Thailand Statistician
April 2024; 22(2): 312-327
http://statassoc.or.th
Contributed paper

A Simple Profile Likelihood-based Confidence Interval for the Risk Ratio in Rare Events Meta-analysis

Patarawan Sangnawakij*
Department of Mathematics and Statistics, Faculty of Science and Technology,
Thammasat University, Pathum Thani, Thailand
*Corresponding author; e-mail: patarawan.s@gmail.com

Received: 31 March 2023
Revised: 19 August 2023
Accepted: 30 August 2023

Abstract

Meta-analysis refers to a quantitative method for performing statistical analysis and summarizing results from independent studies, to draw overall conclusions. When the small number of events in individual studies are observed in one or both treatment groups, the classical meta-analysis can lead to perversion because of data sparsity. In this paper, two confidence intervals for the risk ratio in rare events meta-analysis are proposed. They are derived through the profile likelihood ratio method. An extensive simulation study is performed to evaluate the performance of the proposed estimators. These are compared to the Wald-type and Mantel-Haenzel confidence intervals. By mean of simulations, our confidence interval is found to have a good performance in general cases in the study. It is also robust; in other words, regardless of the number of studies, its simulated coverage probability is close to the specified confidence coefficient with an acceptable average length. Real data analysis on epidemiology and transmission is conducted to assess the computational feasibility of the proposed methods.

Keywords: Multiple studies, interval estimation, likelihood ratio, risk ratio, small events.

1. Introduction

Meta-analysis has become an important statistical tool to summarize and assess the same outcomes from individually multiple independent studies on the same research topic. It is widely used in areas of social science, epidemiology, public health science, and medicine (Borenstein et al., 2009). The outcomes in a meta-analysis can be treated as continuous, counts, or dichotomous data in 2×2 tables. In this paper, we are interested in a meta-analysis setting of binary endpoints with the occurrence of rare events situation. *Rare event* is a small number of events with a low probability of occurrence and very unlikely to observe in a trial. If the trial has a zero event in one group, we refer to it as a single-zero study. If the trial involves zero events in both arms, we refer to it as a double-zero study (Böhning et al., 2015; Wei el al., 2021). The situation in such a case can be happened in areas of data, such as clinical trial, natural phenomena (e.g. major earthquakes and tsunamis), and anthropogenic hazards (Zabriskie et al., 2021). Since some of traditional methods for meta-analysis are not well suited to handle rare outcomes, the approach for dealing with meta-analysing studies with low event rates is challenging and will be taken into a deeper look.

The motivation for this work is given by a real data example in epidemiology. We know that the coronavirus causes Middle East respiratory syndrome (MERS or MERS-CoV), which is the highly

pathogenic and deadly human coronavirus, as well as SARS-CoV and SARS-CoV-2 or COVID-19 (Manmana et al., 2020; Zhu el al., 2020). This virus can be transmitted from person to person, who are in close contact with each other. To avoid this, the Ministry of Public Health (MoPH), health care, and non-health-care (e.g. community) settings have been provided some knowledge, information, and personal data protection policy to avoid person-to-person virus transmission. These are included the use of face masks, eye protection, and physical distancing. In a meta-analytic study referred to Chu et al. (2020), they aimed to investigate the effects of eye protection (face-shield and goggles) on virus transmission. This is because eye protection worn is typically underconsidered, but it may be effective in community settings. Moreover, it is debated in the mainstream media and public health authorities for eye protection on virus transmission in general people. Then, Chu et al. (2020) performed a systematic review (Cooper et al., 2009) on this topic and used four studies published in Years 2016-2019. Table 1 shows the related meta-analytic data used in the previous work. The number of patients with eye protection (treatment group) and without eye protection (control group) that are infected with the MERS-CoV virus, and sample sizes of the patients from each study are reported. Here, one study has double-zero events and two studies have single-zero trial. We can see that all four studies have small events compared to sample sizes in the study. The forest plot with study-specific estimated risk ratios corresponding to the MERS-CoV dataset is displayed in Figure 1. This plot is obtained on performing the inverse variance-weighted average meta-analysis (IVW) using the `metan` command in Stata software (Stata Corp, 2013). We notice that the IVW method excludes the double-zero-event study, which can lead to bias in estimation, especially for small study sizes, as pointed out in Kaul and Diamond (2011). A closer look into these data will be taken again in the numerical application section of this paper.

Table 1 Meta-analytic data on the association of eye protection with risk of MERS-CoV transmission

		Number of event		Sample size	
Study	Year	Eye protection	No eye protection	Eye protection	No eye protection
Alraddadi et al.	2016	1	17	47	165
Ki et al.	2019	0	6	9	64
Kim et al.	2016	0	2	443	294
Ryu et al.	2019	0	0	24	10

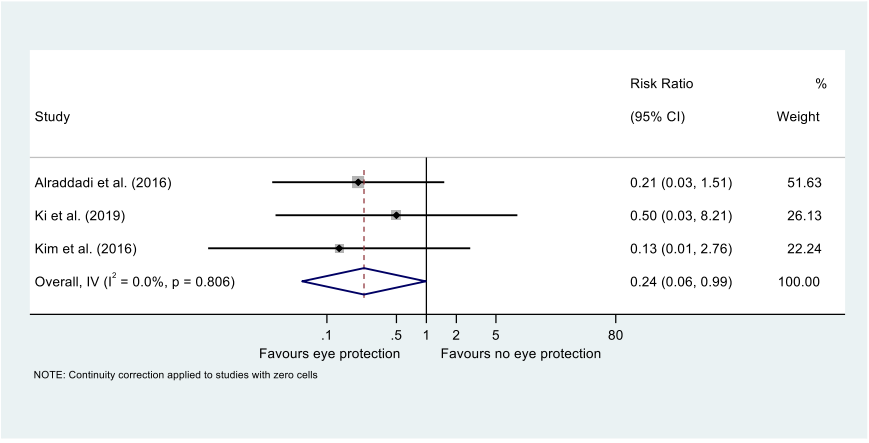

Figure 1 Forest plot of the risk ratio for the association of eye protection to prevent person-to-person transmission of MERS-CoV

We now consider the theoretical part. Summarizing the association between two variables of count outcomes often uses the *risk ratio*. It is the ratio of the probability of an outcome in an treatment to the probability of an outcome in a comparison group. In a meta-analysis of k independent studies, the traditional approach to pool the risk ratio uses the IVW method. This typically requires the number of events in two groups and sample sizes in study i for $i = 1, 2, ..., k$, to calculate the risk or log-risk ratio. According to the IVW approach, the estimated risk ratio calculated on the log-scale is of the form

$$\log(\hat{\theta}_{IVW}) = \frac{\sum_{i=1}^{k} \log(\hat{\theta}_i)/\widehat{Var}(\log(\hat{\theta}_i))}{\sum_{i=1}^{k} 1/\widehat{Var}(\log(\hat{\theta}_i))}, \tag{1}$$

where $\log(\hat{\theta}_i)$ is the log-risk ratio in study i and $\hat{\theta}_i$ is the study-specific risk ratio estimate, computed by $\hat{\theta}_i = x_{i1}n_{i0}/(x_{i0}n_{i1})$. We note that x_{i1} and x_{i0} are the number of events of study i in the treatment and control groups, respectively, and n_{i1} and n_{i0} are the sample sizes of study i in the two groups. Here, $\widehat{Var}(\log(\hat{\theta}_i))$ is the estimated variance of $\log(\hat{\theta}_i)$, derived using the delta method based on the first-order Taylor series. It is given by

$$\widehat{Var}(\log(\hat{\theta}_i)) = \frac{1}{x_{i1}} - \frac{1}{n_{i1}} + \frac{1}{x_{i0}} - \frac{1}{n_{i0}}, \tag{2}$$

where the weights $\widehat{Var}(\log(\hat{\theta}_i)) > 0$. The variance of $\log(\hat{\theta}_{IVW})$ is obtained from the reciprocal of the sum of the weights. For interval estimation, a $(1-\alpha)100\%$ confidence interval for the true log-risk ratio using the Wald-type method is given as

$$CI_{E1} = \log(\hat{\theta}_{IVW}) \pm z_{\alpha/2}\sqrt{1/\sum_{i=1}^{k} \widehat{Var}(\log(\hat{\theta}_i))}, \tag{3}$$

where α is the significance level and $z_{\alpha/2}$ is the $(\alpha/2)100$th upper percentile of a standard normal distribution (Schulze et al., 2003). Anti-logarithm of (1) and (3), the estimated risk ratio is then obtained. Although the risk ratio based on the IVW method is simple and often applied in applications. However, it is important to notice that (1) is undefined for the meta-analysis with zero-event study, as (2) cannot be computed. To address this issue and hold the IVW approach in use, researches suggest excluding the double-zero study before the analysis. Unfortunately, the exclusion of double-zero study which is a set of available real data can bias the treatment effect parameter estimate (Günhan et al., 2020). Number of studies used in this method will be also eliminated, as shown in the forest plot of Figure 1. Alternatively, the value of 0.5 is suggested to add in each cell in the fourfold table of a zero study. It is simple and easy to use; however, adding the corrected value can be affected the bias performance of the estimator in statistical inference of meta-analysis. See more details in Sweeting et al. (2004), Efthimiou el al. (2019), and Bakbergenuly et al. (2020). Instead of using any corrected value, the Mantel-Haenszel (MH) method is introduced for pooling the effect size estimate in homogeneity meta-analysis (Mantel and Haenszel, 1959). Some advantages of the MH method is that it is in the form of a ratio of sum. Hence, it is not sensitive to zero counts in an arm. Moreover, the MH estimator has a closed-form solution which is close to the profile likelihood estimator (Böhning et al., 2008). However, if all studies in either one or both arms have zero events, the MH estimator is undefined. This is a limitation of this method.

In the literature, several statistical methodologies have introduced point estimators for dealing with zero or rare events in the meta-analysis, see for example, Böhning et al. (2022), Noma and Nagashima (2016), and Piaget-Rossel and Taffé (2019). Interval estimation is usually used in applications as well. In this paper, we therefore aim to construct the alternative confidence interval for the risk ratio. These are derived based on the profile likelihood to obtain the Wald-type and likelihood ratio confidence limits. We evaluate the performance of the confidence intervals via simulations, under meta-analytic data in rare-event situations. It will be compared to the well-known IVW and

MH confidence intervals. Finally, the proposed confidence intervals are illustrated with the real-data example motivated at the beginning of this section.

2. Background and Likelihoods

Consider k independent studies in a meta-analysis, we suppose that X_{i1} and X_{i0} are the number of events in study i, for $i = 1, 2, ..., k$, of the treatment and control groups, respectively. Also, we assume that X_{i1} and X_{i0} are two random samples from two Poisson distributions with means $n_{i1}p_{i1}$ and $n_{i0}p_{i0}$, respectively. These are denoted as $X_{ij} \sim Po(n_{ij}p_{ij})$, where n_{ij} are the sample sizes and $p_{ij} \in (0, 1)$ are the event-risk parameters in study i and group j, for $j = 0, 1$. Note that in rare events situation the probability of an event occurring is low, so the event parameter has a small probability value. Under the assumption given above, the likelihood function of p_{i1} and p_{i0}, given observed values $X_{ij} = x_{ij}$ is given by

$$L(p_{i1}, p_{i0}) = \prod_{i=1}^{k} \left[\frac{e^{-n_{i1}p_{i1}}(n_{i1}p_{i1})^{x_{i1}}}{x_{i1}!} \times \frac{e^{-n_{i0}p_{i0}}(n_{i0}p_{i0})^{x_{i0}}}{x_{i0}!} \right]$$

with the log-likelihood

$$\log L(p_{i1}, p_{i0}) = \sum_{i=1}^{k} \left[-n_{i1}p_{i1} + x_{i1} \log(n_{i1}p_{i1}) - n_{i0}p_{i0} + x_{i0} \log(n_{i0}p_{i0}) \right]. \tag{4}$$

The profile likelihood method is then considered. This technique reduces a likelihood to a function of one interested parameter and eliminates the other parameter in the model (Murphy and van der Vaart, 2000; Royston, 2007). The processes used in this work are given as follows. In study i of a meta-analysis, the risk ratio parameter is defined as $\theta_i = p_{i1}/p_{i0}$. According to Böhning et al. (2008), p_{i1} can be rewritten to be $p_{i1} = \theta_i p_{i0}$. If it is replaced in (4), the profile log-likelihood function for θ_i and p_{i0} is then given by

$$\log L(\theta_i, p_{i0}) = \sum_{i=1}^{k} [-(n_{i0} + n_{i1}\theta_i)p_{i0} + x_{i1} \log(\theta_i) + x_i \log(p_{i0})], \tag{5}$$

where $x_i = x_{i1} + x_{i0}$ denotes the total events in study i. Eqn. (5) is used to find the maximum likelihood (ML) estimator for p_{i0}, which is established as $\hat{p}_{i0} = x_i/(n_{i0} + n_{i1}\theta_i)$. Again, $\hat{p}_{i0}$ is replaced in p_{i0} of (5). Hence, we have the profile log-likelihood for θ_i:

$$\log L^*(\theta_i) = \sum_{i=1}^{k} [x_{i1} \log(\theta_i) - x_i \log(n_{i0} + n_{i1}\theta_i)].$$

In homogeneity case, where $\theta_i = \theta$ for all i, the profile log-likelihood for the overall risk ratio (θ) is then given as

$$\log L^*(\theta) = \sum_{i=1}^{k} [x_{i1} \log(\theta) - x_i \log(n_{i0} + n_{i1}\theta)]. \tag{6}$$

To find the ML estimator for θ, log-likelihood (6) is maximized by taking the first-partial derivative with respective to θ. This yields

$$\theta = \frac{\sum_{i=1}^{k} x_{i1}n_{i1}/(n_{i0} + n_{i1}\theta)}{\sum_{i=1}^{k} x_{i0}n_{i0}/(n_{i0} + n_{i1}\theta)}. \tag{7}$$

To solve the maximum profile-likelihood estimate for θ, (7) is used with the iterative method based on the fixed point iteration. This applies the processes given in Algorithm 1.

Algorithm 1: Profile maximum likelihood estimate for θ

Step 0: Given initial value $\theta^{(0)}$ for θ

Step 1: Estimate

$$\theta^{(t+1)} = \frac{\sum_{i=1}^{k} x_{i1} n_{i1}/(n_{i0} + n_{i1}\theta^{(t)})}{\sum_{i=1}^{k} x_{i0} n_{i0}/(n_{i0} + n_{i1}\theta^{(t)})}.$$

Step 2: Repeat Step 1 until $|\theta^{(t+1)} - \theta^{(t)}| < e$.

Note that e is a small error value, say 0.000001, and $\theta^{(t+1)}$ is the estimated value in the $(t+1)$th iteration, where $t = 0, 1, 2, ..., m$ and m is the largest iteration. The maximum profile likelihood estimate obtained from this approach is denoted as $\hat{\theta}_{PL}$.

Another estimator derived based on the basis of the profile likelihood method is explained. This supposes that θ being on the right-hand side of (7) is equal to one (generally, $\theta > 0$). Hence, the explicit formula for θ is simply obtained. It is known as the Mantel-Haenzel (MH) estimator (Mantel and Haenszel, 1959). The MH formula for the risk ratio can be written as

$$\hat{\theta}_{MH} = \frac{\sum_{i=1}^{k} x_{i1} n_{i0}/n_i}{\sum_{i=1}^{k} x_{i0} n_{i1}/n_i},$$

where $n_i = n_{i1} + n_{i0}$. The variance estimator for $\log(\hat{\theta}_{MH})$ derived by Greenland and Robins (Greenland and Robins, 1985) is formulated as

$$\widehat{Var}(\log(\hat{\theta}_{MH})) = \frac{\sum_{i=1}^{k}(x_i n_{i1} n_{i0}/n_i^2 - x_{i1} x_{i0}/n_i)}{\sum_{i=1}^{k}(x_{i1} n_{i0}/n_i)\sum_{i=1}^{k}(x_{i0} n_{i1}/n_i)}.$$

Note that Greenland and Robins (Greenland and Robins, 1985) assumed the random variables followed binomial distributions. Under the large-sample approximation, a $(1-\alpha)100\%$ confidence interval for $\log(\theta)$ using the Wald-type method is therefore given by

$$CI_{E2} = \log(\hat{\theta}_{MH}) \pm z_{\alpha/2}\sqrt{\widehat{Var}(\log(\hat{\theta}_{MH}))}, \tag{8}$$

where $z_{\alpha/2}$ is the $(\alpha/2)100$th upper percentile of a standard normal distribution. The confidence limits for θ can be obtained by anti-logarithm of (8).

3. Alternative Confidence Interval

In this section, the new confidence intervals for the log-risk ratio in meta-analysis are introduced. They are based on the Wald-type and likelihood ratio methods using the basis of profile-likelihood function.

Assuming that $\phi = \log(\theta)$ is the log-risk ratio, which is the parameter of interest. If it is replaced in (6), the profile log-likelihood function for ϕ given the observed value of $X_{ij} = x_{ij}$, for $i = 1, 2, ..., k$ and $j = 0, 1$, is then given by

$$\log L^*(\phi) = \phi \sum_{i=1}^{k} x_{i1} - \sum_{i=1}^{k} x_i \log(n_{i0} + n_{i1} e^{\phi}). \tag{9}$$

Under maximization of $\log L^*(\phi)$ with respect to ϕ, the maximum profile likelihood estimate for ϕ is satisfied

$$\phi = \log\left(\frac{\sum_{i=1}^{k} x_{i1}}{\sum_{i=1}^{k} x_{i1} n_{i1}/(n_{i0} + n_{i1} e^{\phi})}\right).$$

We denote the estimated log-risk ratio from this approach as $\hat{\phi}_{PL}$. Since there has no closed-form solution for $\hat{\phi}_{PL}$, the iterative method is therefore required to find the estimate of ϕ. Based on the normal approximation, the estimated variance of $\hat{\phi}_{PL}$ derived from the inverse of Fisher information using (9) is given by

$$\widehat{Var}(\hat{\phi}_{PL}) = \left(\sum_{i=1}^{k} \frac{x_i n_{i1} n_{i0} e^{\hat{\phi}_{PL}}}{(n_{i0} + n_{i1} e^{\hat{\phi}_{PL}})^2} \right)^{-1},$$

where $e^{\hat{\phi}_{PL}}$ is obtained from the fixed point iteration with

$$\phi^{(t+1)} = \log \left(\frac{\sum_{i=1}^{k} x_{i1}}{\sum_{i=1}^{k} x_{i1} n_{i1} / (n_{i0} + n_{i1} e^{\phi(t)})} \right).$$

We note that $\hat{\phi}_{PL}$ and $\log(\hat{\theta}_{PL})$ are identical (see the result in the application section), so Algorithm 1 can be applied to solve the solution. However, $\hat{\phi}_{PL}$ and $\log(\hat{\theta}_{MH})$ are totally different, although they are derived under the profile likelihood method. According to the Wald-type method with the estimated variance of $\hat{\phi}_{PL}$, a $(1-\alpha)100\%$ confidence interval for ϕ is given as

$$CI_{P1} = \hat{\phi}_{PL} \pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{\phi}_{PL})}, \tag{10}$$

where $z_{\alpha/2}$ is the $(\alpha/2)100$th upper percentile of a standard normal distribution.

Although the Wald method is widely used to construct the confidence interval for parameter, it often performs well when sample sizes or number of studies in meta-analysis are large enough. As noted in Agresti (1990), the likelihood ratio (LR) method could be addressed the problem of small sample sizes. In this section, we introduce the likelihood ratio confidence interval using the profile likelihood function. Here, we first consider the LR statistic:

$$-2 \times \text{profile log-likelihood ratio test.}$$

This has an approximate chi-square distributed with one degree of freedom under the null hypothesis, $H_0 : \phi = \phi_0$, where ϕ_0 is a given value of the log-risk ratio. The sets of overall parameter space and parameter space under H_0 are $\Omega = \{\phi : -\infty < \phi < \infty\}$ and $\omega = \{\phi : \phi = \phi_0\}$, respectively. The profile likelihood ratio statistic is then defined by

$$\Lambda = -2 \log \left(\frac{\sup_{\phi \in \omega} L^*(\phi)}{\sup_{\phi \in \Omega} L^*(\phi)} \right) = -2 \log \left(\frac{L^*(\tilde{\phi})}{L^*(\hat{\phi}_{PL})} \right).$$

This can be also re-written as

$$\Lambda = -2 \left[\log L^*(\tilde{\phi}) - \log L^*(\hat{\phi}_{PL}) \right], \tag{11}$$

where

$$\log L^*(\hat{\phi}_{PL}) = \hat{\phi}_{PL} \sum_{i=1}^{k} x_{i1} - \sum_{i=1}^{k} x_i \log(n_{i0} + n_{i1} e^{\hat{\phi}_{PL}})$$

is the profile log-likelihood function under Ω and

$$\log L^*(\tilde{\phi}) = \phi_0 \sum_{i=1}^{k} x_{i1} - \sum_{i=1}^{k} x_i \log(n_{i0} + n_{i1} e^{\phi_0})$$

is the profile log-likelihood function under H_0. The likelihood ratio statistic given in (11) is an approximately chi-square distribution with one degree of freedom (or $\chi^2_{df=1}$).

In a simple way, the following probability statement is applied to find the lower and upper limits for ϕ. We have that

$$P\left(\chi^2_{\alpha/2,df=1} \leq \Lambda \leq \chi^2_{1-\alpha/2,df=1}\right) = 1-\alpha.$$

Replacing (11) in the above statement, the lower and upper limits for ϕ would be satisfied the two following expressions, respectively,

$$\phi_L = \frac{-0.5 \times \chi^2_{\alpha/2,df=1} + \sum_{i=1}^{k} x_i \log(n_{i0} + n_{i1}e^{\phi_L}) + \hat{\phi}_{PL}\sum_{i=1}^{k} x_{i1} - \hat{A}}{\sum_{i=1}^{k} x_{i1}} \tag{12}$$

and

$$\phi_U = \frac{-0.5 \times \chi^2_{1-\alpha/2,df=1} + \sum_{i=1}^{k} x_i \log(n_{i0} + n_{i1}e^{\phi_U}) + \hat{\phi}_{PL}\sum_{i=1}^{k} x_{i1} - \hat{A}}{\sum_{i=1}^{k} x_{i1}}, \tag{13}$$

where $\hat{A} = \sum_{i=1}^{k} x_i \log(n_{i0} + n_{i1}e^{\hat{\phi}}_{PL})$. $\chi^2_{\alpha/2,df=1}$ and $\chi^2_{1-\alpha/2,df=1}$ are the $(\alpha/2)$100th and $(1-\alpha/2)$100th percentiles of a chi-square distribution with one degree of freedom, respectively. As can be seen from the above expressions, using (12) and (13) may be arduous and hardly control a given confidence probability level, since there has no closed-form solution exists for the interval estimator. Another method which is still based on the likelihood ratio is then applied instead. According to the idea from Doganaksoy (2021), the confidence interval is defined as the minimum and maximum values of a parameter that satisfy a set value of the log-likelihood. So, we use the profile log-likelihood given in (9), and have that

$$\log L^*(\tilde{\phi}) = \kappa, \tag{14}$$

where κ is a constant once data are observed, and κ is given by

$$\kappa = \log L^*(\hat{\phi}_{PL}) - \frac{1}{2}\chi^2_{1-\alpha,df=1}.$$

Therefore, the two endpoints of interval which related to a $(1-\alpha)100\%$ confidence interval for ϕ are obtained by solving (14). Here, we denote it as CI_{P2}. In our computation, the `gosolnp` package in the `Rsolnp` function (Ghalanos and Theussl, 2012) of `R` statistical software is used to calculate the lower and upper limits of CI_{P2}.

4. Simulation Study

A simulation study based on several situations is performed to investigate the performance of the confidence intervals given in the previous section. We generate the data for number of events in the treatment and control groups from two Poisson distributions, where $X_{i1} \sim Po(n_{i1}p_{i1})$ and $X_{i0} \sim Po(n_{i0}p_{i0})$. Number of studies in meta-analysis (k) are 5, 10, 20, and 30, representing small to large study sizes. The sample sizes in two groups are sampled from two uniform distributions: $n_{i1} \sim U(50, 150)$ and $n_{i0} = n_{i1} \times U(r-0.1, r+0.1)$, where the degree of unbalance size (r) is given by 0.5. The log-risk ratio parameter $\phi = \log(p_{i1}/p_{i0})$ is considered in two scenarios: 1) no difference in risk ($\phi = 0$) and an increased risk in the treatment group ($\phi > 0$) and 2) a reduced risk in the treatment group ($\phi < 0$). We set the baseline risk parameter (p_0) as 0.01, 0.05, and 0.10. Hence, the risk in treatment group can be computed by $p_{i1} = p_{i0}e^{\phi}$. Notice that the event occurrence probabilities under these settings are given to be low value, reflecting rare events situations. All parameter settings grouped by the situation of rare events are summarized in Table 2. Then, the 95% confidence intervals for ϕ obtained from the four methods, CI_{E1}, CI_{E2}, CI_{P1}, and CI_{P2}, are computed. Only CI_{E1} requires the continuity correction when single- or double-zero events are observed. Here, we add the value of 0.5 to zero cells in single- and double-zero-event studies, as it is usually done in practice. So that CI_{E1} can be computed and compared to the other interval estimators.

Table 2 Parameter settings in simulation study

Situation	p_{i0}	Scenario 1		Scenario 2	
		ϕ	p_{i1}	ϕ	p_{i1}
Extremely rare event	0.01	0	0.0100	-2	0.0014
		0.5	0.0165	-1.5	0.0022
		1.5	0.0448	-0.5	0.0061
Very rare event	0.05	0	0.0500	-2	0.0068
		0.5	0.0824	-1.5	0.0112
		1.5	0.2241	-0.5	0.0303
Rare event	0.10	0	0.1000	-2	0.0135
		0.5	0.1649	-1.5	0.0223
		1.5	0.4482	-0.5	0.0607

Each situation given in a simulation is repeated 5,000 times. These are implemented in the open-source R statistical platform (R Core Team, 2022). The performance of the confidence interval for the log-risk ratio is evaluated in terms of coverage probability (CP) and expected length (EL) on handling meta-analysis with rare events. In computation by averaging over 5,000 replications, the assessment criteria are approximated by

$$\text{CP} = \frac{\#(L \leq \phi \leq U)}{5,000} \quad \text{and} \quad \text{EL} = \frac{\sum_{h=1}^{5,000}(U_h - L_h)}{5,000},$$

where L and U are the lower and upper limits of parameter ϕ, $\#(L \leq \phi \leq U)$ denotes number of times that ϕ is in between the limits, and h is the h-th iteration. In the decision, the confidence interval which has a coverage probability close to the nominal level of 0.95 with a short expected length is preferred. All this is just another way to say that the confidence interval satisfied this criterion means it can precisely estimate the true parameter.

The coverage probability and interval length of the confidence intervals from simulations under Scenario 1 are shown in Table 3 and Figure 2. It can be seen that the method uses zero-cell corrections or CI_{E1} has low performance in terms of coverage probability in many situations. It cannot control the confidence level to cover the parameter ϕ. Especially, if $\phi \geq 0.5$, coverage probabilities of CI_{E1} are too small, as they are much lower than the target probability level at 0.95. We say that CI_{E1} cannot be used to assess the certainty which a log-risk ratio parameter is in a particular range. Then, we consider the performance of CI_{E2} derived based on MH estimation. CI_{E2} works well to estimate ϕ than CI_{E1}. Coverage probabilities of CI_{E2} are satisfied the target level in many cases when $p_{i0} < 0.1$. However, if p_{i0} is increased, coverage probabilities of CI_{E2} decrease and are lower than 0.95. The latter point is similar to the result obtained from CI_{E1}. Next, the coverage probability of the two methods based on the profile likelihood proposed in this paper is described. CI_{P1} and CI_{P2} have coverage probabilities close to or greater than the nominal coverage probability level at 0.95 in general cases. It seems the coverage probabilities of these two confidence intervals do not depend on ϕ or k. Hence, CI_{P1} and CI_{P2} are more consistent than CI_{E1} and CI_{E2}. Comparing the two proposed interval estimators, CI_{P2} has the simulated coverage probability closer to 0.95 than CI_{P1}. Since CI_{E1} is unsatisfied as noted before, only the coverage probabilities of CI_{E2}, CI_{P1}, and CI_{P2} are displayed by graph, see Figure 2 (a). Clarity, we summarize the best performance of estimators by ordering as CI_{P2}, CI_{P1}, and CI_{E2}. Figure 2 (b) shows the expected lengths of CI_{E2}, CI_{P1}, and CI_{P2}. It can be seen that they are slightly different.

Table 3 Simulated coverage probability and expected length of the four 95% confidence intervals (CIs) for the log-risk ratio ϕ under Scenario 1

			Coverage probability				Expected length			
p_{i0}	ϕ	k	CI_{E1}	CI_{E2}	CI_{P1}	CI_{P2}	CI_{E1}	CI_{E2}	CI_{P1}	CI_{P2}
0.01	0	5	0.9922	0.9848	0.9848	0.9602	2.6927	3.5130	3.5259	3.8558
		10	0.9844	0.9784	0.9784	0.9556	1.8970	2.3605	2.3701	2.4645
		20	0.9722	0.9622	0.9622	0.9528	1.3357	1.5785	1.5855	1.6094
		30	0.9378	0.9632	0.9632	0.9592	1.0911	1.2712	1.2765	1.2889
	0.5	5	0.9678	0.9660	0.9680	0.9612	2.5448	3.2938	3.3044	3.5904
		10	0.9298	0.9702	0.9712	0.9544	1.7757	2.1798	2.1922	2.2713
		20	0.8342	0.9674	0.9686	0.9514	1.2575	1.4914	1.4991	1.5206
		30	0.6888	0.9440	0.9450	0.9490	1.0194	1.1760	1.1827	1.1929
	1.5	5	0.8934	0.9656	0.9686	0.9666	2.2306	3.0162	3.0345	3.2954
		10	0.7634	0.9650	0.9662	0.9554	1.5600	2.0370	2.0464	2.1212
		20	0.4584	0.9542	0.9586	0.9502	1.0936	1.3493	1.3575	1.3751
		30	0.2428	0.9622	0.9622	0.9526	0.8918	1.0866	1.0935	1.1023
0.05	0	5	0.9742	0.9536	0.9588	0.9541	1.4404	1.3909	1.4239	1.4409
		10	0.9576	0.9408	0.9438	0.9542	0.9994	0.9553	0.9789	0.9841
		20	0.9500	0.9466	0.9530	0.9548	0.7013	0.6680	0.6851	0.6868
		30	0.9394	0.9388	0.9486	0.9542	0.5715	0.5446	0.5584	0.5593
	0.5	5	0.9378	0.9398	0.9494	0.9544	1.2947	1.2737	1.3104	1.3245
		10	0.9372	0.9516	0.9594	0.9506	0.9140	0.8924	0.9179	0.9225
		20	0.8752	0.9430	0.9490	0.9504	0.6361	0.6187	0.6369	0.6383
		30	0.8466	0.9560	0.9624	0.9503	0.5194	0.5045	0.5192	0.5200
	1.5	5	0.9116	0.9376	0.9448	0.9525	1.1694	1.1845	1.2235	1.2364
		10	0.8646	0.9524	0.9580	0.9556	0.8156	0.8200	0.8474	0.8515
		20	0.7744	0.9422	0.9552	0.9505	0.5708	0.5730	0.5923	0.5937
		30	0.6534	0.9368	0.9430	0.9496	0.4621	0.4632	0.4791	0.4798
0.10	0	5	0.9498	0.9446	0.9500	0.9512	0.9606	0.9365	0.9850	0.9904
		10	0.9516	0.9454	0.9540	0.9512	0.6741	0.6551	0.6893	0.6911
		20	0.9410	0.9378	0.9504	0.9500	0.4730	0.4594	0.4837	0.4843
		30	0.9356	0.9426	0.9540	0.9544	0.3854	0.3742	0.3941	0.3944
	0.5	5	0.9446	0.9402	0.9522	0.9526	0.8793	0.8675	0.9192	0.9238
		10	0.9110	0.9360	0.9496	0.9490	0.6121	0.6029	0.6397	0.6414
		20	0.8768	0.9370	0.9578	0.9578	0.4308	0.4238	0.4499	0.4504
		30	0.8412	0.9256	0.9394	0.9494	0.3515	0.3457	0.3671	0.3674
	1.5	5	0.8922	0.9370	0.9502	0.9508	0.7814	0.7844	0.8414	0.8454
		10	0.8686	0.9308	0.9464	0.9500	0.5507	0.5527	0.5927	0.5940
		20	0.7872	0.9306	0.9490	0.9496	0.3840	0.3854	0.4138	0.4143
		30	0.7148	0.9192	0.9470	0.9489	0.3144	0.3156	0.3387	0.3390

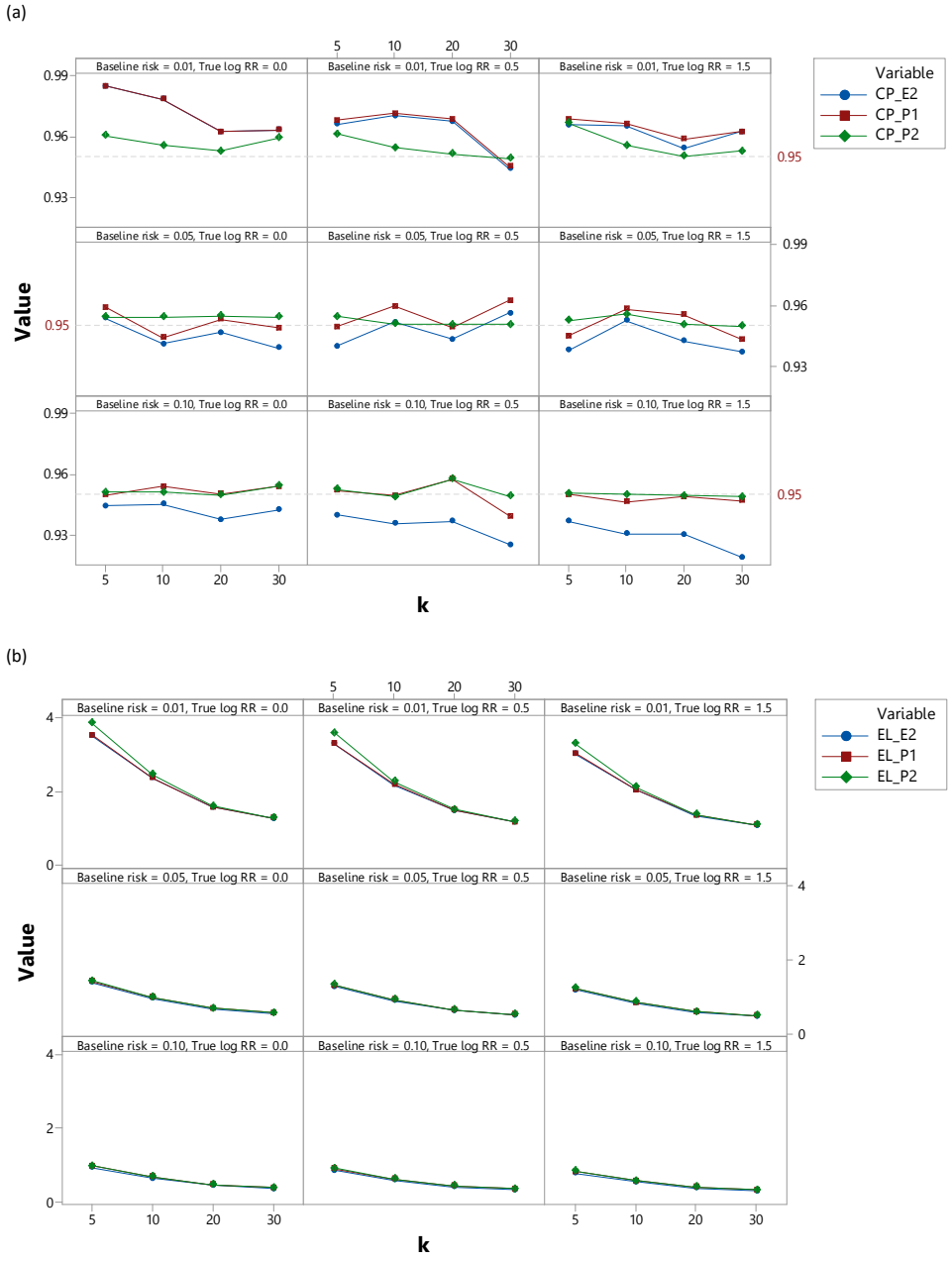

Figure 2 Simulated coverage probability (a) and expected length (b) of the 95% confidence intervals for the log-risk ratio (ϕ), varied on baseline risk parameter (p_{i0}) and number of studies (k) under Scenario 1

Table 4 Simulated coverage probability and expected length of the four 95% confidence intervals (CIs) for the log-risk ratio ϕ under Scenario 2

			Coverage probability				Expected length			
p_{i0}	ϕ	k	CI_{E1}	CI_{E2}	CI_{P1}	CI_{P2}	CI_{E1}	CI_{E2}	CI_{P1}	CI_{P2}
0.01	-2	5	0.9764	0.9912	0.9924	0.9748	3.1208	5.5123	5.5060	6.1977
		10	0.7784	0.9692	0.9708	0.9714	2.1902	4.3790	4.3692	4.8321
		20	0.2566	0.9728	0.9720	0.9656	1.5408	3.0989	3.0957	3.3414
		30	0.0448	0.9700	0.9728	0.9496	1.2571	2.4532	2.4490	2.5690
	-1.5	5	0.9964	0.9956	0.9954	0.9780	3.0568	5.0801	5.0847	5.7458
		10	0.9267	0.9832	0.9836	0.9674	2.1483	3.7377	3.7340	4.1023
		20	0.8928	0.9670	0.9678	0.9539	1.5109	2.4841	2.4855	2.6009
		30	0.7720	0.9656	0.9622	0.9483	1.2328	1.9457	1.9454	1.9845
	-0.5	5	0.9992	0.9966	0.9966	0.9652	2.8611	4.0450	4.0596	4.5451
		10	0.9984	0.9767	0.9695	0.9533	2.0035	2.6469	2.6545	2.7481
		20	0.9976	0.9563	0.9568	0.9488	1.4143	1.7502	1.7564	1.7870
		30	0.9920	0.9585	0.9586	0.9506	1.1535	1.4098	1.4150	1.4304
0.05	-2	5	0.9620	0.9686	0.9680	0.9596	2.2870	2.7253	2.7409	2.8139
		10	0.9202	0.9658	0.9650	0.9532	1.6007	1.8013	1.8122	1.8464
		20	0.8048	0.9652	0.9653	0.9498	1.1248	1.2247	1.2322	1.2438
		30	0.7358	0.9612	0.9656	0.9578	0.9189	0.9958	1.0022	1.0082
	-1.5	5	0.9854	0.9704	0.9712	0.9518	2.0776	2.1614	2.1802	2.2454
		10	0.9357	0.9428	0.9450	0.9539	1.4535	1.4517	1.4653	1.4839
		20	0.9496	0.9426	0.9429	0.9505	1.0222	1.0096	1.0201	1.0261
		30	0.9224	0.9455	0.9458	0.9548	0.8280	0.8127	0.8217	0.8247
	-0.5	5	0.9798	0.9655	0.9604	0.9510	1.6183	1.5394	1.5690	1.5821
		10	0.9826	0.9522	0.9558	0.9502	1.1280	1.0634	1.0849	1.0919
		20	0.9786	0.9424	0.9429	0.9501	0.7904	0.7420	0.7572	0.7595
		30	0.9804	0.9488	0.9456	0.9540	0.6439	0.6037	0.6164	0.6176
0.10	-2	5	0.9540	0.9666	0.9686	0.9497	1.7501	1.8069	1.8304	1.8566
		10	0.9292	0.9588	0.9612	0.9504	1.2173	1.2234	1.2409	1.2528
		20	0.8628	0.9484	0.9492	0.9505	0.8550	0.8488	0.8616	0.8654
		30	0.7830	0.9480	0.9550	0.9500	0.6973	0.6910	0.7012	0.7032
	-1.5	5	0.9642	0.9460	0.9451	0.9539	1.4909	1.4448	1.4760	1.4953
		10	0.9536	0.9582	0.9574	0.9512	1.0381	0.9971	1.0190	1.0252
		20	0.9232	0.9426	0.9540	0.9552	0.7305	0.6987	0.7145	0.7165
		30	0.9146	0.9422	0.9514	0.9507	0.5950	0.5685	0.5814	0.5826
	-0.5	5	0.9660	0.9448	0.9524	0.9478	1.0880	1.0479	1.0921	1.0991
		10	0.9670	0.9404	0.9532	0.9508	0.7562	0.7261	0.7574	0.7597
		20	0.9728	0.9440	0.9534	0.9513	0.5325	0.5107	0.5329	0.5337
		30	0.9650	0.9390	0.9486	0.9480	0.4348	0.4166	0.4346	0.4351

The performance of the four confidence intervals for ϕ based on generated data under Scenario 2 is considered. Table 4 shows that coverage probabilities of CI_{P2} are closer to the target probability level at 0.95 than the other confidence intervals in general cases. The expected lengths of CI_{P2} are slightly greater than those of the comparators. CI_{P1} works well when $p_{i0} > 0.01$. Similarly to the result given in Scenario 1, coverage probabilities of CI_{E1} differ from the nominal coverage probability in many situations. So, CI_{E1} cannot control the confidence level to cover the true parameter for rare events. CI_{E2} has a good performance in terms of coverage probability than CI_{E1}, but its coverage probability is still lower than 0.95 in many cases in the study. The behaviour of CI_{E2}, CI_{P1}, and CI_{P2} are shown by graphs, given in Figure 3.

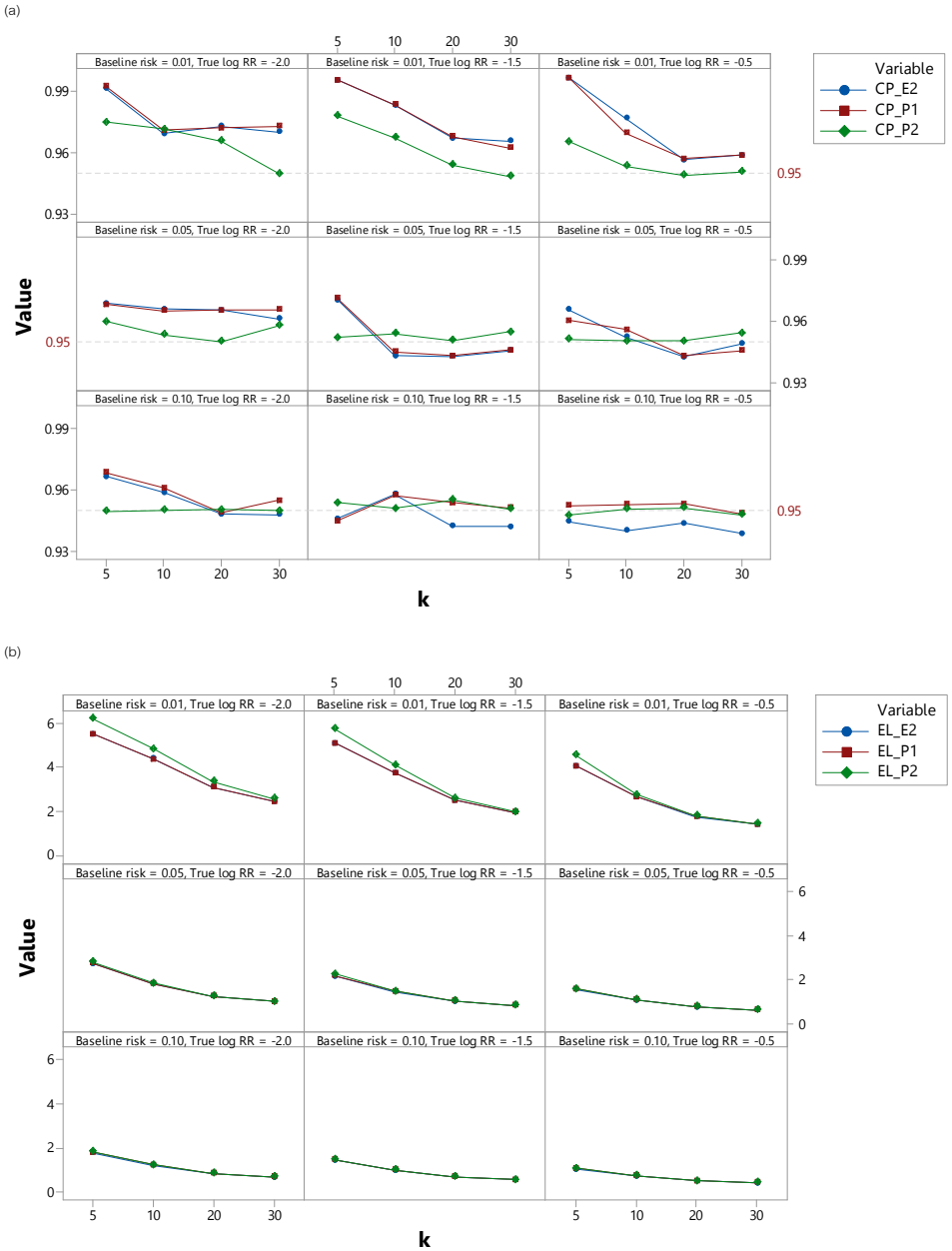

Figure 3 Simulated coverage probability (a) and expected length (b) of the 95% confidence intervals for the log-risk ratio (ϕ), varied on baseline risk parameter (p_{i0}) and number of studies (k) under Scenario 2

As a summary of performance with regard to the coverage probability and length of interval, both scenarios provide the results in the same way. We recommend CI_{P2} as an alternative confidence interval for estimating the log-risk ratio (or risk ratio) in meta-analysis of rare events.

5. Data Application

The data on eye protection for virus transmission as noted in Section 1 were used to illustrate the methods proposed in this paper. They were taken from Chu et al. (2020) and related to the number of MERS-CoV infected patients who used or did not use a barrier to protect infectious virus entering the eye. For this meta-analytic dataset, no statistically significant difference in risk ratios was found across studies (Cochran's Q statistic = 0.43, p-value = 0.81, and I^2 = 0%). Therefore, we did a meta-analysis of associations by pooling risk ratios using the four methods (E1, E2, P1, and P2) derived under the homogeneity assumption. As can be seen from the available data given in Table 1, zero studies are observed in both one and two treatment arms. The traditional IVW method can be used, once the continuity value of 0.5 is added to the studies with zero events before the analysis. Meanwhile, estimation for the risk ratio using the MH and profile likelihood methods can be computed using the original data, without adding any continuity value.

Table 5 Estimated risk ratio and 95% confidence interval using meta-analytic data on the association of eye protection with risk of MERS-CoV transmission

Method	Estimated risk ratio (Log-risk ratio)	95% CI for risk ratio
IVW (E1)	0.2544 (-1.3688)	(0.0665, 0.9724)
MH (E2)	0.1363 (-1.9929)	(0.0204, 0.9064)
Profile likelihood 1 (P1)	0.1199 (-2.1211)	(0.0157, 0.9133)
Profile likelihood 2 (P2)	0.1199 (-2.1211)	(0.0066, 0.5909)

The point estimates and 95% confidence intervals for the risk ratio are shown in Table 5. Since the estimated risk ratios from the existing approaches (E1 and E2) and the proposed methods (P1 and P2) were lower than one, it can be concluded that MERS-CoV virus transmission was lower with eye protection, compared to no eye protection. However, the values of estimated risk ratios from these methods are different. A question arises: what is the estimated value that could be used for this dataset? To answer this point, we refer to the simulation results under Scenario 2. As highlighted in the previous section, the profile likelihood ratio confidence interval (CI_{P2}) can precisely estimate the log-risk ratio in rare events meta-analysis with the suitable coverage probability and expected length. For this application, we therefore interpreted the 95% confidence interval for the risk ratio by CI = (0.0066, 0.5909). The risk in transmission of the virus was 40.91% to 99.34% lower in patients with eye protection compared to patients without eye protection. Note that $\hat{\theta}_{PL} = 0.1199$ and $\log(\hat{\theta}_{PL}) =$ -2.1211.

In the end, Figure 4 shows that our method based on the profile likelihood ratio works well in computation. It can be solved for the lower and upper limits of the log-risk ratio. From the figure, the red-dash lines denote the lower and upper limits, while the black-dash line shows the point estimate of the log-risk ratio for the MERS-CoV data example.

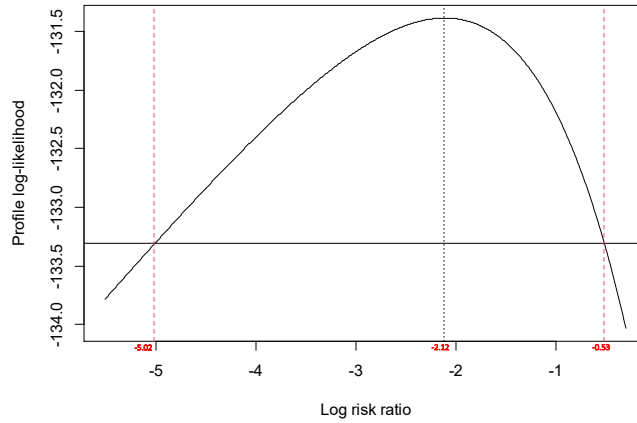

Figure 4 Likelihood ratio plot of 95% profile likelihood ratio confidence interval (CI_{P2}) for the log-risk ratio, using the real-data example

6. Conclusions

In this paper, we develop the two confidence intervals for the risk ratio in meta-analysis for the count outcomes with rare events. The first one is the Wald-type confidence interval using the variance of the profile maximum likelihood estimator derived from the inverse of Fisher information. This extends the idea of constructing the Mantel-Haenzel estimator, which comes from specifying the value of the risk ratio θ on the right-hand side of (7) equal to one. However, in fact, θ can be the positive real value. Moreover, the variance of Mantel-Haenzel estimator derived by Greenland and Robins (1985) uses the random variables which are assumed to have binomial distributions. In the current work, the maximum profile likelihood estimate derived from the profile log-likelihood function for the log-risk ratio is used instated. It does not limit the value of $\theta = 1$. We expect that our approach can be used in wider situations. The second confidence interval is derived based on the likelihood ratio method. Here, the new profile likelihood ratio statistic is proposed. It has an approximate chi-square distribution with one degree of freedom. We use this method because it is noted that the likelihood ratio can be addressed the problem of small data (Agresti, 1990). This could be beneficial for the situation of rare events.

To evaluate the performance of the proposed confidence intervals, an extensive simulation study is conducted. It can be seen that the proposed profile likelihood ratio confidence interval outperforms the confidence intervals based on the Wald and Mantel-Haenzel estimation. In particular, this estimator provides the coverage probability close to the target level with an acceptable short-length interval. It produces consistent estimates, as the performance in terms of coverage probability does not depend on the parameter settings, i.e. number of studies and level of the risk ratio. This means that the profile likelihood ratio confidence interval proposed in this study can be applied in any situation. The Wald-type confidence interval has low performance in estimating the risk ratio in meta-analysis with rare events. This is not a surprise as its method is derived based on the large-sample and normal approximation. So that our work confirms that the Wald method should be avoided for the analysis of rare or zero-study events. The Mantel-Haenzel confidence interval can be used in rare events meta-analysis when the two groups are identical or have a similar risk in very rare or rare events. It is simple to use with the explicit formula. However, we show that its performance is lower than our proposed likelihood ratio confidence interval.

A limitation of our study is that the estimator from the profile likelihood ratio method does not have the explicit formula. Statistical programming is therefore required to find the optimal limits of

the confidence interval. Herein, we suggest the `gosolnp` package in the `R` programming language for solving the problem. It is evidenced by the analysis given in the application section of this paper to ensure convergence to an optimum solution. Finally, it is important to note that the confidence intervals studied in this work are derived based on the homogeneity assumption of the risk ratio. The estimator is then suitable to be used in the fixed effect meta-analysis. However, the effect size estimates can be varied across studies. The random effects model is needed in such a case. In future work, interval estimation on meta-analysis with rare events in heterogeneity situations will be considered and we will take a deeper look into the efficiency of the estimator.

Acknowledgements

My thanks go to the Editor and Referees for the valuable suggestions to improve this paper. The author gratefully acknowledges the financial support provided by Faculty of Science and Technology, Thammasat University, Contract no. SciGR 11/2566.

References

Agresti A. Categorical Data Analysis. New York: John Wiley and Sons; 1990.

Bakbergenuly I, Hoaglin DC,Kulinskaya E. Methods for estimating between-study variance and overall effect in meta-analysis of odds ratios. Res Synth Methods. 2020; 11(3): 426–442.

Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Introduction to Meta-Analysis. Chichester: Wiley; 2009.

Böhning D, Kuhnert R, Rattanasiri S. The Profile Likelihood for Meta-Analytic Studies with Binary Outcome. Boca: Chapman & Hall/CRC; 2008.

Böhning D, Mylona K, Kimber A. Meta-analysis of clinical trials with rare events. Biom J. 2015; 57(4): 633–648.

Böhning D, Sangnawakij P, Holling H. Confidence interval estimation for the Mantel-Haenszel estimator of the risk ratio and risk difference in rare event meta-analysis with emphasis on the bootstrap. J Stat Comput Sim. 2022; 92(6): 1267–1291.

Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. The Lancet. 2020; 395(10242): 1973–1987.

Cooper HM, Hedges LV, Valentine JC. he Handbook of Research Synthesis and Meta-Analysis. New York: Russell Sage Foundation; 2009.

Doganaksoy N. A simplified formulation of likelihood ratio confidence intervals using a novel property. Technometrics. 2021; 63(1): 127–135.

Efthimiou O, Rücker G, Schwarzer G, Higgins JPT, Egger M, Salanti G. Network meta-analysis of rare events using the Mantel-Haenszel method. Stat Med. 2019; 38(16): 2992–3012.

Ghalanos A, Theussl S. Rsolnp: A Package for General Non-Linear Optimization using Augmented Lagrange Multiplier Method. 2012 [cited 2023 Jan 6]; 2(2): 38-48. https://cran.r-project.org/web/packages/Rsolnp/.

Greenland S, Robins JM. Estimation of a common effect parameter from sparse follow-up data. Biometrics. 1985; 41(1): 55–68.

Günhan BK, Röver C, Friede T. Random-effects meta-analysis of few studies involving rare events. Res Synth Methods. 2020; 11(1): 74–90.

Kaul S, Diamond GA. Rosiglitazone and cardiovascular risk: Nissen and Wolski 2010 updated meta-analysis revisited. J Am Coll Cardiol. 2011; 57: 552–E552.

Manmana S, Iamsirithaworn S, Uttayamakul S. Coronavirus disease-19 (COVID-19). J Bamrasnaradura Infect Dis Inst. 2020; 14(2): 124–133.

Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst. 1959; 22(4): 719–748.

Murphy SA, van der Vaart AW. On profile likelihood. J Am Stat Assoc. 2000; 95(450): 449–465.

Noma H, Nagashima K. A note on the Mantel-Haenszel estimators when the common effect assumptions are violated. Epidemiol Methods. 2016; 5(1): 19–35.

Piaget-Rossel R, Taffé P. Meta-analysis of rare events under the assumption of a homogeneous treatment effect. Biom J 2019; 61(6): 1557–1574.

R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2022.

Royston P. Profile likelihood for estimation and confidence intervals. Stata J. 2007; 7(3): 376–387.

Schulze R, Holling H, Böhning D. Meta-Analysis: New Developments and Applications in Medical and Social Sciences. Göttingen: Hogrefe & Huber; 2003.

Stata Corp. Stata Statistical Software, Version 13.1. Texas: College Station; 2013.

Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. Stat Med. 2004; 23(9): 1351–1375.

Wei JJ, Lin EX, Shi JD, Yang K, Hu ZL, Zeng XT, Tong TJ. Meta-analysis with zero-event studies: a comparative study with application to COVID-19 data. Mil Med Res. 2021; 8(41): 1–11.

Zabriskie BN, Corcoran C, Senchaudhuri P. A Comparison of confidence distribution approaches for rare event meta-analysis. Stat Med. 2021; 40(24): 5276–5297.

Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with Pneumonia in China, 2019. N Engl J Med. 2020; 382(8): 727–733.