10-2010

Interval Estimation of Some Epidemiological Measures of Association

Tasneem Zaihra
The College at Brockport, tzaihra@brockport.edu

Sudhir Paul
University of Windsor

Citation/Publisher Attribution:
Zaihra, Tasneem and Paul, Sudhir (2010) "Interval Estimation of Some Epidemiological Measures of Association," The International Journal of Biostatistics: Vol. 6: Iss. 1, Article 35. DOI: 10.2202/1557-4679.1177

The International Journal of Biostatistics

Volume 6, *Issue* 1 2010 *Article* 35

Interval Estimation of Some Epidemiological Measures of Association

Tasneem Zaihra, *University of New Brunswick*
Sudhir Paul, *University of Windsor*

Recommended Citation:
Zaihra, Tasneem and Paul, Sudhir (2010) "Interval Estimation of Some Epidemiological Measures of Association," *The International Journal of Biostatistics*: Vol. 6: Iss. 1, Article 35.
DOI: 10.2202/1557-4679.1177

Interval Estimation of Some Epidemiological Measures of Association

Tasneem Zaihra and Sudhir Paul

Abstract

In epidemiological cohort studies, the probability of developing a disease for individuals in a treatment/intervention group is compared with that of a control group. The groups involve varying cluster sizes, and the binary responses within each cluster cannot be assumed independently. Three major measures of association used to report the efficacy of treatments or effectiveness of public health intervention programs in case of prospective studies are Risk Difference (*RD*), Risk Ratio (*RR*) and Relative Risk Difference (*RED*). The preference of one measure of association over the other in drawing statistical inference depends on design of study. Lui (2004) discusses a number of methods of constructing confidence intervals for each of these measures. Specifically, Lui (2004) discusses four methods for *RD*, four methods for *RR* and three methods for *RED*. For the construction of confidence intervals for *RD*, Paul and Zaihra (2008) compare the four methods discussed by Lui (2004), using extensive simulations with a method based on an estimator of the variance of a ratio estimator by Cochran (1977) and a method based on a sandwich estimator of the variance of the regression estimator using the generalized estimating equations approach of Zeger and Liang (1986). Paul and Zaihra (2008) conclude that the method based on an estimate of the variance of a ratio estimator performs best overall. In this paper, we extend the two new methodologies introduced in Paul and Zaihra (2008) to confidence interval construction of the risk measures *RR* and *RED*. Extensive simulations show that the method based on an estimate of the variance of a ratio estimator performs best overall for constructing confidence interval for the other two risk measures *RR* and *RED* as well. This method involves a very simple variance expression which can be implemented with a very few computer codes. Therefore, it can be considered as an easily implementable alternative for all the three measures of association.

KEYWORDS: clustered data, epidemiological studies, association

Author Notes: This research was partially supported by the Natural Sciences and Engineering Research Council of Canada.

1 Introduction

We often encounter clustered data in many fields, such as epidemiology, preventive medicine, public health and toxicology. Clustered data refers to a set of measurements collected from subjects that are structured in clusters, where a group of related subjects constitutes a cluster, such as a group of students from the same class or rodents in the same litter. For instance, in toxicological studies, a treatment (a stimulus or a control substance) is given to a number of pregnant female animals where each animal is called a litter or cluster. The principal aim of such a study is to determine if the treatment affects the incidence of abnormalities in live foetuses. Another example of clustered data is group randomized trials. Group randomized trials (GRTs) are frequently used in evaluation of public health education and intervention programs. In GRTs or cluster-randomized designs, clusters such as classrooms, schools, clinics, neighborhoods, families or communities are assigned to intervention or control conditions. The outcomes, such as behavior change with respect to employing solar protection (see data in Table 2), are measured on individuals within clusters. In these experiments individuals within the same cluster respond similarly and hence are correlated. Analysis of such data that does not appropriately account for correlation among individuals within clusters leads to erroneous statistical inferences.

The four major measures of association used to report the efficacy of treatments or effectiveness of public health intervention programs are Risk Difference (*RD*), Risk Ratio (*RR*), Relative Risk Difference (*RED*) and Odds Ratio (*OR*). The preference of one measure of association over another in drawing statistical inference depends on the study design. In the case of prospective studies, for instance clinical trials, cohort studies or group randomized trials, etc., risk ratio or relative risk difference is preferable while for case-control retrospective studies, the odds ratio is usually used. As we are dealing with prospective studies we will omit *OR* from further discussion.

The three important measures, risk difference, risk ratio and relative risk difference, having different applications, have been used in the literature to quantify the effect of a suspected risk factor on the probability of developing a given disease (see Lui, 2004). Risk difference is used in public health issues in which the purpose is to measure the magnitude of excess mortality attributed to each disease (see Lui, 2004, chapter 2). Risk ratio is used in toxicological, etiological and cohort studies to quantify the strength of association between a given disease and a suspected risk factor. For example, consider the data in Table 1 from a toxicologic experiment analyzed by Fleiss, Levin and Paik (2003). The data represent the number of pups surviving 21 days of lactation among the number of pups alive four days after the birth from 16 pregnant rats whose diet was chemically treated in one group and was

not treated (control) in the other group. Now, let π_1 and π_0 denote the proportion of pups who would survive 21 lactation days for the treated and the control group respectively. Then the effect of the chemically treated diet can be measured by either the risk difference $RD = \pi_1 - \pi_0$ or the risk ratio (relative risk) $RR = \pi_1/\pi_0$.

Gart and Nam (1988) and Koopman (1984) discussed risk ratio of two binomial proportions for non-clustered data. For other applications and discussions of risk ratio, see Gart (1979) and Morris and Gardner (1988).

Relative difference is used as a measure of impact when, for example, the exposure is preventive. It quantifies the ability of a treatment to reduce the risk of developing an undesirable outcome. It is most popular in public health intervention programs. For further details, see Sheps (1958, 1959), Lui (2004) and Fleiss (1986). Consider the data given in Table 2 (modified from the data given in Table 1.1 of Lui, 2004, p. 7) from an educational intervention program on behavior change with regard to employing solar protection. Now, let π_1 and π_0 denote the proportion of children who employ adequate solar protection in the intervention and the control groups respectively. Then the impact of solar protection can be measured by the relative difference defined by $RED = \frac{\pi_1 - \pi_0}{1-\pi_0}$. As the educational intervention program is expected to motivate (to use solar protection), the value of RED is expected to be positive. See Lui (2004, p. 54-55) for more discussion.

Table 1: Toxicological Data (Weil, 1970). (i) Number of pups surviving 21 lactation days, (ii) Total number of pups alive four days after birth in 16 litters of pregnant rats.

Groups																		
Chemical diet	(i)	12	11	10	9	11	10	10	9	9	5	9	7	10	6	10	7	
	(ii)	12	11	10	9	10	9	9	8	8	4	7	4	5	3	3	0	
Control	(i)	13	12	9	9	8	8	13	12	10	10	9	13	5	7	10	10	
diet	(ii)	13	12	9	9	8	8	12	11	9	9	8	11	4	5	7	7	

Table 2: Radiation Exposure Data (Mayer, 1997). (i) Class Sizes, (ii) Observed number of children with adequate level of solar protection.

Groups																											
Intervention	(i)	3	2	2	5	4	3	1	2	2	2	1	3	1	3	2	2	6	2	4	2	2	2	2	1	1	1	1	1	1
	(ii)	2	1	1	5	3	1	0	0	0	1	0	1	0	1	0	2	6	2	4	1	0	1	1	0	0	1	1	1	1
Control	(i)	2	4	3	2	3	4	4	2	2	3	2	2	4	3	2	3	1	1	2	2	2	3	3	4	1	1	1	1	1
	(ii)	2	4	1	0	3	0	2	1	1	0	0	1	3	0	0	0	0	1	1	0	1	2	1	0	0	0	0	1	1

It is of general interest in epidemiological studies to obtain confidence intervals for one or more of these quantities. The simplest analysis would be based on the assumption that observations within clusters are independent. Such an analysis would bias the inference procedures as observations within clusters are likely correlated. Lui (2001) reviews four methods for constructing confidence intervals

DOI: 10.2202/1557-4679.1177

for risk difference. Paul and Zaihra (2008) propose a new method based on an estimate of the variance of a ratio estimator. Suppose that we independently sample n_i clusters from the ith group, $i = 0,1$, with m_{ij} individuals, $j = 1,\ldots,n_i$. Suppose that x_{ij} is the number of individuals in the jth cluster of the ith group who are exposed to a risk factor. The unbiased estimate of π_i is $\hat{\pi}_i = x_{i.}/m_{i.}$, where $x_{i.} = \sum_{j=1}^{n_i} x_{ij}$ and $m_{i.} = \sum_{j=1}^{n_i} m_{ij}$. This can be written as the ratio of two sample means, $\hat{\pi}_i = \bar{x_i}/\bar{m_i}$, where $\bar{x_i} = x_{i.}/n_i$ and $\bar{m_i} = m_{i.}/n_i$. Then, using a result by Cochran (1977, p. 31) of the estimate of the variance of a ratio estimator, an estimator of the variance of $\hat{\pi}_i$ is

$$v_i = (n_i/(n_i-1)) \sum_{j=1}^{n_i} (x_{ij} - \hat{\pi}_i m_{ij})^2/m_{i.}^2 = n_i/(n_i-1) \sum_{j=1}^{n_i} r_{ij}^2/m_{i.}^2, \qquad (1)$$

where $r_{ij} = x_{ij} - \hat{\pi}_i m_{ij}$. Paul and Zaihra (2008) also develop a sandwich estimator v_{s_i} of the variance of $\hat{\pi}_i$ using the generalized estimating equations approach of Zeger and Liang (1986). They further show a simple relationship between v_i and v_{s_i} as $v_{s_i} = (n_i/(n_i-1)v_i$.

Using the estimator v_i, an approximate $100(1-\alpha)\%$ confidence interval for the risk difference $\Delta = \pi_1 - \pi_0$, proposed by Paul and Zaihra (2008), is $(\hat{\pi_1} - \hat{\pi_0}) \pm Z_{\alpha/2}\sqrt{(v_0 + v_1)}$, where $\hat{\pi_i^*} = (x_{i.} + 0.5)/(m_{i.} + 1)$. The corresponding confidence interval using v_{s_i} as the variance of $\hat{\pi}_i$ is $(\hat{\pi_1} - \hat{\pi_0}) \pm Z_{\alpha/2}\sqrt{(v_{s_0} + v_{s_1})}$. For more details, see Paul and Zaihra (2008). Note that these methods do not assume any specific model for over-dispersion or intra-class correlation. Using an extensive simulation study Paul and Zaihra (2008) show that the former of the above two methods performs better than the latter method and the four methods given by Lui (2001).

Lui, Mayer and Eckhardt (2000) develop (see also Lui, 2004) four asymptotic interval estimators for the risk ratio. An estimate of the risk ratio RR, though biased, is $\hat{RR} = \hat{\pi}_1/\hat{\pi}_0$. Using the delta method, an approximate variance of the estimator $\hat{RR}$ is $var(\hat{RR}) = (\pi_1/\pi_0)^2 \sum_{i=0}^{1} (var(\hat{\pi}_i)/{\pi_i}^2)$. The four procedures developed by Lui et al.(2000), which we review in Section 2, are based on an estimate of the beta-binomial variance $var(\hat{\pi}_i) = \hat{\pi}_i(1-\hat{\pi}_i) f(m_i,\phi_i)/(m_{i.})$, where $f(m_i,\phi_i) = \sum m_{ij}[1 + (m_{ij}-1)\phi_i]/m_{i.}$ and ϕ_i is the cluster specific over-dispersion parameter. Each of the four methods of interval estimators for the risk ratio developed by Lui et al.(2000) is compared, by simulations, using the three estimates of $var(\hat{\pi}_i)$, namely, that used by Lui et al.(2000), v_i and v_{s_i}. Thus, we compare twelve methods and make a recommendation for practical use.

Lui (2004) discusses three methods of constructing confidence interval for the relative risk difference RED. These methods are also based on an estimate of the beta-binomial variance $var(\hat{\pi}_i) = \hat{\pi}_i(1-\hat{\pi}_i) f(m_i,\phi_i)/(m_{i.})$. Each of these three

methods is compared, by simulations, using the three estimates of $var(\hat{\pi}_i)$ discussed above. Thus, here we compare nine methods.

In Section 2 the methods for constructing confidence intervals for RR are given along with results of an extensive simulation study and an example. Section 3 provides the methods for constructing confidence intervals for RED along with some simulation results and an example. A discussion follows in Sections 4.

2 Confidence Interval for the Risk Ratio

2.1 The methods

Lui et al.(2000) evaluated four methods of constructing confidence intervals for the risk ratio. Here we review each of these methods and introduce two other versions, based on v_i and v_{s_i} discussed in Section 1, for each method. Now, an estimate of $var(\hat{RR})$ is $\widehat{var}(\hat{RR}) = (\hat{\pi}_1/\hat{\pi}_0)^2 \sum_{i=0}^{1} (\hat{var}(\hat{\pi}_i)/\hat{\pi}_i^2)$, where $\hat{var}(\hat{\pi}_i)$ is an estimate of $var(\hat{\pi}_i) = \hat{\pi}_i(1-\hat{\pi}_i) f(m_i, \phi_i)/(m_{i.})$.

Thus an approximate $100(1-\alpha)\%$ confidence interval for RR, based on the asymptotic normality of $\hat{RR}$, is given by

$$\left[max\{\hat{RR} - Z_{\alpha/2}\sqrt{\widehat{var}(\hat{RR})}, 0\}, \hat{RR} + Z_{\alpha/2}\sqrt{\widehat{var}(\hat{RR})} \right]. \tag{2}$$

Method M1 (Lui et al. (2000)):

Lui et al. (2000) use an estimate v_{l_i} of the variance of $\hat{\pi}_i$ based on the beta-binomial model using an ANOVA estimate of the beta-binomial over-dispersion parameter as follows:

$v_{l_i} = \hat{var}(\hat{\pi}_i) = \hat{\pi}_i(1-\hat{\pi}_i) f(m_i, \hat{\phi}_i)/(m_{i.})$, where $f(m_i, \hat{\phi}_i) = \sum m_{ij}[1 + (m_{ij} - 1)\hat{\phi}_i]/m_{i.}$, $\hat{\pi}_i = x_{i.}/m_{i.}$ and $\hat{\phi}_i = (BMS_i - WMS_i)/[BMS_i + (m_i^* - 1)WMS_i]$, where $BMS_i = [\sum_j X_{ij}^2/m_{ij} - (\sum_j X_{ij})^2/\sum_j m_{ij}]/(n_i - 1)$ and $WMS_i = [\sum_j X_{ij} - \sum_j X_{ij}^2/m_{ij}]/(\sum_j (m_{ij} - 1))$ are between mean squared and within mean squared errors respectively. The above analysis of variance (ANOVA) type estimate of the intraclass correlation ϕ_i was first proposed by Elston (1977) for correlated continuous data and later used by others, such as, Donner (1981) and Lui et al.(2000).

Method MR1: Use v_i as an estimate of $var(\hat{\pi}_i)$, where

$$v_i = (n_i/(n_i - 1)) \sum_{j=1}^{n_i} (x_{ij} - \hat{\pi}_i m_{ij})^2/m_{i.}^2 = n_i/(n_i - 1) \sum_{j=1}^{n_i} r_{ij}^2/m_{i.}^2, \tag{3}$$

where $r_{ij} = x_{ij} - \hat{\pi}_i m_{ij}$ with $\hat{\pi}_i = (x_{i.} + 0.5)/(m_{i.} + 1)$.

DOI: 10.2202/1557-4679.1177

Method MS1:

Use v_{s_i} as an estimate of $var(\hat{\pi}_i)$, where $v_{s_i} = (n_i/(n_i-1))v_i$.

As the sampling distribution of $\hat{RR}$ can be skewed, the interval estimator (1.2) may not perform well, especially when the number of clusters is small (see Katz, Baptista, Azen and Pike ,1978). To avoid this problem, as an extension of method M1, Lui et al. (2000) propose using logarithmic transformation to improve the normal approximation. Then, using the delta method, an approximate variance of the estimator $log(\hat{RR})$ is $var(log(\hat{RR})) = \sum_{i=0}^{1}(var(\hat{\pi}_i)/{\pi_i}^2)$. Now, an estimate of $var(log(\hat{RR}))$ is $\widehat{var}(log(\hat{RR})) = \sum_{i=0}^{1}(\hat{var}(\hat{\pi}_i)/{\hat{\pi}_i}^2)$. Based on a logarithmic transforation of RR, then, an asymptotic $100(1-\alpha)\%$ confidence interval for the RR is given by

$$\left[(\hat{RR})\exp\left(-Z_{\alpha/2}\sqrt{\widehat{var}(log(\hat{RR}))}\right),(\hat{RR})\exp\left(Z_{\alpha/2}\sqrt{\widehat{var}(log(\hat{RR}))}\right)\right]. \quad (4)$$

Method M2 (Lui et al. (2000)):

Use v_{l_i} for $\hat{var}(\hat{\pi}_i)$ in (1.4).

Method MR2:

Use v_i for $\hat{var}(\hat{\pi}_i)$ in (1.4).

Method MS2:

Use v_{s_i} for $\hat{var}(\hat{\pi}_i)$ in (1.4).

Now, define $Z = \hat{\pi}_1 - RR\hat{\pi}_0$. It can be seen that $E(Z) = 0$ and an estimate of $Var(Z)$ is $\hat{var}(\hat{\pi}_1) + (\hat{RR})^2\hat{var}(\hat{\pi}_0)$. Then, using Fieller's Theorem (see Fieller, 1954, Casella and Berger, 1990) an asymptotic $100(1-\alpha)\%$ confidence interval for RR is obtained by solving the following quadratic equation in RR:

$$A(RR)^2 - 2B(RR) + C \leq 0, \quad (5)$$

where $A = \hat{\pi}_0^2 - Z_{\alpha/2}^2\hat{var}(\hat{\pi}_0)$, $B = \hat{\pi}_1\hat{\pi}_0$ and $C = \hat{\pi}_1^2 - Z_{\alpha/2}^2\hat{var}(\hat{\pi}_1)$. Solving (1.5) an asymptotic $100(1-\alpha)\%$ confidence interval for RR is

$$[RR_1, RR_2], \quad (6)$$

where $RR_1 = max\left\{(B-\sqrt{B^2-AC})/A, 0\right\}$ and $RR_2 = (B+\sqrt{B^2-AC})/A$.

Method M3 (Lui et al. (2000)):

Use v_{l_i} for $\hat{var}(\hat{\pi}_i)$ in (1.6).

Method MR3:

Use v_i for $\hat{var}(\hat{\pi}_i)$ in (1.6).

Method MS3:

Use v_{s_i} for $\hat{var}(\hat{\pi}_i)$ in (1.6).

Note that in order to avoid negative or complex values of RR in the above methods M3, MR3 and MS3, the restrictions $A > 0$ and $B^2 - AC > 0$ need to be imposed.

To reduce possible skewness of the sampling distribution of $(\hat{\pi}_1 - RR\hat{\pi}_0)/\sqrt{\{var(\hat{\pi}_1 - RR\hat{\pi}_0)\}}$, Bailey (1987) proposed considering ${\hat{\pi}_1}^{1/3} - (RR\hat{\pi}_0)^{1/3}$, which after the application of Fieller's Theorem, leads to the approximate $100(1-\alpha)\%$ confidence interval for RR given by

$$\left[max\left\{((B^\dagger - \sqrt{B^{\dagger 2} - A^\dagger C^\dagger})/A^\dagger)^3, 0\right\}, \left\{((B^\dagger + \sqrt{B^{\dagger 2} - A^\dagger C^\dagger})/A^\dagger)^3, \right\}\right] \quad (7)$$

where $A^\dagger = {\hat{\pi}_0}^{2/3} - Z^2_{\alpha/2} v\hat{a}r(\hat{\pi}_0)/9{\hat{\pi}_0}^{4/3}$, $B^\dagger = (\hat{\pi}_1\hat{\pi}_0)^{1/3}$ and $C = {\hat{\pi}_1}^{2/3} - Z^2_{\alpha/2} v\hat{a}r(\hat{\pi}_1){\hat{\pi}_1}^{4/3}$.

Method M4 (Lui et al. (2000)):
Use v_{l_i} for $v\hat{a}r(\hat{\pi}_i)$ in (1.7).
Method MR4:
Use v_i for $v\hat{a}r(\hat{\pi}_i)$ in (1.7).
Method MS4:
Use v_{s_i} for $v\hat{a}r(\hat{\pi}_i)$ in (1.7).

2.2 Simulation study and the results

Now we report on a simulation study conducted to compare the 12 methods discussed in Section 2.1 for the construction of confidence intervals for the risk ratio RR. As mentioned in Section 2.1, Lui et al. (2000) evaluate four methods M1, M2, M3 and M4, and for each method we propose two extensions. All the methods by Lui et al. (2000) are based on an ANOVA-type estimate v_{l_i} (given in Section 2.1) of $var(\hat{\pi}_i)$ and our extensions are based on v_i: the variance of a ratio estimator of $\hat{\pi}_i$ and v_{s_i}: the variance of a sandwich estimator of $\hat{\pi}_i$. Thus method 1 has three versions M1, MR1 and MS1, method 2 has three versions M2, MR2 and MS2, method 3 has three versions M3, MR3 and MS3 and method 4 has three versions M4, MR4 and MS4.

For generating a beta-binomial observation, we first generate a value p from a beta (α, β) distribution. Then, given this value of p, a value y is generated from a $Binomial(m, p)$ distribution, where m is the beta-binomial index. Thus $E(P) = \frac{\alpha}{\alpha+\beta} = \pi$ and $Var(P) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \pi(1-\pi)\phi$ and the unconditional mean and variance of Y are $E(Y) = m\pi$ and $Var(Y) = m\pi(1-\pi)\{1+(m-1)\phi\}$ respectively. The parameter $\phi = \frac{1}{\alpha+\beta+1}$ is the over-dispersion parameter which is also the intra-cluster correlation between two binary observations within the same cluster.

DOI: 10.2202/1557-4679.1177

As in the case of risk difference investigated by Paul and Zaihra (2008), the confidence interval does not exist for some samples if either BMS_i or WMS_i is 0, as the estimate of $\hat{\phi}_i$, used in the methods M1, M2, M3 and M4, is not valid. Also, if $A \leq 0$ or $B^2 - AC \leq 0$ or if $B^{\dagger 2} - A^{\dagger}C^{\dagger} \leq 0$, then confidence intervals by the methods M3 and M4 do not exist.

The number of samples rejected, in general, were small (in 10,000 samples 0 to 1,000 samples were rejected). However, for some small values of n and m and large value of ϕ this number was substantially larger. For example, for $RR = 1$, n=20, m=5, $\pi = .10$ and $\phi = 0.5$ a total of 16,511 samples were rejected before a total of 10,000 good samples could be taken. Table 3 displays the total number of invalid samples for $\pi = .10$, $\phi = 0.1$ and for $\pi = .10$, $\phi = 0.5$ for some configurations of *RR*, *n* and *m*.

Table 3: Number of samples rejected (*NR*) to obtain 10,000 good samples for $\pi = 0.10$, $\phi = 0.1$, $\phi = 0.5$ and some configurations of *RR*, *n*, *m*.

RR	*n*	*m*	*NR* for $\phi = 0.1$	*NR* for $\phi = 0.5$
1	20	5	677	16511
		10	52	7857
		50	0	5357
	30	5	35	2540
		10	2	1618
		50	0	977
	50	5	7	194
		10	2	50
		50	0	15
2	20	5	671	10894
		10	38	7201
		50	5	5370
	30	5	25	2505
		10	9	1609
		50	2	930
	50	5	5	111
		10	3	52
		50	2	20
4	20	5	648	7523
		10	45	5214
		50	15	2490
	30	5	39	1582
		10	10	989
		50	5	171
	50	5	5	525
		10	2	16
		50	0	9

The Coverage probability and average length were based on 10,000 samples in which the confidence interval existed for all methods. The coverage probability for *RR* is equal to the number of times (out of 10,000) that the confidence interval contained the true value of *RR*. For each of the 10,000 samples, the length of the confidence interval was calculated. The average coverage length (average length) is the mean of these 10,000 lengths.

We now compare the coverage probability properties of all the methods. For this we first compare the bias of the three versions of each method. The average bias (coverage probability - 0.95) over all combinations of n and m were plotted against different values of the risk ratio RR. For example, for RR=1, the average bias is the mean of the nine biases for all combinations of n=20, 30, 50 and m=5, 10, 50. The plots for different combinations of π_0 and ϕ are given in Figure 1 to Figure 4 for the four methods.

It appears from the graphs in figures 1, 2, 3 and 4 that, overall, the methods which use the variance of a ratio estimator of π_i, namely, the methods MR1, MR2, MR3 and MR4, have smallest bias. We then compare the average bias of the four methods MR1, MR2, MR3 and MR4 using Figure 5. We see from figure 5 that method MR3, which is based on the variance of ratio estimator v_i of $var(\hat{\pi}_i)$ and Fieller's theorem for the construction of the confidence interval, has the best overall bias property. The next best is that of MR2.

Next, we discuss the properties of all methods in terms of the average coverage length. This property is similar for all three versions of each method. So to save space we give the average length of the confidence intervals for the methods MR1, MR2, MR3 and MR4 in Table 4 for $\pi_0 = .10$, $\phi = .10$ and .50. Results for the other combinations of π_0 and ϕ are similar. In general, MR3 shows larger average length than the other three methods, substantially larger when the sample size is small or the number of clusters is small (see cases n=20, $m_0 = 5, 10$ and n=30, $m_0 = 5$).

DOI: 10.2202/1557-4679.1177

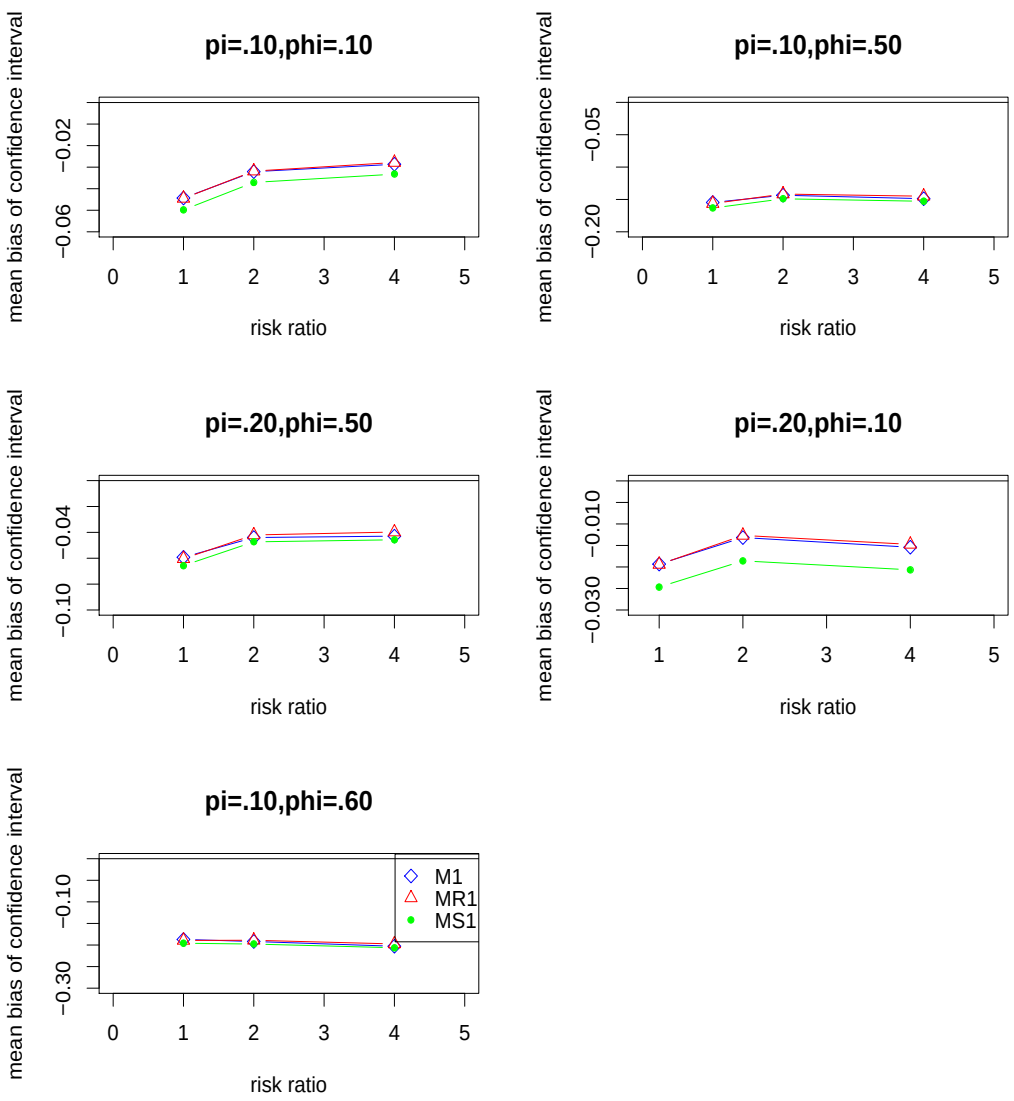

Figure 1: Bias of the confidence intervals of the three versions M1, MR1 and MS1 of method1.

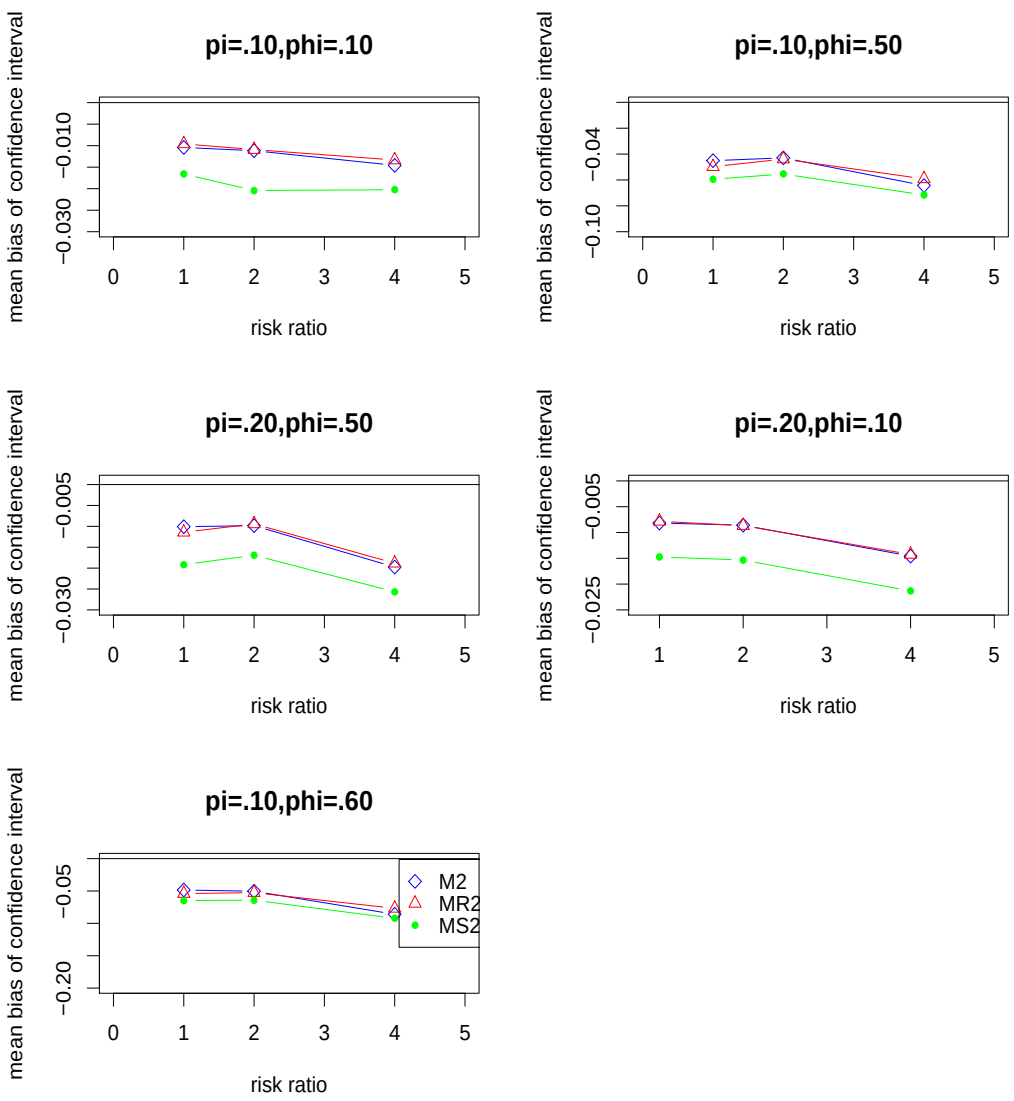

Figure 2: Bias of the confidence intervals of the three versions M2, MR2 and MS2 of method2.

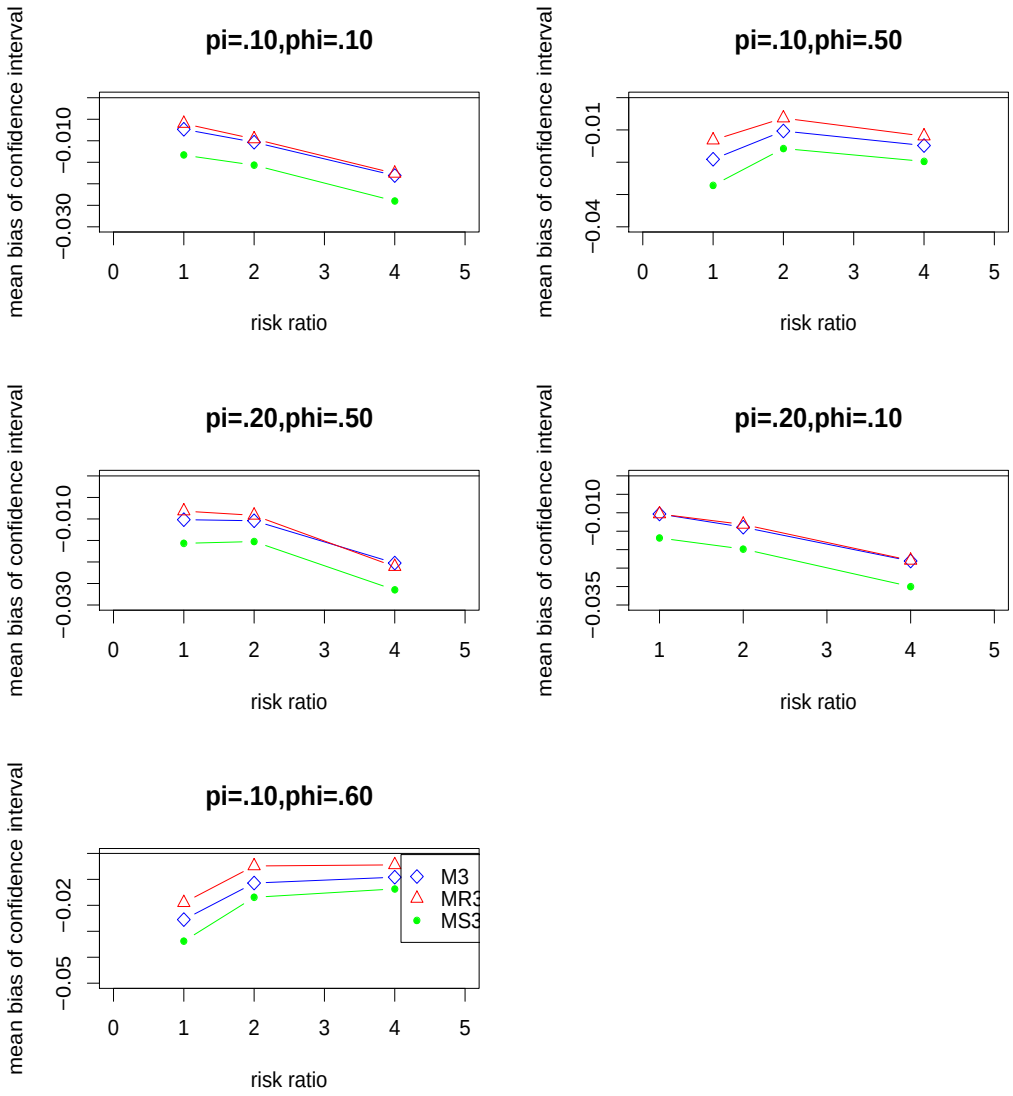

Figure 3: Bias of the confidence intervals of the three versions M3, MR3 and MS3 of method3.

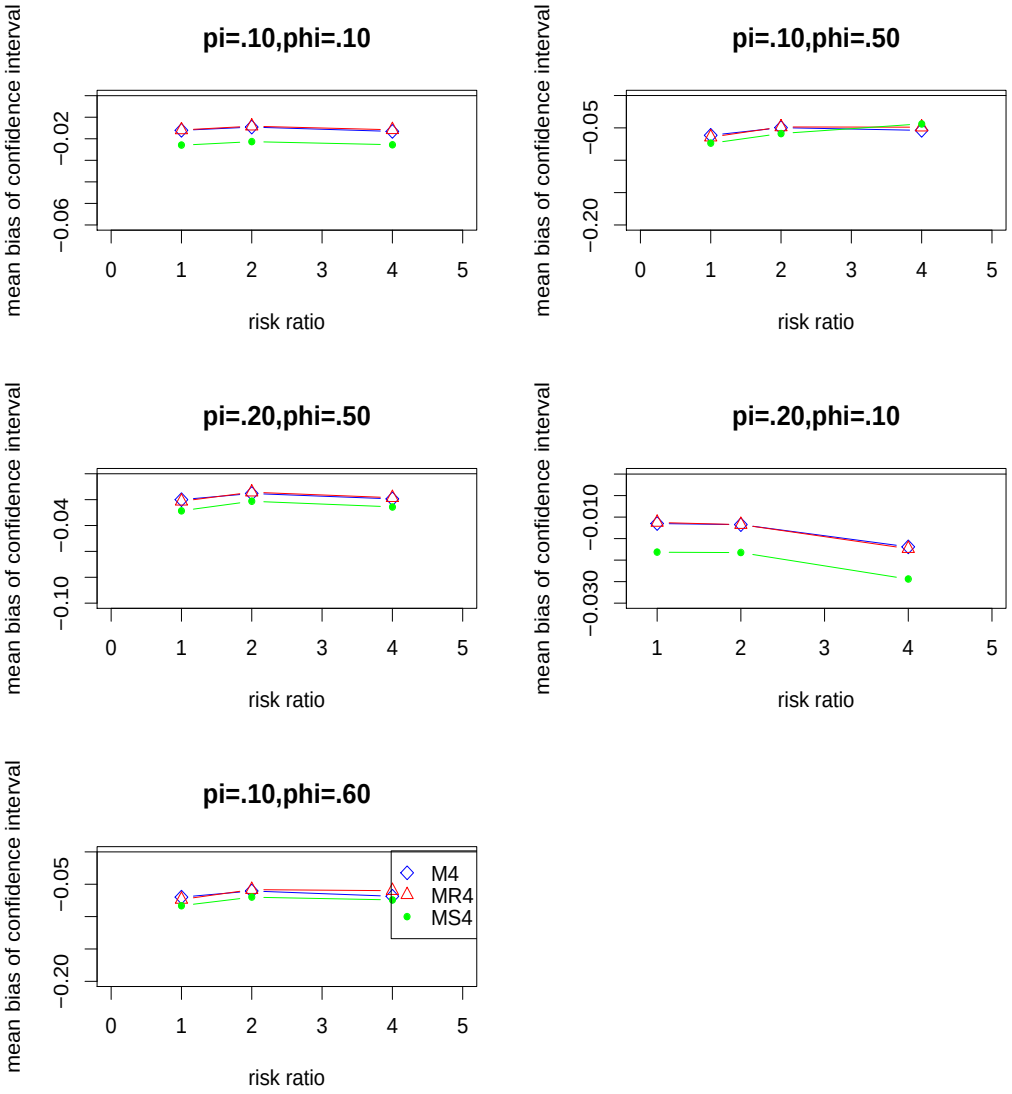

Figure 4: Bias of the confidence intervals of the three versions M4,MR4 and MS4 of method4.

DOI: 10.2202/1557-4679.1177

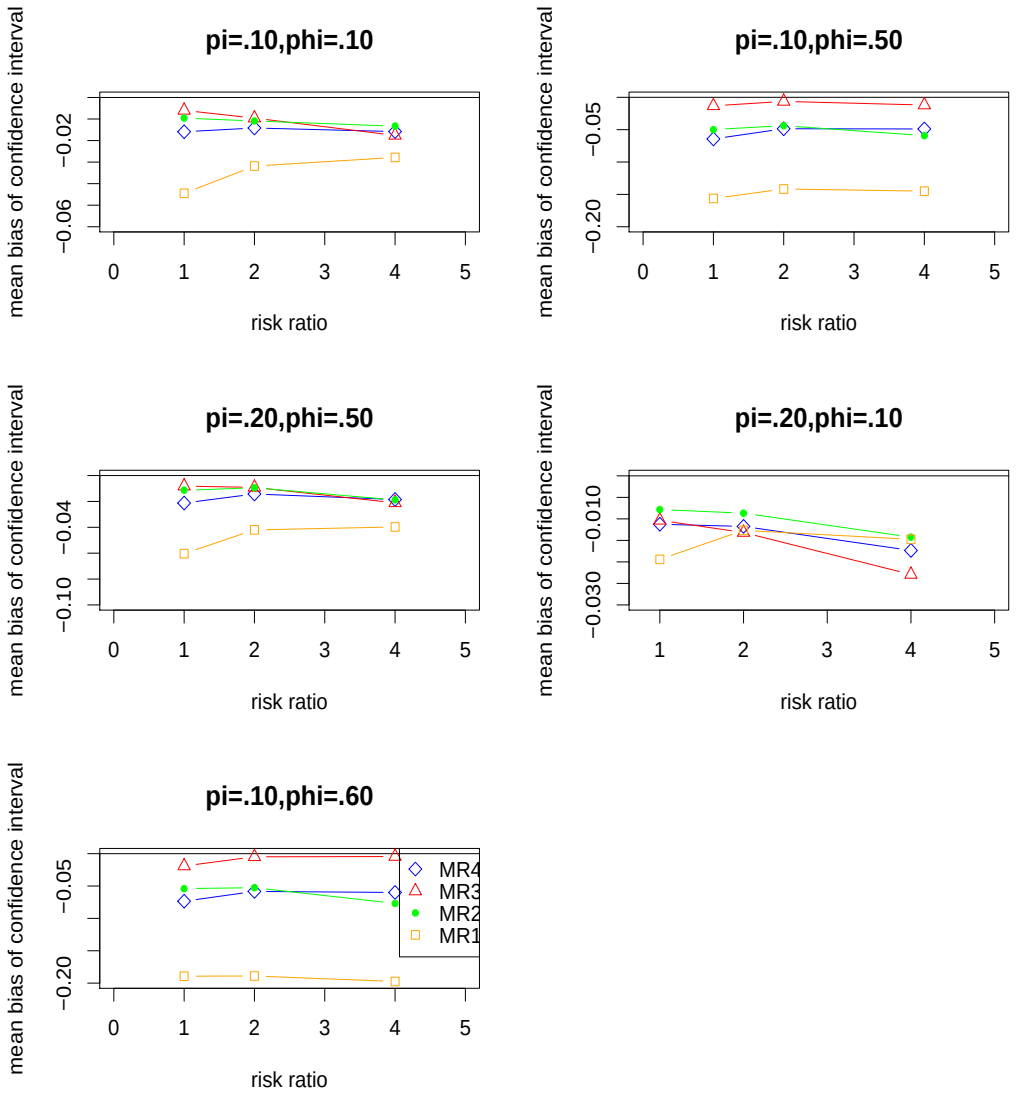

Figure 5: Bias of the confidence intervals by all four methods with variance of ratio estimator

Table 4: The estimated coverage probabilities and average lengths of the confidence intervals (in parenthesis) for the risk ratio by the methods $MR1$, $MR2$, $MR3$, $MR4$; for equal numbers of clusters $n_1 = n_0 = n$ in both groups, mean cluster size $m_0 = 5, 10, 50$; underlying mean probability of response in group 0, $\pi_0 = 0.10$ and $\alpha = 0.05$; based on 10,000 simulations

			ϕ=.1				ϕ=.5			
RR	n	m_0	MR1	MR2	MR3	MR4	MR1	MR2	MR3	MR4
1	20	5	.891	.949	.949	.937	.744	.905	.939	.887
			(2.108)	(2.497)	(6.525)	(2.081)	(2.081)	(2.786)	(21.955)	(2.799)
	20	10	.913	.936	.943	.931	.776	.905	.956	.894
			(1.816)	(2.048)	(3.294)	(2.115)	(2.318)	(3.044)	(16.575)	(3.099)
	20	50	.924	.934	.943	.934	.818	.897	.941	.891
			(1.341)	(1.438)	(1.645)	(1.458)	(2.701)	(3.476)	(19.206)	(3.599)
1	30	5	.915	.943	.942	.935	.840	.935	.957	.927
			(1.813)	(2.043)	(3.459)	(2.113)	(2.274)	(2.859)	(16.347)	(2.966)
	30	10	.924	.942	.946	.939	.855	.927	.952	.921
			(1.442)	(1.562)	(1.071)	(1.591)	(2.360)	(2.919)	(11.706)	(3.047)
	30	50	.931	.941	.946	.939	.882	.930	.951	.924
			(1.086)	(1.137)	(1.227)	(1.146)	(2.427)	(2.953)	(18.693)	(3.099)
1	50	5	.931	.947	.950	.943	.900	.944	.953	.937
			(1.348)	(1.446)	(1.667)	(1.469)	(2.119)	(2.484)	(5.552)	(2.594)
	50	10	.937	.947	.949	.946	.909	.946	.948	.933
			(1.095)	(1.148)	(1.242)	(1.160)	(2.062)	(2.380)	(5.210)	(2.479)
	50	50	.936	.942	.945	.941	.915	.944	.952	.939
			(.830)	(.853)	(.888)	(.857)	(1.909)	(2.166)	(6.749)	(2.228)
2	20	5	.914	.950	.952	.947	.764	.917	.960	.914
			(3.599)	(4.076)	(14.472)	(4.428)	(3.495)	(4.280)	(30.553)	(4.652)
	20	10	.929	.941	.937	.935	.803	.916	.959	.918
			(3.157)	(3.458)	(6.270)	(3.678)	(3.901)	(4.727)	(37.317)	(5.952)
	20	50	.931	.933	.933	.932	.857	.914	.946	.941
			(2.302)	(2.425)	(2.907)	(2.502)	(4.496)	(5.397)	(40.259)	(4.976)
2	30	5	.928	.943	.933	.934	.866	.942	.962	.942
			(3.155)	(3.452)	(6.200)	(3.675)	(3.836)	(4.526)	(30.491)	(5.153)
	30	10	.939	.945	.943	.942	.889	.942	.959	.933
			(2.494)	(2.647)	(3.320)	(2.748)	(4.008)	(4.685)	(29.519)	(5.331)
	30	50	.940	.942	.940	.939	.910	.935	.944	.944
			(1.847)	(1.911)	(2.119)	(1.949)	(4.187)	(4.843)	(19.183)	(4.501)
2	50	5	.935	.946	.942	.943	.921	.951	.951	.938
			(2.326)	(2.451)	(2.958)	(2.533)	(3.680)	(4.150)	(10.708)	(4.269)
	50	10	.943	.943	.941	.941	.927	.943	.944	.938
			(1.877)	(1.943)	(2.170)	(1.988)	(3.557)	(3.964)	(9.099)	(3.913)
	50	50	.942	.942	.945	.942	.931	.944	.939	.926
			(1.412)	(1.441)	(1.520)	(1.458)	(3.333)	(3.667)	(6.087)	(8.062)
4	20	5	.913	.949	.946	.948	.771	.907	.967	.926
			(6.482)	(7.184)	(23.635)	(7.986)	(6.071)	(7.066)	(76.937)	(8.062)
	20	10	.934	.936	.919	.928	.828	.915	.952	.924
			(5.722)	(6.173)	(12.972)	(6.677)	(6.825)	(7.907)	(143.326)	(9.043)
	20	50	.938	.935	.925	.931	.867	.903	.922	.906
			(4.146)	(4.327)	(5.363)	(4.509)	(8.051)	(9.297)	(111.423)	(10.767)
4	30	5	.933	.945	.935	.941	.882	.944	.968	.952
			(5.635)	(6.079)	(11.273)	(6.476)	(6.751)	(7.696)	(79.822)	(8.753)
	30	10	.942	.938	.925	.932	.906	.934	.945	.937
			(4.518)	(4.748)	(6.219)	(4.985)	(7.198)	(8.163)	(94.118)	(9.262)
	30	50	.941	.941	.933	.938	.916	.933	.933	.931
			(3.324)	(3.418)	(3.866)	(3.509)	(8.365)	(9.369)	(55.712)	(9.456)
4	50	5	.942	.944	.935	.942	.928	.948	.941	.945
			(4.214)	(4.401)	(5.472)	(4.593)	(6.669)	(7.364)	(20.550)	(8.159)
	50	10	.941	.942	.935	.939	.930	.942	.931	.937
			(3.344)	(3.440)	(3.896)	(3.533)	(6.342)	(6.935)	(16.404)	(7.604)
	50	50	.944	.942	.939	.941	.933	.942	.927	.935
			(2.544)	(2.587)	(2.768)	(2.626)	(5.991)	(6.487)	(11.208)	(7.038)

DOI: 10.2202/1557-4679.1177

2.3 An example

As an illustrative example we consider the toxicological data discussed in Section 1 and given in Table 1. For these data we obtain $\hat{\pi}_1 = 0.772$, $\hat{\pi}_0 = 0.899$, $\hat{RR} = .859$ and the estimate of the common intraclass correlation $\hat{\phi} = 0.193$. The 95% two sided confidence interval for the risk ratio RR by the four methods $MR1$, $MR2$, $MR3$ and $MR4$ are $(0.700, 1.018)$; $(714, 1.034)$; $(0.703, 1.022)$ and $(0.710, 1.029)$ respectively. The corresponding lengths of the confidence intervals are 0.318, 0.319, 0.319 and 0.318.

To examine the appropriateness of using MR1, MR2, MR3 and MR4 in the particular configuration ($\pi_0 = .899,\ \phi = .193$, RR=.859, n=16, m=9) given by this example, we apply simulation again. When applying MR1, MR2, MR3 and MR4 we obtain estimated coverage probabilities and coverage lengths (in parenthesis) of 95% confidence intervals to be .94(.294), .94(.295), .94(.295) and .94(.295) respectively. These results suggest that all of the above methods of interval estimation should be appropriate for use in this example. Note that all methods MR1, MR2, MR3 and MR4 produce similar confidence intervals and confidence lengths.

3 Confidence Interval for Relative Difference, RED

3.1 The methods

Lui (2004) discuss three methods of constructing confidence intervals for the relative risk difference. As in Section 2, here we review each of these methods and introduce two other versions, based on v_i and v_{s_i}, for each method. To estimate the relative risk difference δ, we substitute $\hat{\pi}_i$ for π_i and obtain the estimator $\hat{\delta} = (\hat{\pi}_1 - \hat{\pi}_0)/(1 - \hat{\pi}_0)$. Using the delta method, an asymptotic variance of the estimator $\hat{\delta}$ is $Var(\hat{\delta}) = \varphi^2 \sum_{i=0}^{1} var(\hat{\pi}_i)/(1-\pi_i)^2$. Now, an estimate of $var(\hat{\delta})$ is $\widehat{var}(\hat{\delta}) = \hat{\varphi}^2 \sum_{i=0}^{1} \hat{var}(\hat{\pi}_i)/(1-\hat{\pi}_i)^2$, where $\hat{\varphi} = (1-\hat{\pi}_1)/(1-\hat{\pi}_0)$. Thus an asymptotic $100(1-\alpha)\%$ confidence interval for the RED is given by

$$\left[max\{\hat{\delta} - Z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\delta})}, 0\}, min\{\hat{\delta} + Z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\delta})}, 1\} \right] \tag{8}$$

Note that to ensure that the confidence interval falls in specified range of δ, we have restrictions on the confidence limits: maximum of $\hat{\delta} - Z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\delta})}$ and 0 for the lower limit and minimum of $\hat{\delta} - Z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\delta})}$ and 1 for the upper limit.

Method RD1 (Lui et al. (2004)):
Use v_{l_i} for $\hat{var}(\hat{\pi}_i)$ in (1.8).
Method RDR1:
Use v_i for $\hat{var}(\hat{\pi}_i)$ in (1.8).
Method RDS1:
Use v_{s_i} for $\hat{var}(\hat{\pi}_i)$ in (1.8).

When both the sample sizes and the probability of positive response π_i are small, the sampling distribution of $\hat{\delta}$ can be skewed and hence the interval estimator (1.8) may not perform well, especially when the number of clusters is small (Katz et al. 1978). To avoid this, Lui (2004) proposes using a logarithmic transformation of $\hat{\varphi} = 1 - \hat{\delta}$ to improve the normal approximation. Then using the delta method and after some algebra, an asymptotic $100(1-\alpha)\%$ confidence interval for δ is given by

$$\left[1 - min\{(\hat{\varphi})\exp\left(Z_{\alpha/2}\sqrt{\widehat{var}(log(\hat{\varphi}))}\right), 1\}, 1 - \hat{\varphi}\exp\left(-Z_{\alpha/2}\sqrt{\widehat{var}(log(\hat{\varphi}))}\right)\right] \quad (9)$$

where $\widehat{var}(\hat{\varphi}) = \sum_{i=0}^{1}(\hat{var}(\hat{\pi}_i)/(1-\hat{\pi}_i)^2)$.
Method RD2 (Lui et al. (2004)):
Use v_{l_i} for $\hat{var}(\hat{\pi}_i)$ in (1.9).
Method RDR2:
Use v_i for $\hat{var}(\hat{\pi}_i)$ in (1.9).
Method RDS2:
Use v_{s_i} for $\hat{var}(\hat{\pi}_i)$ in (1.9).

Now, let $a = \hat{\pi}_1 - \hat{\pi}_0$ and $b = 1 - \hat{\pi}_0$. Further, define $Z = a - REDb$. Now, an estimate of $Var(Z)$ is $\hat{Var}(Z) = v_{11} - 2(\hat{RED})v_{12} + (\hat{RED})^2 v_{22}$, where $v_{11} = \hat{var}(\hat{\pi}_0) + \hat{var}(\hat{\pi}_1)$, $v_{12} = \hat{var}(\hat{\pi}_0)$ and $v_{22} = \hat{var}(\hat{\pi}_0)$. Then, using Fieller's Theorem we obtain an asymptotic $100(1-\alpha)\%$ confidence interval for RED as

$$[RED_1, RED_2], \quad (10)$$

where $RED_1 = max\left\{(B - \sqrt{B^2 - AC})/A, 0\right\}$, $RED_2 = (B + \sqrt{B^2 - AC})/A$, $A = b^2 - v_{22}Z^2_{\alpha/2}$, $B = ab - Z^2_{\alpha/2}v_{22}$ and $C = a^2 - v_{11}Z^2_{\alpha/2}$.
Method RD3 (Lui (2004)):
Use v_{l_i} for $\hat{var}(\hat{\pi}_i)$ in (1.10).
Method RDR3:
Use v_i for $\hat{var}(\hat{\pi}_i)$ in (1.10).
Method RDS3:
Use v_{s_i} for $\hat{var}(\hat{\pi}_i)$ in (1.10).

DOI: 10.2202/1557-4679.1177

3.2 Simulation study and the results

In this section we report on a simulation study conducted to compare the nine methods RD1, RDR1, RDS1, RD2, RDR2, RDS2, RD3, RDR3 and RDS3 discussed in Section 3.1 for the construction of the confidence intervals for the relative difference RED.

As in Section 2.2, we consider equal number of clusters n (=10, 20, 30 and 50) in the two comparison groups, common intraclass correlation coefficient $\phi_1 = \phi_0 = \phi$ and values of π_0 and ϕ as $\pi_0 = .1$ and $\phi = 0.1, .2, .5$. As relative difference is used in cases where the experimental treatment tends to increase the probability of positive response as compared with standard treatment or control, we must have $\pi_1 \geq \pi_0$. So, we consider values of π_1=.15, .20, .25, .3, .4 and .6 producing values of $RED = \frac{\pi_1 - \pi_0}{1-\pi_0}$ =.056, .111, .167, .222, .333 and .555. Further, as in Section 2.2, cluster sizes m_{ij} have been generated from the Poisson distribution with mean m_0=5, 10, 50 with $m_{ij} = 0$ and $m_{ij} = 1$ being excluded. As in the case of RD and RR, a confidence interval does not exist for some samples if either BMS_i or WMS_i is 0, as the estimate of $\hat{\phi}_i$, used in all the methods, is not valid. Further, if $A \leq 0$ or $B^2 - AC \leq 0$, then confidence intervals by the methods RD3, RDR3 and RDS3 do not exist. Here, as compared to Section 2.2, a substantially larger number of samples had to be rejected to produce 10,000 good samples based on which the confidence intervals and average lengths were calculated.

Our simulations show that in terms bias and average length, the properties of the three versions of each method are similar. So, to save space these results are not presented. Further, as in Section 2.2, we study the version based on the variance of a ratio estimator for all three methods, namely, the methods RDR1, RDR2 and RDR3. The results are given Table 5. To save space we only give results for RED=.056, .167 and .333. From the results in Table 5 we see that all the three methods show similar behavior, although method RDR3 seems to have in general, smaller bias and smaller average coverage length. For larger values of ϕ ($\phi = 0.5$) all three methods show under-coverage. For smaller values of ϕ ($\phi = 0.1$) all three methods show under- coverage only when the sample size is small or the number of clusters is small (see for RED=0.056, n=20,30). For larger values of RED there is evidence of over-coverage.

Table 5: The estimated coverage probabilities and average lengths of confidence intervals(in parenthesis) for the relative difference by the methods *RDR*1, *RDR*2, *RDR*3; for equal numbers of clusters $n_1 = n_0 = n$ in both groups, mean cluster size $m_0 = 5, 10, 50$; underlying mean probability of response in group 0, $\pi_0 = 0.10$ and $\alpha = 0.05$; based on 10,000 simulations

			ϕ=.1			ϕ=.5		
RED	n	m_0	RDR1	RDR2	RDR3	RDR1	RDR2	RDR3
.056	20	5	.805	.805	.803	.707	.707	.705
			(.166)	(.160)	(.156)	(.227)	(.211)	(.212)
	20	10	.865	.865	.864	.734	.734	.732
			(.147)	(.142)	(.137)	(.220)	(.206)	(.204)
	20	50	.913	.913	.913	.747	.747	.744
			(.123)	(.120)	(.114)	(.208)	(.196)	(.194)
	30	5	.860	.860	.858	.762	.762	.760
			(.145)	(.141)	(.136)	(.197)	(.186)	(.183)
	30	10	.905	.905	.904	.776	.776	.773
			(.128)	(.125)	(.119)	(.189)	(.180)	(.177)
	30	50	.957	.957	.955	.799	.799	.797
			(.107)	(.106)	(.098)	(.180)	(.172)	(.169)
	50	5	.923	.923	.921	.819	.819	.817
			(.122)	(.120)	(.113)	(.163)	(.157)	(.153)
	50	10	.955	.955	.954	.838	.838	.836
			(.108)	(.106)	(.098)	(.157)	(.153)	(.147)
	50	50	.986	.986	.983	.851	.851	.849
			(.089)	(.089)	(.080)	(.149)	(.145)	(.140)
.111	20	5	.950	.950	.941	.855	.855	.848
			(.214)	(.208)	(.186)	(.284)	(.267)	(.255)
	20	10	.975	.975	.966	.872	.872	.864
			(.187)	(.184)	(.158)	(.272)	(.258)	(.244)
	20	50	.994	.994	.985	.891	.891	.885
			(.155)	(.154)	(.126)	(.261)	(.250)	(.233)
	30	5	.976	.976	.965	.903	.903	.893
			(.185)	(.182)	(.157)	(.248)	(.237)	(.220)
	30	10	.990	.990	.976	.918	.918	.909
			(.162)	(.160)	(.131)	(.239)	(.230)	(.211)
	30	50	.999	.999	.983	.930	.930	.920
			(.132)	(.131)	(.099)	(.227)	(.220)	(.200)
	50	5	.996	.996	.976	.953	.953	.938
			(.153)	(.152)	(.122)	(.209)	(.204)	(.181)
	50	10	.999	.999	.975	.963	.963	.948
			(.131)	(.130)	(.097)	(.200)	(.196)	(.172)
	50	50	1.0	1.0.	.961	.972	.972	.958
			(.104)	(.104)	(.069)	(.191)	(.188)	(.163)
.167	20	5	.988	.988	.964	.932	.932	.913
			(.247)	(.243)	(.196)	(.328)	(.312)	(.281)
	20	10	.996	.996	.970	.943	.943	.922
			(.215)	(.213)	(.163)	(.316)	(.303)	(.269)
	20	50	1.0	1.0	.969	.955	.955	.934
			(.174)	(.174)	(.120)	(.304)	(.292)	(.256)
	30	5	1.0	1.0	.958	.963	.963	.934
			(.211)	(.210)	(.158)	(.288)	(.279)	(.239)
	30	10	1.0	1.0	.957	.963	.963	.934
			(.180)	(.180)	(.125)	(.288)	(.279)	(.239)
	30	50	1.0	1.0	.940	.979	.979	.949
			(.144)	(.145)	(.088)	(.264)	(.258)	(.215)
	50	5	1.0	1.0	.938	.979	.979	.949
			(.169)	(.169)	(.112)	(.264)	(.258)	(.215)
	50	10	1.0	1.0	.917	.972	.972	.945
			(.143)	(.143)	(.085)	(.276)	(.269)	(.227)
	50	50	1.0	1.0	.887	.981	.981	.951
			(.114)	(.114)	(.057)	(.257)	(.257)	(.213)

DOI: 10.2202/1557-4679.1177

3.3 An example

As an illustrative example we consider the data in Table 2 which were analyzed by Paul and Zaihra (2008) for constructing confidence intervals for the risk difference. For these data we obtain $\hat{\pi}_1 = 0.578$, $\hat{\pi}_0 = 0.382$, $\hat{RED} = .317$ and the estimate of the common intraclass correlation $\hat{\phi} = 0.30$. The 95% two sided confidence interval for the RED by the three methods $RDR1$, $RDR2$ and $RDR3$ are $(0.012, .621)$; $(0.000, .563)$ and $(0.000, .590)$ with corresponding lengths of the confidence intervals .609, .563 and .590 respectively.

As in Section 2.3 we examine the appropriateness of using RDR1, RDR2 and RDR3 in the particular configuration ($\pi_0 = .382$, $\phi = .578$, RED=.317, n=29, m=2) given by this example. For this we apply simulation again. When applying RDR1, RDR2 and RDR3 we obtain the estimated coverage probabilities and coverage lengths (in parenthesis) of 95% confidence intervals to be .96(.517), .96(.481) and .90(.442)respectively. These results suggest that RDR1 and RDR2 have smaller bias in the above situation but RDR3 has the smallest coverage length. However, data analysis and simulation from the example failed to identify any consistent behavior of the three methods.

4 Discussion

As in Paul and Zaihra (2008), some very simple methods based on an estimator of the variance of a ratio estimator (see Cochran, 1977) for constructing confidence intervals for the risk ratio and the relative difference have been introduced. These procedures stand out for their computational simplicity.

For constructing confidence intervals for the risk ratio, the method MR3, which is based on the variance of the ratio estimator v_i of $var(\hat{\pi}_i)$ and Fieller's theorem, has the best overall bias property. When the number of clusters is small and the intraclass correlation is .5 or higher, the estimated average length by this method can be substantially higher than those of other interval estimates (Table 4). This is because coefficient $A^{\dagger}$ in the quadratic equations used while applying Fieller's theorem can become quite small in some extreme simulated samples and hence the estimated average length of the resulting interval estimate becomes extremely large. However, the examples discussed in Section 2.3 show that for well behaved data coverage lengths by all methods are similar. Thus, for constructing a confidence interval for the risk ratio, we recommend the method MR3.

For constructing a confidence interval for the relative risk difference, the method RDR3, which is also based on the variance of the ratio estimator v_i of

$var(\hat{\pi}_i)$ and Fieller's theorem, seems to be preferable, as it has, in general, smaller bias and smaller average coverage length.

For studying confidence interval properties of both the risk ratio and the relative risk difference we have conducted simulations using data from a beta-binomial distribution to generate over-dispersed data. However, in order to study the robustness of these procedures it would be interesting to consider other simulation mechanisms, such as, generating data from a probit normal binomial distribution. In other similar studies (see Paul and Islam, 1995) and Paul and Banarje, 1998), similar results were obtained for different procedures irrespective of which over-dispersion mechanism is used to generate data. Whether this holds true in this case is being investigated which will be reported in a future study.

References

Bailey, B. J. R. (1987). Confidence limits to the risk ratio. *Biometrics*, **43**, 201-205.

Casella, G. and Berger, R. L. (1990). *Statistical inference*, Duxbury: Belmont, California.

Cochran, W. G. (1977). *Sampling Techniques,*Wiley: New York, 3rd edition.

Donner, A., Birkett, N., Buck, C. (1981). Randomization by cluster sample size requirements and analysis. *American Journal of Epidemiology*, **14**, 906-914.

Elston, R. C. (1977). Response to query: estimating 'inheritability' of a dichotomous trait. *Biomętrics* **33**, 232-233.

Fieller, E. C. (1954). Some problems in interval estimation. *Journal of Royal statistical Society, series B*, **16**, 175-185.

Fleiss, J.L. (1986). *The Design and Analysis of Clinical Experiments*. Wiley, New York.

Fleiss, J. L., Levin, B. and Paik, M. C. (2003). *Statistical Methods for Rates and Proportions*, Wiley: Hoboken, New Jersey.

Gart, J. J. (1979). Statistical Analyses of the Relative Risk . *Environmental Health Perspeçtives* **32**, 157-167.

Gart, J. J and Nam, J. M. (1988). Approximate interval estimation of the ratio of binomial parameters: A review and corrections for skewness. *Biometrics*, **44**, 323-338.

Katz, D., Baptista,J., Azen, S.P. and Pike, M. C. (1978). Obtaining confidence intervals for the risk ratio in cohort studies. *Biomętrics* **34**, 469-474.

Koopman, P. A. R. (1984). Confidence interval for the ratio of two binomial proportions. *Biomętrics* **40**, 513-517.

DOI: 10.2202/1557-4679.1177

Lui, K. J, Mayer,J. A., and Eckhardt, L. (2000). Confidence intervals for the risk ratio under cluster sampling based on the beta-binomial model. *Statistics in medicine* **19**, 2933-2942.

Lui, K. J. (2001). Interval estimation of simple difference in Dichotomous Data with repeated Measurements. *Biometrical Journal*, **43**, 845-861.

Lui, K. J. (2004). *Statistical Estimation of Epidemiological Risk*. Wiley: CA.

Mayer, J. A, Slymen, D. J, Eckhardt, L., et al. (1997). Reducing ultraviolet radiation exposure in children . *Preventive Medicine*, **26**, 516-522.

Morris, J. A, and Gardner, M. J. (1988). Calculating confidence intervals for relative risks (odds ratios) and standardized ratio and rates. *Statistics in Medicine*, **296**, 1313-1316.

Newcombe, R. G. (1998). Interval estimation for the difference between independent proportions: comparison of eleven methods. *Statistics in Medicine*, **17**, 873-890.

Paul, S. R. (1982). Analysis of proportions of Affected Foetuses in Teratological Experiments. *Biometrics*, **38**, 361-370.

Paul, S. R. and Banerjee, T. (1998). Analysis of two-way layout of count data involving multiple counts in each cell. *American Statistical Association*, **93**, 1419-1429.

Paul, S.R. and Islam, A.S. (1995). Analysis of proportions based on parametric and semi- parametric models. *Biometrics* **51**, 1400-1410.

Paul, S. R and Zaihra, T. (2008). Interval estimation of risk difference for data sampled from clusters. *Statistics in Medicine* **27**, 4207-4220.

Sheps, M.C. (1958). Shall we count for the living or the dead. *New England Journal of Medicine*, **259**, 1210-1214.

Sheps, M.C. (1959). An examination of some methods of comparing several rates or proportions. *Biometrics* **15**, 87-97.

Tarone, R. E. (1979). Testing the goodness of fit of the binomial distribution. *Biometrika* **66**, 585-590.

Weil, C.S. (1970). Selection of valid number of sampling units and a consideration of their combination in toxicological studies involving reproductions, teratogeneses or carcinogenesis. *Food and Cosmetics Toxicology* **8**: 177-182.

Zeger, S. L. and Liang, K. Y. (1986). Longitudinal Data Analysis for discrete and continuous outcomes. *Biometrics*, **42**, 121-130.