MMR MILITARY MEDICAL RESEARCH

METHODOLOGY   **Open Access**

Meta-analysis with zero-event studies: a comparative study with application to COVID-19 data

Jia-Jin Wei[1], En-Xuan Lin[2], Jian-Dong Shi[1], Ke Yang[1], Zong-Liang Hu[3], Xian-Tao Zeng[4] and Tie-Jun Tong[1*]

Abstract

Background: Meta-analysis is a statistical method to synthesize evidence from a number of independent studies, including those from clinical studies with binary outcomes. In practice, when there are zero events in one or both groups, it may cause statistical problems in the subsequent analysis.

Methods: In this paper, by considering the relative risk as the effect size, we conduct a comparative study that consists of four continuity correction methods and another state-of-the-art method without the continuity correction, namely the generalized linear mixed models (GLMMs). To further advance the literature, we also introduce a new method of the continuity correction for estimating the relative risk.

Results: From the simulation studies, the new method performs well in terms of mean squared error when there are few studies. In contrast, the generalized linear mixed model performs the best when the number of studies is large. In addition, by reanalyzing recent coronavirus disease 2019 (COVID-19) data, it is evident that the double-zero-event studies impact the estimate of the mean effect size.

Conclusions: We recommend the new method to handle the zero-event studies when there are few studies in a meta-analysis, or instead use the GLMM when the number of studies is large. The double-zero-event studies may be informative, and so we suggest not excluding them.

Keywords: Continuity correction, Coronavirus disease 2019 data, Meta-analysis, Relative risk, Zero-event studies

Background

Meta-analysis is a statistical method to synthesize evidence from a number of independent studies that addressed the same scientific questions [1, 2]. In clinical studies, experimental data are commonly composed of binary outcomes, and consequently, meta-analyses of binary data have attracted increasing attention in evidence-based medicine [3, 4]. For each study, an effect size is reported to quantify the treatment effect by comparing the event probabilities between the treatment group and the control group, including the odds ratio (OR), the relative risk (RR), and the risk difference (RD). In meta-analysis, when the study-specific effect size is estimated based on a two-by-two contingency table, the zero-event problem in one or both groups frequently occurs, which may cause an unexpected calculation complication in the statistical inference of the effect size. If the study involves a zero event in one group, we refer to it as a *single-zero-event study*; and if the study involves zero events in both groups, we refer to it as a *double-zero-event study* [5]. Vandermeer et al. [6] and Kuss [7] applied random sampling techniques and found that 30% of meta-analyses from the 500 sampled Cochrane reviews included one or more single-zero-event studies, while 34% of the reviews involved at least one meta-analysis with a double-zero-event study.

*Correspondence: tongt@hkbu.edu.hk
[1]Department of Mathematics, Hong Kong Baptist University, Hong Kong, China
Full list of author information is available at the end of the article

BMC

As a recent example, Chu et al. [8] conducted several meta-analyses to evaluate the effectiveness of physical distancing, face masks, and eye protection on the spread of three coronaviruses, which caused severe acute respiratory syndrome (SARS), Middle East respiratory syndrome (MERS) or coronavirus disease 2019, also known as COVID-19 [9, 10]. Specifically, they considered RR as the effect size and applied the random-effects model to pool the observed effect sizes with an inverse-variance weight assigned to each study [11, 12]. As a result, for their meta-analysis on physical distancing, they concluded that the risk of infection will be significantly decreased with a further physical distance. We note, however, that there are 8 single-zero-event studies and 7 double-zero-event studies among a total of 32 studies. In particular for the 7 studies on COVID-19 data, 4 of them are single-zero-event studies and 2 of them are double-zero-event studies. To escape the zero-event problem, Chu et al. [8] excluded the double-zero-event studies from their meta-analyses, which, however, may introduce an estimation bias to the overall effect size [7]. More recently, Xu et al. [13] revisited 442 meta-analyses with or without the double-zero-event studies, and then by a comparative study, they concluded that the double-zero-event studies do contain valuable information and should not be excluded from the meta-analysis.

Inspired by the aforementioned examples, we provide a selective review on the existing methods for meta-analysis that can handle the zero-event studies. For ease of presentation, we will mainly focus on the random-effects model with RR as the effect size, whereas the same comparison also applies to OR and RD. For more details on meta-analysis of OR and RD with the zero-event studies, one may refer to [7] and the references therein, in which the author discussed the methods applicable to all the three effect sizes as well as some methods only applicable to one of them. For a given study, we let n_1 be the number of samples in the treatment group with X_1 being the number of events, and n_2 be the number of samples in the control group with X_2 being the number of events. Let also X_1 follow a binomial distribution with parameters n_1 and $p_1 > 0$, and X_2 follow a binomial distribution with parameters n_2 and $p_2 > 0$. We further assume that X_1 and X_2 are independent of each other. Then to estimate RR $= p_1/p_2$, the maximum likelihood estimator is known as

$$\widehat{\text{RR}} = \frac{X_1/n_1}{X_2/n_2} = \frac{X_1 n_2}{X_2 n_1} \qquad (1)$$

Note that $\widehat{\text{RR}}$ is often right-skewed. To derive the statistical inference on RR, researchers frequently apply the log scale so that the resulting estimator can be more normally distributed. Specifically by Agresti [14], the approximate variance of $\ln\left(\widehat{\text{RR}}\right)$ is

$$\text{var}\left[\ln\left(\widehat{\text{RR}}\right)\right] \approx \frac{1}{X_1} - \frac{1}{n_1} + \frac{1}{X_2} - \frac{1}{n_2} \qquad (2)$$

By (1) and (2), when there are zero events in one or both groups, the classic method for estimating RR suffers from the zero-event problem and will no longer be applicable.

To have a valid estimate of RR, originated from Haldane [15], one often recommends to add 0.5 to the counts of events and non-events if some count is zero [16, 17]. This method is referred to a correction method and has been extensively used in meta-analysis to deal with the zero-event studies. For further developments on the continuity correction, one may refer to Sweeting et al. [18], Carter et al. [19], and the references therein. On the other side, there are also statistical models without the continuity correction to handle meta-analysis with the zero-event studies, such as the generalized linear mixed models [4, 20, 21].

The remainder of this paper is organized as follows. In "Methods with the continuity correction" section, we first review the random-effects model and the existing methods with the continuity correction, and then propose a new method of the continuity correction for estimating RR. In "The generalized linear mixed models" section, we review the generalized linear mixed models for meta-analysis. In "Simulation studies" section, we conduct simulation studies to evaluate the performance of the reviewed methods and our new method. In "Application to COVID-19 data" section, we apply all the well performed methods to a recent meta-analysis on COVID-19 data for further evaluation of their performance. We then conclude the paper in "Discussion" and "Conclusions" sections with some interesting findings, and provide the supplementary materials in the Appendix.

Methods
Methods with the continuity correction
Suppose that there are k studies in the meta-analysis, and y_i for $i = 1, \ldots, k$ are the observed effect sizes for each study. By DerSimonian and Laird [22], the random-effects model can be expressed as

$$y_i = \theta + \zeta_i + \epsilon_i \qquad (3)$$

where θ is the mean effect size, ζ_i are the deviations of each study from θ, and ϵ_i are the sampling errors. We further assume that ζ_i are independent and identically distributed random variables from $N(0, \tau^2)$, ϵ_i are independent random errors from $N(0, \sigma_i^2)$, and that they are independent of each other. In addition, τ^2 is referred to as the between-study variance, and σ_i^2 are referred to as the within-study variances.

For the random-effects model in (3), by the inverse-variance method the mean effect size θ can be estimated by

$$\hat{\theta} = \frac{\sum_i w_i^* y_i}{\sum_i w_i^*} \qquad (4)$$

where $w_i^* = 1/\left(\sigma_i^2 + \tau^2\right)$ are the weights assigned to each individual study [23]. In meta-analysis, the within-study variances σ_i^2 are routinely estimated by the variances of the observed effect sizes, denoted by $var(y_i)$. While for the between-study variance, DerSimonian and Laird [22] proposed the method of moments estimator as

$$T^2 = \frac{Q - k + 1}{C} \qquad (5)$$

where $Q = \sum_i w_i \left(y_i - \sum_i w_i y_i / \sum_i w_i\right)^2$ is known as the Q statistic, and $C = \sum_i w_i - \sum_i w_i^2 / \sum_i w_i$ with $w_i = 1/\sigma_i^2$ for $i = 1, \dots, k$.

We note, however, that the random-effects model may suffer from the zero-event problem. Taking RR as an example, if we apply the random-effects model for meta-analysis, then the effect sizes y_i will be the observed $\ln(RR)$ values. Now for estimating $\ln(RR)$, if we plug in $\widehat{RR}$ from formula (1) directly, then $\ln\left(\widehat{RR}\right)$ will not be well defined when the studies involve the zero events, and so is for the variance estimate of $\ln\left(\widehat{RR}\right)$ in formula (2). Consequently, without a valid estimate of the effect size and of its within-study variance, the random-effects model cannot be applied to estimate the mean effect size by the inverse-variance method. This shows that a correction on $\widehat{RR}$ is often desired in meta-analysis with some studies involving zero events.

Existing methods with the continuity correction
Let $c_1 > 0$ and $c_2 > 0$ be two values for the continuity correction. To overcome the zero-event problem, one common approach is to estimate p_1 by $(X_1 + c_1)/(n_1 + 2c_1)$ and estimate p_2 by $(X_2 + c_2)/(n_2 + 2c_2)$. Plugging them into (1) and (2), we have

$$\widetilde{RR}\left(c_1, c_2\right) = \frac{X_1 + c_1}{n_1 + 2c_1} \cdot \frac{n_2 + 2c_2}{X_2 + c_2} \qquad (6)$$

Accordingly, the 95% confidence interval (CI) of RR is

$$\exp\left\{\ln\left(\widetilde{RR}\left(c_1, c_2\right)\right) \pm 1.96 \sqrt{\frac{1}{X_1 + c_1} - \frac{1}{n_1 + 2c_1} + \frac{1}{X_2 + c_2} - \frac{1}{n_2 + 2c_2}}\right\} \qquad (7)$$

For the values of c_1 and c_2 in (6), there are mainly three suggestions in the literature that are widely used for the random-effects meta-analysis.

(i) When $c_1 = c_2 = 0.5$, it yields the Haldane estimator [15] as

$$\widetilde{RR}_{Haldane} = \begin{cases} \dfrac{X_1 + 0.5}{n_1 + 1} \cdot \dfrac{n_2 + 1}{X_2 + 0.5} & X_1 = 0 \text{ or } n_1, X_2 = 0 \text{ or } n_2, \\ \dfrac{X_1 n_2}{n_1 X_2} & \text{otherwise} \end{cases} \qquad (8)$$

(ii) When $c_1 = n_1/(n_1 + n_2)$ and $c_2 = n_2/(n_1 + n_2)$, it yields the TACC estimator [18] as

$$\widetilde{RR}_{TACC} = \begin{cases} \dfrac{X_1 + c_1}{n_1 + 2c_1} \cdot \dfrac{n_2 + 2c_2}{X_2 + c_2} & X_1 = 0 \text{ or } n_1, X_2 = 0 \text{ or } n_2, \\ \dfrac{X_1 n_2}{X_2 n_1} & \text{otherwise} \end{cases} \qquad (9)$$

For the balanced case when $n_1 = n_2$, the TACC estimator is equivalent to the Haldane estimator. Also to implement this estimator, one may apply *metabin* in the R package "meta" with the setting incr = "TACC" [24].

(iii) When $c_1 = c_2 = 1$, it yields the Carter estimator [19] as

$$\widetilde{RR}_{Carter} = \frac{X_1 + 1}{n_1 + 2} \cdot \frac{n_2 + 2}{X_2 + 1} \qquad (10)$$

Besides the continuity correction methods in family (6), another alternative is to estimate p_1 by $(X_1 + c_1)/(n_1 + c_1)$ and estimate p_2 by $(X_2 + c_2)/(n_2 + c_2)$. Then with $c_1 = c_2 = 0.5$, it yields the Pettigrew estimator [25] as

$$\widetilde{RR}(0.5, 0.5) = \frac{X_1 + 0.5}{n_1 + 0.5} \cdot \frac{n_2 + 0.5}{X_2 + 0.5} \qquad (11)$$

and the 95% CI of RR as

$$\exp\left\{\ln[\widetilde{RR}(0.5, 0.5)] \pm 1.96 \sqrt{\frac{1}{X_1 + 0.5} - \frac{1}{n_1 + 0.5} + \frac{1}{X_2 + 0.5} - \frac{1}{n_2 + 0.5}}\right\} \qquad (12)$$

Moreover, to avoid a zero standard error, Hartung and Knapp [26] suggested not to correct X_1 and X_2 when $X_1 = n_1$ and $X_2 = n_2$.

A hybrid method with the continuity correction
Note that the existing methods are all constructed to first estimate p_1 and p_2, and then take their ratio as an estimate of RR $= p_1/p_2$. Nevertheless, noting that p_2 is in the denominator rather than in the numerator, inverting an optimal estimate for p_2 may not necessarily yield an optimal estimate for $1/p_2$. In this section, we propose a hybrid method that is to estimate p_1 and $1/p_2$ directly, and then take their product to estimate RR.

For the estimation of p_1, we show in Appendix 1 that the mean squared error (MSE) of $(X_1 + c_1)/(n_1 + 2c_1)$ is smaller than the MSE of $(X_1 + c_1)/(n_1 + c_1)$ in most settings. We thus consider to apply $(X_1 + c_1)/(n_1 + 2c_1)$

to estimate p_1 in RR. While to estimate the reciprocal of p_2, one may consider $(n_2 + 2c_2)/(X_2 + c_2)$ as in (6). Or instead, another option can be to consider $(n_2 + c_2)/(X_2 + c_2)$ as originated in (11), see also [27] and [28] for more discussion. And if we take the latter one, then a hybrid estimator of RR can be constructed as

$$\widehat{\text{RR}}(c_1, c_2) = \frac{X_1 + c_1}{n_1 + 2c_1} \cdot \frac{n_2 + c_2}{X_2 + c_2} \qquad (13)$$

For the optimal values of c_1 and c_2 in (13), our simulation studies in Appendices 2 and 3 show that $c_1 = 0.5$ and $c_2 = 0.5$ are among the best options. In view of this, our new hybrid estimator is taken as follows:

$$\widehat{\text{RR}}(0.5, 0.5) = \frac{X_1 + 0.5}{n_1 + 1} \cdot \frac{n_2 + 0.5}{X_2 + 0.5} \qquad (14)$$

whereas the 95% CI of RR is given as

$$\exp\left\{ \ln\left(\widehat{\text{RR}}(0.5, 0.5)\right) \pm 1.96 \sqrt{\frac{1}{X_1 + 0.5} - \frac{1}{n_1 + 1} + \frac{1}{X_2 + 0.5} - \frac{1}{n_2 + 0.5}} \right\} \qquad (15)$$

Comparison of the continuity correction methods

In this section, we conduct a numerical study to compare the finite sample performance of the existing and new methods. For ease of presentation, we refer to the confidence intervals associated with (8), (9), (10), (11) and (14) as the Haldane interval, the TACC interval, the Carter interval, the Pettigrew interval, and the hybrid interval, respectively.

To generate the data, we let $p_2 = 0.05$, 0.15, 0.85 or 0.95, and $p_1 = p_2 \times \text{RR}$ with RR ranging from 0.2 to min$\{5, 1/p_2\}$. We also consider different combinations of the sample sizes. For the sake of brevity, only the results for balanced samples with $n_1 = n_2 = 10$ or 50 are presented, whereas the results for the unbalanced samples are postponed to Appendix 4. Recall that the Haldane and TACC intervals are the same when $n_1 = n_2$, and we thus present the results for the Haldane interval only. With $N = 100,000$ repetitions for each setting, we generate random numbers from the binomial distributions with parameters (p_1, n_1) and (p_2, n_2) to yield the estimates of RR and their CIs. We then compute the frequencies of the true RR falling in the CIs as the coverage probability estimates. Moreover, the expected lengths of the CIs on the log scale are computed by $N^{-1} \sum_{s=1}^{N} (\ln(\text{UL}_s) - \ln(\text{LL}_s))$, where UL_s and LL_s are the upper and lower limits of the sth CI.

For $p_2 = 0.05$ or 0.15, the top four panels of Figs. 1 and 2 show that the Haldane interval is the most conservative interval in most settings, and it provides the longest expected lengths compared to the other three intervals. The Carter interval may have downward spikes in the left or right tail, although it leads to the shortest expected lengths. We also note that the simulation results of the

Pettgrew interval and the hybrid interval are nearly the same. Their coverage probabilities and expected lengths are intermediate between those of the other two intervals in most settings.

From the bottom four panels of Figs. 1 and 2 with $p_2 = 0.85$ or 0.95, it is evident that the Haldane interval has a satisfactory performance in most settings with the coverage probabilities around the nominal level. In contrast, the Carter interval fails to provide enough large coverage probabilities in most settings, so does the Pettgrew interval when n_1 and n_2 are small. Note also that the coverage probabilities of the hybrid interval are comparable to the Haldane interval as long as p_2 is not extremely large. Moreover, the hybrid interval yields shorter expected lengths than the Haldane interval.

To sum up, when p_2 is small, the Pettgrew interval and the hybrid interval are less conservative than the Haldane interval in most settings. While for large p_2, the Haldane interval and the hybrid interval perform better than the Pettgrew interval in terms of coverage probability. In addition, the expected lengths of the hybrid interval are always shorter than the Haldane interval. This shows that the hybrid interval can serve as a good alternative for the interval estimation of RR.

The generalized linear mixed models

The generalized linear mixed models (GLMMs) are extensions of the generalized linear model, which include both the fixed and random effects as linear predictors [14]. Different types of the GLMMs have been proposed in the literature including a few reviews and comparison studies [4, 29]. Among the existing models, the bivariate GLMM has been well recognized and being recommended for estimating RR in meta-analysis [20].

Let p_{i1} and p_{i2} be the event probabilities in the treatment and control groups of the ith study, respectively. The bivariate GLMM is represented as

$$\begin{aligned} g(p_{i1}) &= \Omega_1 + \zeta_{i1} \\ g(p_{i2}) &= \Omega_2 + \zeta_{i2} \end{aligned} \qquad (16)$$

where $g(\cdot)$ is the link function, Ω_1 and Ω_2 are the fixed effects, and the random effects are given by

$$\begin{pmatrix} \zeta_{i1} \\ \zeta_{i2} \end{pmatrix} \overset{\text{ind}}{\sim} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_1^2 & \rho\tau_1\tau_2 \\ \rho\tau_1\tau_2 & \tau_2^2 \end{pmatrix} \right]$$

The mean effect size based on model (16) was defined as

$$\text{RR}_{\text{GLMM}} = \frac{E(p_1)}{E(p_2)} = \frac{\int_{-\infty}^{\infty} g^{-1}(\Omega_1 + t) \tau_1^{-1} \phi(t/\tau_1) \, dt}{\int_{-\infty}^{\infty} g^{-1}(\Omega_2 + t) \tau_2^{-1} \phi(t/\tau_2) \, dt} \qquad (17)$$

where $E(p_1)$ and $E(p_2)$ are the mean event probabilities in the control and treatment groups, $g^{-1}(\cdot)$ is the inverse function of the link, and $\phi(\cdot)$ is the probability density function of the standard normal distribution
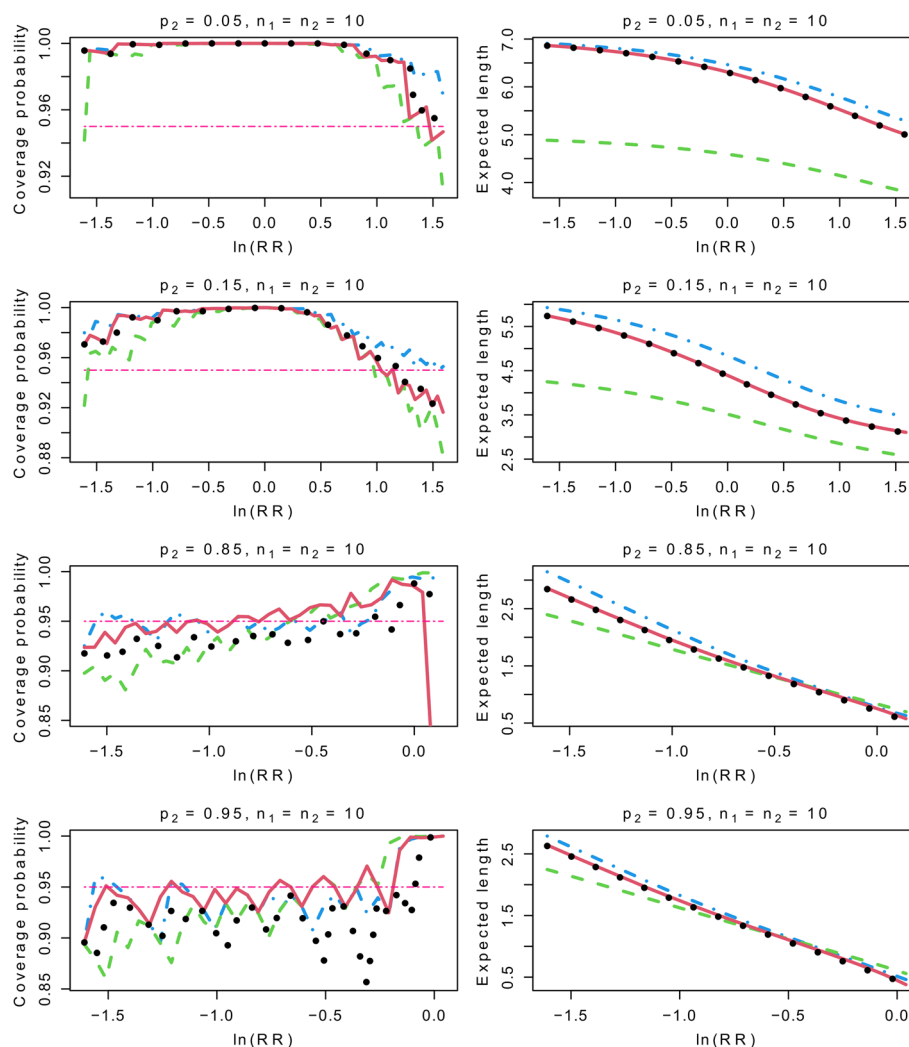
Fig. 1 Comparison of the four CIs of RR with $p_2 = 0.05$, 0.15, 0.85 or 0.95, and $n_1 = n_2 = 10$. The dot-dashed lines represent the simulation results of the Haldane interval, the dashed lines represent the simulation results of the Carter interval, the dotted lines represent the simulation results of the Pettigrew interval, and the solid lines represent the simulation results of the hybrid interval. CI: Confidence interval, RR: Relative risk

[30]. For the logit link, Zeger et al. [31] proposed an approximate formula $E(p_j) \approx \text{expit}\left(\Omega_j / \sqrt{1 + C^2 \tau_j^2}\right)$ with $C = 16\sqrt{3}/(15\pi)$. For the probit link, $E(p_j) = \Phi\left(\Omega_j / \sqrt{1 + \tau_j^2}\right)$, where $j = 1$ or 2, and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. While for the other links, there does not exist a closed form of formula (17) and so a numerical approximation is often needed [32].

For the parameter estimation in model (16), Jackson et al. [4] provided a detailed introduction for the implementation based on the R package "lme4" in their model 6. Alternatively, one may also apply the function *meta.biv* in the R package "altmeta" maintained by Lin and Chu [33], in which the 95% CI of RR can be derived by the bootstrap resampling method.

Results

Simulation studies

In this section, we compare the performance of the reviewed methods on handling meta-analysis with the zero-event studies, including the continuity correction methods and the generalized linear mixed models. Among the existing continuity correction methods, we note that the Haldane and TACC estimators are comparable and among the best when estimating the mean effect size, in contrast to the other two methods including the Carter and Pettigrew estimators. Hence, for the sake of brevity, we only present the results of the Haldane and TACC estimators in the main text but provide the simulation results for all four methods in Appendix 5. Besides the Haldane and TACC estimators, we also consider the newly introduced hybrid estimator
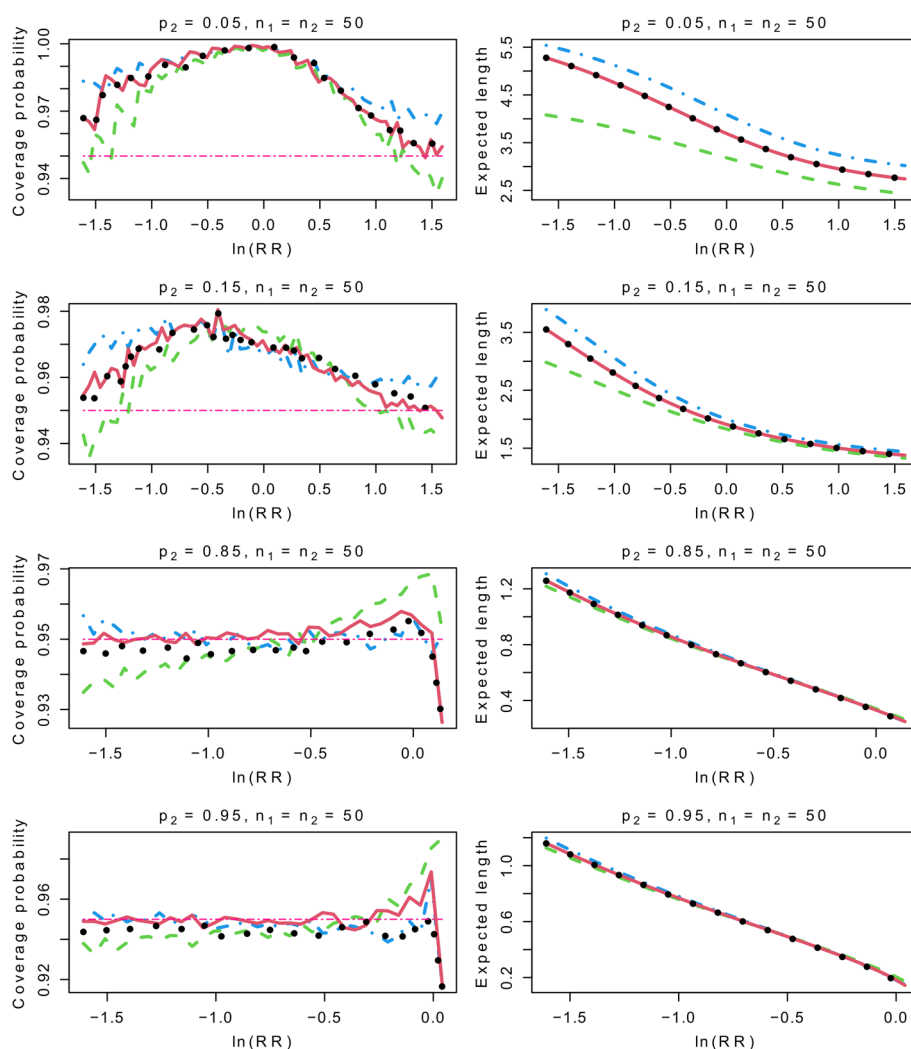
Fig. 2 Comparison of the four CIs of RR with $p_2 = 0.05$, 0.15, 0.85 or 0.95, and $n_1 = n_2 = 50$. The dot-dashed lines represent the simulation results of the Haldane interval, the dashed lines represent the simulation results of the Carter interval, the dotted lines represent the simulation results of the Pettigrew interval, and the solid lines represent the simulation results of the hybrid interval. CI: Confidence interval, RR: Relative risk

and the GLMM with the logit link for further comparison.

To conduct the meta-analysis, we consider $k = 3$, 6 and 12 as three different numbers of studies. Also by (3), we let $\theta = \ln(RR)$ be the mean effect size that ranges from $\ln(0.2)$ to $\ln(5)$, and then generate the random effects ζ_i from $N(0, \tau^2)$ with $\tau^2 = 0.25$ or 1. Next, we randomly generate n_{i2} from the log-normal distribution based on the assumption that $\ln(n_{i2}) \overset{\text{ind}}{\sim} N(3.35, 1.00)$ [34]. It is also assumed by [34] that the ratios between n_{i1} and n_{i2} follow the uniform distribution with values from 0.84 to 2.04. In addition, we generate the event probabilities of the control group p_{i2} from the uniform distribution with values from 0.01 to $\min\{0.99, 1/\exp(\theta)\}$. Then accordingly, the event probabilities of the treatment group are given by $p_{i1} = \exp(\theta + \zeta_i) p_{i2}$, where $\exp(\theta + \zeta_i) p_{i2} \geqslant 1$ will be

discarded. Finally, we generate X_{i1} and X_{i2} from the binomial distributions with parameters (n_{i1}, p_{i1}) and (n_{i2}, p_{i2}), respectively. Note that the data will be re-generated if the number of events or non-events in one group are both zero. Finally, with $N = 10,000$ repetitions for each setting, we compute the mean squared errors (MSEs) between the estimated RR and the true RR to evaluate the accuracy of the methods.

From the top two panels of Fig. 3, it is evident that the three continuity correction methods perform much better than the GLMM in nearly all settings when k is small. Moreover, the hybrid estimator is consistently better than the Haldane and TACC estimators. The middle two panels show that, when k is moderate, the three continuity correction methods still perform better than the GLMM in most settings. Finally, the bottom two panels indicate
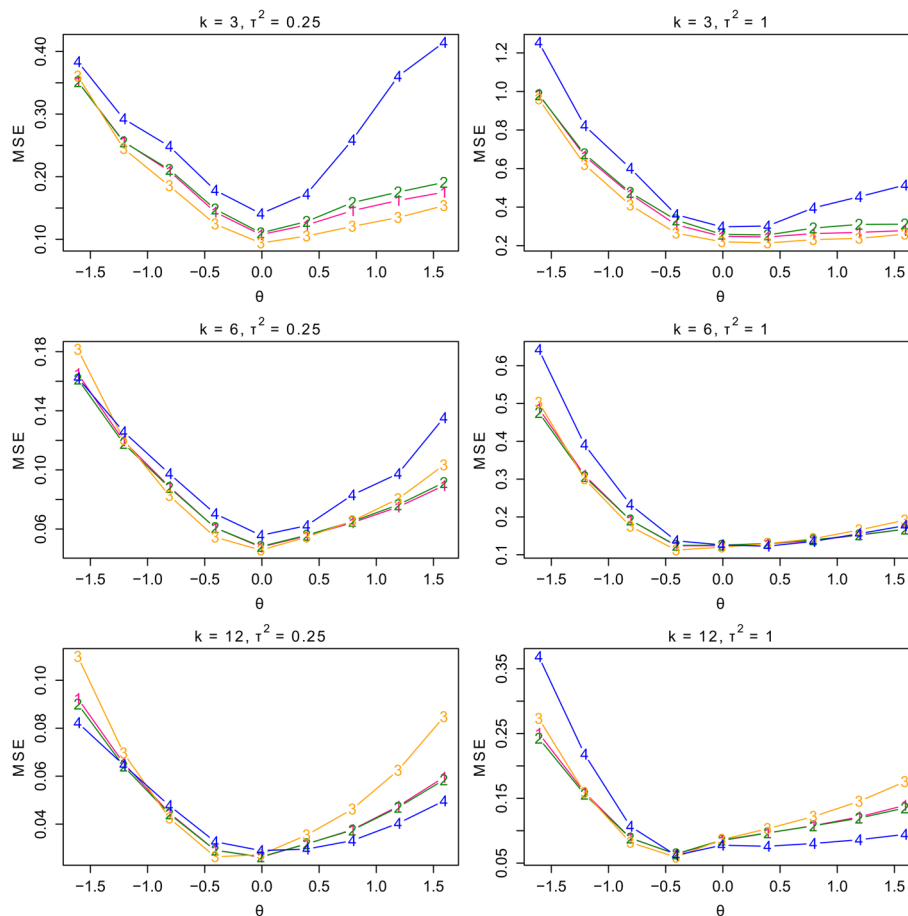
Fig. 3 Comparison of the four methods with $k = 3, 6$ or 12, $\tau^2 = 0.25$ or 1. "1" represents the results of the random-effects model with the Haldane estimator, "2" represents the results of the random-effects model with the TACC estimator, "3" represents the results of the random-effects model with the hybrid estimator, and "4" represents the results of the GLMM. TACC: Treatment arm continuity correction, GLMM: Generalized linear mixed model, MSE: Mean squared error

that the GLMM performs the best in most settings when k is large. To conclude, the accuracy of the different methods depends on the number of studies. In particular, for meta-analysis with few studies, the random-effects model with the hybrid estimator is more reliable for handling the zero-event studies than the other methods; and for meta-analysis with large studies, we recommend the GLMM to handle the random-effects meta-analysis.

Application to COVID-19 data

As mentioned earlier, Chu et al. [8] conducted a systematic review that revealed the connections of physical distancing, face masks, and eye protection with the transmission of SARS, MERS, and COVID-19. It is noteworthy that their analytical results have attracted more and more attention. As an evidence, their paper has received a total of 1236 citations in Google Scholar as of 16 March 2021. In this section, we propose to reanalyze COVID-19 data and compare the performance of the different methods with or without the double-zero-event studies, including the Haldane estimator, the TACC estimator, the hybrid estimator, and the GLMMs.

Note that the treatment group represents a further physical distance and the control group represents a shorter physical distance. As shown in the top panel of Fig. 4, [8] applied the random-effects model with the Haldane estimator and removed the double-zero-event studies from their meta-analysis. The overall effect size of 0.15 with the 95% CI being $[0.03, 0.73]$ indicates that the infection risk will be significantly reduced with a further physical distance. The middle panel of Fig. 4 reports that the random-effects model with the TACC estimator yields the overall effect size of 0.12 with the 95% CI being $[0.03, 0.50]$. Moreover, the bottom panel of Fig. 4 shows that the random-effects model with the hybrid estimator yields the overall effect size of 0.13 with the 95% CI being $[0.03, 0.72]$. Note also that the study-specific CIs here are always narrower than the CIs in the top panel, which
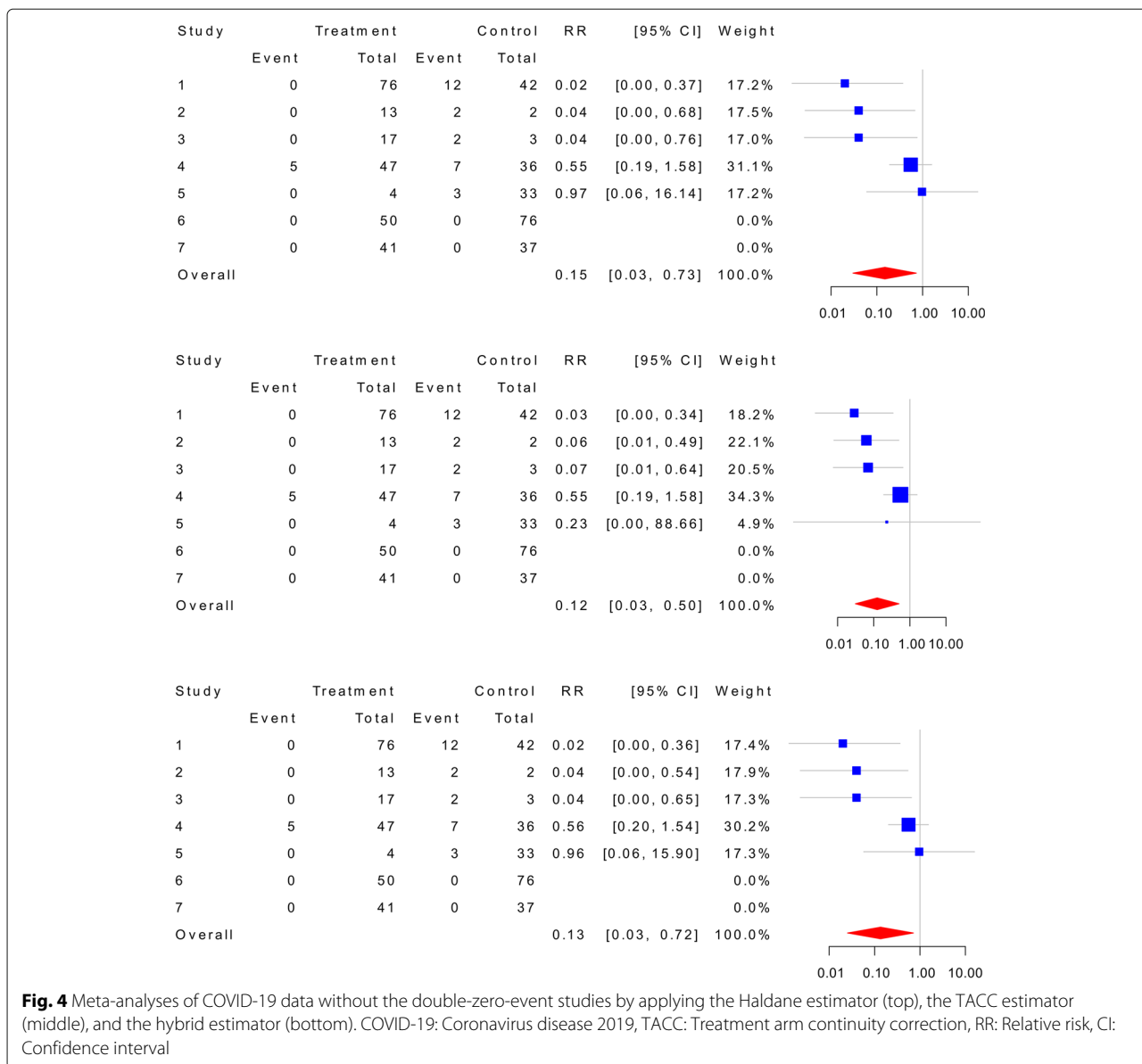
Fig. 4 Meta-analyses of COVID-19 data without the double-zero-event studies by applying the Haldane estimator (top), the TACC estimator (middle), and the hybrid estimator (bottom). COVID-19: Coronavirus disease 2019, TACC: Treatment arm continuity correction, RR: Relative risk, CI: Confidence interval

coincides with the simulation results that the expected lengths of the CI associated with the hybrid estimator are shorter than the Haldane estimator. In addition, the GLMM in (16) does not provide the estimates of the study-specific effect sizes, so the results are listed as follows. By the bootstrap resampling with 1000 replicates, the GLMM with the logit link yields the overall effect size of 0.20 with the 95% bootstrap CI being [0.05, 0.55]. Also, the GLMM with the probit link yields the overall effect size of 0.18 with the 95% CI being [0.04, 0.55].

To reanalyze COVID-19 data, we now include the double-zero-event studies. The top panel of Fig. 5 shows that the random-effects model with the Haldane estimator yields the overall effect size of 0.22 with 95% CI being [0.06, 0.82]. The middle panel of Fig. 5 presents that

the random-effects model with the TACC estimator provides the overall effect size of 0.18 with the 95% CI being [0.06, 0.57]. While for the hybrid estimator, it is shown by the bottom panel that the overall effect size is 0.21 with 95% CI being [0.05, 0.81]. At last, the GLMM with the logit link provides the overall effect size of 0.29 with the 95% CI being [0.10, 0.64], and the GLMM with the probit link provides the overall effect size of 0.28 with the 95% CI being [0.10, 0.56].

Discussion
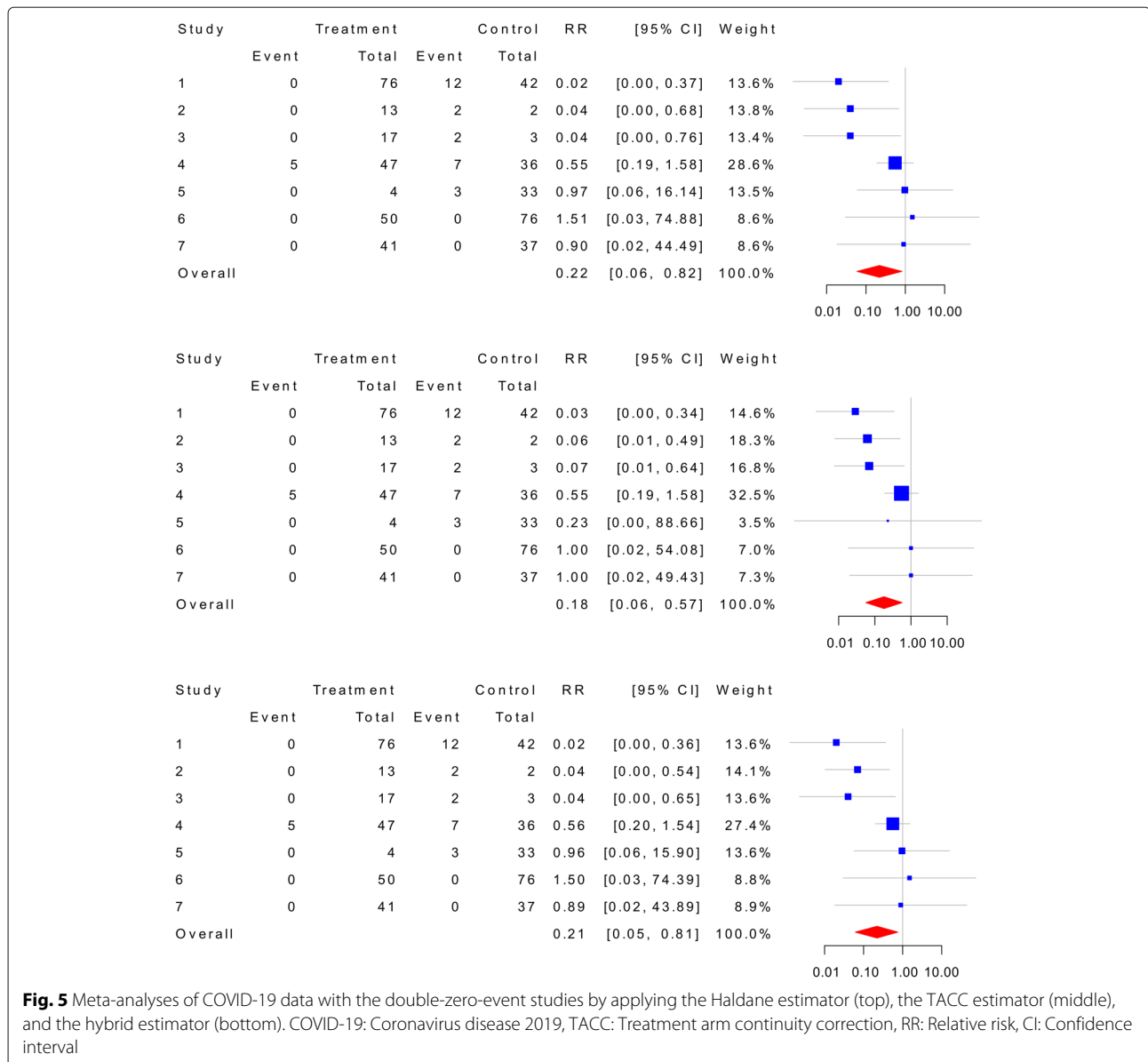
To handle the zero-event studies in meta-analysis of binary data, researchers often apply the random-effects model with the continuity correction, or instead, the GLMMs. From the simulation results, we note that the

Study		Treatment		Control	RR	[95% CI]	Weight	
	Event	Total	Event	Total				
1	0	76	12	42	0.02	[0.00, 0.37]	13.6%	
2	0	13	2	2	0.04	[0.00, 0.68]	13.8%	
3	0	17	2	3	0.04	[0.00, 0.76]	13.4%	
4	5	47	7	36	0.55	[0.19, 1.58]	28.6%	
5	0	4	3	33	0.97	[0.06, 16.14]	13.5%	
6	0	50	0	76	1.51	[0.03, 74.88]	8.6%	
7	0	41	0	37	0.90	[0.02, 44.49]	8.6%	
Overall					0.22	[0.06, 0.82]	100.0%	

0.01 0.10 1.00 10.00

Study		Treatment		Control	RR	[95% CI]	Weight	
	Event	Total	Event	Total				
1	0	76	12	42	0.03	[0.00, 0.34]	14.6%	
2	0	13	2	2	0.06	[0.01, 0.49]	18.3%	
3	0	17	2	3	0.07	[0.01, 0.64]	16.8%	
4	5	47	7	36	0.55	[0.19, 1.58]	32.5%	
5	0	4	3	33	0.23	[0.00, 88.66]	3.5%	
6	0	50	0	76	1.00	[0.02, 54.08]	7.0%	
7	0	41	0	37	1.00	[0.02, 49.43]	7.3%	
Overall					0.18	[0.06, 0.57]	100.0%	

0.01 0.10 1.00 10.00

Study		Treatment		Control	RR	[95% CI]	Weight	
	Event	Total	Event	Total				
1	0	76	12	42	0.02	[0.00, 0.36]	13.6%	
2	0	13	2	2	0.04	[0.00, 0.54]	14.1%	
3	0	17	2	3	0.04	[0.00, 0.65]	13.6%	
4	5	47	7	36	0.56	[0.20, 1.54]	27.4%	
5	0	4	3	33	0.96	[0.06, 15.90]	13.6%	
6	0	50	0	76	1.50	[0.03, 74.39]	8.8%	
7	0	41	0	37	0.89	[0.02, 43.89]	8.9%	
Overall					0.21	[0.05, 0.81]	100.0%	

0.01 0.10 1.00 10.00

Fig. 5 Meta-analyses of COVID-19 data with the double-zero-event studies by applying the Haldane estimator (top), the TACC estimator (middle), and the hybrid estimator (bottom). COVID-19: Coronavirus disease 2019, TACC: Treatment arm continuity correction, RR: Relative risk, CI: Confidence interval

performance of the different methods depends on the number of studies. For meta-analysis with few studies, the random-effects model with the continuity correction is able to perform better than the GLMM, especially the hybrid continuity correction. We also note that the hybrid continuity correction can yield a reliable confidence interval for a single RR. Although the continuity correction does show some advantages, it should be used with caution since an arbitrary correction may lead to a bias or even reverse the result of a meta-analysis, especially when the numbers of samples in the two groups are fairly unbalanced [7, 13]. When the number of studies is large, the GLMM is preferable to the random-effects model with the continuity correction. In other words, the performance of the GLMM relies on a sufficient number of studies [35].

Also as shown in Ju et al. [34], the GLMM also requires enough total events in the two groups, e.g., larger than 10.

Besides the random-effects model we have compared, it is noteworthy that there are also other models for meta-analysis that can handle the zero-event studies including, for example, the beta-binomial model [36–38]. Most meta-analyses with rare events have a small degree of heterogeneity, and so the common-effect model may be more suitable than the random-effects model [39]. In addition, Li and Rice [40] showed that the fixed-effects model can also provide an accurate CI for meta-analysis of OR with the zero-event studies. Apart from that, it is also noteworthy that the fixed-effects model can serve as a convincing model for meta-analysis with few studies [12, 41–43]. As a future work, it can be interesting to investigate the best

model for meta-analysis with few studies which include the zero-event studies as well.

For the double-zero-event studies in meta-analysis, we have shown by reanalyzing COVID-19 data that they do impact the estimate of the mean effect size, and so they may not be uninformative. As noted by Friedrich et al. [44], including the double-zero-event studies moves the mean effect size estimate toward the direction of the null hypothesis. If one arbitrarily excluded the informative double-zero-event studies, there would be a risk of overstating the treatment effect such that the conclusion would be less reliable. As recommended by the literature [7, 13] and the references therein, we suggest including the double-zero-event studies in meta-analysis.

Apart from model comparison, the selection of effect sizes has attracted more and more attention in the literature. In particular, there is a recent debate on the choice of RR or OR in clinical epidemiology, in which a number of important properties of RR or OR together with their pros and cons were discussed including, for example, portability and collapsibility [45–47]. In view of this, we have also analyzed COVID-19 data with OR being the effect size and present the results in Appendix 6 with R code in Appendix 7. To handle the zero-event studies, we apply four methods that have been reviewed in this paper, including Haldane's continuity correction, TACC, the GLMM, and the empirical continuity correction proposed by Sweeting et al. [18]. For more techniques on meta-analysis of OR with the zero-event studies, one may refer to [4, 7, 18, 29, 34] and the references therein.

Conclusions

In this paper, we revisited the existing methods that are widely used to handle the zero-event problem in meta-analysis of binary data, in particular with RR as the effect size which is also known as the risk ratio. For the methods with the continuity correction, we reviewed four existing estimators of RR and also introduced a new hybrid estimator with their applications to the random-effects model. Apart from those, the GLMM was also included which is a state-of-the-art method without the continuity correction. By a comparative study and also a real data analysis on COVID-19 data, we found that the random-effects model with the hybrid estimator can serve as a more reliable method for handling the zero-event studies when there are few studies in a meta-analysis, and recommend using the GLMM when the number of studies is large. This paper also provides a useful addition to Chu et al. [8], and meanwhile, it also calls for further observational studies in this field.

Abbreviations
OR: Odds ratio; RR: Relative risk; RD: Risk difference; SARS: Severe acute respiratory syndrome; MERS: Middle East respiratory syndrome; COVID-19: Coronavirus disease 2019; CI: Confidence interval; MSE: Mean squared error;

GLMMs: Generalized linear mixed models; TACC: Treatment arm continuity correction

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s40779-021-00331-6.

Additional file 1: Supplementary information: Appendix

Authors' contributions
TJT, JJW, EXL, and XTZ reviewed the literature and designed the methods. TJT, JJW, and JDS conducted the simulation studies. TJT, JJW, KY, and ZLH conducted the experiments and analyzed the real data. All authors contributed to the manuscript preparation. All authors read and approved the final manuscript.

Availability of data and materials
Not applicable.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Department of Mathematics, Hong Kong Baptist University, Hong Kong, China. [2] Shenzhen Research Institute of Big Data, Shenzhen, China. [3] College of Mathematics and Statistics, Shenzhen University, Shenzhen, China. [4] Center for Evidence-Based and Translational Medicine, Zhongnan Hospital of Wuhan University, Wuhan, China.

References
1. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to Meta-Analysis. Chichester, UK: John Wiley & Son; 2011.
2. Ma LL, Wang YY, Yang ZH, Huang D, Weng H, Zeng XT. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? Mil Med Res. 2020;7:7.
3. Davey J, Turner RM, Clarke MJ, Higgins JPT. Characteristics of meta-analyses and their component studies in the cochrane database of systematic reviews: a cross-sectional, descriptive analysis. BMC Med Res Methodol. 2011;11:160.
4. Jackson D, Law M, Stijnen T, Viechtbauer W, White IR. A comparison of seven random–effects models for meta-analyses that estimate the summary odds ratio. Stat Med. 2018;37(7):1059–85.
5. Ren Y, Lin L, Lian Q, Zou H, Chu H. Real-world performance of meta-analysis methods for double-zero-event studies with dichotomous outcomes using the cochrane database of systematic reviews. J Gen Intern Med. 2019;34(6):960–8.

6. Vandermeer B, Bialy L, Hooton N, Hartling L, Klassen TP, Johnston BC, Wiebe N. Meta-analyses of safety data: a comparison of exact versus asymptotic methods. Stat Methods Med Res. 2009;18(4):421–32.

7. Kuss O. Statistical methods for meta-analyses including information from studies without any events–add nothing to nothing and succeed nevertheless. Stat Med. 2015;34(7):1097–116.

8. Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ, study authors C-SURGES. 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):1973–87.

9. Jin YH, Cai L, Cheng ZS, Cheng H, Deng T, Fan YP, et al. A rapid advice guideline for the diagnosis and treatment of 2019 novel coronavirus (2019-ncov) infected pneumonia (standard version). Mil Med Res. 2020;7:4.

10. Jin YH, Zhan QY, Peng ZY, Ren XQ, Yin XT, Cai L, et al. Chemoprophylaxis, diagnosis, treatments, and discharge management of COVID-19: An evidence-based clinical practice guideline (updated version). Mil Med Res. 2020;7:41.

11. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta–analysis. Res Synth Methods. 2010;1(2):97–111.

12. Lin E, Tong T, Chen Y, Wang Y. Fixed-effects model: the most convincing model for meta-analysis with few studies. Preprint at https://arxiv.org/abs/2002.04211. 2020.

13. Xu C, Li L, Lin L, Chu H, Thabane L, Zou K, Sun X. Exclusion of studies with no events in both arms in meta-analysis impacted the conclusions. J Clin Epidemiol. 2020;123:91–9.

14. Agresti A. Categorical Data Analysis, 2nd Edition. Hoboken: John Wiley & Son; 2003.

15. Haldane JB. The estimation and significance of the logarithm of a ratio of frequencies. Ann Hum Genet. 1956;20(4):309–11.

16. Schwarzer G. meta: An r package for meta-analysis. R News. 2007;7:40–5.

17. Weber F, Knapp G, Ickstadt K, Kundt G, Glass Ä. Zero–cell corrections in random–effects meta–analyses. Res Synth Methods. 2020;11(6):913–9.

18. 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):1351–75.

19. Carter RE, Lin Y, Lipsitz SR, Newcombe RG, Hermayer KL. Relative risk estimated from the ratio of two median unbiased estimates. J Royal Stat Soc: Ser C Appl Stat. 2010;59(4):657–71.

20. Chu H, Nie L, Chen Y, Huang Y, Sun W. Bivariate random effects models for meta-analysis of comparative studies with binary outcomes: methods for the absolute risk difference and relative risk. Stat Methods Med Res. 2012;21(6):621–33.

21. Chen Y, Hong C, Ning Y, Su X. Meta–analysis of studies with bivariate binary outcomes: a marginal beta–binomial model approach. Stat Med. 2016;35(1):21–40.

22. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7(3):177–88.

23. Laird NM, Mosteller F. Some statistical methods for combining experimental results. Int J Technol Assess Heal Care. 1990;6(1):5–30.

24. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evid Based Ment Health. 2019;22(4):153–60.

25. Pettigrew HM, Gart JJ, Thomas DG. The bias and higher cumulants of the logarithm of a binomial variate. Biometrika. 1986;73(2):425–35.

26. Hartung J, Knapp G. A refined method for the meta–analysis of controlled clinical trials with binary outcome. Stat Med. 2001;20(24):3875–89.

27. Fattorini L. Applying the Horvitz-Thompson criterion in complex designs: a computer-intensive perspective for estimating inclusion probabilities. Biometrika. 2006;93(2):269–78.

28. Seber GAF. Statistical Models for Proportions and Probabilities. Heidelberg: Springer; 2013.

29. Bakbergenuly I, Kulinskaya E. Meta-analysis of binary outcomes via generalized linear mixed models: a simulation study. BMC Med Res Methodol. 2018;18(1):70.

30. McCullagh P. Sampling bias and logistic models. J R Stat Soc Ser B Stat Methodol. 2008;70(4):643–77.

31. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. Biometrics. 1988;44(4):1049–60.

32. Lin L, Chu H. Meta-analysis of proportions using generalized linear mixed models. Epidemiology. 2020;31(5):713–7.

33. Lin L, Chu H. altmeta: Alternative Meta-Analysis Methods. 2020. https://CRAN.R-project.org/package=altmeta.

34. Ju J, Lin L, Chu H, Cheng LL, Xu C. Laplace approximation, penalized quasi-likelihood, and adaptive gauss-hermite quadrature for generalized linear mixed models: towards meta-analysis of binary outcome with sparse data. BMC Med Res Methodol. 2020;20(1):152.

35. Gronsbell J, Hong C, Nie L, Lu Y, Tian L. Exact inference for the random–effect model for meta–analyses with rare events. Stat Med. 2020;39(3):252–64.

36. Sarmanov O. Generalized normal correlation and two-dimensional fréchet classes. Sov Math Dokl. 1966;7:596–9.

37. Chen Y, Luo S, Chu H, Su X, Nie L. An empirical Bayes method for multivariate meta-analysis with an application in clinical trials. Commun Stat Theory Methods. 2014;43(16):3536–51.

38. Luo S, Chen Y, Su X, Chu H. mmeta: an R package for multivariate meta-analysis. J Stat Softw. 2014;56(11):11.

39. Jia P, Lin L, Kwong JSW, Xu C. Many meta-analyses of rare events in the cochrane database of systematic reviews were underpowered. J Clin Epidemiol. 2021;131:113–22.

40. Li QK, Rice K. Improved inference for fixed–effects meta–analysis of 2×2 tables. Res Synth Methods. 2020;11(3):387–96.

41. Bender R, Friede T, Koch A, Kuss O, Schlattmann P, Schwarzer G, Skipka G. Methods for evidence synthesis in the case of very few studies. Res Synth Methods. 2018;9(3):382–92.

42. Rice K, Higgins JP, Lumley T. A re–evaluation of fixed effect(s) meta–analysis. J R Stat Soc Ser A Stat Methodol. 2018;181(1):205–27.

43. Yang K, Kwan HY, Yu Z, Tong T. Model selection between the fixed-effects model and the random-effects model in meta-analysis. Stat Interface. 2020;13(4):501–10.

44. Friedrich JO, Adhikari NK, Beyene J. Inclusion of zero total event trials in meta-analyses maintains analytic consistency and incorporates all available data. BMC Med Res Methodol. 2007;7:5.

45. Doi SA, Furuya-Kanamori L, Xu C, Lin L, Chivese T, Thalib L. Questionable utility of the relative risk in clinical research: a call for change to practice. J Clin Epidemiol. 2020. https://doi.org/10.1016/j.jclinepi.2020.08.019.

46. Xiao M, Chen Y, Cole SR, MacLehose R, Richardson D, Chu H. Is OR "portable" in meta-analysis? Time to consider bivariate generalized linear mixed model. 2020. Preprint at https://www.medrxiv.org/content/10.1101/2020.11.05.20226811v1.

47. Doi SA, Furuya-Kanamori L, Xu C, Chivese T, Lin L, Musa OA, Hindy G, Thalib L, Harrell Jr FE. The OR is "portable" but not the RR: time to do away with the log link in binomial regression. J Clin Epidemiol. 2021. https://doi.org/10.13140/RG.2.2.31631.10407.

Publisher's Note