ARTICLE TEMPLATE

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

ARTICLE HISTORY

Compiled August 9, 2021

ABSTRACT

This paper takes a deeper look into uncertainty assessment of the Mantel-Haenszel estimator. We look at the risk ratio and the risk difference as the parameters of interest. In the homogeneity case, all developed confidence intervals for these parameters behave acceptably, even in the rare events situation. For heterogeneity, we demonstrate that the non-parametric bootstrap approach provides confidence interval estimates for the risk difference with acceptable coverage, depending on the number of studies. For the risk ratio, the situation is more complex as typically distributions for the log-relative risk are considered. The Mantel-Haenszel estimator overestimates the expected value of the distribution of the log-relative risk whatever it may be. However, if we consider as true value the estimand of Mantel-Haenszel estimator, reasonable coverage probabilities can be achieved with the bootstrap. A source for the occurrence of this problem can be seen in the fact that the moments of a non-linearly transformed relative risk variable are not equal to the non-linearly transformed moments of the respective relative risk variable. If the transformation is concave, as a consequence of Jensen's inequality, the mean of the log-relative risk will be at most as large as the logarithm of the mean of relative risk, constituting the overestimation bias of the Mantel-Haenszel estimator. Of course, these issues disappear in the homogeneity case as the relative risk is constant across the studies.

KEYWORDS

Bootstrap; estimand; Mantel-Haenszel estimator; meta-analysis, rare events

1. Introduction and motivation

The paper considers the following situation in meta-analysis. In k independent studies, counts of events are observed in an intervention and control group. This setting can be described by a count random variable or number of events Y_{ij}. The index i indicates the study i, for $i = 1, 2, ..., k$, where k denotes the number of available studies. Also, $j = 1$ denotes an intervention group and $j = 0$ a control group. The mean of Y_{ij} is given by $E(Y_{ij}) = \lambda_{ij}P_{ij}$, where λ_{ij} denotes the event occurrence risk and P_{ij} stands for the person-time at risk which is considered as non-random and reduces to the number at risk n_{ij}, if all members in study i share the same person-time. In many research studies, we are interested often in settings where the probability of no events is large. So that low frequency counts such as 0, 1, or 2 are observed. The Poisson assumption comes into play in this connection, which assumes that

$$Y_{ij} \sim Po(\lambda_{ij}P_{ij}), \tag{1}$$

where $Po(\theta)$ denotes the Poisson distribution with the density $e^{-\theta}\theta^{y_{ij}}/y_{ij}!$ for count $y_{ij} = 0, 1, 2, ...$ Under assumption (1), the variance of Y_{ij} is given as $\mathrm{Var}(Y_{ij}) = \lambda_{ij}P_{ij}$. We will make occasionally use of (1), but will always say when we do so.

Let us first illustrate the setting with an example as follows. The data on a systematic review of the effectiveness of prophylactic antibiotic treatment on infectious complications in women undergoing caesarean delivery are used. These were originally published by Smaill and Hofmeyr [1], and mentioned again in Cooper et al. [2]. The data include 61 studies with counts of occurrence of wound infection as outcome in women undergoing caesarean delivery. The intervention group uses prophylactic antibiotics, whereas the control group is placebo or no prophylactic antibiotics. The data show sample sizes with an average of 80 persons per trial in the treatment group and 63 persons per trial in the control. The occurrence of wound infection is observed relatively rarely. Many of the component studies are small in size and there are zero events in each of two arms, especially in the treatment. The entire dataset can be found in the appendix. The forest plot of the estimated risk ratios obtained from the

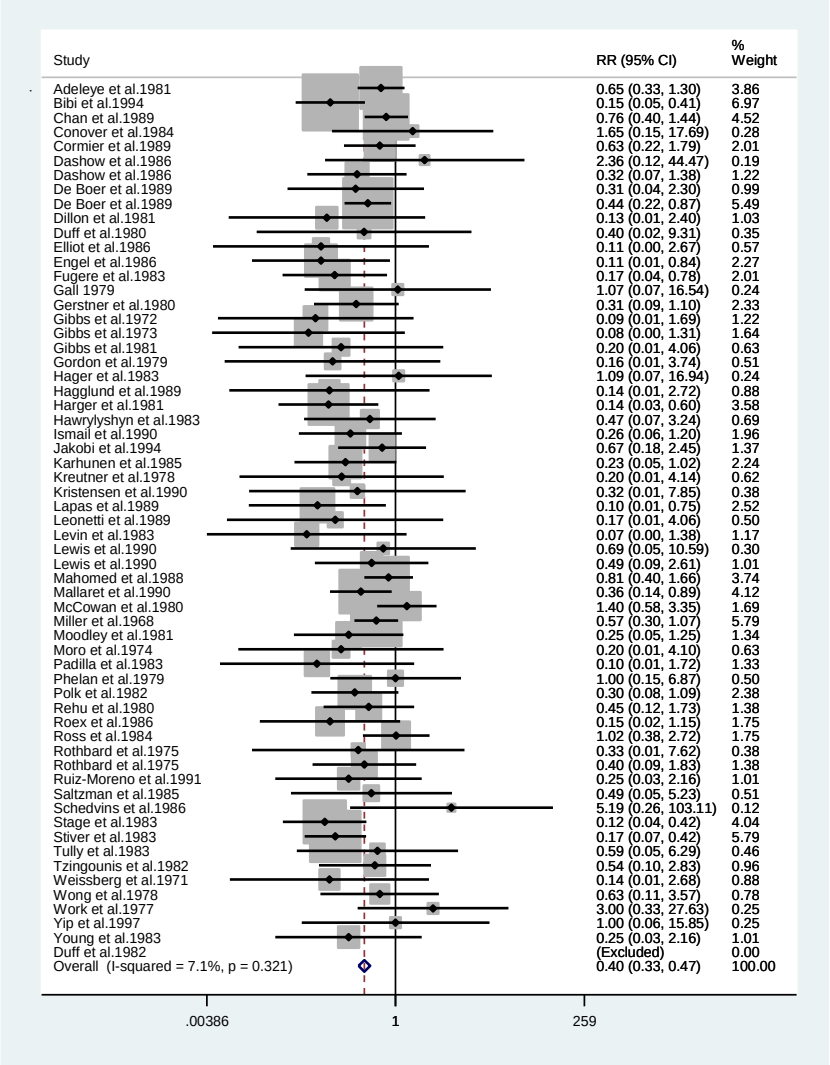

Figure 1. Forest plot of meta-analytic data on the prophylactic antibiotics in caesarean section.

METAN package in Stata [3] is shown in Figure 1. In this case, we point out that if the inverse variance-weighted average method [4] (the method that requires the study-specific estimates and its standard error) is used, in fact, the double-zero or single-zero study would have to be excluded before the analysis because the risk ratio estimate is undefined. Thus, some available studies will be ignored. When a statistical package is applied, zero cell frequencies will be replaced by the value 0.5, also called a continuity correction. This shows a disadvantage of the traditional method, as it can introduce bias in estimation [5].

To address these problems, several statistical methods have been introduced to estimate the effect parameters. Examples are the maximum likelihood method, the Peto method, the median unbiased estimator method, and regression analysis in generalized linear models [6]. Our interest in this paper is the Mantel-Haenszel (MH) approach [7,8] which is widely used in applications, such as in epidemiologic and clinical studies. For a more general reference, see Jewell [9] and Landis et al. [10]. The MH estimator uses a weighting procedure different from the inverse variance-weighted average method. Furthermore, it has a number of benefits as will be outlined below. For estimating the risk ratio (RR), the estimator from the MH method is given by

$$\widehat{RR}_{MH} = \frac{\sum_i Y_{i1} P_{i0}/P_i}{\sum_i Y_{i0} P_{i1}/P_i} = \frac{V}{W}, \tag{2}$$

where $P_i = P_{i0} + P_{i1}$, for $i = 1, 2, ..., k$. We note that the estimand is the risk ratio[1]. A careful definition of the estimand is needed. If $\lambda_{i0} = \lambda_0$ and $\lambda_{i1} = \lambda_1$ for all i, then the risk ratio is simply λ_1/λ_0. However, if λ_{i0} arise from a distribution with mean Λ_0 and, potentially, also λ_{i1} arise from a distribution with mean Λ_1, then we can define the risk ratio as Λ_1/Λ_0. We leave it at this point not further specified which form these distributions will have, but will return to this issue later-on. If both distributions concentrate all mass on one point λ_1 and λ_0, respectively, the risk ratio is λ_1/λ_0, and we are in the situation of *homogeneity*. We are mentioning this here to be clear what *the estimand* is when we are in the general situation of non-homogeneity

[1]Some authors carefully distinguish between the *risk ratio*, when the ratio considered relates to risks, and the *rate ratio*, when the ratio considered relates to rates. Here we uniquely speak of risk ratios even though the ratio involves rates, e.g. person-times.

or *heterogeneity* [4]. Note that the MH estimator eliminates the baseline parameters and concentrates on estimating an average risk ratio.

The MH estimator has several advantages:

- it is not sensitive towards the occurrence of zero counts and is always defined unless only zeros occur in one group,
- it is in the form of a ratio of sums (and not a sum of ratios) and we will exploit this nature in the following,
- it is quite closely related to profile maximum likelihood estimation [11].

We like to point out here that, in meta-analysis in general, one can investigate the behavior of an estimator with respect to the sample size within a study. Another investigation could be related to the number of studies. We are in meta-analysis where the number of studies is rather small, typically below 100, often much lower than that. Our interest is here on rare events studies. Hence our work can be seen as a study in the small sample and small event behavior and we will focus in particular on the bootstrap approach for these reasons. However, we like to mention that little work has been done on the asymptotic behavior of the MH estimator such as by Noma and Nagashima [12], who look also at the case when the common effect assumption (homogeneity of effect) is violated. Furthermore, it is reported in Bakbergenuly and Kulinskaya [13] and Bakbergenuly et al. [14] that the MH estimator experiences bias under effect heterogeneity.

In applications of meta-analysis using count outcome data, the MH method has become the favourite tool. See, for example, [15–21]. But also in theoretical developments the MH method has been considered, for example in network meta-analysis [22–24]. This evidence indicates that the MH procedure is a standard method widely used in meta-analysis. We therefore take here an in-depth look at the uncertainty assessment for this estimator.

2. Variance and confidence interval estimation of the Mantel-Haenszel estimator of the relative risk

2.1. Formula-based method

In the following, we give a brief review of available methods for variance estimation of the MH estimator of the relative risk and, often associated with this, confidence interval estimation. One of the benefits of the ratio structure of the MH estimator presented in (2) is that the variances of V and W are easily available:

$$\sigma_V^2 = \mathrm{Var}(V) = \sum_i \frac{P_{i0}^2}{P_i^2}\mathrm{Var}(Y_{i1}) = \sum_i \frac{P_{i0}^2}{P_i^2}\lambda_{i1}P_{i1}, \tag{3}$$

where we have used the Poisson assumption (1) at the second equality in (3) which we will also keep for the remainder of this section. The latter can be estimated by $\hat{\sigma}_V^2 = \sum_i \frac{P_{i0}^2}{P_i^2}Y_{i1}$. Similarly, we can estimate $\sigma_W^2 = \mathrm{Var}(W)$ by $\hat{\sigma}_W^2 = \sum_i \frac{P_{i1}^2}{P_i^2}Y_{i0}$. Note that both V and W are not sensitive to the occurrence of zeros.

It is now straightforward to use the *delta method* to derive the variance of $\log \widehat{RR}_{MH}$, the natural logarithm of the MH estimator of relative risk. The delta method provides

$$\mathrm{Var}(\log V - \log W) \approx \frac{\sigma_V^2}{[E(V)]^2} + \frac{\sigma_W^2}{[E(W)]^2}, \tag{4}$$

which can be estimated by

$$\frac{\sum_i \frac{P_{i0}^2}{P_i^2}Y_{i1}}{V^2} + \frac{\sum_i \frac{P_{i1}^2}{P_i^2}Y_{i0}}{W^2}. \tag{5}$$

From (5) assuming approximate normality, it is easy to derive an approximate two-sided $(1-\alpha)100\%$ confidence interval for the log-relative risk using the Wald-type method, namely

$$\log(V/W) \pm z\sqrt{\frac{\sum_i \frac{P_{i0}^2}{P_i^2}Y_{i1}}{V^2} + \frac{\sum_i \frac{P_{i1}^2}{P_i^2}Y_{i0}}{W^2}}, \tag{6}$$

where $z = \Phi^{-1}(1-\alpha/2)$ with $\Phi(\cdot)$ being the cumulative distribution function of the standard normal distribution and α is the significance level. Taking anti-log's of (6) will provide a confidence interval for the true parameter RR.

Another approach is in the spirit of *Fieller's method* [25] to construct the confidence interval for the relative risk. Based on this approach, we rewrite the ratio V/W as a linear combination of V and W. In doing so, we look at

$$Z = V - RR \times W$$

and note that $E(Z) = 0$, as $RR = E(V)/E(W)$, and $\mathrm{Var}(Z) = \sigma_V^2 + RR^2\sigma_W^2$, as V and W are independent. Hence, an approximating standard normal distribution is considered for $Z/\sqrt{\mathrm{Var}(Z)}$. As consequence, a confidence interval can be constructed via

$$P\left(-z \leq \frac{V - RR \times W}{\sqrt{\sigma_V^2 + RR^2\sigma_W^2}} \leq z\right) \approx 1-\alpha. \tag{7}$$

The inequalities inside the brackets of the probability statement can be equivalently written as

$$-z\sqrt{\sigma_V^2 + RR^2\sigma_W^2} \leq V - RR \times W \leq z\sqrt{\sigma_V^2 + RR^2\sigma_W^2}, \tag{8}$$

or

$$\frac{V}{W} - z\frac{\sqrt{\sigma_V^2 + RR^2\sigma_W^2}}{W} \leq RR \leq \frac{V}{W} + z\frac{\sqrt{\sigma_V^2 + RR^2\sigma_W^2}}{W},$$

leading to the final confidence interval by replacing RR with $\widehat{RR}_{MH}$ under the root

$$\widehat{RR}_{MH} \pm z\widehat{RR}_{MH}\sqrt{\hat{\sigma}_V^2/V^2 + \hat{\sigma}_W^2/W^2}. \tag{9}$$

The approach does not directly provide an estimate of the variance of $\widehat{RR}_{MH}$, but

from (9) it seems natural to suggest

$$\widehat{RR}^2_{MH}[\hat{\sigma}^2_V/V^2 + \hat{\sigma}^2_W/W^2] \tag{10}$$

as an estimator of the variance of $\widehat{RR}_{MH}$. To compare (10) with (5), we need to go back to the log-scale. Using again the delta method, we find

$$\text{Var}(\log \widehat{RR}_{MH}) \approx \frac{\text{Var}(\widehat{RR}_{MH})}{\widehat{RR}^2_{MH}} = \frac{\hat{\sigma}^2_V}{V^2} + \frac{\hat{\sigma}^2_W}{W^2}, \tag{11}$$

which is identical to (5). Hence, both approaches are closely related.

In addition, there is a different way to utilize (8). If we divide the inequality chain by W and add RR on all sides the expression, we achieve

$$RR - z\sqrt{\sigma^2_V + RR^2\sigma^2_W}/W \leq V/W \leq RR + z\sqrt{\sigma^2_V + RR^2\sigma^2_W}/W. \tag{12}$$

This can be written as

$$L(RR) \leq \widehat{RR}_{MH} \leq U(RR), \tag{13}$$

where $L(RR) = RR - z\sqrt{\sigma^2_V + RR^2\sigma^2_W}/W$ and $U(RR) = RR + z\sqrt{\sigma^2_V + RR^2\sigma^2_W}/W$. Now, $L(RR)$ and $U(RR)$ are strictly monotone increasing functions of RR, so that inverse functions $L^{-1}(RR)$ and $U^{-1}(RR)$ exist. Hence, (13) can be written equivalently as

$$U^{-1}(\widehat{RR}_{MH}) \leq RR \leq L^{-1}(\widehat{RR}_{MH}), \tag{14}$$

which defines the confidence interval as $(U^{-1}(\widehat{RR}_{MH}), L^{-1}(\widehat{RR}_{MH}))$. This method of construction is sometimes also called the *pivotal method*. In practice, we need to find the two solutions of the quadratic equation in RR

$$W^2(RR - \widehat{RR}_{MH})^2/z^2 = \sigma^2_V + RR^2\sigma^2_W$$

and the two solutions define then the confidence interval for RR.

Typically, in practice the estimate of variance of the MH estimator suggested by Greenland and Robins [26] is used. The estimated variance obtained from *Greenland and Robins's method* is given as

$$\widehat{\mathrm{Var}}(\log \widehat{RR}_{MH}) = \frac{\sum_i P_{i0}P_{i1}Y_i/P_i^2}{VW}, \tag{15}$$

where $Y_i = Y_{i0} + Y_{i1}$ and $P_i = P_{i0} + P_{i1}$. Greenland and Robins pointed out that the formula will be only valid under homogeneity of the effect parameter:

> All the variance formulas presented here were derived under the assumption of homogeneity across strata of the effect parameter.

Another approach is utilizing the simple form of the *profile likelihood* function in this case. The model, allowing for baseline heterogeneity, is given as

$$\log E(Y_{ij}) = \log P_{ij} + \alpha_i + \beta \times j, \tag{16}$$

for $i = 1, 2, ..., k$ and $j = 0, 1$, the latter indicating control or intervention. The coefficient $\beta = \log RR$ is the log-relative risk. The log-likelihood is then provided as (up to an additive constant not involving the parameters)

$$\log L = -\sum_i e^{\alpha_i} P_{i0}(1 + R_i e^{\beta}) + \sum_i Y_i \alpha_i + \sum_i Y_{i1}\beta, \tag{17}$$

where $R_i = P_{i1}/P_{i0}$. The log-likelihood (17) is maximized for fixed the parameter of interest β by $e^{\hat{\alpha}_i} = Y_i/[P_{i0}(1 + R_i e^{\beta})]$. We achieve the profile log-likelihood function

$$\ell(\beta) = \sum_i Y_{i1}\beta - \sum_i Y_i \log(1 + R_i e^{\beta}). \tag{18}$$

The profile likelihood needs to be maximized numerically, but in the case of balanced studies ($R_i = 1$) there is a closed-form solution which corresponds to the MH estimator. The profile log-likelihood can now be used to derive a variance estimate by means of the

estimated Fisher information, which is given as the negative of the second derivative of (18):

$$\hat{I}(\beta) = \sum_i Y_i \frac{R_i e^{\beta}}{(1 + R_i e^{\beta})^2}. \tag{19}$$

We now proceed as follows. As we expect that the MH estimator and the profile maximum likelihood estimator are close as point out in [11], we then suggest to use

$$\hat{I}(\log \widehat{RR}_{MH})^{-1} = \left(\sum_i Y_i \frac{R_i \widehat{RR}_{MH}}{(1 + R_i \widehat{RR}_{MH})^2} \right)^{-1} \tag{20}$$

as an estimate for the variance of $\log \widehat{RR}_{MH}$. All the variance estimators for the risk ratio provided here are closely related as the following result shows, which is straightforward to prove.

Theorem 2.1. *If all studies are balanced, i.e.* $P_{i1} = P_{i0}$ *for all studies* $i = 1, 2, ..., k$, *then* $(5) = (15) = (20) = 1/\sum_i Y_{i1} + 1/\sum_i Y_{i0}$.

Another method of confidence interval construction is the *method of variance estimates recovery* (MOVER). For a brief overview see Newcombe [27] or Donner and Zhou [28]. The method works for a parameter difference $\theta = \theta_1 - \theta_2$ as follows. Let L_i be the lower confidence interval and U_i the upper confidence interval limit for θ_i $(i = 1, 2)$ at level $1 - \alpha$, then $L = \hat{\theta} - \sqrt{(\hat{\theta}_1 - L_1)^2 + (U_2 - \hat{\theta}_2)^2}$ and $U = \hat{\theta} + \sqrt{(\hat{\theta}_1 - U_1)^2 + (L_2 - \hat{\theta}_2)^2}$ provide a two-sided $(1 - \alpha)100\%$ confidence interval for θ. Here, $\hat{\theta}_i$ are estimates of θ_i for $i = 1, 2$ and $\hat{\theta} = \hat{\theta}_1 - \hat{\theta}_2$ is the estimator of θ. To apply the method for our case, we use $\hat{\theta}_1 = \log V$ and $\hat{\theta}_2 = \log W$ as we can safely assume that both are non-zero. If we assume that L_1 and U_1 are constructed as

$$\log V \pm z \sqrt{\frac{\sum_i \frac{P_{i0}^2}{P_i^2} Y_{i1}}{V^2}}$$

(see also (6)) and that L_2 and U_2 are constructed as

$$\log W \pm z\sqrt{\frac{\sum_i \frac{P_{i1}^2}{P_i^2} Y_{i0}}{W^2}}.$$

Then, the MOVER interval for the log-relative risk corresponds to

$$\log \widehat{RR}_{MH} \pm z\sqrt{\frac{\sum_i \frac{P_{i0}^2}{P_i^2} Y_{i1}}{V^2} + \frac{\sum_i \frac{P_{i1}^2}{P_i^2} Y_{i0}}{W^2}},$$

which is the confidence interval we obtain on the basis of (6). Of course, if the construction of the individual limits differs from the above construction, the MOVER method will lead to a different interval.

2.2. Bootstrap method

A further way of constructing confidence intervals for the log-relative risk uses the bootstrap. This method is a simple, but powerful statistical tool firstly described by Efron [29]. It is also widely used in applications. In practice, we take B, say B = 1,000, samples $(Y_{i1}^*, Y_{i0}^*, P_{i1}^*, P_{i0}^*)$ with replacement of size k from the original $(Y_{i1}, Y_{i0}, P_{i1}, P_{i0})$. For each of these bootstrap samples, the estimator $\log \widehat{RR}_{MH_b}^*$ is determined, where $b = 1, 2, ..., B$. Then, a $(1-\alpha)100\%$ *bootstrap percentile* confidence interval is calculated by

$$\left(\log \widehat{RR}_{MH}^*(\alpha/2), \log \widehat{RR}_{MH}^*(1-\alpha/2)\right),$$

where $\log \widehat{RR}_{MH}^*(\alpha/2)$ is the $(\alpha/2)$th percentile of the bootstrap estimates and α is the significant level. Furthermore, a $(1-\alpha)100\%$ *bootstrap variance* confidence interval for the log-relative risk is given by

$$\log \widehat{RR}_{MH}^* \pm z\sqrt{\mathrm{Var}(\log \widehat{RR}_{MH}^*)},$$

where $\log \widehat{RR}^*_{MH} = \sum_{b=1}^{B} \log \widehat{RR}^*_{MH_b}/B$ and $\mathrm{Var}(\log \widehat{RR}^*_{MH}) = \sum_{b=1}^{B}(\log \widehat{RR}^*_{MH_b} - \log \widehat{RR}^*_{MH})/B$ is the bootstrap variance. The confidence interval for the relative risk is then found by taking the anti-log.

We also look at the *bias corrected and accelerated percentile interval* (BC_a) [30,31]. It is a distribution-free method used to construct bootstrap confidence intervals. This is similar to the percentile bootstrap method as noted before, but corrects for bias and skewness in the distribution of bootstrap estimates. The idea of BC_a interval uses percentiles of bootstrap distribution depended on estimating the two parameters: acceleration parameter, namely $\hat{a}$, and bias-correction coefficient, namely $\hat{z}_0$ [32]. Considering the log-relative risk, a $(1-\alpha)100\%$ BC_a confidence interval is given by

$$\left(\log \widehat{RR}^*_{MH}(\alpha_1), \log \widehat{RR}^*_{MH}(\alpha_2)\right),$$

where $\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{\alpha/2}}{1-\hat{a}(\hat{z}_0+z_{\alpha/2})}\right)$, $\alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{1-\alpha/2}}{1-\hat{a}(\hat{z}_0+z_{1-\alpha/2})}\right)$, and $z_{\alpha/2}$ is the $(\alpha/2)100$th percentile of the standard normal distribution. The bias-correction factor is computed from

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#(\log \widehat{RR}^*_{MH_b} < \log \widehat{RR}_{MH})}{B}\right),$$

where $\#(\log \widehat{RR}^*_{MH_b} < \log \widehat{RR}_{MH})$ is the number of bootstrap replications that provide $\log \widehat{RR}^*_{MH_b}$ less than the MH log-relative risk of the original sample. Furthermore, the acceleration is given by

$$\hat{a} = \frac{\sum_{i=1}^{k}(\log \widehat{RR}_{MH(.)} - \log \widehat{RR}_{MH_i})^3}{6\left[\sum_{i=1}^{k}(\log \widehat{RR}_{MH(.)} - \log \widehat{RR}_{MH_i})^2\right]^{3/2}},$$

where $\widehat{RR}_{MH_i}$ is the mean estimate that excludes the i-th data point (jackknife estimate) and $\widehat{RR}_{MH(.)}$ is the mean of $\widehat{RR}_{MH_i}$ [33,34]. In computation, the BC_a interval is simply estimated using the `boot` or `bootstrap` packages of `R` (https://www.r-project.org/).

3. Variance estimation for the Mantel-Haenszel estimator of the risk difference

The risk difference (RD) is one of effect sizes alternatively used in meta-analysis of count outcomes. However, and we point this out right at the beginning, the risk difference is not invariant w.r.t. change of duration across studies. To illustrate we consider a setting with event risk of 0.1 in the intervention group and 0.05 in the control group. In a unit time, with 100 persons at risk we would expect 10 events in the intervention and 5 events in the control group, leading to a risk difference of 0.05. If duration time doubles, we would expect 20 events in the intervention group and 10 in the control group giving a risk difference of 0.1. Hence, the risk difference appears heterogeneous although this is clearly not the case. This problem does not occur with the risk ratio and might be one of the reasons why the latter is more popular in practice. So, when working with the risk difference it seems sensible to assume that all studies have comparable duration and we drop the concept of person-time.

In this section, we assume that Y_{ij} are independent binomial distributions with sample sizes n_{ij} and success or event probabilities λ_{ij}, for $i = 1, 2, ..., k$ and $j = 0, 1$. Again, $j = 1$ represents the intervention group and $j = 0$ the control group. The probability mass function of Y_{ij} is given as

$$\binom{n_{ij}}{y_{ij}} \lambda_{ij}^{y_{ij}} (1 - \lambda_{ij})^{n_{ij} - y_{ij}},$$

where count $y_{ij} = 0, 1, ..., n_{ij}$. The mean and variance of Y_{ij} are $E(Y_{ij}) = n_{ij}\lambda_{ij}$ and $\text{Var}(Y_{ij}) = n_{ij}\lambda_{ij}(1 - \lambda_{ij})$, respectively.

For a given study i, Y_{i1} and Y_{i0} denote the number of persons with an event out of n_{i1} in the intervention and n_{i0} in the control group, respectively. In addition, there are the associated risks λ_{i1} and λ_{i0}. The true effect for the risk difference is therefore denoted as $RD = \Lambda_1 - \Lambda_0$, where $-1 < RD < 1$. In general, RD can be homogeneous or heterogeneous across studies. To capture heterogeneity, we denote with Λ_1 the expected value of the distribution leading to unobserved values $\lambda_{11}, \lambda_{21}, ..., \lambda_{k1}$, similarly for Λ_0. If both distributions concentrate all mass on one point λ_1 and λ_0, respectively,

$RD = \lambda_1 - \lambda_0$, and we are in the situation of homogeneity [35].

To estimate the parameter from sparse data or low frequency counts, the MH risk difference estimator is suggested, and given as

$$\widehat{RD}_{MH} = \frac{\sum_i (Y_{i1} n_{i0} - Y_{i0} n_{i1})/n_i}{\sum_i n_{i1} n_{i0}/n_i} = \frac{\sum_i (Y_{i1} n_{i0} - Y_{i0} n_{i1})/n_i}{\sum_i w_i}, \tag{21}$$

where $w_i = n_{i1} n_{i0}/n_i$ are the MH weights and $n_i = n_{i1} + n_{i0}$ [7,36]. This estimator enjoys the same properties as the MH estimator of the risk ratio. $\widehat{RD}_{MH}$ is unbiased (not only asymptotically unbiased) as

$$E(\widehat{RD}_{MH}) = \frac{\sum_i (\Lambda_1 n_{i0} n_{i0} - \Lambda_0 n_{i0} n_{i1})/n_i}{\sum_i n_{i1} n_{i0}/n_i} = \Lambda_1 - \Lambda_0.$$

Furthermore, an advantage of this estimator is that it does not rely on the study-specific effect size estimates. $\widehat{RD}_{MH}$ is therefore a suitable estimator for meta-analytic settings where double-zero or single-zero studies occur, unless all studies are zero in one arm [6].

In practice, the variance estimator for $\widehat{RD}_{MH}$ based on *Greenland and Robins's method* [26] is applied. Following this method, two sample proportions of λ_{i1} and λ_{i0}, estimated by $\hat{\lambda}_{i1} = Y_{i1}/n_{i1}$ and $\hat{\lambda}_{i0} = Y_{i0}/n_{i0}$, are used. Taking the variance of (21), the estimated variance according to Greenland and Robins ($\widehat{\mathrm{Var}}_{gr}(\widehat{RD}_{MH})$) is therefore given as

$$\widehat{\mathrm{Var}}_{gr}(\widehat{RD}_{MH}) = \frac{\sum_i w_i^2 [Y_{i1}(n_{i1} - Y_{i1})/n_{i1}^3 + Y_{i0}(n_{i0} - Y_{i0})/n_{i0}^3]}{(\sum_i n_{i1} n_{i0}/n_i)^2}, \tag{22}$$

where $w_i = n_{i1} n_{i0}/n_i$. However, the Greenland and Robins's variance estimator is only consistent when sample size n_i for each component study becomes large [37]. An alternative method for estimating the variance of $\widehat{RD}_{MH}$ has been suggested by Sato [38]. The variance estimator obtained from the latter is noted to be consistent when the number of studies become large and while the within-study data might remain sparse [39].

Based on the method of Sato, the construction of the variance of $\widehat{RD}_{MH}$ is explained

as follows. Let $\lambda_{i1} = RD + \lambda_{i0}$ and $\lambda_{i0} = \lambda_{i1} - RD$. Since Y_{i1} and Y_{i0} are independent, the variance of the numerator of (21) is given by

$$\mathrm{Var}(\sum_i (Y_{i1}n_{i0} - Y_{i0}n_{i1})/n_i) = \sum_i \mathrm{Var}((Y_{i1}n_{i0} - Y_{i0}n_{i1})/n_i). \tag{23}$$

From (23), the term $\mathrm{Var}((Y_{i1}n_{i0} - Y_{i0}n_{i1})/n_i)$ is further developed. We obtain two different expressions:

$$\frac{RD[n_{i0}^2(n_{i1} - Y_{i1}) - n_{i1}^2(n_{i0} - Y_{i0})] + Y_{i0}n_{i0}(n_{i1} - Y_{i1}) + Y_{i1}n_{i1}(n_{i0} - Y_{i0})}{n_i^2}$$

and

$$\frac{RD(Y_{i0}n_{i1}^2 - Y_{i1}n_{i0}^2) + Y_{i1}n_{i0}(n_{i0} - Y_{i0}) + Y_{i0}n_{i1}(n_{i1} - Y_{i1})}{n_i^2}.$$

Next, these two formulas are averaged, so that we have

$$\begin{aligned}\mathrm{Var}((Y_{i1}n_{i0} - Y_{i0}n_{i1})/n_i) &= \frac{RD[Y_{i0}n_{i1}^2 - Y_{i1}n_{i0}^2 + n_{i1}n_{i0}(n_{i0} - n_{i1})/2]}{n_i^2} \\ &\quad + \frac{Y_{i1}(n_{i0} - Y_{i0}) + Y_{i0}(n_{i1} - Y_{i1})}{2n_i}.\end{aligned} \tag{24}$$

It is now easy to derive the variance estimator of $\widehat{RD}_{MH}$ using (23) and (24) by substituting $\widehat{RD}_{MH}$ into RD. Therefore, the variance estimator suggested by Sato (or $\widehat{\mathrm{Var}}_{st}$) is given as

$$\widehat{\mathrm{Var}}_{st}(\widehat{RD}_{MH}) = \frac{\widehat{RD}_{MH}\sum_i S_i + \sum_i T_i}{(\sum_i n_{i1}n_{i0}/n_i)^2}, \tag{25}$$

where $S_i = [Y_{i0}^2 n_{i1}^2 - Y_{i1}n_{i0}^2 + n_{i1}n_{i0}(n_{i0} - n_{i1})/2]/n_i^2$, $T_i = [Y_{i1}(n_{i0} - Y_{i0}) + Y_{i0}(n_{i1} - Y_{i1})]/(2n_{i1}n_{i0})$, and n_{i1} and n_{i0} must be greater than one. The two-sided $(1-\alpha)100\%$ confidence interval for RD based on the Wald-type method using the estimated vari-

ance of Sato is of the form

$$\widehat{RD}_{MH} \pm z\sqrt{\widehat{\text{Var}}_{st}(\widehat{RD}_{MH})}.$$

The performance of the confidence intervals given in Sections 2 and 3 will be evaluated using simulations. Since, in fact homogeneity or heterogeneity situation can be assumed, they are studied in the next section.

4. Simulation study

4.1. MH log-relative risk

The study investigates the performance of the Wald-type confidence intervals using the variances of $\log \widehat{RR}_{MH}$ derived based on the delta method (dt), the pivotal method (pivot), the Greenland and Robins's method (gr), and the profile likelihood method (pl). These are compared to the bootstrap percentile (bp), the bootstrap variance (bv), and the bias-corrected and accelerated bootstrap percentile (bca) confidence intervals. In this simulation, the data for the log-relative risks from k studies, $\log RR_i$, were generated from an $N(\mu, \tau^2)$, where the true generating parameter $\mu = \log RR$ = -0.5, 0, 0.5, and variance $\tau^2 = 0.5$. The number of studies k were set as 10, 20, 30, 50, and 100. The number of events in the treatment group Y_{i1} and in the comparison group Y_{i0} were generated from $Po(\lambda_{i1}P_{i1})$ and $Po(\lambda_{i0}P_{i0})$, respectively, where $\lambda_{i0} = 0.05$, corresponding to a rare events setting as also noted in [6], and λ_{i1} was calculated by $RR_i \times \lambda_{i0}$, for $i = 1, 2, ..., k$. We also looked at the case $\lambda_{i0} = 0.01$ with results presented in the appendix, as Table A1. The person-time P_{i1} were sampled from a uniform distribution, $U(50, 100)$, and $P_{i0} = d \times P_{i1}$, where $d \sim U(0.9, 1)$, the degree of within-study-imbalance. Note that when $\tau^2 = 0$, we are in the situation of homogeneity.

Each scenario was then repeated H = 5,000 times using `R` [40]. A large number of H was used to eliminate only random error due to simulation. The performance of the 95% confidence interval for the parameter of interest was investigated using the coverage probability. Here, our focus in performance of the confidence interval is that the coverage probability on average is close to the nominal level of 0.95.

4.1.1. Simulation results of $\log \widehat{RR}_{MH}$ *for estimating the generating parameter* (μ)

Table 1 shows the values of $\log \widehat{RR}_{MH}$ on average under meta-analysis for the rare events setting. In the *heterogeneity* case, $\log \widehat{RR}_{MH}$ obtained from the formula and bootstrap methods are greater than the generating parameter μ in all cases considered. It can be also seen that the coverage probabilities of all confidence intervals are lower than the target probability of 0.95. The confidence intervals using the four variance formulas provide the coverage probability lower than the bootstrap confidence intervals in all situations. In conclusion, the results indicate that in the heterogeneity case $\log \widehat{RR}_{MH}$ overestimates μ. As can be also seen in Figure 2 for a case when $k = 100$, μ is given as -0.5 but the estimated values from both methods (formula-based and bootstrap) are approximately equal to -0.25. Furthermore, when $\mu = 0$, the estimated values are 0.25. These lead to the Wald-type and bootstrap confidence intervals having a low coverage probability. In contrast, the point and interval estimates obtained from the variance formulas under *homogeneity* situations provide a good estimation for the log-relative risk. Those of the bootstrap methods perform well when $k > 30$.

To consider a different situation from the previous simulation, we evaluate the performance of the MH log-relative risk under heterogeneity by generating the data from a two-component mixture distribution. The binary indicator for the mixture component is sampled form a Bernoulli distribution with event probability 0.5. The marginal mean for the mixture on the log-scale is then given as $\mu = 0.5 \log RR_1 + 0.5 \log RR_2$, where RR_i is the relative risk in component i, for $i = 1, 2$. If the outcome of the Bernoulli experiment is 1 then $\log RR_1$ = -0.5 is used, and $\log RR_2 = 0.5$ is used otherwise. Note that for this case $\mu = 0$ and $\exp(\mu) = 1$. On the RR-scale we yield $0.5 \exp(-0.5) + 0.5 \exp(0.5) = 1.1276$ and this is what the MH-estimator estimates as can be seen in column eight of Table 2 (upper part) as $\log 1.1276 = 0.1201$.

Whereas in first scenario the risk ratio is 1 on the log-scale, we consider another setting where it is also 1 but on the RR-scale. In this scenario, if the outcome of the Bernoulli experiment is 1, we use $RR_1 = 0.5$, else $RR_2 = 1.5$. Thus, the marginal mean of the mixture on the RR-scale is given by $0.5RR_1 + 0.5RR_2$, which takes the value 1 for the given values. This is what the MH estimator estimates as $\log 1 = 0$ (see again column eight of Table 2, lower part). However, on the log-scale we yield

Table 1. The mean of log-MH estimate and coverage probability (CP) of the 95% confidence intervals for the true generating log-relative risk (μ) using variance formula-based and bootstrap methods, when generating data under the normal distribution.

		Formula-based method					Bootstrap method			
μ	k	$\log \widehat{RR}_{MH}$	CP_{dt}	CP_{pivot}	CP_{gr}	CP_{pl}	$\log \widehat{RR}_{MH}$	CP_{bp}	CP_{bv}	CP_{bca}
Heterogeneity case										
-0.5	10	-0.2842	0.7394	0.7504	0.7396	0.7394	-0.3139	0.8608	0.8568	0.8442
	20	-0.2625	0.6410	0.6548	0.6416	0.6410	-0.2785	0.8240	0.8140	0.8004
	30	-0.2541	0.5608	0.5760	0.5606	0.5608	-0.2651	0.7818	0.7690	0.7390
	50	-0.2613	0.4492	0.4622	0.4490	0.4492	-0.2681	0.6868	0.6744	0.6146
	100	-0.2544	0.2194	0.2274	0.2194	0.2194	-0.2580	0.4230	0.4140	0.3848
0	10	0.2173	0.7178	0.7112	0.7182	0.7178	0.1945	0.8620	0.8594	0.8394
	20	0.2369	0.5954	0.5914	0.5948	0.5954	0.2243	0.8208	0.8100	0.7826
	30	0.2394	0.5208	0.5180	0.5212	0.5210	0.2308	0.7608	0.7582	0.7066
	50	0.2424	0.3706	0.3680	0.3702	0.3706	0.2370	0.6336	0.6284	0.5986
	100	0.2500	0.1492	0.1486	0.1492	0.1492	0.2472	0.3484	0.3468	0.3246
0.5	10	0.7265	0.6628	0.6462	0.6622	0.6628	0.7090	0.8454	0.8414	0.8380
	20	0.7353	0.5736	0.5572	0.5732	0.5732	0.7249	0.8140	0.8072	0.7668
	30	0.7451	0.4654	0.4514	0.4654	0.4654	0.7379	0.7418	0.7354	0.7092
	50	0.7434	0.3244	0.3140	0.3242	0.3242	0.7388	0.6034	0.6028	0.5634
	100	0.7477	0.1214	0.1180	0.1214	0.1214	0.7453	0.3234	0.3220	0.2888
Homogeneity case										
-0.5	10	-0.5056	0.9508	0.9462	0.9504	0.9508	-0.5138	0.8950	0.9100	0.9074
	20	-0.5062	0.9522	0.9492	0.9522	0.9522	-0.5104	0.9250	0.9322	0.9286
	30	-0.5035	0.9514	0.9500	0.9514	0.9512	-0.5062	0.9320	0.9378	0.9424
	50	-0.5034	0.9498	0.9466	0.9490	0.9499	-0.5051	0.9366	0.9400	0.9382
	100	-0.5011	0.9550	0.9544	0.9548	0.9550	-0.5019	0.9496	0.9520	0.9470
0	10	-0.0041	0.9532	0.9514	0.9532	0.9532	-0.0039	0.8994	0.9110	0.9024
	20	-0.0017	0.9532	0.9528	0.9532	0.9532	-0.0013	0.9286	0.9316	0.9314
	30	0.0025	0.9538	0.9534	0.9542	0.9538	0.0027	0.9368	0.9410	0.9382
	50	0.0001	0.9498	0.9496	0.9498	0.9498	0.0003	0.9376	0.9384	0.9416
	100	-0.0006	0.9516	0.9512	0.9514	0.9516	-0.0006	0.9428	0.9438	0.9440
0.5	10	0.5096	0.9522	0.9466	0.9522	0.9522	0.5156	0.8978	0.9096	0.9062
	20	0.5006	0.9512	0.9518	0.9512	0.9512	0.5034	0.9300	0.9358	0.9250
	30	0.5016	0.9544	0.9540	0.9544	0.9544	0.5035	0.9380	0.9404	0.9350
	50	0.4993	0.9474	0.9484	0.9470	0.9474	0.5005	0.9408	0.9414	0.9380
	100	0.5015	0.9526	0.9520	0.9526	0.9526	0.5021	0.9464	0.9480	0.9438

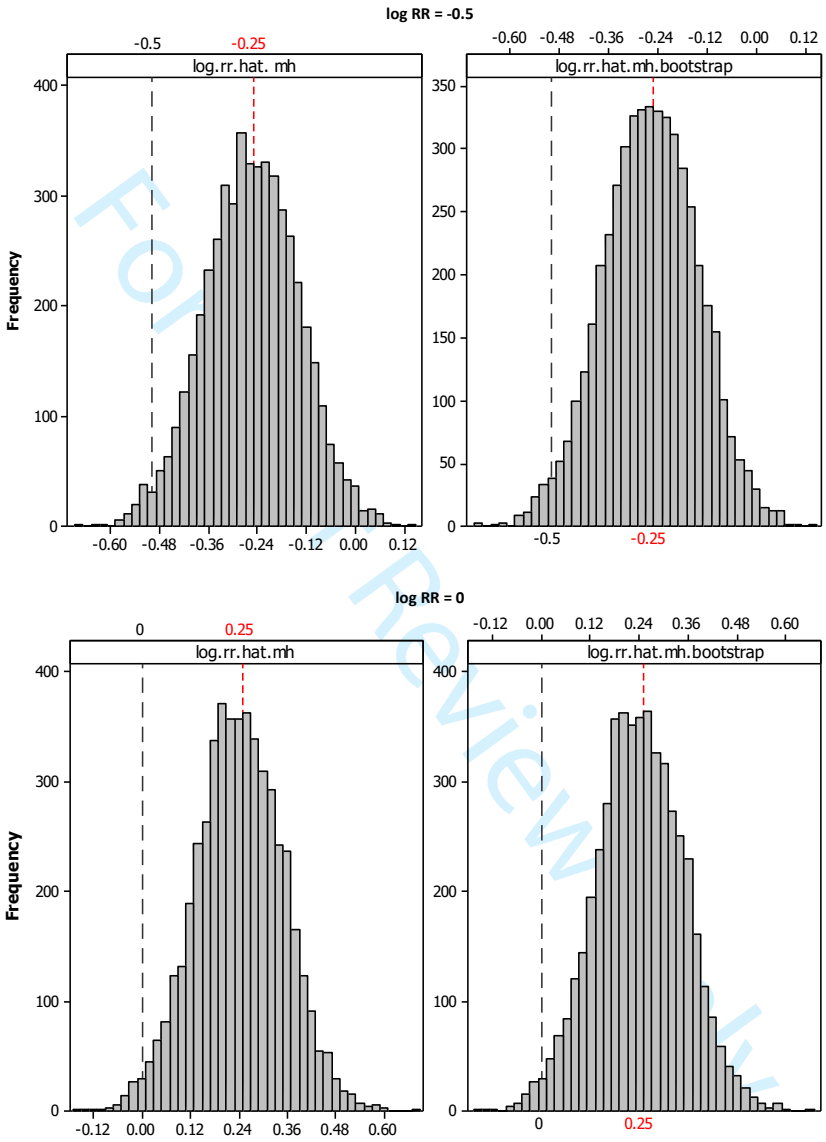

Figure 2. Plots of simulated MH log-relative risk from formula-based method (left) and bootstrap method (right) under heterogeneity case, when $\log RR_i \sim N(\mu = \log RR, \tau^2 = 0.5)$ and $k = 100$.

Table 2. The mean of log-MH estimate and coverage probability (CP) of the 95% confidence intervals for the true generating log-relative risk (μ) using variance formula-based and bootstrap methods, when generating data under the two-component mixture case.

			Formula-based method					Bootstrap method		
μ	k	$\log \widehat{RR}_{MH}$	CP_{dt}	CP_{pivot}	CP_{gr}	CP_{pl}	$\log \widehat{RR}_{MH}$	CP_{bp}	CP_{bv}	CP_{bca}
Case 1: True $\log RR = 0.5(-0.5)+0.5(0.5) = 0$										
0	10	0.1134	0.8676	0.8626	0.8678	0.8676	0.1044	0.8744	0.8798	0.8746
	20	0.1132	0.8318	0.8288	0.8314	0.8318	0.1089	0.8790	0.8778	0.8836
	30	0.1122	0.8072	0.8042	0.8074	0.8072	0.1091	0.8714	0.8694	0.8652
	50	0.1168	0.7416	0.7386	0.7414	0.7416	0.1150	0.8282	0.8244	0.8318
	100	0.1205	0.5966	0.5952	0.5966	0.5966	0.1196	0.7004	0.7002	0.7162
Case 2: True $\log RR = 0.5 \log 0.5 + 0.5 \log 1.5 =$ -0.14										
-0.14	10	-0.0150	0.8546	0.8528	0.8534	0.8540	-0.0275	0.8750	0.8782	0.8776
	20	-0.0059	0.8074	0.8068	0.8072	0.8070	-0.0122	0.8708	0.8686	0.8694
	30	-0.0040	0.7624	0.7652	0.7628	0.7624	-0.0082	0.8456	0.8432	0.8446
	50	-0.0037	0.6898	0.6908	0.6902	0.6898	-0.0060	0.7950	0.7894	0.7926
	100	-0.0019	0.5178	0.5194	0.5180	0.5178	-0.0031	0.6488	0.6442	0.6468

$\mu = (0.5 \log RR_1 + 0.5 \log RR_2) = -0.1438$, which again shows the underestimation fact relative to the value of zero which the MH-estimate estimates on the RR-scale for this case.

Let us return to the case when $\log RR_i \sim N(\mu, \tau^2)$, where $\mu = \log RR$. In other words, the log-relative risk is normally distributed which is a quite common assumption often met in mixed modelling. In this case, we point out again that in the heterogeneity situation $\widehat{RR}_{MH}$ is a biased estimator for RR or $\exp(\mu)$. This can be explained using a special case as follows. From (2), when the person-times are balanced within studies, we have $E(\widehat{RR}_{MH}) \approx E(RR)$. However, in fact the log-relative risk is a random variable having a normal distribution with mean μ and variance τ^2. This means that the relative risk has a log-normal distribution, denoted as $RR_i \sim LN(\mu, \tau^2)$, where $E(RR_i) = \exp(\mu + \tau^2/2)$. Therefore, $\widehat{RR}_{MH}$ will estimate the parameter $\exp(\mu + \tau^2/2)$, instead of $\exp(\mu)$ as we might have expected. An evidence for confirming this hypothesis is shown in Figure 2. Here, we investigate what $\log \widehat{RR}_{MH}$ is estimating. As we have seen in the balanced case, the asymptotic expected value of $\widehat{RR}_{MH}$ is $\exp(\mu + \tau^2/2)$. For example, when $\mu = 0$ and $\tau^2 = 0.5$, the logarithm of the asymptotic expected value of $\widehat{RR}_{MH}$ is $0.5/2 = 0.25$, not equal to $\mu = 0$.

4.1.2. Simulation results of $\log \widehat{RR}_{MH}$ *for estimating the estimand (θ)*

As we noted in the above section, it is not appropriate to consider $\log \widehat{RR}_{MH}$ as an estimator of μ, if there is heterogeneity. The performance of this estimator including confidence intervals was therefore investigated using simulations once again. However,

in this part the *estimand* of the MH estimator, denoted by θ, was used as the true value and we consider only the heterogeneity situation. When the log-relative risk is generated under the normal distribution, the estimand is $\theta = \exp(\mu + \tau^2/2)$, where μ = -0.5, 0, and 0.5, and $\tau^2 = 0.5$. Under the two-component mixture distribution of the log-relative risk, the estimand is computed by $w \times RR_1 + (1 - w) \times RR_2$, where the event probability from the Bernoulli distribution is $w = 0.5$, RR_1 and RR_2 are corresponded to the previous simulation settings of the heterogeneity case. The main results are described in the following. Tables 3 and 4 show the coverage probabilities of the confidence intervals based on the two main methods under the normal and mixture distributions, respectively. It can be seen that the values of $\log \widehat{RR}_{MH}$ are close to θ, particularly when $k > 30$. The coverage probabilities of the confidence intervals considered here are greater than those when we used the true value as μ (see subsection 4.1.1). In more detail, the bootstrap intervals provide coverage probabilities greater than the Wald-type confidence intervals. Especially, the bootstrap variance confidence interval outperforms the other bootstrap intervals under comparison.

The bias of the MH estimator has been mentioned before such as in the case of the MH estimator for the odds ratio (see Bakbergenuly et al. [13] and Bakbergenuly and Kulinskaya [14]). The deeper reason for this overestimation bias of the MH estimator in the case of heterogeneity stems from Jensen's inequality [41]. If there is heterogeneity in the relative risk, then RR will have a distribution and it follows from the concavity of the logarithm that

$$\log E(RR) \geq E(\log RR).$$

In the case that $\log RR_i \sim N(\mu, \tau^2)$ we have $E(\log RR) = \mu$, but what we estimate is $\log E(RR)$ which equals $\mu + \tau^2/2$. Hence, an overestimation bias occurs. We could correct for this overestimation bias if we would need to have an estimate for τ^2, but this is usually not available without further modelling assumptions and additional computational expense. From our perspective, it seems more fair to consider the performance of the MH estimator in terms for what it estimates. This precisely means to consider the true heterogeneity on the RR-scale and not on the log-scale. So, for

Table 3. The mean of log-MH estimate and coverage probability (CP) of the 95% confidence intervals for the *estimand* of MH estimator for the log-relative risk (θ) using variance formula-based and bootstrap methods, when generating data under the normal distribution with heterogeneity.

		Formula-based method					Bootstrap method			
θ	k	$\log \widehat{RR}_{MH}$	CP_{dt}	CP_{pivot}	CP_{gr}	CP_{pl}	$\log \widehat{RR}_{MH}$	CP_{bp}	CP_{bv}	CP_{bca}
-0.25	10	-0.2955	0.8422	0.8360	0.8420	0.8422	-0.3262	0.8878	0.9101	0.8987
	20	-0.2727	0.8226	0.8214	0.8224	0.8226	-0.2891	0.9054	0.9146	0.9062
	30	-0.2617	0.8272	0.8230	0.8268	0.8270	-0.2727	0.9218	0.9286	0.9256
	50	-0.2555	0.8222	0.8212	0.8222	0.8222	-0.2624	0.9322	0.9334	0.9304
	100	-0.2555	0.8186	0.8176	0.8186	0.8186	-0.2590	0.9318	0.9358	0.9338
0.25	10	0.2196	0.8098	0.8088	0.8104	0.8098	0.1974	0.8982	0.8997	0.8987
	20	0.2345	0.8000	0.8012	0.7994	0.7998	0.2220	0.9090	0.9182	0.9118
	30	0.2395	0.8018	0.8028	0.8022	0.8018	0.2310	0.9240	0.9284	0.9250
	50	0.2457	0.8054	0.8046	0.8052	0.8054	0.2401	0.9358	0.9400	0.9364
	100	0.2490	0.7986	0.7988	0.7986	0.7986	0.2461	0.9356	0.9386	0.9404
0.75	10	0.7296	0.7826	0.7904	0.7822	0.7826	0.7120	0.8980	0.8992	0.8986
	20	0.7341	0.7770	0.7802	0.7778	0.7774	0.7236	0.9114	0.9186	0.9112
	30	0.7406	0.7806	0.7786	0.7806	0.7806	0.7333	0.9246	0.9284	0.9238
	50	0.7463	0.7740	0.7770	0.7744	0.7742	0.7417	0.9336	0.9348	0.9346
	100	0.7467	0.7664	0.7666	0.7664	0.7664	0.7443	0.9402	0.9416	0.9402

Table 4. The mean of log-MH estimate and coverage probability (CP) of the 95% confidence intervals for the *estimand* of MH estimator for the log-relative risk (θ) using variance formula-based and bootstrap methods, under the two-component mixture case with heterogeneity effect.

		Formula-based method					Bootstrap method			
θ	k	$\log \widehat{RR}_{MH}$	CP_{dt}	CP_{pivot}	CP_{gr}	CP_{pl}	$\log \widehat{RR}_{MH}$	CP_{bp}	CP_{bv}	CP_{bca}
Case 1: Estimand of MHE $= \log(0.5\exp(-0.5)+0.5\exp(-0.5)) = 0.12$										
0.12	10	0.1046	0.8996	0.8962	0.8992	0.8996	0.0958	0.8984	0.9120	0.9008
	20	0.1145	0.9028	0.9010	0.9028	0.9028	0.1101	0.9254	0.9316	0.9330
	30	0.1169	0.9010	0.8994	0.9014	0.9010	0.1141	0.9342	0.9366	0.9380
	50	0.1191	0.9030	0.9026	0.9030	0.9030	0.1174	0.9448	0.9464	0.9486
	100	0.1184	0.9012	0.9016	0.9012	0.9012	0.1175	0.9506	0.9522	0.9510
Case 2: Estimand of MHE $= \log(0.5\times 0.5+0.5\times 1.5) = 0$										
0	10	-0.0116	0.9012	0.8976	0.9010	0.9012	-0.0244	0.8980	0.9136	0.9066
	20	-0.0014	0.8996	0.8988	0.8998	0.8996	-0.0074	0.9290	0.9332	0.9350
	30	-0.0072	0.8910	0.8890	0.8908	0.8910	-0.0113	0.9314	0.9378	0.9362
	50	-0.0044	0.8950	0.8952	0.8954	0.8950	-0.0069	0.9440	0.9466	0.9492
	100	-0.0004	0.8904	0.8902	0.8904	0.8904	-0.0017	0.9408	0.9438	0.9430

example, if the heterogeneity distribution of the risk ratio is normal on the log-scale it is only appropriate to evaluate the performance of the MH estimator on the associated distribution on the RR-scale. We return to these issues in the discussion.

4.2. MH risk difference

To investigate the performance of the MH estimator for the true risk difference (RD) under rare events, simulation settings were set as follows. Under heterogeneity, we generated an indicator variable for the two-component mixture distribution from a Bernoulli with event probability $w = 0.25$ and 0.75. If the outcome of Bernoulli experiment was 1 we set $RD_i = a$, otherwise $RD_i = b$, for $i = 1, 2, ..., k$. To avoid cases with negative treatment probability, we used $a = 0.04$ and $b = -0.04$. The risk for the control group was given as $\lambda_{i0} = 0.05$, and the risk for the treatment group was calculated by $\lambda_{i1} = RD_i + \lambda_{i0}$, where RD_i depended on the result of a Bernoulli experiment. We set λ_{i0} and λ_{i1} as noted above for obtaining rare events. Then, the true risk difference was estimated by $RD = w \times a + (1 - w) \times b$, which were -0.02 or 0.02. For homogeneity situation, $RD = 0.04$ was considered without using an indicator. The number under risk were generated as unbalanced, with n_{i1} sampled from a discrete $U(50, 100)$ and $n_{i0} = d \times n_{i1}$ with $d \sim U(0.9, 1)$. We then generated Y_{i1} and Y_{i0} from two binomial distributions, $B(n_{i1}, \lambda_{i1})$ and $B(n_{i0}, \lambda_{i0})$, respectively. Furthermore, we looked at $\lambda_{i0} = 0.01$, where the results are presented in Table A2 of the appendix.

The performance of the 95% Wald-type confidence intervals for RD based on Greenland and Robins's variance (gr), Sato's variance (st), and the bootstrap percentile (bp) and bootstrap variance (bv) confidence intervals are presented in Table 5. It can be concluded that $\widehat{RD}_{MH}$ from the formula-based and bootstrap methods work well and provide the estimated values close to RD, even in the case of heterogeneity. However, the coverage probabilities of the confidence intervals based on the two formula variances in heterogeneity cases are lower than the nominal level of 0.95. The bootstrap confidence intervals have the coverage probability greater in this case, but they cover the true parameter with a satisfied probability when $k > 30$. Again for homogeneity case, the confidence intervals based on the two variance formulas perform well in terms of coverage probability. This finding is also supported by Klingenberg [37] and Lui [39],

Table 5. The mean of MH estimate and coverage probability (CP) of the 95% confidence interval for the risk difference using variance formula-based and bootstrap methods.

		Formula-based method			Bootstrap method		
RD	k	$\widehat{RD}_{MH}$	$CP_{rd.st}$	$CP_{rd.gr}$	$\widehat{RD}_{MH}$	$CP_{rd.bp}$	$CP_{rd.bv}$
				Heterogeneity case			
-0.02	10	-0.0197	0.8148	0.8102	-0.0197	0.8774	0.8822
	20	-0.0202	0.8164	0.8116	-0.0202	0.9188	0.9180
	30	-0.0201	0.8136	0.8164	-0.0201	0.9294	0.9282
	50	-0.0198	0.8206	0.8138	-0.0198	0.9426	0.9414
	100	-0.0200	0.8218	0.8044	-0.0200	0.9422	0.9416
0.02	10	0.0201	0.8454	0.8506	0.0201	0.8972	0.8976
	20	0.0200	0.8482	0.8524	0.0200	0.9254	0.9242
	30	0.0198	0.8492	0.8576	0.0198	0.9354	0.9366
	50	0.0200	0.8514	0.8552	0.0200	0.9398	0.9402
	100	0.0200	0.8472	0.8510	0.0200	0.9424	0.9440
				Homogeneity case			
0.04	10	0.0399	0.9494	0.9436	0.0399	0.8958	0.9044
	20	0.0397	0.9498	0.9492	0.0397	0.9260	0.9296
	30	0.0401	0.9470	0.9520	0.0401	0.9328	0.9360
	50	0.0401	0.9550	0.9482	0.0401	0.9426	0.9456
	100	0.0400	0.9532	0.9428	0.0400	0.9490	0.9500

and makes sense as these variance estimators are constructed under homogeneity of effect.

5. Case study

5.1. The effectiveness of prophylactic antibiotics in caesarean section (continued)

We now illustrate the various estimators at the example of a case study. A systematic review of the effectiveness of prophylactic antibiotic treatment on infectious complications in women undergoing caesarean delivery [2] as noted in the first section of this paper is used. We note again that many of the component studies in the data were small in size and there were zero events in each of two arms, especially in the treatment group. If the inverse variance-weighted average method is used, the double-zero or single-zero study would have to be excluded before the analysis because the risk ratio is undefined. However, the MH meta-analysis and bootstrap methods can be used when those zero-event studies are included.

Using the data given in Table B1 of the appendix, we first focus on checking the homogeneity of the risk ratio. The chi-squared test on testing the hypothesis of homogeneity gives a p-value of 0.32 and Higgins's I^2 is 7.1%. This indicates no empirical

evidence of a difference of risk ratios between studies. In our analysis using 61 studies, the estimated risk ratio from the MH estimator and bootstrap method equal 0.3864 and 0.3854, respectively. For the risk difference, the MH and bootstrap estimates are -0.0575 and -0.0574, respectively. The 95% confidence intervals for the parameter of interest and the variance estimates of the MH estimator are also computed, and given in the first part of Table 6. There is a clear indication of a significant difference between the two therapies. Overall, a women undergoing caesarean delivery has lower risk for infectious complications if in the prophylactic antibiotic treatment group relative to being in placebo or in the no prophylactic antibiotic treatment group.

Under this data set, the meta-analysis based on the MH method and the bootstrap can be computed. They provide the summary results in the same way, even though rare events are available. The MH estimate and the confidence intervals for the risk ratio obtained from the formula-based and bootstrap methods do not much differ. These confirm the simulation results under homogeneity situations presented in Section 4.

5.2. The association of MERS exposure proximity with infection

In the second example, we use the meta-analytic data on the effects of physical distance 1 metre (m) or more (exposure) and less than 1 m (non-exposure) on Middle East respiratory syndrome (MERS) transmission reported in Chu et al. [42]. The data given in Table B2 of the appendix include eight studies with confirmed cases of MERS comparing distances between people and MERS infected patients of 1 m or larger with smaller distances. The results given in Figure 3 show that there is a significant difference of risk ratios between studies (or heterogeneity effect) with Higgins's I^2 of 74.9% and p-value of 0.008. In our analysis, the estimated MH risk ratio and risk difference based on 8 studies are 0.1569 and -0.1197, respectively. The bootstrap values for the MH relative risk and risk difference are 0.1699 and -0.1079, respectively. We also estimate the 95% confidence intervals for the risk ratio and risk difference using several methods. They are presented in the second part of Table 6. From the results, the confidence intervals for the risk difference are slightly different. However, the confidence intervals based on the variance formulas for the log-relative risk are quite different from the bootstrap confidence intervals. Given the results in Section 4 we give more trust

Table 6. The estimated variance of the estimators and 95% confidence intervals for the relative risk and risk difference using two meta-analytic data examples.

Approach	$\widehat{\text{Var}}(\log \widehat{RR}_{MH})$	Confidence limits for RR	$\widehat{\text{Var}}(\widehat{RD}_{MH})$	Confidence limits for RD
Example 1: Prophylactic antibiotics in caesarean section				
Delta	0.0089	(0.3212, 0.4649)	-	-
Greenland & Robins	0.0089	(0.3214, 0.4647)	2.8×10^{-5}	(-0.0679, -0.0470)
Profile likelihood	0.0088	(0.3215, 0.4645)	-	-
Pivot method	-	(0.3187, 0.4626)	-	-
Sato	-	-	2.9×10^{-5}	(-0.0681, -0.0468)
Bootstrap percentile	-	(0.2958, 0.4861)	-	(-0.0727, -0.0427)
Bootstrap variance	0.0167	(0.3003, 0.4945)	5.9×10^{-5}	(-0.0725, -0.0423)
BCa bootstrap	-	(0.2929, 0.4827)	-	(-0.0746, -0.0436)
Example 2: Association of MERS exposure proximity with infection				
Delta	0.2586	(0.0579, 0.4250)	-	-
Greenland & Robins	0.1419	(0.0750, 0.3282)	0.0009	(-0.1815, -0.0134)
Profile likelihood	0.1896	(0.0668, 0.3682)	-	-
Pivot	-	(0.0159, 0.3707)	-	-
Sato	-	-	0.0009	(-0.1784, -0.0610)
Bootstrap percentile	-	(0.0499, 0.5875)	-	(-0.1815, -0.0134)
Bootstrap variance	0.3997	(0.0493, 0.5879)	0.0023	(-0.2012, -0.0145)
BCa bootstrap	-	(0, 0.5453)	-	(-0.1968, -0.0217)

to the bootstrap intervals, in particular the bootstrap percentile and the bootstrap variance interval.

6. Discussion

The Mantel-Haenszel (MH) approach can be viewed as a model-free summary tool in meta-analysis of count data. It can be used whether there is homogeneity or not in the effect measure interest. We have seen that in the case of heterogeneity in the risk ratio certain issues arise which cannot easily be coped with in the bootstrap approach, but rather needed to be decided which scale (log-scale or not) for the risk ratio should be used. We have investigated interval estimation for the MH estimator when outcomes are counts of events and focused on rare events leading to low count values. The MH estimators for the relative risk ($\widehat{RR}_{MH}$) and risk difference ($\widehat{RD}_{MH}$) are widely applied in this situation. To their advantage, they are not sensitive to zero-event studies. To compute the variance of the relative risk estimator which is usually done on the log-scale and the variance of the risk difference estimator, several methods are introduced and applied in interval estimation including the Wald-type confidence intervals. Furthermore, we study the bootstrap confidence intervals. The performance of the confidence intervals was assessed using simulations with various settings.

Under *homogeneity*, $\log \widehat{RR}_{MH}$ and the confidence intervals for $\log RR$ constructed

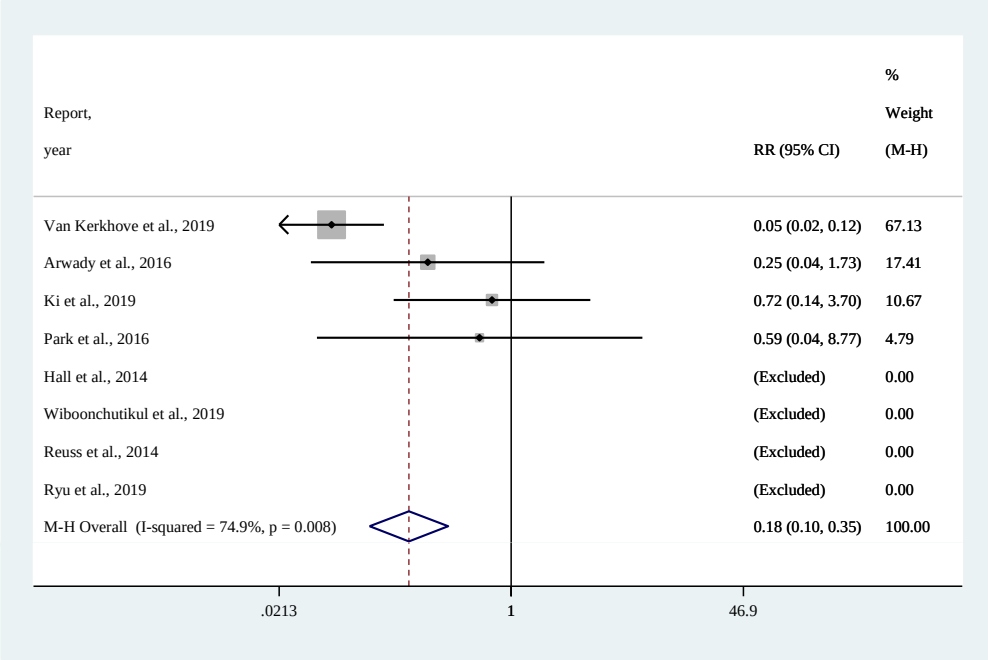

Figure 3. Forest plot of meta-analytic data on the association of MERS exposure proximity with infection.

based on the Wald-type method using the standard variance formulas considered in this work perform well. There is no significant bias in estimation. The coverage probabilities are very close to the target level. The bootstrap confidence intervals also satisfy the nominal coverage probability when number of studies are large. Furthermore, $\widehat{RD}_{MH}$ and interval estimation for RD using the variance formulas provide satisfactory results under this assumption. However, the results change when *heterogeneity* in the effect is assumed. $\log \widehat{RR}_{MH}$ overestimates and is a biased estimator for $\log RR$. Also, we have given explanations why this occurs. The Wald-type confidence intervals have a low performance in terms of coverage probability. We have then concluded that confidence interval for $\log RR$ constructed using the variance estimator derived under the assumption of homogeneity should not be used if heterogeneity presents. In contrast, $\widehat{RD}_{MH}$ can be used to estimate the true RD unbiasedly, whether there is heterogeneity or not. The bootstrap confidence intervals for RD are able to provide acceptable coverage as we have shown in simulation study.

For the risk ratio, in the setting of heterogeneity things are more complex as we have seen. However, we focus here again on some of the issues. If we are willing to assume

that the log-relative risk is normally distributed across studies, then the MH estimator of relative risk will overestimate the true relative risk. Correct bootstrap inference could be constructed if the confidence interval could calibrate by correcting for the overestimation bias. This, however, requires an estimate of the variance of the latent distribution of the log-relative risk which will require additional model assumptions and computational effort. Also, in that case the bootstrap approach would no longer be required. Alternatively, one might consider accepting a normal distribution (or any other) of the relative risk scale in which case the bootstrap inference will deliver correct inference. However, working on the relative risk scale with continuous distributions seems less acceptable than doing so on the log-relative risk scale.

As a further alternative, we might think of the distribution of relative risk in a non-parametric way on the relative risk scale. Again, here the bootstrap approach will provide correct inference. The question, however, remains if one willing to accept that non-symmetric weights are treated symmetrically. To illustrate, consider the two-component mixture which gives equal weights to sub-populations with mean relative risks 0.5 and 1.5, respectively. The marginal mean of the relative risk would be 1, and this is also what the MH estimator would estimate. However, most epidemiologists will likely feel that the symmetric effect for 0.5 on the relative risk scale should be 2 giving a marginal relative risk mean of 1 on the log-relative risk scale $0.5 \log 0.5 + 0.5 \log 2 = 0$. However, the MH estimate will deliver a mean of 1.25 in this case, instead of 1 as it is on the log-relative risk scale.

Ultimately, it is the question of which scale we are willing to accept for the relative risk and which distribution for it. These problems do not exist for the risk difference as we are working on a unique scale and this might be considered as an advantage.

Acknowledgements

The authors would like to thank the Editor, Associate Editor, and referees for reviewing the manuscript and providing valuable comments. The work has been partially funded by the German Research Foundation HO1286/16-1.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

[1] Smaill F, Hofmeyr GJ. Antibiotic prophylaxis for cesarean section. Cochrane Database Syst Rev. 2002;3:1–71.

[2] Cooper NJ, Abrams KR, Sutton AJ, et al. A Bayesian methods for Markov modelling in cost-effectiveness analysis: an application to taxane use in advanced breast cancer. J R Stat Soc Ser A Stat Soc. 2003;166(3):389–405.

[3] Palmer T, Sterne J. Meta-Analysis in Stata: An Updated Collection from the Stata Journal. TX: Stata Press; 2015.

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

[5] 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.

[6] 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.

[7] 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.

[8] Robins J, Greenland S, Breslow NE. A general estimator for the variance of the Mantel-Haenszel odds ratio. Am J Epidemiol. 1986;124(5):719–723.

[9] Jewell NP. Statistics for Epidemiology. Boca Raton: CRC Press; 2004.

[10] Landis JR, Sharp TJ, Kuritz SJ, et al. Mantel-Haenszel Methods. Chichester: John Wiley & Sons; 2005.

[11] Böhning D, Sangnawakij P. The identity of two meta-analytic likelihoods and the ignorability of double-zero studies. Biostatistics. 2020;kxaa004:1–7.

[12] 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.

[13] Bakbergenuly I, Kulinskaya E. Beta-binomial model for meta-analysis of odds ratios. Stat Med. 2017;36(11):1715–1734.

[14] 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.

[15] Li L, Wang X. Meta-analysis of rare binary events in treatment groups with unequal variability. Stat Methods Med Res. 2019;28(1):263–274.

[16] Huang L, Yin Y, Yang L, et al. Comparison of antibiotic therapy and appendectomy for acute uncomplicated appendicitis in children: a meta-analysis. JAMA Pediatr. 2017;171(5):426–434.

[17] Cui Y, Li C, Xu Z, et al. Robot-assisted versus conventional laparoscopic operation in anus-preserving rectal cancer: a meta-analysis. Ther Clin Risk Manag. 2017;13:1247–1257.

[18] Gottlieb M, Holladay D, Peksa GD. Ultrasound-assisted lumbar punctures: a systematic review and meta-analysis. Acad Emerg Med. 2019;26(1):85–96.

[19] Kirk AHP, Yang J, Sim WC, et al. Systematic review of the effect of topical application of human breast milk on early umbilical cord separation. J Obstet Gynecol Neonatal Nurs. 2019;48(2):121–130.

[20] Webster KE, Hewett TE. What is the evidence for and validity of return-to-sport testing after anterior cruciate ligament reconstruction surgery? a systematic review and meta-analysis. Sports Med. 2019;49(6):917–929.

[21] Facchinetti G, D'Angelo D, Piredda M, et al. Continuity of care interventions for preventing hospital readmission of older people with chronic diseases: a meta-analysis. Int J Nurs Stud. 2020;101:1–10.

[22] Senn S, Gavini F, Magrez D, et al. Issues in performing a network meta-analysis. Stat Methods Med Res. 2012;22(2):169–189.

[23] Efthimiou O, Rücker E, Schwarzer G, et al. Network meta-analysis of rare events using the Mantel-Haenszel method. Stat Med. 2019;38(3):2992–3012.

[24] Rücker G, Nikolakopoulou A, Papakonstantinou T, et al. The statistical importance of a study for a network meta-analysis estimate. BMC Med Res Methodol. 2020;20(190):1–13.

[25] Fieller EC. The distribution of the index in a bivariate normal distribution. Biometrika. 1932;24(3):428–440.

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

[27] Newcombe RG. MOVER-R confidence intervals for ratios and products of two independently estimated quantities. Stat Methods Med Res. 2016;25(5):1774–1778.

[28] Donner A, Zou GY. Closed-form confidence intervals for functions of the normal mean and standard deviation. Stat Methods Med Res. 2010;21(4):347–359.

[29] Efron B. Bootstrap methods: another look at the jackknife. Ann Stat. 1979;7(1):1–26.

[30] Efron B, Tibshirani R. An Introduction to the Bootstrap. New York: Chapman & Hall; 1993.

[31] Efron B. Better bootstrap confidence intervals. J Am Stat Assoc. 1987;82(397):171–185.

[32] Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? a practical guide for medical statisticians. Stat Med. 2000;19(9):1141–1164.

[33] DiCiccio TJ, Efron B. Bootstrap confidence intervals. Stat Sci. 1996;11(3):189–212.

[34] DelosReyes JM. Estimation of correlation confidence intervals via the bootstrap: non-normal distributions [masters thesis]. Norfolk (VA): Old Dominion University; 2019.

[35] Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. Philadelphia: Lippincott Williams & Wilkins; 2008.

[36] Cochran W. Some methods for strengthening the common χ^2 tests. Biometrics. 1954;10(4):417–451.

[37] Klingenberg B. A new and improved confidence interval for the Mantel-Haenszel risk difference. Stat Med. 2014;33(17):2968–2983.

[38] Sato T. On the variance estimator for the Mantel-Haenszel risk difference. Biometrics. 1989;45(4):1323–1324.

[39] Lui KJ. Statistical Estimation of Epidemiological Risk. Chichester: John Wiley & Sons; 2004.

[40] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Austria: Vienna; 2019.

[41] Dragomir SS. Some refinements of Jensen's inequality. J Math Anal Appl. 1992;168(2):518–522.

[42] Chu DK, Akl EA, Duda S, et al. 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. Lancet. 2020;395(10242):1973–1987.

Appendices

Appendix A. Additional simulation result

According to parameter settings given in the simulation section, the coverage probabilities of the 95% confidence intervals under $\lambda_{i0} = 0.01$ are given in the following. It

is important to note that the results obtained from setting $\lambda_{i0} = 0.05$, presented in Section 4, and $\lambda_{i0} = 0.01$ are in the same way. The performance of estimator obtained from the bootstrap method is better than that of the formula-based method.

Table A1. The mean of log-MH estimate and coverage probability (CP) of the 95% confidence intervals for the true generating log-relative risk (μ) using variance formula-based and bootstrap methods, when generating data under the normal distribution ($\lambda_{i0} = 0.01$).

		Formula-based method					Bootstrap method			
μ	k	$\log \widehat{RR}_{MH}$	CP_{dt}	CP_{pivot}	CP_{gr}	CP_{pl}	$\log \widehat{RR}_{MH}$	CP_{bp}	CP_{bv}	CP_{bca}
Heterogeneity case										
-0.5	10	-0.2809	0.6482	0.6433	0.6482	0.6482	-0.3155	0.8874	0.8608	0.8612
	20	-0.2882	0.5740	0.5802	0.5742	0.5740	-0.3161	0.8030	0.8170	0.7996
	30	-0.2564	0.5244	0.5440	0.5214	0.5244	-0.2738	0.7776	0.7816	0.7824
	50	-0.2614	0.4842	0.5046	0.4840	0.4842	-0.2717	0.6538	0.6546	0.6492
	100	-0.2529	0.3660	0.3850	0.3658	0.3660	-0.2576	0.5572	0.5500	0.5542
0	10	0.2281	0.6122	0.6045	0.5918	0.5924	0.2060	0.8780	0.8986	0.8886
	20	0.2426	0.5478	0.5344	0.5488	0.5482	0.2382	0.8748	0.8954	0.8866
	30	0.2466	0.5142	0.5026	0.5034	0.5140	0.2431	0.8636	0.8738	0.8708
	50	0.2452	0.4542	0.4466	0.4540	0.4542	0.2428	0.7266	0.7332	0.7278
	100	0.2498	0.3110	0.3054	0.3108	0.3110	0.2474	0.5076	0.5112	0.5040
0.5	10	0.7232	0.6076	0.5721	0.5988	0.5980	0.7630	0.8644	0.8972	0.8648
	20	0.7371	0.5360	0.4958	0.5358	0.5362	0.7663	0.8718	0.8976	0.8802
	30	0.7489	0.4974	0.4648	0.4970	0.4972	0.7531	0.8526	0.8680	0.8482
	50	0.7546	0.4234	0.3898	0.4243	0.4234	0.7565	0.7940	0.8118	0.7990
	100	0.7586	0.2534	0.2288	0.2531	0.2532	0.7534	0.4506	0.4628	0.4628
Homogeneity case										
-0.5	10	-0.5411	0.9792	0.9574	0.9692	0.9732	-0.5580	0.8992	0.9078	0.9026
	20	-0.5226	0.9628	0.9592	0.9621	0.9613	-0.5582	0.9184	0.9458	0.9364
	30	-0.5104	0.9544	0.9475	0.9536	0.9544	-0.5275	0.9284	0.9506	0.9432
	50	-0.5008	0.9556	0.9514	0.9556	0.9552	-0.5211	0.9384	0.9488	0.9472
	100	-0.5046	0.9562	0.9512	0.9558	0.9562	-0.5087	0.9456	0.9502	0.9514
0	10	0.0112	0.9686	0.9620	0.9686	0.9683	0.0131	0.8918	0.9198	0.9058
	20	0.0072	0.9542	0.9484	0.9534	0.9542	0.0091	0.9202	0.9430	0.9354
	30	-0.0004	0.9526	0.9488	0.9524	0.9526	0.0008	0.9296	0.9418	0.9428
	50	0.0056	0.9508	0.9494	0.9506	0.9508	0.0065	0.9394	0.9488	0.9424
	100	0.0053	0.9574	0.9562	0.9576	0.9571	0.0057	0.9490	0.9508	0.9502
0.5	10	0.5129	0.9644	0.9583	0.9640	0.9644	0.5603	0.8970	0.9230	0.9000
	20	0.5018	0.9552	0.9474	0.9550	0.9545	0.5305	0.9252	0.9470	0.9394
	30	0.5079	0.9496	0.9478	0.9496	0.9492	0.5243	0.9270	0.9402	0.9398
	50	0.5021	0.9496	0.9498	0.9494	0.9496	0.5136	0.9388	0.9438	0.9474
	100	0.5013	0.9532	0.9484	0.9530	0.9532	0.5065	0.9456	0.9476	0.9480

Table A2. The mean of MH estimate and coverage probability (CP) of the 95% confidence interval for the risk difference using variance formula-based and bootstrap methods ($\lambda_{i0} = 0.01$).

		Formula-based method			Bootstrap method		
RD	k	$\widehat{RD}_{MH}$	$CP_{rd.st}$	$CP_{rd.gr}$	$\widehat{RD}_{MH}$	$CP_{rd.bp}$	$CP_{rd.bv}$
Heterogeneity case							
-0.02	10	-0.0204	0.8243	0.8219	-0.0204	0.8774	0.8715
	20	-0.0201	0.8245	0.8227	-0.0201	0.9188	0.9057
	30	-0.0203	0.8259	0.8229	-0.0203	0.9294	0.9316
	50	-0.0199	0.8237	0.8221	-0.0199	0.9426	0.9332
	100	-0.0198	0.8293	0.8263	-0.0198	0.9422	0.9448
0.02	10	0.0199	0.8345	0.8309	0.0199	0.8972	0.8852
	20	0.0198	0.8343	0.8317	0.0198	0.9254	0.9100
	30	0.0198	0.8397	0.8377	0.0198	0.9354	0.9348
	50	0.0200	0.8447	0.8423	0.0200	0.9398	0.9408
	100	0.0199	0.8433	0.8407	0.0199	0.9424	0.9446
Homogeneity case							
0.04	10	0.0399	0.9440	0.9420	0.0399	0.8954	0.8956
	20	0.0400	0.9500	0.9484	0.0400	0.9258	0.9256
	30	0.0399	0.9522	0.9500	0.0399	0.9386	0.9414
	50	0.0399	0.9550	0.9534	0.0399	0.9448	0.9470
	100	0.0399	0.9472	0.9456	0.0399	0.9410	0.9432

Appendix B. Data set

Two meta-analytic data sets used in the application of this paper are given in Tables B1 and B2 as follows.

Table B1. Meta-analytic data on prophylactic antibiotics treatment and placebo (no prophylactic antibiotics) in caesarean section.

Report, year	Treatment event/ size	Placebo event/ size	Report, year	Treatment event/ size	Placebo event/ size
Adeleye et al., 1981	11/ 58	14/ 48	Leonetti et al., 1989	0/ 100	1/ 50
Bibi et al., 1994	4/ 133	28/ 136	Levin et al., 1983	0/ 85	3/ 43
Chan et al., 1989	27/ 299	12/ 101	Lewis et al., 1990	1/ 36	1/ 25
Conover et al., 1984	2/ 68	1/ 56	Lewis et al., 1990	2/ 76	4/ 75
Cormier et al., 1989	5/ 55	8/ 55	Mahomed et al., 1988	12/ 115	15/ 117
Dashow et al., 1986	3/ 100	0/ 33	Mallaret et al., 1990	6/ 136	16/ 130
Dashow et al., 1986	4/ 183	3/ 44	McCowan et al., 1980	9/ 35	7/ 38
De Boer et al., 1989	1/ 11	5/ 17	Miller et al., 1968	13/ 150	23/ 150
De Boer et al., 1989	10/ 80	21/ 74	Moodley et al., 1981	2/ 40	4/ 20
Dillon et al., 1981	0/ 46	4/ 55	Moro et al., 1974	0/ 74	2/ 74
Duff et al., 1980	0/ 26	1/ 31	Padilla et al., 1983	0/ 34	5/ 37
Duff et al., 1982	0/ 42	0/ 40	Phelan et al., 1979	2/ 61	2/ 61
Elliot et al., 1986	0/ 119	1/ 39	Polk et al., 1982	3/ 146	9/ 132
Engel et al., 1986	1/ 50	9/ 50	Rehu et al., 1980	4/ 88	4/ 40
Fugere et al., 1983	2/ 60	6/ 30	Roex et al., 1986	1/ 64	7/ 65
Gall, 1979	1/ 46	1/ 49	Ross et al., 1984	7/ 57	7/ 58
Gerstner et al., 1980	3/ 53	9/ 50	Rothbard et al., 1975	0/ 16	1/ 16
Gibbs et al., 1972	0/ 33	4/ 28	Rothbard et al., 1975	2/ 31	6/ 37
Gibbs et al., 1973	0/ 34	6/ 34	Ruiz-Moreno et al., 1991	1/ 50	4/ 50
Gibbs et al., 1981	0/ 50	2/ 50	Saltzman et al., 1985	1/ 50	2/ 49
Gordon et al., 1979	0/ 78	1/ 36	Schedvins et al., 1986	2/ 26	0/ 27
Hager et al., 1983	1/ 43	1/ 47	Stage et al., 1983	3/ 133	12/ 66
Hagglund et al., 1989	0/ 80	3/ 80	Stiver et al., 1983	6/ 244	17/ 117
Harger et al., 1981	2/ 196	14/ 190	Tully et al., 1983	1/ 52	2/ 61
Hawrylyshyn et al., 1983	2/ 124	2/ 58	Tzingounis et al., 1982	2/ 46	4/ 50
Ismail et al., 1990	2/ 74	8/ 78	Weissberg et al., 1971	0/ 40	3/ 40
Jakobi et al., 1994	4/ 167	5/ 140	Wong et al., 1978	2/ 48	3/ 45
Karhunen et al., 1985	2/ 75	9/ 77	Work et al., 1977	3/ 40	1/ 40
Kreutner et al., 1978	0/ 48	2/ 49	Yip et al., 1997	1/ 160	1/ 160
Kristensen et al., 1990	0/ 102	1/ 99	Young et al., 1983	1/ 50	4/ 50
Lapas et al., 1989	1/ 50	10/ 50			

Table B2. Meta-analytic data on the association of Middle East respiratory syndrome (MERS) exposure proximity with infection.

Report, year	Treatment event ($>$ 1 m)	Treatment size	Control event ($\leq$ 1 m)	Control size
Van Kerkhove et al., 2019	8	774	11	54
Arwady et al., 2016	1	10	8	20
Ki et al., 2019	2	29	4	42
Park et al., 2016	0	3	5	25
Hall et al., 2014	0	5	0	43
Wiboonchutikul et al., 2019	0	16	0	22
Reuss et al., 2014	0	12	0	69
Ryu et al., 2019	0	7	0	27