Research Article

Statistics in Medicine

Received 11 December 2008, Accepted 28 August 2009 Published online 23 October 2008 in Wiley Interscience

(www.interscience.wiley.com) DOI: 10.1002/sim.3744

A comparison of methods for the construction of confidence interval for relative risk in stratified matched-pair designs

Nian-Sheng Tang,[a]*† Hui-Qiong Li[a] and Man-Lai Tang[b]

A stratified matched-pair study is often designed for adjusting a confounding effect or effect of different trails/centers/groups in modern medical studies. The relative risk is one of the most frequently used indices in comparing efficiency of two treatments in clinical trials. In this paper, we propose seven confidence interval estimators for the common relative risk and three simultaneous confidence interval estimators for the relative risks in stratified matched-pair designs. The performance of the proposed methods is evaluated with respect to their type I error rates, powers, coverage probabilities, and expected widths. Our empirical results show that the percentile bootstrap confidence interval and bootstrap-resampling-based Bonferroni simultaneous confidence interval behave satisfactorily for small to large sample sizes in the sense that (i) their empirical coverage probabilities can be well controlled around the pre-specified nominal confidence level with reasonably shorter confidence widths; and (ii) the empirical type I error rates of their associated test statistics are generally closer to the pre-specified nominal level with larger powers. They are hence recommended. Two real examples from clinical laboratory studies are used to illustrate the proposed methodologies. Copyright © 2009 John Wiley & Sons, Ltd.

Keywords: bootstrap-resampling method; confidence interval; relative risk; stratified matched-pair designs

1. Introduction

Matched-pair designs are often employed to increase the efficiency of treatment comparison in clinical trials and laboratory studies. As an example, we consider a study in paratuberculosis, a debilitating disease of ruminants caused by Mycobacterium paratuberculosis [1]. The enzyme-linked immunosorbent assay (ELISA) is regarded as the most acceptable method for serodiagnosis, while the dot immunobinding assay (DIA) is regarded as a new, simple, rapid, and inexpensive alternative for screening the paratuberculosis infected cattle. To compare the accuracy of the two serological methods, sera were grouped on the basis of their fecal culture results. The positive and negative culture groups were taken to be two different strata. DIA and ELISA were used to test bovine sera for paratuberculosis antibody. Data are presented in the following table in which sensitivity and specificity are those correct diagnostic results with the corresponding positive and negative cultures, respectively, and δ_1 and δ_2 are, respectively, the relative risks for sensitivities and specificities of DIA and ELISA.

[a]*Department of Statistics, Yunnan University, Kunming 650091, People's Republic of China*

[b]*Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong*

**Correspondence to: Nian-Sheng Tang, Department of Statistics, Yunnan University, Kunming 650091, People's Republic of China.*

†*E-mail: nstang@ynu.edu.cn*

Contract/grant sponsor: NSFC; contract/grant numbers: 10561008, 10761011

Contract/grant sponsor: Ph.D. Special Scientific Research Foundation of Chinese University; contract/grant number: 20060673002

Contract/grant sponsor: New Century Excellent Talents in Universities; contract/grant number: 2008PY036

Contract/grant sponsor: Research Grant Council of the Hong Kong Special Administrative Region; contract/grant numbers: HKBU261007, HKBU261508

Testing bovine sera for Mycobacterium paratuberculosis by ELISA or DIA

Positive culture results (+)					Negative culture result (−)				
		ELISA					ELISA		
		+	−	Sum			−	+	Sum
DIA	+	105	9	114	DIA	−	130	5	135
	−	5	49	54		+	8	18	26
	Sum	110	58	168		Sum	138	23	161

Sensitivity:
DIA = 114/168 = 0.68
ELISA = 110/168 = 0.65
$\hat{\delta}_1 = 1.05$

Specificity:
DIA = 135/161 = 0.84
ELISA = 138/161 = 0.86
$\hat{\delta}_2 = 0.97$

It is of interest to see if there is a difference between δ_1 and δ_2, which could indicate an effect of different culture results on the relative risk. If there is no difference between δ_1 and δ_2, it is also of interest to see if there is a difference between the diagnosis efficiency of DIA and that of ELISA.

Statistical inference for the above two questions of interest is usually conducted by hypothesis testing and confidence interval construction. Although they are closely related, hypothesis testing focuses on a single priori hypothesis. In contrast, a confidence interval can avoid such a problem by providing a range of plausible parameter values. For a given investigation, the confidence interval reveals both the magnitude and the precision of the estimated parameter value, whereas the p value obtained from hypothesis test confounds the two aspects of the data. Hence, confidence interval construction in practical applications is preferred.

Confidence interval construction for paired binary data has received considerable attention in experimental trials or epidemiological studies in past years. May and Johnson [2] and Lui [3] studied the performance of Wald-type confidence interval, Quesenberry and Hurst confidence interval (see, Quesenberry and Hurst [4]), and likelihood-ratio-test-based confidence interval for the differences in correlated binary proportions on the basis of asymptotic theories. Their empirical results showed that the latter two perform well for moderate to large sample sizes (e.g. $\geqslant$30). Newcombe [5, 6] proposed 10 confidence intervals for the difference between binomial proportions in paired-design studies, and investigated the performance of these confidence intervals via simulation studies. Tango [7] presented the profile-likelihood-based confidence interval and the score-test-based confidence interval for risk difference in paired-design studies on the basis of asymptotic theories, and some empirical results given in Tango [8] and Newcombe [9] indicated that the profile-likelihood-based confidence interval and the score-test-based confidence interval perform satisfactorily in large-sample designs. Tang *et al.* [10] noted that Tango's [7] confidence intervals perform unsatisfactorily in small-sample designs. Therefore, Tang *et al.* [10] developed an exact unconditional confidence interval and an approximate unconditional confidence interval for proportion difference in small-sample paired studies. However, all the above cited works are confined to a single 2×2 table, and there is little work for confidence interval construction in multiple 2×2 tables. Therefore, the main purpose of this article is to construct confidence intervals for the common relative risk in stratified matched-pair designs based on the profile-likelihood-ratio test, the Cochran statistic, the weighted-least-squares (WLS) statistic, the score statistic, bootstrap-resampling method, and simultaneous confidence intervals for the relative risks based on Bonferroni method and bootstrapling method.

The remainder of this paper is organized as follows. Ten confidence interval estimates for the relative risk in stratified matched-pair designs are presented in Section 2. In Section 3, simulation studies are conducted to investigate the performance of the proposed test statistics and their test-based confidence intervals in terms of their type I error rates, powers, empirical coverage probabilities, and expected interval widths. Real examples from a clinical study and a paratuberculosis study are used to illustrate our proposed methodologies in Section 4. Concluding remarks are given in Section 5.

2. Confidence interval estimators

Consider a stratified matched-pair design in which two diagnostic tests (one is a new diagnostic test and the other is the standard one) are, respectively, conducted for the same n_j subjects in the jth stratum ($j=1,2,\ldots,J$). Let x_{11j}, x_{10j}, x_{01j}, and x_{00j} be the observed numbers of pairs (1,1), (1,0), (0,1), and (0,0) in the jth stratum, respectively; and let p_{11j}, p_{10j}, p_{01j}, p_{00j} be their corresponding probabilities. The data structure for the jth stratum can be summarized in the following table:

Data structure for the jth stratum in a stratified matched-pair design

	Standard diagnostic test		
New diagnostic test	Positive (1)	Negative (0)	Total
Positive (1)	$x_{11j}(p_{11j})$	$x_{10j}(p_{10j})$	$x_{1.j}(p_{1j})$
Negative (0)	$x_{01j}(p_{01j})$	$x_{00j}(p_{00j})$	$x_{0.j}(q_{1j})$
Total	$x_{.1j}(p_{0j})$	$x_{.0j}(q_{0j})$	$n_j(1.0)$

Here, $0<p_{ikj}<1$ $((i,k)=(1,1),(1,0),(0,1),(0,0)$, and $j=1,2,\cdots,J)$, and $p_{11j}+p_{01j}=p_{0j}$, $p_{11j}+p_{10j}=p_{1j}$, $p_{10j}+p_{00j}=q_{0j}$, $p_{01j}+p_{00j}=q_{1j}$, $p_{0j}+q_{0j}=1.0$, $p_{1j}+q_{1j}=1.0$; $x_{11j}+x_{01j}=x_{\cdot 1j}$, $x_{10j}+x_{00j}=x_{\cdot 0j}$, $x_{11j}+x_{10j}=x_{1\cdot j}$, $x_{01j}+x_{00j}=x_{0\cdot j}$, $x_{\cdot 1j}+x_{\cdot 0j}=n_j$, and $x_{1\cdot j}+x_{0\cdot j}=n_j$. The probability density function of $(x_{11j},x_{10j},x_{01j})$ for a given value of n_j in the jth stratum is given by

$$f(x_{11j},x_{10j},x_{01j}|p_{11j},p_{10j},p_{01j},n_j)=\frac{n_j!}{x_{11j}!x_{10j}!x_{01j}!(n_j-x_{11j}-x_{10j}-x_{01j})!}$$
$$\times p_{11j}^{x_{11j}}p_{10j}^{x_{10j}}p_{01j}^{x_{01j}}(1-p_{11j}-p_{10j}-p_{01j})^{n_j-x_{11j}-x_{10j}-x_{01j}}$$

2.1. Confidence intervals for the common relative risk across J strata

In this subsection, we assume a common relative risk between two marginal probabilities across the J strata, i.e. $\delta=p_{1j}/p_{0j}$ for $j=1,2,\ldots,J$. Under this assumption, we have

$$p_{11j}=p_{0j}-p_{01j},\quad p_{10j}=(\delta-1)p_{0j}+p_{01j}\quad\text{and}\quad p_{00j}=1-\delta p_{0j}-p_{01j}$$

The log-likelihood of the observed frequencies $\{(x_{11j},x_{10j},x_{01j},x_{00j}):j=1,\ldots,J\}$ is given by

$$l(\boldsymbol{p}_0,\delta)=\sum_{j=1}^{J}\{x_{11j}\log(p_{0j}-p_{01j})+x_{10j}\log[(\delta-1)p_{0j}+p_{01j}]+x_{01j}\log(p_{01j})+x_{00j}\log(1-\delta p_{0j}-p_{01j})\}+C \tag{1}$$

where $\boldsymbol{p}_0=(p_{01},\ldots,p_{0J},p_{011},\ldots,p_{01J})$, and C is a constant that does not depend on parameters δ, p_{0j}, and p_{01j} $(j=1,2,\ldots,J)$.

2.1.1. Confidence interval based on the profile-likelihood-ratio test. The log-profile-likelihood for the observed frequencies $\{(x_{11j},x_{10j},x_{01j},x_{00j}):j=1,\ldots,J\}$ is given by

$$l(\tilde{\boldsymbol{p}}_0,\delta)=\sum_{j=1}^{J}\{x_{11j}\log(\tilde{p}_{0j}-\tilde{p}_{01j})+x_{10j}\log[(\delta-1)\tilde{p}_{0j}+\tilde{p}_{01j}]+x_{01j}\log(\tilde{p}_{01j})+x_{00j}\log(1-\delta\tilde{p}_{0j}-\tilde{p}_{01j})\}+C \tag{2}$$

where $\tilde{\boldsymbol{p}}_0=(\tilde{p}_{01},\tilde{p}_{02},\ldots,\tilde{p}_{0J},\tilde{p}_{011},\tilde{p}_{012},\ldots,\tilde{p}_{01J})$ in which $\tilde{p}_{0j}$ and $\tilde{p}_{01j}$ are the constrained maximum likelihood estimates (CMLEs) of p_{0j} and p_{01j} for a given value of parameter δ $(j=1,\ldots,J)$. Maximizing the log-profile-likelihood yields

$$\tilde{p}_{01j}=\{-b_j+(b_j{}^2-4a_jc_j)^{1/2}\}/(2a_j),\quad \tilde{p}_{0j}=[(x_{11j}+x_{10j}+x_{01j})/n_j-\tilde{p}_{01j}]/\delta \tag{3}$$

where $a_j=n_j(1+\delta)$, $b_j=(x_{01j}+x_{11j})\delta^2-(x_{11j}+x_{10j}+2x_{01j})$, and $c_j=x_{01j}(1-\delta)(x_{11j}+x_{10j}+x_{01j})/n_j$ (see Appendix A).

Let $\hat{\boldsymbol{p}}_0$ be the value of $\tilde{\boldsymbol{p}}_0$ evaluated at $\delta=\hat{\delta}$, where $\hat{\delta}$ is the MLE of δ for the log-profile-likelihood given in Equation (2), which is the solution to the following equation:

$$\frac{\partial l}{\partial\delta}=\sum_{j=1}^{J}\left[\frac{x_{10j}}{(\delta-1)\tilde{p}_{0j}+\tilde{p}_{01j}}-\frac{x_{00j}}{1-\delta\tilde{p}_{0j}-\tilde{p}_{01j}}\right]=0$$

Clearly, there are no explicit solutions for $\hat{p}_{0j}$, $\hat{p}_{01j}$, and $\hat{\delta}$. Moreover, $\hat{\boldsymbol{p}}_0$ and $\hat{\delta}$ are just the unconstrained MLEs of $\boldsymbol{p}_0$ and δ for the log-likelihood given in Equation (1). That is, $\hat{\boldsymbol{p}}_0$ and $\hat{\delta}$ are the solutions to the following equations:

$$\frac{\partial l}{\partial p_{0j}}=0,\quad \frac{\partial l}{\partial p_{10j}}=0,\quad \frac{\partial l}{\partial\delta}=0$$

It is rather difficult to get the closed solutions of the above joint equations. Therefore, the well-known Fisher scoring iterative algorithm is employed to obtain $\hat{p}_{0j}$, $\hat{p}_{01j}$, and $\hat{\delta}$.

The profile-likelihood ratio statistic is given by

$$T_p=2[l(\hat{\boldsymbol{p}}_0,\hat{\delta})-l(\tilde{\boldsymbol{p}}_0,\delta)]$$

which is asymptotically distributed as chi-square distribution with one degree of freedom. Therefore, the approximate $(1-\alpha)100\%$ profile-likelihood-based confidence interval for δ is given by $[\delta_L,\delta_U]$ where $0<\delta_L<\delta_U$ are the smaller and the larger roots of δ in the following equation:

$$2[l(\hat{\boldsymbol{p}}_0,\hat{\delta})-l(\tilde{\boldsymbol{p}}_0,\delta)]=\chi^2_{1,\alpha} \tag{4}$$

where $\chi^2_{1,\alpha}$ is the upper α-percentile of central χ^2 distribution with one degree of freedom. In fact, δ_L and δ_U can be obtained via the bisection searching algorithm.

2.1.2. Confidence interval based on Cochran statistic. For $j=1,2,\ldots,J$, it is easily shown that the naive estimates of p_{01j}, p_{1j}, p_{0j}, and $\delta_j=p_{1j}/p_{0j}$ are given by $\breve{p}_{01j}=x_{01j}/n_j$, $\breve{p}_{1j}=(x_{11j}+x_{10j})/n_j$, $\breve{p}_{0j}=(x_{11j}+x_{01j})/n_j$, and $\breve{\delta}_j=\hat{p}_{1j}/\hat{p}_{0j}$, respectively. By delta method, mean and variance of $\hat{\delta}_j$ can be approximated by δ_j and $\sigma_j^2=(2p_{01j}+p_{0j}(\delta_j-1))\delta_j/(n_jp_{0j}^2)$, respectively. σ_j^2 has the following expression under the assumption that $\delta_1=\delta_2=\cdots=\delta_J=\delta$:

$$\sigma_j^2=(2p_{01j}+p_{0j}(\delta-1))\delta/(n_jp_{0j}^2)$$

In this case, σ_j^2 can be estimated by $\tilde{\sigma}_j^2=(2\tilde{p}_{01j}+\tilde{p}_{0j}(\delta-1))\delta/(n_j\tilde{p}_{0j}^2)$, where $\tilde{p}_{0j}$ and $\tilde{p}_{01j}$ are given in Equation (3). Thus, the Cochran statistic for testing $H_0:\delta=\delta_0$ versus $H_1:\delta\neq\delta_0$ based on Mantel–Haenszel (MH) weights can be expressed as

$$T_c(\delta_0)=\sum_{j=1}^{J}w_j(\breve{\delta}_j-\delta_0)/\sqrt{\left\{\sum_{j=1}^{J}w_j^2(2\tilde{p}_{01j}+\tilde{p}_{0j}(\delta_0-1))\delta_0/(n_j\tilde{p}_{0j}^2)\right\}}$$

which is asymptotically distributed as a standard normal distribution under H_0, where $w_j=n_j/\sum_{k=1}^{J}n_k$, and $\tilde{p}_{01j}$ and $\tilde{p}_{0j}$ are given in Equation (4) by replacing δ by δ_0. Thus, the approximate $(1-\alpha)100$ per cent Cochran-test-based confidence interval is given by $[\delta_{cl},\delta_{cu}]$, where δ_{cl} and δ_{cu} are the smaller and the larger roots of the following equation:

$$T_c^2(\delta)=\chi_{1,\alpha}^2 \tag{5}$$

A MH-type Wald statistic for testing $H_0:\delta=\delta_0$ versus $H_1:\delta\neq\delta_0$ can be expressed as

$$T_W(\delta_0)=\sum_{j=1}^{J}w_j(\breve{\delta}_j-\delta_0)/\sqrt{\left\{\sum_{j=1}^{J}w_j^2(2\breve{p}_{01j}+\breve{p}_{0j}(\delta_0-1))\delta_0/(n_j\breve{p}_{0j}^2)\right\}}$$

which is asymptotically distributed as a standard normal distribution under H_0. Again, the approximate $(1-\alpha)100$ per cent confidence interval based on MH-type Wald statistic is given by $[\delta_{wl},\delta_{wu}]$, where δ_{wl} and δ_{wu} are the smaller and the larger roots of the following equation:

$$T_W^2(\delta)=\chi_{1,\alpha}^2 \tag{6}$$

2.1.3. Confidence interval based on the WLS. Following the arguments of Fleiss [11], the commonly used WLS estimator for δ is given by

$$\breve{\delta}_{\mathrm{wls}}=\frac{\sum_{j=1}^{J}\breve{W}_j\breve{\delta}_j}{\sum_{j=1}^{J}\breve{W}_j}$$

where $\breve{W}_j=1/\breve{\mathrm{Var}}(\breve{\delta}_j)=(x_{11j}+x_{01j})^3/[(x_{11j}+x_{10j})(x_{10j}+x_{01j})]$. Thus, a statistic for testing $H_0:\delta=\delta_0$ versus $H_1:\delta\neq\delta_0$ based on the WLS method can be expressed as

$$T_{wls}=\sqrt{\left\{\sum_{j=1}^{J}\breve{W}_j\right\}}(\breve{\delta}_{\mathrm{wls}}-\delta_0)$$

which is asymptotically distributed as a standard normal distribution when $\min\{n_1,\ldots,n_J\}$ is sufficiently large. Therefore, an asymptotic $(1-\alpha)100$ per cent confidence interval for δ based on the WLS estimator is given by

$$\left[\max\left\{\breve{\delta}_{\mathrm{wls}}-z_{\alpha/2}\Big/\sqrt{\left(\sum_{j=1}^{J}\breve{W}_j\right)},0\right\},\ \breve{\delta}_{\mathrm{wls}}+z_{\alpha/2}\Big/\sqrt{\left(\sum_{j=1}^{J}\breve{W}_j\right)}\right] \tag{7}$$

where $z_{\alpha/2}$ is the upper $\alpha/2$-percentile point of the standard normal distribution.

2.1.4. Confidence interval based on score statistic. Following the arguments of Tang *et al.* [12], it follows from Equation (1) that the score function of log-likelihood with respect to δ under H_0 is

$$S(\tilde{\boldsymbol{p}}_0)=\sum_{j=1}^{J}\left\{\frac{x_{10j}+x_{11j}-(x_{11j}+x_{01j})\delta_0}{\delta_0(2\tilde{p}_{01j}+\tilde{p}_{0j}(\delta_0-1))}\right\}$$

where $\tilde{p}_{0j}$ and $\tilde{p}_{01j}$ are given in Equation (3). Variance of $S(\tilde{\boldsymbol{p}}_0)$ under H_0 can be estimated by (see Appendix B)

$$\widehat{\mathrm{Var}}_0\{S(\tilde{\boldsymbol{p}}_0)\}=\sum_{j=1}^{J}\frac{n_j}{\delta_0(2\tilde{p}_{01j}+\tilde{p}_{0j}(\delta_0-1))}$$

Following the arguments of Nam [1], it follows from the above equations that the score statistic for testing hypothesis $H_0:\delta=\delta_0$ against $H_1:\delta\neq\delta_0$ is given by

$$T_S(\delta_0)=\sum_{j=1}^{J}\left\{\frac{x_{10j}+x_{11j}-(x_{11j}+x_{01j})\delta_0}{\delta_0(2\tilde{p}_{01j}+\tilde{p}_{0j}(\delta_0-1))}\right\}\Bigg/\left\{\sum_{j=1}^{J}\frac{n_j}{\delta_0(2\tilde{p}_{01j}+\tilde{p}_{0j}(\delta_0-1))}\right\}^{1/2}$$

which is asymptotically distributed as a standard normal distribution under H_0. Therefore, the approximate $(1-\alpha)100$ per cent confidence limits for δ based on the score statistic can be obtained by solving the following equation:

$$T_S(\delta)=\pm z_{\alpha/2} \tag{8}$$

where the plus and the minus signs correspond to the lower limit δ_{ls} and the upper limit δ_{su}, respectively. These two limits can be easily obtained by secant method (see, e.g. Tango [7]).

2.1.5. Confidence intervals based on bootstrap-resampling method. When sample sizes (i.e. n_j) are small, confidence interval estimates based on large sample theories may not be reliable. In these cases, confidence intervals based on bootstrap-resampling method are usually recommended [13]. Given the observed data $\{(x_{11j},x_{10j},x_{01j},x_{00j}):j=1,\ldots,J\}$, we can obtain the maximum likelihood estimates $\hat{\delta}$ and $\hat{\boldsymbol{p}}_0$ of parameters δ and $\boldsymbol{p}_0$ via the method introduced in Section 2.1.1 (or 2.1.2). Based on the estimates $\hat{\delta}$ and $\hat{p}_0$, we can generate Bootstrap data $\{(x^*_{11j},x^*_{10j},x^*_{01j},x^*_{01j}):j=1,\ldots,J\}$ via the following distribution:

$$(x^*_{11j},x^*_{10j},x^*_{01j},x^*_{01j})\sim\text{Multinomial}(n_j;\hat{p}_{0j}-\hat{p}_{01j},(\hat{\delta}-1)\hat{p}_{0j}+\hat{p}_{01j},\hat{p}_{01j},1-\hat{\delta}\hat{p}_{0j}-\hat{p}_{01j})$$

for $j=1,\ldots,J$. For each generated bootstrap sample $Y^*_{\text{obs}}=\{(x^*_{11j},x^*_{10j},x^*_{01j},x^*_{01j}):j=1,\ldots,J\}$, we can calculate the MLEs $\hat{\delta}^*$ and $\hat{\boldsymbol{p}}^*_0$ of parameters δ and $\boldsymbol{p}_0$. Independently repeating the above process G times, we obtain G bootstrap replications $\{\hat{\delta}^*_g\}_{g=1}^G$ of the MLE of parameter δ via the method introduced in Section 2.1.1. Consequently, the standard error, se(δ), of δ can be estimated by the sample standard deviation of the G replications, i.e.

$$\hat{\text{se}}(\hat{\delta})=\sqrt{\left\{\frac{1}{G-1}\sum_{g=1}^{G}[\hat{\delta}^*_g-(\hat{\delta}^*_1+\cdots+\hat{\delta}^*_G)/G]^2\right\}}$$

If $\{\hat{\delta}^*_g\}_{g=1}^G$ is approximately normally distributed, a $(1-\alpha)100$ per cent bootstrap confidence interval for δ is given by

$$[\hat{\delta}-z_{1-\alpha/2}\hat{\text{se}}(\hat{\delta}),\hat{\delta}+z_{1-\alpha/2}\hat{\text{se}}(\hat{\delta})] \tag{9}$$

which is referred as the simple bootstrap confidence interval.

Alternatively, if $\{\hat{\delta}^*_g\}_{g=1}^G$ is not-normally distributed, a $100(1-\alpha)$ per cent bootstrap confidence interval for δ can be obtained by $[\delta_{Bl},\delta_{Bu}]$, where δ_{Bl} and δ_{Bu} are the $\alpha/2$ and $(1-\alpha/2)$ quantiles of $\{\hat{\delta}^*_g\}_{g=1}^G$, which is referred as the percentile bootstrap confidence interval.

2.2. Simultaneous confidence intervals for relative risks

In Section 2.1, we discuss confidence interval construction for the common relative risk across J strata on the basis of assumption that the relative risks for J strata are equal to some unknown fixed value. Here, one natural question is whether the assumption holds or not. It is well known that multiple comparison can be used to answer this question when $J\geqslant 2$. In particular, we can use $(1-\alpha)$-level simultaneous confidence intervals to analyze multiple comparison of J relative risks $\delta_j=p_{1j}/p_{0j}$ for $j=1,\ldots,J$ as the comparison procedure cannot only answer whether $\delta_j=\delta$ holds or not but also let one know which $j\in\{1,\ldots,J\}$ satisfying $\delta_j>\delta$ and which $j\in\{1,\ldots,J\}$ satisfying $\delta_j<\delta$. Therefore, we shall investigate the simultaneous confidence intervals for J relative risks based on the following methods.

2.2.1. Simultaneous confidence interval for relative risks based on Bonferroni method. For each j ($j=1,2,\ldots,J$), it follows from the arguments of Section 2.1.2 that as $n_j\to+\infty$, we have

$$T_j(\delta_j)=\frac{\breve{\delta}_j-\delta_j}{\breve{\sigma}_j}\xrightarrow{L}N(0,1)$$

where $\breve{\delta}_j=(x_{11j}+x_{10j})/(x_{11j}+x_{01j})$, $\breve{\sigma}_j^2=(x_{10j}+x_{01j})(x_{11j}+x_{10j})/(x_{11j}+x_{01j})^3$, and $\xrightarrow{L}$ denotes convergence in distribution. An approximate $100(1-\alpha)$ per cent confidence interval for δ_j ($j=1,2,\ldots,J$) is given by

$$[\max\{\breve{\delta}_j-z_{1-\alpha/2}\breve{\sigma}_j,0\},\breve{\delta}_j+z_{1-\alpha/2}\breve{\sigma}_j]$$

Their joint coverage probability is less than $1-\alpha$. That is, the above joint confidence intervals cannot attain the pre-specified confidence level nor they cannot control familywise error rate (FWE), which is defined as FWE=Pr (at least one interval

does not contain the true parameter) $\triangleq\alpha$. Here, the simple simultaneous confidence intervals for $\delta_1,\ldots,\delta_J$ defined as $\Pr(\delta_j\in[\max\{\check{\delta}_j-t\check{\sigma}_j,0\},\check{\delta}_j+t\check{\sigma}_j]$ for all $j=1,2,\ldots,J)=1-\alpha$ can be used to control the FWE. For more theories on construction of simple simultaneous confidence intervals, one can consult Roy and Bose [14]. In order to construct the simple simultaneous confidence intervals with the given confidence level $1-\alpha$, it is necessary to find the value of t. However, the exact value of t is difficult to find. The approximate procedure based on Bonferroni's inequality is adopted here to obtain conservative simultaneous confidence intervals. The approximate $100(1-\alpha)$ per cent Bonferroni simultaneous (BS) confidence intervals for $\delta_1,\ldots,\delta_J$ are given by

$$[\max\{\check{\delta}_j-z_{1-\alpha/(2J)}\check{\sigma}_j,0\},\check{\delta}_j+z_{1-\alpha/(2J)}\check{\sigma}_j]\quad\text{for } j=1,2,\ldots,J \tag{10}$$

Again, for each j ($j=1,2,\ldots,J$), it follows from the arguments of Section 2.1.2 that as $n_j\to+\infty$, we have

$$T_j^0(\delta_j)=\frac{\sqrt{n_j/\delta_j}(\check{\delta}_j-\delta_j)\tilde{p}_{0j}}{\sqrt{2\tilde{p}_{01j}+\tilde{p}_{0j}(\delta_j-1)}}\xrightarrow{L}N(0,1)$$

where $\tilde{p}_{0j}$ and $\tilde{p}_{01j}$ can be computed via Equation (3) by using δ_j to replace δ, which yields that the approximate $100(1-\alpha)$ per cent modified Bonferroni simultaneous (MBS) confidence intervals for δ_j ($j=1,\ldots,J$) are given by $[\delta_{lj},\delta_{uj}]$ for $j=1,\ldots,J$, where δ_{lj} and δ_{uj} are the roots of the following equation:

$$T_j^0(\delta_j)=\pm z_{1-\alpha/(2J)} \tag{11}$$

in which the plus and the minus signs correspond to δ_{lj} and δ_{uj}, respectively.

2.2.2. Simultaneous confidence intervals for relative risks based on bootstrap-resampling method. When sample sizes (i.e. n_j) are small, the BS confidence intervals may be unreliable and conservative. In these cases, simultaneous confidence intervals based on bootstrap-resampling method are usually recommended [15]. Given the observed data $\{(x_{11j},x_{10j},x_{01j},x_{00j}):j=1,\ldots,J\}$, we can obtain the maximum likelihood estimates $\hat{\delta}$ and $\hat{\boldsymbol{p}}_0$ of parameters δ and $\boldsymbol{p}_0$ via the method introduced in Section 2.1.2. Based on the estimates $\hat{\delta}$ and $\hat{\boldsymbol{p}}_0$, we can generate Bootstrap data via the following distribution:

$$(x_{11j}^*,x_{10j}^*,x_{01j}^*,x_{00j}^*)\sim\text{Multinomial}(n_j;\hat{p}_{0j}-\hat{p}_{01j},(\hat{\delta}-1)\hat{p}_{0j}+\hat{p}_{01j},\hat{p}_{01j},1-\hat{\delta}\hat{p}_{0j}-\hat{p}_{01j})$$

for $j=1,2,\ldots,J$. For each generated bootstrap sample $Y_{\text{obs}}^*=\{(x_{11j}^*,x_{10j}^*,x_{01j}^*,x_{00j}^*):j=1,\ldots,J\}$, we can compute values of statistic $T_j(\delta_j)$ for $j=1,\ldots,J$ and denote $t^*=\max_{1\leqslant j\leqslant J}|T_j(\delta_j)|$. Independently repeating the above process G times, we obtain G bootstrap replications $\{t_g^*\}_{g=1}^G$ of t^*. Consequently, the estimated value of critical value is $\hat{z}_{(\alpha)}$, which is the $1-\alpha$ empirical quantile of the G values $\{t_g^*\}_{g=1}^G$. The $(1-\alpha)100$ per cent bootstrap-resampling-based simultaneous confidence intervals for δ_j ($j=1,2,\ldots,J$) are thus given by

$$[\max\{\check{\delta}_j-\hat{z}_{(\alpha)}\check{\sigma}_j,0\},\check{\delta}_j+\hat{z}_{(\alpha)}\check{\sigma}_j] \tag{12}$$

3. Simulation studies

In this section, we investigate the performance of various confidence interval estimators (i.e. C=Cochran; MH=Mantel Haenszel; WLS=weighted-least-squares; PL=profile-likelihood ratio; S=score; SB=simple bootstrap; PB=quantile bootstrap; BS=Bonferroni simultaneous; MBS=modified Bonferroni simultaneous; and BBS=bootstrap Bonferroni simultaneous) of δ and δ_j proposed in Section 2 for two strata. The simulation settings cover both balanced (e.g. $n_1=n_2=10$, 20, 30, and 50) and imbalanced (i.e. $n_1=10$ and $n_2=30$, and $n_1=30$ and $n_2=50$) designs. For each pair of given (n_1,n_2), we consider the following probability settings: (i) $q_{11}=0.13$, $q_{12}=0.033$, $p_{011}=0.05$, $p_{012}=0.01$; (ii) $q_{11}=0.13$, $q_{12}=0.033$, $p_{011}=0.1$, $p_{012}=0.032$; (iii) $q_{11}=0.32$, $q_{12}=0.16$, $p_{011}=0.2$, $p_{012}=0.08$; and (iv) $q_{11}=0.32$, $q_{12}=0.16$, $p_{011}=0.3$, $p_{012}=0.13$. The nominal level is chosen to be 0.05.

To investigate the performance of various procedures for the common relative risk, i.e. $\delta_j=\delta_0$ for $j=1$ and 2, we consider $\delta_0=1.5$ and 3.0 for (i) and (ii), and $\delta_0=0.9$, 1.5, and 3.0 for (iii) and (iv). It is noteworthy that all confidence interval estimators considered in this manuscript are test-based confidence intervals. First, we would like to investigate the type I error rate (with null hypotheses $H_0:\delta_j=\delta_0$ for $j=1$, 2) and power (with alternative hypotheses $H_A:\delta_j=\delta_A=1.0$ for $j=1,2$) performance for their respective test statistics for relative risk. For each given setting $(n_1,n_2,q_{11},q_{12},p_{011},p_{012},\delta)$, $M=5000$ data vectors $\boldsymbol{x}^{(m)}=(x_{111}^m,x_{101}^m,x_{011}^m,x_{001}^m,\ldots,x_{11J}^m,x_{10J}^m,x_{01J}^m,x_{00J}^m)$ are generated with δ being δ_0 for type I error rate investigation and with δ being δ_A for power investigation. Note that when $x_{11j}=x_{10j}=x_{01j}=0$, the Cochran and WLS statistics are undefined. In these cases, we apply the commonly used adjustment for sparse data in contingency table analysis by adding 0.5 to each cell. The type I error rate and power for a given test, say T, at the 0.05 level are simply estimated by (the number of rejections of H_0 by test T at the 0.05 level in 5000 replications) with $\delta=\delta_0$ and $\delta=\delta_A$, respectively. The 95 per cent confidence interval for the type I error rate with its true nominal level value being 0.05 is given by $(0.05-2\sqrt{0.95(1-0.95)/5000},0.05+2\sqrt{0.95(1-0.95)/5000})=(0.0438,0.0562)$. The estimate type I error rates (with null hypotheses $H_0:\delta_j=\delta_0$ for $j=1$, 2) and power (with alternative hypotheses $H_A:\delta_j=\delta_A=1.0$

Table I. Empirical type I error rates (per cent) of various tests based on 5000 trials with $\alpha=5$ per cent for two strata.

n_1	n_2	q_{11}	q_{12}	p_{011}	p_{012}	δ_0	C	MH	WLS	PL	S	SB	PB	BS	MBS	BBS
10	10	0.13	0.033	0.05	0.01	1.5	9.3	12.6	19.5	2.2	2.4	4.7	5.3	17.6	10.3	4.9
						3.0	7.7	10.1	18.1	3.8	3.2	4.6	4.4	14.0	6.8	4.6
				0.1	0.032	1.5	10.1	12.6	16.5	3.3	3.9	4.3	4.8	14.5	11.6	5.1
						3.0	6.9	8.2	9.3	3.6	3.8	4.8	5.1	9.9	7.5	5.5
		0.32	0.16	0.2	0.08	0.9	6.0	6.1	6.1	3.7	3.9	5.4	5.8	6.9	6.3	5.3
						1.5	12.9	13.6	15.8	3.1	3.3	4.8	5.0	10.5	8.4	4.8
						3.0	9.5	10.8	16.8	7.9	6.6	5.8	5.3	13.4	9.8	5.4
				0.3	0.13	0.9	6.1	6.9	11.1	3.9	4.3	4.6	4.8	7.3	6.0	4.8
						1.5	11.4	14.5	15.9	3.8	4.0	5.1	4.8	14.1	12.4	5.7
						3.0	13.9	16.8	17.7	4.4	4.6	4.8	5.0	13.6	11.8	4.7
20	20	0.13	0.033	0.05	0.01	1.5	7.3	8.4	11.8	5.5	5.4	5.3	5.0	11.5	9.8	5.8
						3.0	9.5	12.4	14.1	4.2	4.6	5.0	5.1	11.5	9.8	5.8
				0.1	0.032	1.5	9.3	8.7	10.2	5.5	5.3	4.8	5.2	9.2	7.9	5.2
						3.0	10.5	11.9	15.0	7.5	6.6	5.9	5.4	12.4	17.3	4.7
		0.32	0.16	0.2	0.08	0.9	4.3	5.7	9.7	3.1	3.9	4.6	4.9	6.6	5.8	5.0
						1.5	8.1	9.6	11.8	3.4	3.2	5.3	5.8	10.4	7.7	5.8
						3.0	5.9	6.5	15.0	4.8	5.0	5.8	5.1	12.3	6.4	4.9
				0.3	0.13	0.9	4.8	5.8	10.4	5.5	5.3	5.3	5.0	11.7	6.0	5.2
						1.5	9.7	7.5	9.0	3.4	4.0	4.9	5.1	7.2	5.9	5.3
						3.0	11.3	12.6	12.7	6.9	6.5	5.9	5.5	12.1	9.8	5.4
30	30	0.13	0.033	0.05	0.01	1.5	7.5	8.1	12.6	5.4	5.0	5.5	5.3	10.7	9.6	4.7
						3.0	12.9	9.5	14.2	5.9	5.7	5.8	5.5	10.8	7.0	4.6
				0.1	0.032	1.5	7.7	8.3	10.3	4.7	4.9	5.8	5.0	9.8	6.0	5.5
						3.0	13.6	8.0	13.0	6.6	6.3	6.0	5.8	10.6	6.5	4.9
		0.32	0.16	0.2	0.08	0.9	4.6	5.5	7.1	4.8	5.0	5.4	5.0	5.3	5.0	5.1
						1.5	5.8	6.4	7.8	5.7	5.7	4.9	5.2	9.5	6.3	4.8
						3.0	6.6	7.1	7.2	5.3	5.1	5.3	4.8	7.6	6.5	5.7
				0.3	0.13	0.9	5.2	5.8	9.7	5.3	5.4	5.7	5.0	9.5	5.7	4.8
						1.5	7.1	6.6	9.2	5.4	5.8	4.9	4.8	8.6	6.4	5.1
						3.0	7.7	8.4	10.5	7.6	7.0	5.4	5.3	12.3	7.5	5.3
50	50	0.13	0.033	0.05	0.01	1.5	5.3	6.0	6.3	5.5	5.3	5.3	5.0	5.9	5.5	5.5
						3.0	5.9	5.8	6.1	5.5	5.5	5.4	5.2	6.0	6.0	4.9
				0.1	0.032	1.5	5.0	5.7	5.4	4.8	4.4	5.0	5.0	5.8	5.7	5.1
						3.0	5.8	6.8	9.6	3.9	4.3	5.4	4.9	6.2	5.8	4.7
		0.32	0.16	0.2	0.08	0.9	4.6	4.9	5.8	5.0	5.0	5.4	4.9	7.0	6.0	5.2
						1.5	5.7	6.7	8.4	5.5	5.0	5.4	4.9	7.0	6.0	5.2
						3.0	5.0	5.0	5.3	4.4	4.6	4.9	5.1	5.5	5.3	5.1
				0.3	0.13	0.9	4.7	5.8	6.0	5.1	5.1	5.4	5.2	4.4	5.5	4.5
						1.5	5.8	6.8	7.7	4.0	4.4	5.0	4.9	6.6	5.9	5.2
						3.0	4.7	5.3	5.4	4.8	5.0	5.1	5.1	6.0	5.9	5.2
10	30	0.13	0.033	0.05	0.01	1.5	8.9	9.8	14.7	4.1	4.3	5.5	4.8	13.1	10.3	5.3
						3.0	6.6	8.3	12.8	4.7	5.0	5.1	4.9	9.9	7.4	5.0
				0.1	0.032	1.5	9.7	10.5	12.4	5.4	5.2	4.9	4.8	10.6	6.8	5.1
						3.0	6.6	8.3	13.6	5.3	4.9	5.1	5.4	10.7	7.1	5.8
		0.32	0.16	0.2	0.08	0.9	7.6	7.3	7.8	6.4	6.0	5.8	5.3	6.4	6.9	5.0
						1.5	11.3	10.9	10.5	6.3	5.9	5.8	5.4	12.1	9.8	5.5
						3.0	8.8	9.6	15.7	5.7	5.5	5.3	5.1	12.9	7.4	5.7
				0.3	0.13	0.9	6.3	6.9	7.1	5.8	5.3	5.3	5.0	6.6	6.0	4.9
						1.5	10.9	9.8	13.7	7.9	7.6	5.9	5.7	9.8	8.7	5.2
						3.0	12.3	9.8	13.7	7.9	7.6	5.9	5.7	9.8	8.7	5.2
30	50	0.13	0.033	0.05	0.01	1.5	4.9	6.6	8.9	4.6	4.8	5.0	5.0	8.9	7.2	4.9
						3.0	11.1	7.4	10.5	5.4	5.2	5.2	5.6	9.3	6.0	5.6
				0.1	0.032	1.5	5.8	6.0	6.5	4.9	5.0	5.1	5.0	6.1	5.8	5.0
						3.0	5.9	6.1	7.1	3.8	4.6	5.4	4.9	7.3	6.3	5.4
		0.32	0.16	0.2	0.08	0.9	4.3	4.5	5.0	4.0	4.8	5.1	4.7	5.8	5.1	5.1

Table I. *Continued*.

n_1	n_2	q_{11}	q_{12}	p_{011}	p_{012}	δ_0	C	MH	WLS	PL	S	SB	PB	BS	MBS	BBS
						1.5	5.8	5.5	6.0	4.9	5.0	5.1	4.9	6.1	5.7	5.0
						3.0	5.5	5.1	5.8	4.2	4.5	4.8	4.6	6.4	5.9	5.3
				0.3	0.13	0.9	4.5	4.9	5.8	4.1	4.4	5.3	5.1	4.1	4.6	4.6
						1.5	5.9	7.0	7.0	4.0	4.3	4.5	4.7	7.7	5.6	4.9
						3.0	5.7	6.9	9.7	5.7	5.5	5.8	5.4	8.3	6.5	5.8

Table II. Empirical powers (per cent) of various tests based on 5000 trials with $\alpha=5$ per cent and $H_A:\delta_i=\delta_A=1.0$ for two strata.

n_1	n_2	q_{11}	q_{12}	p_{011}	p_{012}	δ_0	C	MH	WLS	PL	S	SB	PB	BS	MBS	BBS
10	10	0.13	0.033	0.05	0.01	1.5	99.4	99.7	100.0	99.3	99.4	99.4	99.4	100.0	99.3	99.4
						3.0	96.7	96.4	100.0	95.4	95.5	96.5	96.4	100.0	96.2	96.4
				0.1	0.032	1.5	99.9	99.9	99.9	99.0	99.2	99.4	99.2	99.9	99.9	99.2
						3.0	97.9	98.1	99.5	97.9	98.0	98.4	98.2	100.0	99.1	98.7
		0.32	0.16	0.2	0.08	0.9	97.9	95.8	100.0	92.1	92.3	93.7	94.5	99.3	97.2	93.1
						1.5	98.6	98.4	99.2	95.7	96.0	97.1	98.0	99.0	98.2	97.0
						3.0	98.4	98.1	98.5	94.9	95.1	95.0	94.9	99.4	97.3	94.2
				0.3	0.13	0.9	89.4	89.9	91.1	90.4	90.8	91.6	92.0	91.7	88.1	91.0
						1.5	96.0	95.7	96.3	96.5	96.5	96.0	96.1	97.9	96.2	96.2
						3.0	92.7	93.5	94.3	91.5	92.0	94.2	94.6	98.0	95.7	95.1
20	20	0.13	0.033	0.05	0.01	1.5	100.0	100.0	99.9	99.0	99.1	100.0	100.0	100.0	100.0	99.4
						3.0	99.4	98.1	100.0	98.5	98.7	99.1	99.4	100.0	98.7	98.9
				0.1	0.032	1.5	93.9	94.3	95.7	92.5	92.7	92.1	93.0	94.7	93.1	91.7
						3.0	91.1	92.4	90.8	91.1	91.5	92.8	93.4	100.0	97.2	94.0
		0.32	0.16	0.2	0.08	0.9	88.2	89.5	90.6	89.4	91.0	91.4	92.0	88.0	97.8	91.6
						1.5	96.1	97.8	96.6	92.7	93.0	93.4	94.1	98.5	96.2	93.7
						3.0	95.9	96.4	97.1	87.6	88.4	90.5	91.6	98.3	97.0	92.1
				0.3	0.13	0.9	87.1	88.2	89.3	84.2	84.6	85.1	84.9	90.6	88.7	84.3
						1.5	92.7	90.1	95.5	92.6	93.0	93.4	93.7	97.7	96.0	93.2
						3.0	95.4	94.9	97.1	87.6	88.0	89.4	90.0	98.1	96.9	90.7
30	30	0.13	0.033	0.05	0.01	1.5	100.0	100.0	99.8	97.4	97.5	97.8	98.1	98.6	98.4	98.0
						3.0	98.4	99.1	100.0	97.4	97.0	96.7	97.2	100.0	99.4	97.7
				0.1	0.032	1.5	95.8	96.2	97.4	90.7	91.0	93.1	93.6	97.9	97.0	92.9
						3.0	94.0	94.4	98.6	83.6	84.1	83.4	84.0	95.1	93.8	85.1
		0.32	0.16	0.2	0.08	0.9	97.9	98.4	97.8	93.8	94.0	94.7	95.2	97.2	96.4	95.0
						1.5	90.6	91.1	91.4	89.9	90.2	91.3	91.0	89.6	90.2	90.4
						3.0	95.6	96.0	95.7	87.8	88.1	89.6	90.4	96.2	95.8	90.1
				0.3	0.13	0.9	88.1	98.7	89.2	84.4	85.0	86.2	86.7	89.4	88.7	87.0
						1.5	82.2	88.9	93.9	90.2	90.5	91.4	91.8	96.5	94.7	92.4
						3.0	94.9	95.6	95.8	89.2	89.4	90.2	90.7	96.7	95.3	91.3
50	50	0.13	0.033	0.05	0.01	1.5	97.3	98.4	98.9	93.7	94.0	94.6	95.2	95.4	94.3	95.0
						3.0	98.2	98.9	99.8	93.1	93.5	93.2	93.4	98.1	97.5	93.2
				0.1	0.032	1.5	96.9	97.5	97.8	93.8	93.0	94.2	94.6	97.1	95.7	94.5
						3.0	94.7	94.9	95.1	86.3	86.4	85.1	86.0	95.4	94.2	86.5
		0.32	0.16	0.2	0.08	0.9	87.6	88.3	89.4	82.9	83.1	84.2	84.5	88.5	86.4	84.1
						1.5	92.4	93.6	95.4	86.3	86.5	87.0	87.2	93.7	91.5	85.1
						3.0	94.8	95.5	96.1	91.8	92.0	92.5	92.7	95.0	83.9	92.6
				0.3	0.13	0.9	86.2	87.2	88.5	84.2	84.5	84.6	84.3	85.1	85.9	84.5
						1.5	92.7	93.4	95.5	86.5	86.7	87.1	87.5	93.7	93.4	87.4
						3.0	93.5	94.0	94.3	88.2	88.5	89.2	89.8	97.0	93.7	89.5
10	30	0.13	0.033	0.05	0.01	1.5	96.7	95.2	99.8	94.6	94.5	95.2	95.7	99.8	97.4	95.3
						3.0	98.3	97.9	100.0	96.5	96.8	96.9	97.4	99.5	96.8	97.0
				0.1	0.032	1.5	97.7	98.4	98.7	98.1	98.0	98.2	98.5	99.7	98.5	98.4

Table II. *Continued*.

n_1	n_2	q_{11}	q_{12}	p_{011}	p_{012}	δ_0	C	MH	WLS	PL	S	SB	PB	BS	MBS	BBS
						3.0	94.3	93.7	99.4	97.2	97.4	97.9	98.2	100.0	98.6	98.1
		0.32	0.16	0.2	0.08	0.9	91.9	92.0	92.2	90.4	90.7	91.3	92.0	92.7	90.9	91.6
						1.5	96.4	95.2	95.7	95.9	96.0	96.3	96.8	98.6	97.4	96.0
						3.0	95.4	90.1	97.0	95.3	94.9	95.4	95.8	98.6	96.7	95.6
				0.3	0.13	0.9	94.7	91.8	92.2	90.6	91.0	90.8	91.2	91.4	94.9	92.1
						1.5	96.1	96.0	94.3	94.7	94.7	95.0	95.4	96.2	95.8	95.2
						3.0	97.2	95.8	95.3	94.8	95.0	95.3	95.6	97.8	96.2	95.4
30	50	0.13	0.033	0.05	0.01	1.5	96.9	97.4	97.6	94.5	94.9	95.1	95.3	98.1	97.8	95.0
						3.0	100.0	99.8	100.0	94.2	94.5	96.2	96.8	100.0	98.4	96.3
				0.1	0.032	1.5	97.1	96.5	97.1	92.8	93.0	93.7	94.6	99.5	96.2	94.7
						3.0	96.9	96.9	96.9	89.5	90.2	92.3	93.1	96.2	95.4	92.9
		0.32	0.16	0.2	0.08	0.9	89.5	90.2	91.2	86.1	86.3	87.1	87.5	88.3	87.8	87.3
						1.5	96.2	95.8	95.4	91.7	92.3	93.4	94.2	97.9	96.2	92.9
						3.0	96.2	96.5	94.7	86.6	87.0	88.2	88.9	96.9	95.7	88.6
				0.3	0.13	0.9	90.8	92.4	91.3	85.2	85.0	87.6	88.3	89.2	88.9	88.1
						1.5	94.0	93.7	95.0	91.5	91.8	92.1	93.4	97.9	95.8	93.5
						3.0	96.2	95.6	95.0	88.1	88.7	89.3	90.2	97.3	95.3	89.8

for $j=1$, 2) of various tests for relative risk are reported in Tables I and II, respectively. Following Tang *et al.* [10], we define a test to be *liberal* if its empirical type I error rate >0.0562 for $\alpha=0.05$, to be *conservative* if its empirical type I error rate <0.0438 for $\alpha=0.05$, to be *robust* otherwise. According to the results, we have the following observations:

(a) Asymptotic tests based on C (i.e. Cochran), MH (i.e. Mantel Haenszel), WLS (i.e. weighted-least-squares), BS (i.e. Bonferroni simultaneous), and MBS (i.e. modified Bonferroni simultaneous) can be regarded as liberal. In the worst cases, their empirical type I error rates could be greater than 0.10.
(b) Asymptotic tests based on PL (i.e. profile-likelihood ratio) and S (i.e. Score) are occasionally liberal in small-sample designs (e.g. $n_1=n_2\leqslant 30$). They become robust when sample sizes increase (e.g. $n_1=n_2\geqslant 50$, and $n_1=30$ and $n_2=50$).
(c) All bootstrapped tests (i.e. simple bootstrap, quantile bootstrap, and BBS) can be classified to be robust.
(d) Given that the type I error rates are controlled around the pre-assigned nominal level, all bootstrap tests yield the largest powers and they are hence recommended in practice.

Second, we investigate the performance of various confidence intervals for common relative risk (i.e. $\delta_j=\delta_0$, $j=1$, 2) with respect to their coverage probabilities and confidence interval widths. For this purpose, we adopt the same configuration settings for type I error rate assessment. Let $[l(\boldsymbol{x}^{(m)}), u(\boldsymbol{x}^{(m)})]$ be any common relative risk confidence interval for δ (i.e. C, MH, WLS, PL, S, SB, and PB) or $[l_j(\boldsymbol{x}^{(m)}), u_j(\boldsymbol{x}^{(m)})]$ be any simultaneous relative risk confidence interval for δ_j, $j=1$, 2 (i.e. BS, MBS, and BBS), their corresponding empirical coverage probabilities and expected widths are estimated by

(i) *Coverage probability*

$$\frac{1}{M}\sum_{m=1}^{M} I\{\delta_0\in[l(\boldsymbol{x}^{(m)}), u(\boldsymbol{x}^{(m)})]\}\left(\text{or}\quad \frac{1}{M}\sum_{m=1}^{M} I\{\delta_j\in[l_j(\boldsymbol{x}^{(m)}), u_j(\boldsymbol{x}^{(m)})]: j=1,\ldots,J\}\right)\quad\text{and}$$

(ii) *Expected interval width*

$$\frac{1}{M}\sum_{m=1}^{M}[u(\boldsymbol{x}^{(m)})-l(\boldsymbol{x}^{(m)})]\left(\text{or}\ \frac{1}{M}\sum_{m=1}^{M}\left[\min_j\{u_j(\boldsymbol{x}^{(m)})\}-\max_j\{l_j(\boldsymbol{x}^{(m)})\}\right]\right)$$

The corresponding results are reported in Tables III and IV. In addition, we also investigate the performance of the three simultaneous confidence intervals (i.e. BS, MBS, and BBS) when the relative risks are different in the two strata and the results are reported in Tables V and VI. For the number of replications being $M=5000$, it is noted that an estimated empirical coverage probability is not significantly different from the given coverage level (i.e. 0.95 in our case) if it lies within $(0.95-2\sqrt{0.95(1-0.95)/5000}, 0.95+2\sqrt{0.95(1-0.95)/5000})=(0.9438, 0.9562)$. According to the simulation results, we have the following observations:

(a) Confidence intervals based on the score statistic (i.e. S) and the profile-likelihood test statistic (i.e. PL) produce similar coverage probabilities, and generally more preferable than those asymptotic confidence intervals (i.e. C, MH, WLS, BS, and MBS) for even small sample sizes in the sense that the coverage probabilities of the former are closer to pre-given confidence level than those of the latter.

Table III. Empirical coverage probabilities (per cent) for various 95 per cent confidence intervals for common relative risk δ for two strata.

n_1	n_2	q_{11}	q_{12}	p_{011}	p_{012}	δ_0	C	MH	WLS	PL	S	SB	PB	BS	MBS	BBS
10	10	0.13	0.033	0.05	0.01	1.5	90.7	87.4	80.5	97.8	97.6	95.3	94.7	82.4	89.7	95.1
						3.0	92.3	89.9	81.9	96.2	96.8	95.4	95.6	86.0	93.2	95.4
				0.1	0.032	1.5	89.1	87.4	83.5	96.6	96.1	95.7	95.2	85.5	88.4	94.9
						3.0	93.1	91.8	90.7	96.4	96.2	95.2	94.9	90.1	92.5	94.5
		0.32	0.16	0.2	0.08	0.9	94.0	93.9	93.9	96.3	96.1	94.6	94.2	93.1	93.7	94.7
						1.5	87.1	86.4	84.2	96.9	96.7	95.2	95.0	89.5	91.6	95.2
						3.0	90.5	89.2	83.2	92.1	93.4	94.2	94.7	86.6	90.2	94.6
				0.3	0.13	0.9	93.9	93.1	88.9	96.1	95.7	95.4	95.2	92.7	94.0	95.2
						1.5	88.6	85.4	84.1	96.2	96.0	94.9	95.2	85.9	87.6	94.3
						3.0	86.1	83.2	82.3	95.6	95.4	95.2	95.0	86.4	88.2	95.3
20	20	0.13	0.033	0.05	0.01	1.5	92.7	91.6	88.2	94.5	94.6	94.7	95.0	88.5	93.2	95.0
						3.0	90.5	87.6	85.9	95.8	95.4	95.0	94.9	88.5	90.2	94.2
				0.1	0.032	1.5	90.7	91.3	89.8	94.5	94.7	95.2	94.8	90.8	92.1	94.8
						3.0	89.5	88.1	85.0	92.5	93.4	94.1	94.6	87.6	92.7	95.3
		0.32	0.16	0.2	0.08	0.9	95.7	94.3	90.3	96.9	96.1	95.4	95.1	93.4	94.2	95.0
						1.5	91.9	90.4	88.2	96.6	96.8	94.7	94.2	89.6	92.3	94.2
						3.0	94.1	93.5	85.0	95.2	95.0	94.2	94.9	87.7	93.6	95.1
				0.3	0.13	0.9	95.2	94.2	89.6	94.5	94.7	94.7	95.0	88.3	94.0	94.8
						1.5	90.3	92.5	91.0	96.6	96.0	95.1	94.9	92.8	94.1	94.7
						3.0	88.7	87.4	87.3	93.1	93.5	94.1	94.5	87.9	90.2	94.6
30	30	0.13	0.033	0.05	0.01	1.5	92.5	91.9	87.4	94.6	95.0	94.5	94.7	89.3	90.4	95.3
						3.0	87.1	90.5	85.8	94.1	94.3	94.2	94.5	89.2	93.0	95.4
				0.1	0.032	1.5	92.3	91.7	89.7	95.3	95.1	94.2	95.0	90.2	94.0	94.5
						3.0	86.4	92.0	87.0	93.4	93.7	94.0	94.2	89.4	93.5	95.1
		0.32	0.16	0.2	0.08	0.9	95.4	94.5	92.9	95.2	95.0	94.6	95.0	94.7	95.0	94.9
						1.5	94.2	93.6	92.2	94.3	94.3	95.1	94.8	90.5	93.7	95.2
						3.0	93.4	92.9	92.8	94.7	94.9	94.7	95.2	92.4	93.5	94.3
				0.3	0.13	0.9	94.8	94.2	90.3	94.7	94.6	94.3	95.0	90.5	94.3	95.2
						1.5	92.9	93.4	90.8	94.6	94.2	95.1	95.2	91.4	93.6	94.9
						3.0	92.3	91.6	89.5	92.4	93.0	94.6	94.7	88.7	92.5	94.7
50	50	0.13	0.033	0.05	0.01	1.5	94.7	94.0	93.7	94.5	94.7	94.7	95.0	94.1	94.5	94.5
						3.0	94.1	94.2	93.9	94.5	94.5	94.6	94.8	94.0	94.0	95.1
				0.1	0.032	1.5	95.0	94.3	94.6	95.2	95.6	95.0	95.0	94.2	94.3	94.9
						3.0	94.2	93.2	90.4	96.1	95.7	94.6	95.1	93.8	94.2	95.3
		0.32	0.16	0.2	0.08	0.9	95.4	95.1	94.2	95.0	95.0	94.9	95.1	94.8	94.5	95.0
						1.5	94.3	93.3	91.6	94.5	95.0	94.6	95.1	93.0	94.0	94.8
						3.0	95.0	95.0	94.7	95.6	95.4	95.1	94.9	94.5	94.7	94.9
				0.3	0.13	0.9	95.3	94.2	94.0	94.9	94.9	94.6	94.8	95.6	94.5	95.5
						1.5	94.2	93.2	92.3	96.0	95.6	95.0	95.1	93.4	94.0	94.3
						3.0	95.3	94.7	94.6	95.2	95.0	94.9	94.9	94.0	94.1	94.8
10	30	0.13	0.033	0.05	0.01	1.5	91.1	90.2	85.3	95.9	95.7	94.5	95.2	86.9	89.7	94.7
						3.0	93.4	91.7	87.2	95.3	95.0	94.9	95.1	90.1	92.6	95.0
				0.1	0.032	1.5	90.3	89.5	87.6	94.8	94.8	95.1	95.2	89.4	93.2	94.9
						3.0	93.4	91.7	86.6	94.7	95.1	94.9	94.6	89.3	92.9	94.2
		0.32	0.16	0.2	0.08	0.9	92.4	92.7	92.2	93.6	94.0	94.2	94.7	93.6	93.1	95.0
						1.5	88.7	89.1	89.5	93.7	94.1	94.2	94.6	87.9	90.2	94.5
						3.0	91.2	90.4	84.7	94.3	94.5	94.7	94.9	87.1	92.6	94.3
				0.3	0.13	0.9	93.7	93.1	92.9	94.2	94.7	94.7	95.0	93.4	94.0	95.1
						1.5	89.1	90.2	89.5	96.4	96.2	95.5	95.2	88.9	93.6	94.7
						3.0	87.7	90.2	86.3	92.1	92.4	94.1	94.3	90.2	91.3	94.8
30	50	0.13	0.033	0.05	0.01	1.5	95.1	93.4	91.1	95.4	95.2	95.0	95.0	91.1	92.8	95.1
						3.0	88.9	92.6	89.5	94.6	94.8	94.8	94.4	90.7	94.0	94.4
				0.1	0.032	1.5	94.2	94.0	93.5	95.1	95.0	94.9	95.0	93.9	94.2	95.0
						3.0	94.1	93.9	92.9	96.2	95.4	94.6	95.1	92.7	93.7	94.6

Table III. *Continued*.

n_1	n_2	q_{11}	q_{12}	p_{011}	p_{012}	δ_0	C	MH	WLS	PL	S	SB	PB	BS	MBS	BBS
		0.32	0.16	0.2	0.08	0.9	95.7	95.4	95.0	96.0	95.2	94.9	95.3	94.2	94.9	94.9
						1.5	94.2	94.5	94.0	95.1	95.0	94.9	95.1	93.9	94.3	95.0
						3.0	94.5	94.9	94.2	95.8	95.5	95.2	95.4	93.6	94.1	94.7
				0.3	0.13	0.9	95.5	95.1	94.2	95.9	95.6	94.7	94.9	95.9	95.4	95.4
						1.5	94.1	93.0	93.0	96.0	95.7	95.5	95.3	92.3	94.4	95.1
						3.0	94.3	93.1	90.3	94.3	94.5	94.2	94.6	91.7	93.5	94.2

Table IV. Expected confidence width for various 95 per cent confidence intervals for common relative risk δ for two strata.

n_1	n_2	q_{11}	q_{12}	p_{011}	p_{012}	δ_0	C	MH	WLS	PL	S	SB	PB	BS	MBS	BBS
10	10	0.13	0.033	0.05	0.01	1.5	1.375	1.010	1.009	1.344	1.342	1.201	1.194	1.324	1.316	1.187
						3.0	1.974	1.752	1.823	1.921	1.913	1.623	1.619	1.885	1.876	1.702
				0.1	0.032	1.5	1.654	1.520	1.401	1.530	1.541	1.413	1.406	1.542	1.433	1.422
						3.0	2.519	2.403	2.468	2.435	2.431	2.326	2.318	2.501	2.494	2.433
		0.32	0.16	0.2	0.08	0.9	0.674	0.662	0.651	0.644	0.637	0.636	0.631	0.668	0.659	0.629
						1.5	1.674	1.521	1.481	1.552	1.543	1.465	1.458	1.534	1.546	1.451
						3.0	2.505	2.478	2.454	2.461	2.452	2.431	2.425	2.468	2.473	2.429
				0.3	0.13	0.9	0.769	0.750	0.758	0.741	0.735	0.705	0.694	0.754	0.747	0.701
						1.5	2.107	1.994	1.717	1.993	1.984	1.975	1.967	2.129	2.116	1.970
						3.0	2.761	2.757	2.724	2.618	2.611	2.437	2.431	2.768	2.780	2.440
20	20	0.13	0.033	0.05	0.01	1.5	1.044	0.976	0.747	0.951	0.948	0.921	0.914	0.978	0.964	0.927
						3.0	2.538	2.579	3.005	2.533	2.527	2.514	2.506	2.616	2.601	2.524
				0.1	0.032	1.5	1.028	1.017	0.803	1.019	1.014	1.006	0.994	1.023	1.015	1.020
						3.0	2.703	2.686	3.056	2.438	2.431	2.220	2.196	3.004	2.749	2.225
		0.32	0.16	0.2	0.08	0.9	0.821	0.715	0.726	0.405	0.411	0.421	0.421	0.735	0.714	0.430
						1.5	1.378	1.354	1.045	1.365	1.360	1.332	1.326	1.309	1.297	1.341
						3.0	2.894	2.783	2.798	2.396	2.391	2.377	2.368	2.778	2.761	2.372
				0.3	0.13	0.9	0.527	0.482	0.365	0.513	0.510	0.433	0.419	0.354	0.397	0.428
						1.5	1.502	1.489	1.193	1.421	1.416	1.388	1.376	1.493	1.487	1.392
						3.0	2.907	2.765	2.432	2.584	2.572	2.446	2.430	2.758	2.749	2.580
30	30	0.13	0.033	0.05	0.01	1.5	0.895	0.873	0.591	0.741	0.743	0.716	0.709	0.749	0.730	0.711
						3.0	2.916	2.714	2.373	2.937	2.920	2.367	2.361	3.190	2.954	2.352
				0.1	0.032	1.5	0.984	0.869	0.656	0.820	0.814	0.704	0.682	0.835	0.817	0.693
						3.0	1.841	1.963	2.509	1.832	1.829	1.824	1.810	2.000	1.942	1.815
		0.32	0.16	0.2	0.08	0.9	0.334	0.321	0.274	0.318	0.314	0.234	0.218	0.204	0.275	0.206
						1.5	1.430	1.395	0.846	1.014	0.992	0.840	0.829	1.044	0.967	0.831
						3.0	2.830	2.997	3.052	2.492	2.487	2.476	2.462	2.920	2.889	2.489
				0.3	0.13	0.9	0.422	0.334	0.259	0.297	0.290	0.257	0.248	0.303	0.298	0.250
						1.5	1.140	1.184	0.959	1.153	1.142	0.956	0.940	1.179	1.133	0.948
						3.0	1.865	1.976	2.001	1.976	1.970	1.858	1.844	1.985	1.899	1.840
50	50	0.13	0.033	0.05	0.01	1.5	0.585	0.562	0.469	0.498	0.488	0.453	0.441	0.589	0.557	0.432
						3.0	1.776	1.795	1.821	1.648	1.641	1.622	1.618	2.405	1.902	1.625
				0.1	0.032	1.5	0.814	0.786	0.502	0.791	0.786	0.713	0.704	0.723	0.718	0.709
						3.0	1.841	1.865	1.893	1.901	1.885	1.830	1.814	2.446	2.104	1.824
		0.32	0.16	0.2	0.08	0.9	0.238	0.235	0.141	0.234	0.223	0.139	0.120	0.164	0.152	0.127
						1.5	0.890	0.876	0.664	0.885	0.880	0.656	0.649	0.806	0.812	0.642
						3.0	1.930	2.057	2.020	2.384	2.368	1.914	1.902	1.998	2.032	1.911
				0.3	0.13	0.9	0.304	0.289	0.205	0.276	0.261	0.211	0.203	0.239	0.228	0.204
						1.5	1.042	0.994	0.746	0.945	0.932	0.742	0.724	0.912	0.876	0.733
						3.0	2.060	2.245	2.080	2.704	2.686	2.214	2.201	2.527	2.693	2.207
10	30	0.13	0.033	0.05	0.01	1.5	0.842	0.776	0.694	0.837	0.811	0.768	0.750	0.808	0.784	0.761
						3.0	2.496	2.465	2.895	2.277	2.261	2.242	2.219	2.502	2.473	2.233

Table IV. *Continued*.

n_1	n_2	q_{11}	q_{12}	p_{011}	p_{012}	δ_0	C	MH	WLS	PL	S	SB	PB	BS	MBS	BBS
				0.1	0.032	1.5	0.866	0.828	0.749	0.765	0.743	0.732	0.715	0.867	0.819	0.719
						3.0	2.800	2.914	3.019	2.418	2.401	2.396	2.381	2.614	2.562	2.387
		0.32	0.16	0.2	0.08	0.9	0.290	0.288	0.286	0.259	0.250	0.247	0.231	0.255	0.262	0.249
						1.5	1.204	1.193	0.958	1.189	1.175	0.950	0.932	1.089	0.999	0.940
						3.0	3.211	3.498	3.586	2.899	2.901	2.884	2.865	3.362	3.194	2.877
				0.3	0.13	0.9	0.346	0.327	0.276	0.308	0.302	0.270	0.244	0.310	0.291	0.255
						1.5	1.535	1.326	1.074	1.282	1.266	1.068	1.042	1.228	1.197	1.051
						3.0	2.915	2.891	2.762	2.836	2.820	2.754	2.316	2.936	2.755	2.586
30	50	0.13	0.033	0.05	0.01	1.5	0.799	0.754	0.509	0.767	0.760	0.498	0.487	0.625	0.576	0.511
						3.0	2.530	2.432	2.013	2.551	2.498	1.987	1.976	2.515	2.240	2.112
				0.1	0.032	1.5	0.679	0.664	0.552	0.635	0.628	0.567	0.549	0.670	0.658	0.550
						3.0	2.565	2.377	2.111	2.584	2.576	2.119	2.007	2.611	2.489	2.114
		0.32	0.16	0.2	0.08	0.9	0.224	0.198	0.145	0.244	0.230	0.164	0.148	0.164	0.153	0.150
						1.5	0.741	0.724	0.712	0.831	0.824	0.710	0.702	0.853	0.819	0.708
						3.0	2.244	2.219	2.321	2.116	2.110	2.008	1.992	2.506	2.389	2.017
				0.3	0.13	0.9	0.245	0.230	0.212	0.225	0.220	0.208	0.191	0.226	0.213	0.198
						1.5	0.897	0.862	0.798	0.873	0.854	0.786	0.743	0.941	0.892	0.761
						3.0	2.895	2.666	2.338	2.238	2.230	2.224	2.187	2.714	2.576	2.193

Table V. Empirical coverage probabilities (per cent) for 95 per cent simultaneous confidence intervals of δ_j for two strata.

n_1	n_2	q_{11}	q_{12}	p_{011}	p_{012}	δ_1	δ_2	BS	MBS	BBS
10	10	0.13	0.033	0.05	0.01	0.9	1.5	87.1	90.4	95.7
						0.9	3.0	83.8	85.6	94.7
						1.5	3.0	81.9	82.4	94.2
				0.1	0.032	0.9	1.5	81.4	83.6	94.0
						0.9	3.0	88.7	89.2	94.6
						1.5	3.0	85.4	86.2	94.3
		0.32	0.16	0.2	0.08	0.9	1.5	92.8	93.5	95.6
						0.9	3.0	80.7	82.6	94.0
						1.5	3.0	90.6	92.9	95.4
				0.3	0.13	0.9	1.5	93.8	94.2	95.0
						0.9	3.0	84.1	86.9	94.8
						1.5	3.0	92.2	93.4	95.3
30	30	0.13	0.033	0.05	0.01	0.9	1.5	85.5	86.2	94.2
						0.9	3.0	86.8	87.5	95.1
						1.5	3.0	88.1	89.2	94.9
				0.1	0.032	0.9	1.5	79.4	83.4	93.9
						0.9	3.0	85.6	87.2	94.6
						1.5	3.0	88.5	88.9	94.8
		0.32	0.16	0.2	0.08	0.9	1.5	83.5	87.4	94.2
						0.9	3.0	91.4	92.6	95.0
						1.5	3.0	91.7	93.2	95.4
				0.3	0.13	0.9	1.5	79.5	83.6	94.3
						0.9	3.0	85.8	87.9	95.3
						1.5	3.0	89.4	92.5	94.9
50	50	0.13	0.033	0.05	0.01	0.9	1.5	92.3	93.5	94.2
						0.9	3.0	92.7	93.0	94.1
						1.5	3.0	93.4	93.6	94.5
				0.1	0.032	0.9	1.5	91.3	92.4	94.0
						0.9	3.0	90.6	91.4	93.9
						1.5	3.0	91.9	92.5	94.3

Table V. *Continued.*

n_1	n_2	q_{11}	q_{12}	p_{011}	p_{012}	δ_1	δ_2	BS	MBS	BBS
				0.2	0.08	0.9	1.5	91.3	92.4	94.0
						0.9	3.0	94.0	94.2	95.4
						1.5	3.0	93.9	94.0	94.8
				0.3	0.13	0.9	1.5	97.3	97.2	95.5
						0.9	3.0	92.7	93.6	94.4
						1.5	3.0	90.7	91.8	94.0
10	30	0.13	0.033	0.05	0.01	0.9	1.5	83.4	85.2	94.8
						0.9	3.0	83.1	85.0	94.0
						1.5	3.0	85.1	87.2	94.3
				0.1	0.032	0.9	1.5	81.1	83.2	95.1
						0.9	3.0	80.7	81.9	94.5
						1.5	3.0	85.1	87.0	94.7
		0.32	0.16	0.2	0.08	0.9	1.5	83.7	85.6	94.2
						0.9	3.0	86.4	87.3	94.0
						1.5	3.0	84.2	85.9	94.7
				0.3	0.13	0.9	1.5	85.3	86.9	94.6
						0.9	3.0	81.5	83.7	94.2
						1.5	3.0	91.3	93.4	95.1
30	50	0.13	0.033	0.05	0.01	0.9	1.5	85.7	87.6	94.0
						0.9	3.0	85.0	86.7	94.5
						1.5	3.0	84.2	86.1	94.0
				0.1	0.032	0.9	1.5	81.9	84.6	94.6
						0.9	3.0	82.3	85.2	94.1
						1.5	3.0	82.5	84.9	94.9
		0.32	0.16	0.2	0.08	0.9	1.5	84.3	87.9	94.8
						0.9	3.0	81.2	83.6	94.2
						1.5	3.0	86.4	87.8	94.7
				0.3	0.13	0.9	1.5	86.1	88.5	94.7
						0.9	3.0	83.5	85.7	94.0
						1.5	3.0	83.6	86.2	94.3

(b) The MBS confidence intervals are better than their original BS confidence intervals in the sense that the coverage probabilities of the former are closer to pre-specified confidence level than those of the latter.
(c) All confidence intervals perform satisfactorily for large sample sizes (e.g. $n_1 = n_2 \geqslant 50$).
(d) Confidence intervals based on bootstrap-resampling method (i.e. SB, PB, and BBS) outperform the other confidence intervals in the sense that the coverage probabilities of the formers are closer to pre-given confidence level while their confidence widths are reasonably shorter.

4. Real examples

In this section, two real examples are used to illustrate the proposed methods. The first example is taken from a clinical study of several radio allegro sorbent test (RAST) methods (Garcia *et al.* [16]) and has even been analyzed by Tang *et al.* [17]. Briefly, 30 positive control sera (serum samples from penicillin allergic subjects with a positive clinical history and a positive penicillin skin test) and 30 negative control sera (sera from subjects with no history of penicillin allergy and negative skin test) were tested for benzylpenicilloyl determinant (BPO)-specific immunoglobulin E (IgE) antibodies by RAST using different conjugates couple to the solid phase. Here, the standard procedure is benzylpenicillin conjugated to human serum albumin (HSA) and the new procedure is benzylpenicillin conjugated to an aminospacer (SP). The above data structure is summarized in Table VII. Consider the positive and negative control sera as two different strata. Sensitivity and specificity are those correct diagnostic results with the corresponding positive and negative control sera, respectively.

SP is preferable to HSA (e.g. 86.7 per cent versus 63.3 per cent) in positive control serum, and SP is also better than HSA in negative control serum. The relative risks in positive and negative control sera are $\delta_1 = 1.368$ and $\delta_2 = 1.16$, respectively. To investigate whether the relative risk in positive control serum equals to that in negative control serum, we compute the 95 per cent BS confidence interval via Equation (10), MBS confidence interval via Equation (11) by using secant method (see, e.g. Tango [7]) and Bonferroni simultaneous (BBS) confidence interval based on bootstrap-resampling method via Equation (12) with $G = 1000$. For the positive control serum, confidence interval estimators of δ_1 for BS, MBS, and BBS methods are given

Table VI. Expected confidence width for 95 per cent simultaneous confidence intervals of δ_j for two strata.

n_1	n_2	q_{11}	q_{12}	p_{011}	p_{012}	δ_1	δ_2	BS	MBS	BBS
10	10	0.13	0.033	0.05	0.01	0.9	1.5	0.240	0.237	0.224
						0.9	3.0	0.442	0.457	0.416
						1.5	3.0	2.175	2.158	2.149
				0.1	0.032	0.9	1.5	0.305	0.319	0.311
						0.9	3.0	0.362	0.375	0.351
						1.5	3.0	2.355	2.342	2.327
		0.32	0.16	0.2	0.08	0.9	1.5	0.855	0.860	0.843
						0.9	3.0	1.052	1.111	1.032
						1.5	3.0	3.240	3.248	3.116
				0.3	0.13	0.9	1.5	1.146	1.152	1.132
						0.9	3.0	1.391	1.405	1.364
						1.5	3.0	2.148	2.167	2.130
30	30	0.13	0.033	0.05	0.01	0.9	1.5	0.119	0.134	0.107
						0.9	3.0	0.373	0.384	0.354
						1.5	3.0	1.552	1.572	1.442
				0.1	0.032	0.9	1.5	0.135	0.152	0.124
						0.9	3.0	0.236	0.256	0.227
						1.5	3.0	1.075	1.123	1.054
		0.32	0.16	0.2	0.08	0.9	1.5	0.506	0.521	0.498
						0.9	3.0	0.680	0.697	0.662
						1.5	3.0	2.263	2.257	2.243
				0.3	0.13	0.9	1.5	0.677	0.684	0.656
						0.9	3.0	0.915	0.926	0.907
						1.5	3.0	2.576	2.555	2.541
50	50	0.13	0.033	0.05	0.01	0.9	1.5	0.112	0.125	0.104
						0.9	3.0	0.354	0.368	0.332
						1.5	3.0	2.107	2.126	2.099
				0.1	0.032	0.9	1.5	0.216	0.234	0.207
						0.9	3.0	0.409	0.416	0.400
						1.5	3.0	1.465	1.478	1.435
				0.2	0.08	0.9	1.5	0.438	0.447	0.413
						0.9	3.0	0.711	0.735	0.686
						1.5	3.0	2.311	2.343	2.302
				0.3	0.13	0.9	1.5	0.534	0.555	0.510
						0.9	3.0	0.986	1.009	0.975
						1.5	3.0	2.439	2.454	2.422
10	30	0.13	0.033	0.05	0.01	0.9	1.5	0.259	0.271	0.232
						0.9	3.0	0.501	0.523	0.488
						1.5	3.0	2.264	0.280	0.241
				0.1	0.032	0.9	1.5	0.170	0.194	1.154
						0.9	3.0	0.207	0.245	0.191
						1.5	3.0	1.688	1.692	1.667
		0.32	0.16	0.2	0.08	0.9	1.5	0.484	0.496	0.472
						0.9	3.0	0.671	0.695	0.641
						1.5	3.0	2.542	2.558	2.523
				0.3	0.13	0.9	1.5	0.681	0.698	0.662
						0.9	3.0	1.025	1.346	1.019
						1.5	3.0	2.502	2.543	2.477
30	50	0.13	0.033	0.05	0.01	0.9	1.5	0.267	0.288	0.243
						0.9	3.0	0.456	0.467	0.442
						1.5	3.0	1.987	2.005	1.972
				0.1	0.032	0.9	1.5	0.423	0.431	0.418
						0.9	3.0	0.332	0.366	0.319
						1.5	3.0	1.245	1.263	1.231

Table VI. *Continued*.

n_1	n_2	q_{11}	q_{12}	p_{011}	p_{012}	δ_1	δ_2	BS	MBS	BBS
		0.32	0.16	0.2	0.08	0.9	1.5	0.592	0.600	0.584
						0.9	3.0	0.670	0.683	0.654
						1.5	3.0	1.198	1.204	1.175
				0.3	0.13	0.9	1.5	0.296	0.312	0.289
						0.9	3.0	0.301	0.325	0.294
						1.5	3.0	1.436	1.452	1.417

Table VII. Observed frequencies for BPO-HSA and BPO-SP for the positive and negative control groups.

		BPO-HSA	
Penicillin allergy		Consistent	Inconsistent
+BPO-SP	Consistent	17	9
	Inconsistent	2	2
−BPO-SP	Consistent	24	5
	Inconsistent	1	0

by [0.910, 1.826], [0.940, 1.80], [0.970, 1.622], respectively. For the negative control serum, the corresponding confidence interval estimators of δ_2 are [0.924, 1.397], [0.910, 1.370], [0.940, 1.410], respectively. These results show that we cannot reject $\delta_1 = \delta_2$ at 5 per cent significance level as all simultaneous confidence intervals for δ_1 and δ_2 contain some common values. Thus, it may be reasonable for us to assume that the relative risks in the two sera are identical. Under this assumption, we calculate the 95 per cent confidence intervals for a common relative risk δ based on the log-profile-likelihood method (see, e.g. Equation (4) by using the bisection searching algorithm), the Cochran statistic method (see, e.g. Equation (5) by using the bisection searching algorithm), the MH-type Wald statistic method (see, e.g. Equation (6) by using the bisection searching algorithm), the WLS statistic method (see, e.g. Equation (7)), and the score statistic method (see, e.g. Equation (8) by using the secant method), which are given by [1.06, 1.54], [1.09, 1.64], [1.11, 1.60], [1.06, 1.58], and [1.06, 1.51], respectively. In addition, we also compute the 95 per cent simple bootstrap confidence interval (see, e.g. Equation (9)) and percentile bootstrap confidence interval (see, e.g. the next paragraph after Equation (9)) of δ based on 1000 bootstrap data sets, which are given by [1.01, 1.553] and [1.06, 1.46], respectively. Because the lower limits of all these confidence intervals are larger than 0.9, applying these confidence interval estimators leads to the same conclusion that SP is non-inferiority to HSA based on the pre-specified acceptance value $\delta_0 = 0.9$ at the one-sided 0.025 significance level, which is consistent with that given in Tang *et al.* [17].

The second example is taken from paratuberculosis data analysis introduced in Section 1. In this example, $x_{111} = 105$, $x_{101} = 9$, $x_{011} = 5$, $x_{001} = 49$, $n_1 = 168$ $x_{112} = 130$, $x_{102} = 5$, $x_{012} = 8$, $x_{002} = 18$, and $n_2 = 161$. Based on these data, we obtain the following results that the relative risk of DIA and ELISA for sensitivities is $\hat{\delta}_1 = 1.05$, which implies that DIA is slightly preferable to ELISA in terms of sensitivity, while the relative risk of DIA and ELISA for specificities is $\hat{\delta}_2 = 0.97$ which indicates that ELISA is a little better than DIA in terms of specificity. Now we want to know whether DIA is non-inferiority to ELISA for screening the suspected paratuberculosis for the infected cattle in positive and negative culture groups. In this case, we regard the positive and negative culture groups as two different strata. We calculate the 95 per cent BS confidence interval, MBS confidence interval, and Bonferroni simultaneous (BBS) confidence interval based on bootstrap-resampling method with $G = 1000$. For the positive culture group, the confidence intervals of δ_1 for BS, MBS, and BBS methods are given by [0.96, 1.11], [0.93, 1.09], [0.92, 1.06], respectively. For the negative culture group, the corresponding confidence intervals of δ_2 are given by [0.92, 1.06], [0.95, 1.06], [0.94, 1.06], respectively. As all simultaneous confidence intervals for δ_1 and δ_2 include some common values, we cannot reject $\delta_1 = \delta_2$ at the 5 per cent significance level. Thus, it is reasonable for us to consider the case that $\delta_1 = \delta_2$. For this case, we calculate the 95 per cent confidence intervals of a common relative risk δ based on the log-profile-likelihood statistic method, the Cochran statistic method, the MH-type Wald statistic method, the WLS statistic method, and the score statistic method by using those methods described in the first example, which are given by [0.96, 1.06], [0.97, 1.06], [0.98, 1.06], [0.96, 1.04], and [0.96, 1.06], respectively. In addition, we also compute the 95 per cent simple bootstrap confidence interval and percentile bootstrap confidence interval of δ based on 5000 bootstrap data sets, which are given by [0.96, 1.04] and [0.95, 1.02], respectively. We notice that all these confidence intervals are entirely to the right of $\delta = 0.9$, which indicates that DIA is non-inferiority to ELISA based on the pre-specified acceptance marginal value $\delta_0 = 0.9$ at the 0.025 significance level (see Bickel and Doksum [18], p. 247). This result is consistent with that of Nam [1].

5. Conclusion

We consider the construction of confidence intervals and simultaneous confidence intervals for relative risks in stratified matched-pair studies, and propose 10 confidence interval estimators of relative risks based on asymptotic theories and Bootstrap resampling method. Various simulation studies from balanced to imbalanced designs are conducted to investigate the performance of the preceding proposed confidence intervals with respect to their coverage probabilities and confidence widths. Among all the confidence interval estimators under investigation, simultaneous confidence interval based on bootstrap resampling method and percentile confidence interval based on the bootstrap sampling method behave satisfactorily in the sense that their coverage probabilities are very close to the pre-specified confidence level while their expected confidence widths are reasonably short for small to large sample sizes.

Although we only considered confidence intervals of relative risk in stratified matched-pair studies with one stratification factor, the theoretics and methods developed for relative risk can be extended to measures such as proportion difference and the odds ratio, and the situation in which there are two or more stratification factors. These generalizations will be of great research interest. Also, in some multi-center clinical trials with center being strata or observational studies, the stratum may be a random variable. In this case, it is necessary to develop some new theories and methods to obtain the efficient confidence interval estimators. To save space, we only discuss the fixed effect stratification case in this paper. We shall discuss the case that the stratum is a random sample in another paper.

We have also done comparisons for designs with larger number of strata (e.g. $J \geqslant 3$), and the results for larger number of strata are similar to those for two strata. Therefore, we omit their results in this article and the results can be available upon request. Also, we have prepared Matlab programs which implement the proposed methods and they are available upon request.

APPENDIX A

A.1. Derivation of the CMLEs $\tilde{p}_{01j}$

Differentiating l given in (1) with respect to p_{01j} and p_{0j} yields

$$\frac{\partial l}{\partial p_{0j}} = \frac{x_{11j}}{p_{0j}-p_{01j}} + \frac{(\delta-1)x_{10j}}{(\delta-1)p_{0j}+p_{01j}} - \frac{x_{00j}\delta}{1-\delta p_{0j}-p_{01j}}$$

$$\frac{\partial l}{\partial p_{01j}} = -\frac{x_{11j}}{p_{0j}-p_{01j}} + \frac{x_{10j}}{(\delta-1)p_{0j}+p_{01j}} - \frac{x_{00j}}{1-\delta p_{0j}-p_{01j}} + \frac{x_{01j}}{p_{01j}}$$

Thus, the CMLEs $\tilde{p}_{0j}$ and $\tilde{p}_{01j}$ of p_{0j} and p_{01j} ($j=1,\ldots,J$) are the solutions of the following equations: $\partial l/\partial p_{0j}=0$ and $\partial l/\partial p_{01j}=0$. Therefore, $\tilde{p}_{01j}$ and $\tilde{p}_{0j}$ of p_{01j} and p_{0j} for a given value of parameter δ can be obtained by solving following equations:

$$\frac{x_{01j}}{p_{01j}} + \frac{x_{10j}}{(\delta-1)p_{0j}+p_{01j}} = \frac{x_{11j}}{p_{0j}-p_{01j}} + \frac{x_{00j}}{1-\delta p_{0j}-p_{01j}}$$

$$\frac{x_{01j}}{p_{01j}} + \frac{x_{10j}\delta}{(\delta-1)p_{0j}+p_{01j}} = \frac{x_{00j}(1+\delta)}{1-\delta p_{0j}-p_{01j}}$$

Following the derivations of Tang *et al.* [6], $\tilde{p}_{01j}$ is the larger root of the following quadratic equation: $f(x)=a_jx^2+b_jx+c_j=0$, which is given by $\tilde{p}_{01j}=\{-b_j+(b_j{}^2-4a_jc_j)^{1/2}\}/(2a_j)$, where

$$a_j=n_j(1+\delta), \quad b_j=(x_{01j}+x_{11j})\delta^2-(x_{11j}+x_{10j}+2x_{01j})$$

and

$$c_j=x_{01j}(1-\delta)(x_{11j}+x_{10j}+x_{01j})/n_j$$

for $j=1,2,\ldots,J$. Also, we have

$$\tilde{p}_{0j}=[(x_{11j}+x_{10j}+x_{01j})/n_j-\tilde{p}_{01j}]/\delta$$

A.2. Derivation of variance of $\partial l/\partial\delta$

It is easily shown from the log-likelihood given in Equation (1) that

$$I_{11} \triangleq E(-\partial^2 l/\partial\delta^2)=n_j\left[\frac{1}{(\delta_0-1)p_{0j}+p_{01j}}+\frac{1}{1-\delta_0 p_{0j}-p_{01j}}\right]$$

$$I_{12} \triangleq E(-\partial^2 l/\partial\delta\partial p_{0j})=n_j\left[\frac{\delta_0-1}{(\delta_0-1)p_{0j}+p_{01j}}+\frac{\delta_0}{1-\delta_0 p_{0j}-p_{01j}}\right]$$

$$I_{13} \triangleq E(-\partial^2 l/\partial\delta\partial p_{01j}) = I_{11}$$

$$I_{22} \triangleq E(-\partial^2 l/\partial p_{0j}^2) = n_j\left[\frac{1}{p_{0j}-p_{01j}} + \frac{(\delta_0-1)^2}{(\delta_0-1)p_{0j}+p_{01j}} + \frac{\delta_0^2}{1-\delta_0 p_{0j}-p_{01j}}\right]$$

$$I_{23} \triangleq E(-\partial^2 l/\partial p_{0j}\partial p_{01j}) = -\frac{n_j}{p_{0j}-p_{01j}} + I_{12}$$

$$I_{33} \triangleq E(-\partial^2 l/\partial p_{01j}^2) = n_j\left(\frac{1}{p_{0j}-p_{01j}} + \frac{1}{p_{01j}}\right) + I_{11}$$

Then, it follows from the general results of efficient scores (Rao [19]) that the variance of $\partial l_j/\partial\beta$ under $\delta=\delta_0$ can be shown to be

$$\mathrm{Var}\left(\left.\frac{\partial l_j}{\partial\delta}\right|_{p_{01j}=\tilde{p}_{01j},p_{0j}=\tilde{p}_{0j},\delta=\delta_0}\right) = \frac{n_j}{\delta_0[2\tilde{p}_{01j}+\tilde{p}_{0j}(\delta_0-1)]}$$

As the stratum is independent, variance of $\partial l/\partial\delta$ can be given by

$$\mathrm{Var}\left(\left.\frac{\partial l}{\partial\beta}\right|_{p_{01}=\tilde{p}_{01},p_0=\tilde{p}_0,\delta=\delta_0}\right) = \sum_{j=1}^{J}\frac{n_j}{\delta_0[2\tilde{p}_{01j}+\tilde{p}_{0j}(\delta_0-1)]}$$

Acknowledgements

Dr N. S. Tang's work was supported by grants from the NSFC (10561008, 10761011), the Ph.D. Special Scientific Research Foundation of Chinese University (20060673002), and by program for New Century Excellent Talents in Universities (NCET-07-0737) and 2008PY036. Dr M. L. Tang's research was fully supported by two grants from the Research Grant Council of the Hong Kong Special Administrative Region (Project Nos. HKBU261007 and HKBU261508).

References

1. Nam J. Non-inferiority of new procedure to standard procedure in stratified matched-pair design. *Biometrical Journal* 2006; **48**:966–977.
2. May WL, Johnson WD. Confidence intervals for differences in correlated binary proportions. *Statistics in Medicine* 1997; **16**:2127–2136.
3. Lui KJ. Comment on confidence intervals for differences in correlated binary proportions. *Statistics in Medicine* 1998; **17**:2017–2020.
4. Quesenberry CP, Hurst DC. Large sample simultaneous confidence intervals for multinomial proportions. *Technometrics* 1964; **6**:191–195.
5. Newcombe RG. Improved confidence intervals for the difference between binomial proportions based on paired data. *Statistics in Medicine* 1998; **17**:2635–2650.
6. Newcombe RG. Reply to comment on improved confidence intervals for the difference between binomial proportions based on paired data. *Statistics in Medicine* 1999; **18**:3513.
7. Tango T. Equivalence test and confidence interval for the difference in proportions for the paired-sample design. *Statistics in Medicine* 1998; **17**:891–908.
8. Tango T. Confidence intervals for differences in correlated binary proportions. *Statistics in Medicine* 2000; **19**:133–139.
9. Newcombe RG. Confidence intervals for the mean of a variable taking the values 0, 1 and 2. *Statistics in Medicine* 2003; **22**:2085–2086.
10. Tang ML, Tang NS, Chan ISF. Confidence interval construction for proportion difference in small-sample paired studies. *Statistics in Medicine* 2005; **24**:3565–3579.
11. Fleiss JL. *Statistical Methods for Rates and Proportions* (2nd edn). Wiley: New York, 1981.
12. Tang NS, Tang ML, Chan ISF. On tests of equivalence via non-unity relative risk for matched-pair design. *Statistics in Medicine* 2003; **22**:1217–1233.
13. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. Chapman & Hall, CRC: Boca Raton, 1993.
14. Roy SN, Bose RC. Simultaneous confidence interval estimation. *Annals of Mathematical Statistics* 1953; **24**:513–536.
15. Westfall PH, Young SS. *Resampling-based Multiple Testing*: *Examples and Methods for p-value Adjustment*. Wiley: New York, 1993.
16. Garcia JJ, Blanca M, Moreno F, Vega JM, Mayorga C, Fernandez J, Juarez C, Romano A, De Ramon E. Determination of IgE antibodies to the benzylpenicilloyl determinant: a comparison of the sensitivity and specificity of three radio allergo sorbent test methods. *Journal of Clinical Laboratory and Analysis* 1997; **11**:251–257.
17. Tang NS, Tang ML, Wang SF. Sample size determination for matched-pair equivalence trials using rate ratio. *Biostatistics* 2007; **8**:625–631.
18. Bickel PJ, Doksum KA. *Mathematical Statistics*: *Basic Ideas and Selected Topics* (2nd edn). Prentice-Hall: NJ, 2001.
19. Rao CR. *Linear Statistical Inference and its Application* (2nd edn). Wiley: New York, 1985.