0.177	0.190
			b	94.6	94.8	95.0	94.7	94.5	94.9	94.9	96.4	0.176	0.176	0.177	0.178	0.177	0.178	0.178	0.191
		B	a	94.4	94.7	94.9	94.5	94.5	95.0	95.1	96.3	0.176	0.177	0.178	0.180	0.181	0.181	0.181	0.194
			b	94.5	94.8	95.0	94.3	94.3	94.7	94.5	96.3	0.183	0.183	0.184	0.184	0.183	0.185	0.184	0.197
	0.2	A	a	94.5	94.7	94.9	94.4	94.5	95.0	95.1	96.4	0.171	0.172	0.173	0.174	0.175	0.175	0.176	0.189
			b	94.3	94.6	94.8	94.5	94.6	94.8	94.7	96.4	0.172	0.172	0.173	0.173	0.172	0.174	0.173	0.187
		B	a	94.6	94.8	95.0	94.6	94.7	94.9	94.9	96.5	0.178	0.179	0.179	0.180	0.180	0.181	0.181	0.194
			b	94.4	94.6	94.8	94.8	94.8	95.0	94.8	96.2	0.178	0.179	0.179	0.180	0.178	0.181	0.179	0.192

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

Table 12: Comparisons of eight interval estimation methods for eight strata.

Case	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	$GW1$	$GW2$	$AW1$	$AW2$	MSC
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	0.147
			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	0.152
		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	0.150
			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	0.156
	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	0.149
			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	0.151
		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	0.153
			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	0.156
	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	0.149
			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	0.148
		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	0.153
			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	0.152
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	0.114
			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	0.118
		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	0.116
			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	0.121
	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	0.116
			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	0.117
		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	0.119
			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	0.121
	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	0.115
			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	0.115
		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	0.118
			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	0.116
III	0	A	a	94.7	95.0	95.1	94.6	94.5	96.8	96.4	96.3	0.107	0.108	0.109	0.117	0.115	0.118	0.116	0.124
			b	94.9	95.0	95.2	94.4	94.6	95.8	95.2	96.6	0.116	0.117	0.117	0.122	0.119	0.123	0.119	0.128
		B	a	94.4	94.7	95.0	94.3	94.4	95.8	95.4	96.3	0.112	0.113	0.114	0.120	0.118	0.121	0.118	0.127
			b	94.8	94.9	95.1	94.8	94.7	95.7	94.9	96.5	0.121	0.122	0.122	0.126	0.122	0.127	0.123	0.132
	0.1	A	a	94.5	94.7	94.9	94.8	94.6	96.2	95.8	96.2	0.112	0.113	0.114	0.120	0.117	0.120	0.118	0.126
			b	94.8	95.1	95.2	94.7	94.6	95.7	95.0	96.4	0.117	0.117	0.118	0.121	0.118	0.122	0.119	0.128
		B	a	94.5	94.6	94.8	94.5	94.7	95.6	95.1	96.1	0.117	0.118	0.118	0.123	0.120	0.124	0.121	0.129
			b	94.8	94.9	95.1	94.9	95.0	95.8	95.2	96.5	0.122	0.122	0.122	0.126	0.122	0.127	0.122	0.131
	0.2	A	a	94.3	94.5	94.5	94.6	94.8	96.0	95.3	96.4	0.114	0.114	0.115	0.119	0.116	0.120	0.117	0.126
			b	95.0	95.2	95.3	94.4	94.3	95.4	94.6	96.7	0.114	0.115	0.115	0.118	0.114	0.119	0.115	0.124
		B	a	94.9	95.1	95.3	94.5	94.6	95.5	94.8	96.8	0.118	0.119	0.119	0.123	0.120	0.124	0.120	0.129
			b	94.5	94.7	94.9	95.0	95.0	96.0	95.3	96.3	0.118	0.120	0.120	0.123	0.118	0.124	0.119	0.128

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