Hybrid-Based Confidence Intervals for the Risk Ratio in the Analysis of Correlated Binary Data

Gopal Nath[*] , Suojin Wang[†] , and Krishna Saha[‡§]

Abstract

Correlated binary data often arise in epidemiological cohort studies. The risk ratio (RR) is one of three major useful measures of association for summarizing the results from such epidemiological cohort studies. In applications, the RR and its complement, the percentage reduction in risk, have a direct interpretation. This usually measures the relative change in the epidemiological risk due to the application of the treatment. Standard approaches for estimating RR available in software packages may lead to biased inferences when applied to a correlated binary data. In this paper, we develop some simple and efficient inference procedures for estimating RR based on a hybrid method introduced by Zou (2008) using four existing interval methods for a single proportion for correlated binary data. A simulation study is conducted to investigate the performance of the proposed methods, and an application to a toxicological study is used to illustrate the proposed methods.

Key Words: correlated binary data, confidence interval, coverage probability, expected length, risk ratio

1. Introduction

Correlated binary data frequently arise in a wide range of biomedical applications. For instance, consider a toxicological study originally studied by Paul (1982). In this study, the data refer to litters of varying sizes, each litter having a number of abnormalities due to a control group and low dose. As seen, individuals within the same litter respond alike and hence are correlated. The number of clusters for this study is moderate with the cluster sizes ranging 1 to 12. Main purpose of such a study is to determine if the treatment affects the incidence of abnormalities in live foetuses. In order to determine this inference problem, some important measures of association such as risk difference (RD), risk ratio (RR) and relative risk difference (RRD) can be used (see Lui, 2004). The preference of one measure of association over another in drawing statistical inference depends on the study design. RD 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). RR is used in toxicological, etiological and cohort studies to quantify the strength of association between a given disease and a suspected risk factor. Although some of the inference procedures for RD and RR have been developed for correlated binary data, little attention is paid to extending numerous inference procedures available in literature for a single proportion case. In this paper, we focus on minimizing this gap by extending some of the recommended procedures for a single proportion to the ratio of the proportions in two treatment groups.

[*]Department of Mathematics and Statistics, Murray State University, Faculty Hall, Murray, KY 42071, USA

[†]Department of Statistics, Texas A&M University, College Station, TX, USA

[‡]Department of Mathematical Sciences, Central Connecticut State University, 1615 Stanley Street, New Britain, CT 06050, USA

[§]**Correspondence**

Let $\phi_i, i = 1, 2$ be the proportion of abnormalities in live foetuses who received the ith treatment. Then estimating RR ($\eta = \phi_1/\phi_2$) will determine whether the treatment affects the incidence of abnormalities in live foetuses. We, therefore, need to develop some efficient inference procedures for estimating $\eta = \phi_1/\phi_2$ for such a design. In this paper, we construct several explicit asymptotic two-sided confidence intervals (CIs) for $\eta = \phi_1/\phi_2$ using the method of variance of estimates recovery (MOVER) proposed by Zou and Donner (2008) and also known as the square-and-add method introduced by Newcombe (1998). The basic idea is to recover variance estimates required for the proportion ratio from the confidence limits for single proportions. The CI estimators for a single proportion, which are incorporated with the MOVER, will include the CIs proposed by Donner and Klar (1993), Lee and Dubin (1994), Jun and Ahn (2000), and Saha et al. (2016).

2. Confidence Intervals for the Ratio of Two Individual Proportions

Suppose that we independently sample m_i clusters from the ith treatment, $i = 1, 2$. Let n_{ij} be the number of individuals in the jth cluster, $j = 1, \ldots, m_i$, who received the ith treatment. Furthermore, suppose that Y_{ij} of the n_{ij} individuals are total successes by the ith treatment. Under the usual assumption $Y_{ij}|p_{ij}$ follows binomial (n_{ij}, p_{ij}), where p_{ij} is the probability that an individual in the jth cluster was cured by the ith treatment. We further assume that the binomial probability p_{ij} is a random variable having mean ϕ_i and variance $\phi_i(1-\phi_i)\theta_i$. The unconditional mean and variance of Y_{ij} are then $n_{ij}\phi_i$ and $n_{ij}\phi_i(1-\phi_i)[1+(n_{ij}-1)\theta_i]$, respectively. Note that the parameter ϕ_i is the proportion of an individual who received the ith treatment and the parameter θ_i is the common intraclass correlation between the binary observations within each cluster in the ith group. In this article, the parameter of interest is $\eta = \phi_1/\phi_2$. In particular, we would like to construct explicitly simple but efficient confident interval procedures of η as follows.

2.1 Hybrid Method

Basically, the hybrid method is known as the method of variance estimates recovery (MOVER) proposed by Zou and Donner (2008) and also known as the square-and-add method introduced by Newcombe (1998). Using this approach, two separate confidence intervals for the two individual success rates are combined to construct a single confidence interval for the ratio of two success rates, $\eta = \phi_i/\phi_2$. In order to construct a confidence interval for η, first consider a $100(1-\alpha)\%$ CI for $\phi_1 - \phi_2$, where ϕ_1 and ϕ_2 denote any two parameters of interest. Let $\hat{\phi}_1$ and $\hat{\phi}_2$ be two estimates of ϕ_1 and ϕ_2, respectively. By the Central Limit Theorem, a $100(1-\alpha)\%$ CI for $\phi_1 - \phi_2$ is given by (L^*, U^*), where

$$L^* = \hat{\phi}_1 - \hat{\phi}_2 - z_{\alpha/2}\sqrt{\text{var}(\hat{\phi}_1) + \text{var}(\hat{\phi}_2)} \text{ and } U^* = \hat{\phi}_1 - \hat{\phi}_2 + z_{\alpha/2}\sqrt{\text{var}(\hat{\phi}_1) + \text{var}(\hat{\phi}_2)}.$$

However, this procedure performs well only when sample sizes are sufficiently large or when the sampling distributions of $\hat{\phi}_i$ $(i = 1, 2)$ are close to normal distribution. From the above equations, it can be shown that L^* and U^* can be regarded as the minimum and maximum parameter values that satisfy

$$\frac{[(\hat{\phi}_1 - \hat{\phi}_2) - L^*]^2}{\text{var}(\hat{\phi}_1) + \text{var}(\hat{\phi}_2)} = z_{\alpha/2} \text{ and } \frac{[U^* - (\hat{\phi}_1 - \hat{\phi}_2)]^2}{\text{var}(\hat{\phi}_1) + \text{var}(\hat{\phi}_2)} = z_{\alpha/2},$$

respectively. Suppose a $100(1-\alpha)\%$ CI for ϕ_i is $(l_i, u_i), i = 1, 2$, where $l_i = \hat{\phi}_i - z_{\alpha/2}\sqrt{\text{var}(\hat{\phi}_i)}$ implies $\widehat{\text{var}}(\hat{\phi}_i) = (\hat{\phi}_i - l_i)^2/z^2_{\alpha/2}$ under $\phi_i \approx l_i$. Similarly, $u_i = \hat{\phi}_i +$

$z_{\alpha/2}\sqrt{\text{var}(\hat{\phi}_i)}$ implies $\widehat{\text{var}}(\hat{\phi}_i) = (u_i - \hat{\phi}_i)^2/z^2_{\alpha/2}$ under $\phi_i \approx u_i$. Based on the possible values (l_1, u_1) of ϕ_1 and (l_2, u_2) of ϕ_2, the values closest to the minimum L and maximum U are $l_2 - u_1$ and $u_2 - l_1$, respectively. As a result, for setting L with $\phi_2 \approx l_2$ and $\phi_1 \approx u_1$, we have $\text{var}(\hat{\phi}_1) + \text{var}(\hat{\phi}_2) = (u_1 - \hat{\phi}_1)^2/z^2_{\alpha/2} + (\hat{\phi}_2 - l_2)^2/z^2_{\alpha/2}$, which gives

$$L^* = \hat{\phi}_1 - \hat{\phi}_2 - \sqrt{(\hat{\phi}_2 - l_2)^2 + (u_1 - \hat{\phi}_1)^2}. \tag{1}$$

Similarly, we have

$$U^* = \hat{\phi}_1 - \hat{\phi}_2 + \sqrt{(u_2 - \hat{\phi}_2)^2 + (\hat{\phi}_1 - l_1)^2}. \tag{2}$$

Now, let (L, U) be the $(1-\alpha)100\%$ confidence interval for $\eta = \phi_i/\phi_2$, that is,

$$P(L \leq \phi_1/\phi_2 \leq U) = 1 - \alpha.$$

Equivalently,

$$P(\phi_1 - U\phi_2 \leq 0 \leq \phi_2 - L\phi_1) = 1 - \alpha.$$

For fixed L and U, we apply (1) to $\phi_2 - L\phi_1$ and (2) to $\phi_1 - U\phi_2$ and by setting $L^* = 0$ and $U^* = 0$, we obtain the $(1-\alpha)100\%$ confidence interval for $\eta = \phi_i/\phi_2$ as

$$L = \frac{\hat{\phi}_1\hat{\phi}_2 - \sqrt{(\hat{\phi}_1\hat{\phi}_2)^2 - u_2(2\hat{\phi}_1 - l_1)l_1(2\hat{\phi}_2 - u_2)}}{u_2(2\hat{\phi}_2 - u_2)} \tag{3}$$

and

$$U = \frac{\hat{\phi}_1\hat{\phi}_2 + \sqrt{(\hat{\phi}_1\hat{\phi}_2)^2 - u_1(2\hat{\phi}_1 - u_1)l_2(2\hat{\phi}_2 - l_2)}}{l_2(2\hat{\phi}_2 - l_2)}, \tag{4}$$

where $\hat{\phi}_1$ and $\hat{\phi}_2$ are the estimates of ϕ_1 and ϕ_2, respectively.

It is easily seen that to obtain a $100(1-\alpha)\%$ MOVER based confidence interval for $\eta = \phi_1/\phi_2$ using Equations (3) and (4), one needs two separate $100(1-\alpha)\%$ confidence intervals: (l_1, u_1) for ϕ_1 and (l_2, u_2) for ϕ_2. Saha, Miller and Wang (2015) investigated the problem of confidence intervals for a single proportion for clustered binary data. Based on their analysis, they recommended the Wilson score and the profile likelihood methods. There are three other approaches available in literature to obtain the confidence intervals (l_i, u_i) for ϕ_i, which are included here as well. The formulae for these confidence intervals are briefly discussed as follows.

2.1.1 The Wilson score interval

The natural estimator of ϕ_i $(i = 1, 2)$ can easily be obtained as the overall sample proportion $\hat{\phi}_i = /n_{i.}$, where $Y_{i.} = \sum_j^{m_i} Y_{ij}$ and $n_{i.} = \sum_j^{m_i} n_{ij}$. The variance of $\hat{\phi}_i$ is given by $\text{Var}(\hat{\phi}_i) = \phi_i(1-\phi_i)\xi_i/n_{i.}$, where $\xi_i = \sum n_{ij}[1 + (n_{ij} - 1)\theta_i]/n_{i.}$. Using the central limit theorem, it can be shown that $n_{i.}^{1/2}(\hat{\phi}_i - \phi_i)/\sqrt{\phi_i(1-\phi_i)\hat{\xi}_i}$ converges in distribution to the standard normal distribution as $k \to \infty$, where $\hat{\xi}_i$ is obtained by replacing θ_i by its estimate $\hat{\theta}_i$. Then, the approximate $100(1-\alpha)\%$ Wilson confidence interval for ϕ_i is the root of the quadratic equation

$$P(n_{i.}(\hat{\phi}_i - \phi_i)^2/[\phi_i(1-\phi_i)\hat{\xi}_i] \leq z^2_{\alpha/2}) = 1 - \alpha.$$

After some straightforward algebra, it can be obtained as

$$\text{WI}: \quad (l_i, u_i) = \tilde{\phi}_i \pm \frac{z_{\alpha/2}}{\tilde{n}_{i.}} \sqrt{n_{i.}\hat{\phi}_i(1-\hat{\phi}_i)\hat{\xi}_i + \frac{\hat{\xi}_i^2 z_{\alpha/2}^2}{4}},$$

where

$$\tilde{\phi}_i = \frac{n_{i.}\hat{\phi}_i + 0.5\hat{\xi}_i z_{\alpha/2}^2}{n_{i.} + \hat{\xi}_i z_{\alpha/2}^2} = \frac{Y_{i.} + 0.5\hat{\xi}_i z_{\alpha/2}^2}{n_{i.} + \hat{\xi}_i z_{\alpha/2}^2} \text{ and } \tilde{n}_{i.} = n_{i.} + \hat{\xi}_i z_{\alpha/2}^2.$$

It is worthwhile to note here that for non-clustered data when there is no cluster effect, that is, $\theta_i = 0$ (or $\xi_i = 1$) the same intervals are produced (see, for example, Newcombe, 1998). The estimate $\hat{\theta}_i$ can be obtained using the analysis of variance (ANOVA) method, which is given by $\hat{\theta}_i^a = (BMS_i - WMS_i)/[BMS_i + (n_i^* - 1)WMS_i]$, where $BMS_i = [\sum_j Y_{ij}^2/n_{ij} - (\sum_j Y_{ij})^2/\sum_j n_{ij}]/(k-1)$ and $WMS_i = [\sum_j Y_{ij} - \sum Y_{ij}^2/n_{ij}]/\sum_j(n_{ij} - 1)$ are the between mean-squared and within mean-squared errors, respectively, and $n_i^* = [(\sum_j n_{ij})^2 - \sum_j n_{ij}^2]/[(k-1)\sum_j n_{ij}]$. Therefore, one can obtain Wilson CIs for ϕ_i $(i = 1, 2)$ using the above interval (l_i, u_i) by substituting ANOVA estimate of θ_i in the equation for $\hat{\xi}_i$ above. We denote this interval as HB_1.

2.1.2 The CI based on PL

A profile likelihood based confidence interval approach has been shown to provide accurate results when computing confidence limits for a single proportion (Newcombe, 1998) or the difference between two proportions (Pradhan et al., 2014) in the case of non-clustered binary data. Let $l(\phi, \psi)$ be the log-likelihood function, where ϕ is the parameter of interest and ψ is the nuisance parameter. Also, let $l_p(\phi) = l(\phi, \hat{\psi}(\phi))$ be the profile likelihood for ϕ, where $\hat{\psi}(\phi)$ is obtained from the reduced model with respect to ψ keeping ϕ fixed. Then the approximate $100(1-\alpha)\%$ profile likelihood (PL) based confidence interval for ϕ is given by

$$\{\phi : l_p(\phi) \geq l(\hat{\phi}, \hat{\psi}) - \frac{1}{2}\chi_{1,\alpha}^2\}, \tag{5}$$

where $\hat{\phi}$ and $\hat{\psi}$ are the estimates of ϕ and ψ in the full model and $\chi_{1,\alpha}^2$ is the $100(1-\alpha)$ percentile of a chi-squared distribution with one degree of freedom. Due to a superior model for clustered binary data, we consider the beta-binomial model in order to obtain the PL based confidence interval for ϕ. Consequently, the log-likelihood function for the beta-binomial model, apart from a constant, is given by

$$l(\phi, \psi) = \sum_{i=1}^{m} \left[\sum_{j=0}^{y_i - 1} ln\{(1-\psi)\phi + j\psi\} + \sum_{j=0}^{n_i - y_i - 1} ln\{(1-\phi)(1-\psi) + j\psi\} - \sum_{j=0}^{n_i - 1} \{(1-\psi) + j\psi\} \right].$$

The maximum likelihood estimates $\hat{\phi}$ and $\hat{\psi}$ of ϕ and ψ are obtained by solving the following estimating equations

$$\frac{\partial l}{\partial \phi} = \sum_{i=1}^{m} \{ \sum_{r=0}^{y_i - 1} \frac{1-\psi}{(1-\psi)\phi_i + r\psi} - \sum_{r=0}^{n_i - y_i - 1} \frac{1-\psi}{(1-\psi)(1-\phi_i) + r\psi} \} = 0$$

and

$$\frac{\partial l}{\partial \psi} = \sum_{i=1}^{m} \{ \sum_{r=1}^{y_i - 1} \frac{-\phi_i + r}{(1-\psi)\phi_i + r\psi} + \sum_{r=0}^{n_i - y_i - 1} \frac{-(1-\phi_i) + r}{(1-\psi)(1-\phi_i) + r\psi} - \sum_{r=0}^{n_i - 1} \frac{r-1}{(1-\psi) + r\psi} \} = 0$$

simultaneously. Finally, the interval limits can be obtained by finding the two roots of the above equation (5), one in the interval $(0, \hat{\phi})$ and the other in the interval $(\hat{\phi}, 1)$, using either the bisection method or Brent's method.

2.1.3 Confidence Interval Based on Donner and Klar Approach

As we discussed earlier, the natural estimator of ϕ_i is $\hat{\phi}_i = Y_{i.}/n_{i.}$, where $Y_{i.} = \sum_{j=1}^{m_i} Y_{ij}$ and $n_{i.} = \sum_{j=1}^{m_i} n_{ij}$. Donner and Klar (1993) also used the natural estimator $\hat{\phi}_i$ $(i = 1, 2)$, but they estimated the variance of $\hat{\phi}_i$ by incorporating the dependence of the responses within each subject which is given by

$$\widehat{\text{var}}_{DK}(\hat{\phi}_i) = \frac{\hat{\phi}_i(1-\hat{\phi}_i)[1 + (\sum_{j=1}^{m_i} n_{ij}^2/n_{i.} - 1)\hat{\psi}_i]}{n_{i.}},$$

where $\hat{\psi}_i$ is the ANOVA estimate of ψ_i. It can then be shown that $(\hat{\phi}_i - \phi_i)/\sqrt{\widehat{\text{var}}_{DK}(\hat{\phi}_i)}$ is asymptotically $N(0, 1)$ as $m_i \to \infty$. Thus, an approximate $100(1-\alpha)\%$ confidence interval for ϕ_i is obtained as

$$L_{DK} = \hat{\phi}_i - z_{1-\alpha/2}\sqrt{\widehat{\text{var}}_{DK}(\hat{\phi}_i)}$$

and

$$U_{DK} = \hat{\phi}_i + z_{1-\alpha/2}\sqrt{\widehat{\text{var}}_{DK}(\hat{\phi}_i)}.$$

2.1.4 Confidence Interval Based on Lee and Dubin Approach

Lee and Dubin (1994) proposed a weighted estimator of ϕ_i $(i = 1, 2)$ by assigning equal weights to the clusters regardless of the cluster sizes, which is given by $\hat{\phi}_i^{ew} = \sum_{j=1}^{m_i} \hat{\phi}_{ij}/m_i$, where $\hat{\phi}_{ij} = Y_{ij}/n_{ij}$ $(j = 1, \ldots, m_i; i = 1, 2)$ is the estimator of ϕ_{ij} from the ith subject. Then the variance estimate of $\hat{\phi}_i^{ew}$ is obtained as $\widehat{\text{var}}_{DK}(\hat{\phi}_i^{ew}) = \sum_{j=1}^{m_i}(\hat{\phi}_{ij} - \hat{\phi}_i^{ew})^2/[m_i(m_i-1)]$. It follows that $((\hat{\phi}_i^{ew} - \phi_i)/\sqrt{\widehat{\text{var}}_{LD}(\hat{\phi}_i^{ew})}$ is asymptotically $N(0, 1)$ as $m_i \to \infty$. Thus, an approximate $100(1-\alpha)\%$ confidence interval for ϕ_i is obtained as

$$L_{LD} = \hat{\phi}_i^{ew} - z_{1-\alpha/2}\sqrt{\widehat{\text{var}}_{LD}(\hat{\phi}_1^{ew})}$$

and

$$U_{LD} = \hat{\phi}_i^{ew} + z_{1-\alpha/2}\sqrt{\widehat{\text{var}}_{LD}(\hat{\phi}_1^{ew})}.$$

2.1.5 Confidence Interval Based on Jun and Ahn Approach

Jun and Ahn (2000) used another weighted estimator of ϕ_i $(i = 1, 2)$ by minimizing the variance of the estimator of ϕ_i $(i = 1, 2)$. This weighted estimator of ϕ_i $(i = 1, 2)$ may be interpreted as a generalized estimating equation estimator using the true exchangeable correlation structure, and is given by $\hat{\phi}_i^w = \sum_{j=1}^{m_i} w_{ij}\hat{\phi}_{ij}$, where $\hat{\phi}_{ij} = Y_{ij}/n_{ij}$ and $w_{ij} = [n_{ij}\{1 + (n_{ij}-1)\hat{\psi}_i\}^{-1}]/\sum_{j=1}^{m_i}[n_{ij}\{1 + (n_{ij}-1)\hat{\psi}_i\}^{-1}]$ $(j = 1, \ldots, m_i; i = 1, 2)$ are the weights obtained by minimizing the variance of $\hat{\phi}_i^w$ based on the true exchangeable correlation structure. Note that $\hat{\psi}_i$ $(i = 1, 2)$ is the ANOVA estimator of the intraclass correlation coefficients which can provide negative estimates when the true intraclass correlation coefficients are close to 0 and the number of clusters is small. In such a case, we truncate the ANOVA estimate of ψ_i at 0 since the true intraclass correlation coefficient ψ_i should be nonnegative under the exchangeable condition. The variance estimate of $\hat{\phi}_i^w$ can be obtained as $\widehat{\text{var}}_{JA}(\hat{\phi}_i^w) = \hat{\phi}_i^w(1-\hat{\phi}_i^w)/\sum_{j=1}^{m_i}[n_{ij}\{1 + (n_{ij}-1)\hat{\psi}_i\}^{-1}]$. Then an approximate $100(1-\alpha)\%$ confidence interval for ϕ_i is obtained as

$$L_{JA} = \hat{\phi}_i^w - z_{1-\alpha/2}\sqrt{\widehat{\text{var}}_{JA}(\hat{\phi}_i^w)}$$

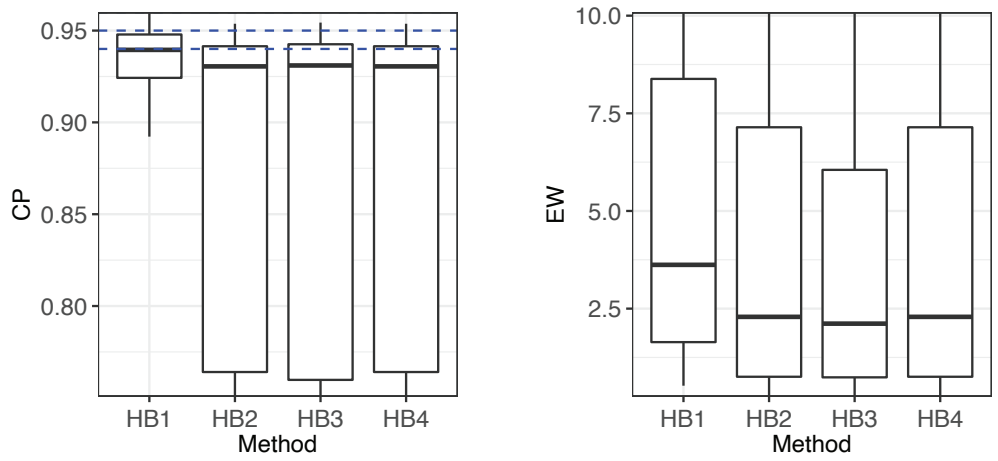
and

$$U_{JA} = \hat{\phi}_i^w + z_{1-\alpha/2}\sqrt{\widehat{\text{var}}_{JA}(\hat{\phi}_i^w)}.$$

3. Simulations

In this section, we investigate the performance of the small and moderate sample behavior of the proposed methods in terms of observed coverage probability and average interval length using the pre-assigned confidence level of 95%.

We considered the number of clusters $k = 20, 30, 50$ with mean cluster size, $m = 10, 50, 100$, and the response probability $\phi_1 = 0.2$ and $\eta = 1, 2, 4$. Based on historical data in biomedical applications, the intraclass correlation coefficients for two groups were set as: (0.1, 0.1) , (0.5, 0.5), (0.1, 0.25), (0.25, 0.5). We generated data Y_{ij} based on the beta-binomial distribution and generated 10,000 data sets for each assessment.



(a) Covearage Probability (CP)

(b) Expected Width (EW)

Figure 1: The coverage probability and the expected width of the 95% nominal confidence interval for $\eta = \phi_1/\phi_2$ are shown based on all methods discussed in Section 2. Boxplots were constructed using all combinations of parameters discussed in Section 3.

The observed coverage probability (CP) and the expected interval length (EL) for two-sided confidence intervals (l_j, u_j) for $\eta = \phi_1/\phi_2$ were obtained by

$$\text{CP} = \frac{\sum_{t=1}^{10000} I(l_t \leq \eta \leq u_t)}{10000} \quad \text{and} \quad \text{EL} = \frac{\sum_{t=1}^{10000}(u_t - l_t)}{10000},$$

where $I = 1$ if $l_t \leq \eta \leq u_t$, and I = 0, otherwise. The results are reported in Figure 1

from which we make the following observations:

- The CPs of HB2, HB3, and HB4 are nearly identical, while HB1 has reasonability close to the nominal level.
- HB1 provides better coverage than the other three methods and maintains coverage close to the nominal level.
- All four methods tend to have similar ELs; however, the HB3 tend to have smaller ELs compared to the other methods.

4. Example of A Toxicological Study

We revisit the example of estimating the risk ratio $\eta = \phi_1/\phi_2$ to determine whether the treatment affects the incidence of abnormalities in live foetuses. Some summary statistics

for the data set are presented in Table 1. From Table 1, we see that the mean cluster sizes are almost the same and the estimated probabilities for low and control dose groups are also the same, that is, the estimated risk ratio is almost 1. The distributions of cluster-level proportions for both treatment groups are shown in Figure 2 which indicates that the distributions are highly skewed.

Table 1: Summary statistics for the data set in a toxicological study

Toxicological Treatments	# of subjects	# of clusters	mean cluster size	success rate
Control, C	215	27	7.962	0.134
Low Dose, L	133	19	7.00	0.135

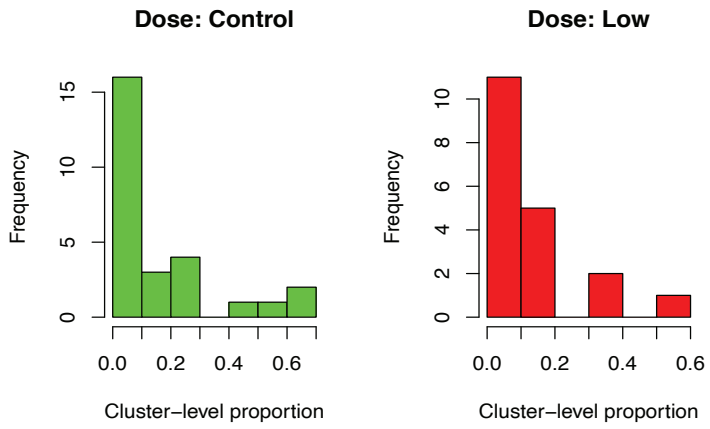


Figure 2: The distributions of cluster-level proportions for both treatment groups in a toxicological study.

The 95% confidence intervals for the risk ratio $\eta = \phi_1/\phi_2$ obtained using the proposed methods are given in Table 2. It is seen from Table 2 that all four confidence intervals include zeros, showing that there are no statistical significance that the treatment affects the incidence of abnormalities in the live foetuses. As expected due to positive correlation within each litter, our proposed method HB1 shows the shorter width compared to the methods discussed here.

5. Conclusion

This paper proposed four methods to construct the confidence intervals for the success ratio $\eta = \phi_1/\phi_2$ for a correlated binary data based on the hybrid procedure using the two separate CIs for a single proportion. The results of a simulation study suggest that the proposed HB1 method generally perform well as its observed CPs are close to the nominal coverage level. Although HB3 generally has shorter ELs in most of the data scenarios considered here, it suffers seriously to maintain the expected coverage probability. We,

Table 2: The 95% confidence intervals for ϕ_1/ϕ_2 obtained using the HB1, HB2, HB3 and HB4 methods.

Method	Lower Limit	Upper Limit	Comparison Length
HB1	0.462	2.151	1.689
HB2	0.402	2.509	2.107
HB3	0.476	3.392	2.916
HB4	0.446	2.758	2.312

therefore, recommend the HB1 procedure for the risk ratio $\eta = \phi_1/\phi_2$ for a correlated binary data from epidemiological cohort studies or similar fields.

Acknowledgements

This paper was supported in part by a CSU-AAUP University research grant.

REFERENCES

Cooper, M. R., Dear, K. B. G., McIntyre, 0. R., Ozer, H., Ellerton, J., Cannellos, G., Duggan, B., and Schiffer, C. (1993), "A Randomized Clinical TrialCcomparing Melphalan/Prednisone with and without a-2b Interferon in Newly-Diagnosed Patients with Multiple Myeloma: A Cancer and Leukemia Group B Study". *Journal of Clinical Oncology*, **11**, 155-160.

Donner, A., and Klar, N. (1993). "Confidence interval construction for effect measures arising from cluster randomization trials". *Journal of Clinical Epidemiology*, **46**, 123–131.

Donner, A. and Klar, N. (2000), "Design and Alalysis of Cluster Randomized Trials in Health Research". Arnold, New York, USA.

G. Y. Zou ,A. (2007) "Donner, Construction of confidence limits about effect measures: A general approach". *Statistics in Medicine*, **27**, 1693-1702.

Jung, S. H., & Ahn, C. (2000). "Estimation of response probability in correlated binary data: A new approach". *Drug Information Journal*, **34**, 599–604.

Kupper, L. L., Portier, C., Hogan, M. D., and Yamamoto, E. (1986). "The Impact of Litter Effects on Dose-Response Modeling in Teratology". *Biometrics*, **42**, 85-98.

Lee, E. W. and Dubin, N. (1994). "Estimation and sample size considerations for clustered binary responses". *Statistics in Medicine* **13** , 1241-1252

Saha, K. K. and Wang, S., and Miller, D. "A Comparison of Some Approximate Confidence Intervals for a Single Proportion for Clustered Binary Outcome Data". *International Journal of Biostatistics*, bf 12, 1-24.

Saha, K. K. and Wang, S. "Confidence Intervals for the Difference in the Success Rates of Two Treatments in the Analysis of Correlated Binary Responses". *Biometrical Journal*, **61**, 983-1002.

Venzon, D. J. and Moolgavkar, S. H. (1988), "A method for computing profilelikelihood-based confidence intervals". *Applied Statistics*, **37**, 87-94.

Zaihra, Tasneem and Paul, Sudhir (2010), "Interval Estimation of Some Epidemiological Measures of Association," *The International Journal of Biostatistics* **6**, Iss. 1.