

# Synthesis of evidence from zero-events studies: A comparison of one-stage framework methods

Chang Xu<sup>1,2</sup>  | Luis Furuya-Kanamori<sup>3</sup> | Lifeng Lin<sup>4</sup> 

<sup>1</sup>Department of Population Medicine, College of Medicine, Qatar University, Doha, Qatar

<sup>2</sup>Chinese Evidence-based Medicine Center, West China Hospital, Sichuan University, Chengdu, China

<sup>3</sup>UQ Centre for Clinical Research, Faculty of Medicine, University of Queensland, Brisbane, Australia

<sup>4</sup>Department of Statistics, Florida State University, Tallahassee, Florida, USA

## Correspondence

Chang Xu, Department of Population Medicine, College of Medicine, Qatar University, Al Jamiaa Street, P. O. Box 2713, Doha, Qatar.  
Email: xuchang2016@runbox.com

## Funding information

Australian National Health and Medical Research Council Fellowship, Grant/Award Number: APP1158469

## Abstract

In evidence synthesis, dealing with zero-events studies is an important and complicated task that has generated broad discussion. Numerous methods provide valid solutions to synthesizing data from studies with zero-events, either based on a frequentist or a Bayesian framework. Among frequentist frameworks, the one-stage methods have their unique advantages to deal with zero-events studies, especially for double-arm-zero-events. In this article, we give a concise overview of the one-stage frequentist methods. We conducted simulation studies to compare the statistical properties of these methods to the two-stage frequentist method (continuity correction) for meta-analysis with zero-events studies when double-zero-events studies were included. Our simulation studies demonstrated that the generalized estimating equation with unstructured correlation and beta-binomial method had the best performance among the one-stage methods. The random intercepts generalized linear mixed model showed good performance in the absence of obvious between-study variance. Our results also showed that the continuity correction with inverse-variance heterogeneous (IVhet) analytic model based on the two-stage framework had good performance when the between-study variance was obvious and the group size was balanced for included studies. In summary, the one-stage framework has unique advantages to deal with studies with zero events and is not susceptible to group size ratio. It should be considered in future meta-analyses whenever possible.

## KEYWORDS

beta-binomial model, generalized estimating equation, generalized linear mixed model, meta-analysis, zero-events study

## Highlights

### What is already known

- Dealing with studies with zero events in both arms has raised broad discussions in the community of meta-analysis methodology; the routine scheme

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Research Synthesis Methods* published by John Wiley & Sons Ltd.

is to remove them from the meta-analysis or, in rare situations, to use a continuity correction to produce rough estimates.

- Methods based on the one-stage framework have a unique advantage to deal with studies with zero events in both arms.

#### What is new

- Among many one-stage methods that can be used to deal with studies with zero events in both arms, the generalized estimating equation (GEE) with unstructured correlation and beta-binomial (BB) method generally have the best performance.
- The one-stage meta-analytic methods are not susceptible to group size ratio; even for extremely unbalanced studies, they generally perform well.

#### Potential impact for Research Synthesis Methods readers outside the authors' field

- Our study provide evidence that current routine methods for dealing with studies with no events in both arms in meta-analysis may not be the optimal choice; at least, the one-stage methods should be considered.
- Whenever possible, systematic review authors should consider the GEE method with unstructured correlation to deal with zero-events problems when there is small between-study variance in a meta-analysis, and the BB method when there is considerable between-study variance.

## 1 | INTRODUCTION

In evidence synthesis, dealing with zero-events studies is an important and complicated task and has generated broad discussion.<sup>1–8</sup> Zero-events occur when the risk of events is low and/or the sample size is small, frequently seen with safety outcomes. For studies with zero-events in a single arm, there is a unanimous agreement that such studies should be incorporated into meta-analysis, and many well-established methods (e.g., Peto's odds ratio, continuity correction) are available to achieve this.<sup>9–11</sup> However, for studies with zero-events in both arms (known as double-arm-zero-events studies), controversy remains whether such studies should be incorporated into meta-analysis or not. Researchers who tend to discard double-arm-zero-events studies from a meta-analysis may argue that such studies are non-informative and they add nothing to the pooled effect sizes of relative measures, such as odds ratio (OR) and risk ratio (RR).<sup>12</sup> Although this assumption is supported by the theory of conditional likelihood, it is problematic and has been criticized by the following three arguments.

1. Meta-analysis is a process to synthesize all available evidence for studies on the same topic for a comprehensive inference, and double-arm-zero-events studies are an important source of evidence.

2. Double-arm-zero-events studies are not necessarily “non-informative.” For example, for studies with balanced sample sizes, zero-events in both arms indicate there are comparable treatment effects; excluding them may lead to under- or over-estimation of the effects.<sup>4</sup> Xu et al have shown that excluding studies with no events in both arms can result in the change of the direction of ORs, confidence intervals (CIs), and significance of *p*-values.<sup>13</sup>
3. From the statistical point of view, whether double-arm-zero-events studies contribute information for the inference depends on the methods (along with its assumptions) and the effect measure chosen. Based on the one-stage methods, the effect and variance of such studies could be defined, and they contribute information for the inference. On this basis, we believe that these studies should not be simply discarded.

There are now many methods that provide valid solutions to synthesizing data from double-arm-zero-events studies. They are well-suited for analysis of various effect measures, including both relative and absolute measures.<sup>14–21</sup> These advances make it possible to solve the problem of zero-events in meta-analysis, and they improve decision-making. Kuss<sup>4</sup> has summarized 11 methods for meta-analysis to include information from studies with no events in both arms without a continuity correction. These methods are either based on a

two-stage frequentist framework, for example, the arcsine difference<sup>1</sup> and MH risk difference,<sup>22</sup> or a one-stage frequentist framework, for example, the generalized linear mixed model (GLMM)<sup>23,24</sup> and beta-binomial model (BB).<sup>5,25,26</sup> In addition, Bayesian methods were also proposed to deal with studies with no events.<sup>27,28</sup> The advantages of the Bayesian methods have been widely recognized and documented. With proper prior distributions for the event risks and sample distributions for the events, Bayesian methods are expected to produce satisfactory effect estimates.<sup>27,28</sup> However, the major limitation of the Bayesian methods is that the choice of prior distributions may greatly influence the results (at least for those with double-arm-zero-events studies), resulting in uncertainty of the inference.<sup>13,29</sup> The one-stage frequentist framework methods do not suffer from such a problem because there is no need to borrow strength from prior information or post hoc continuity corrections. Moreover, the one-stage methods are generally easier to implement by a wide range of statistical software than the Bayesian methods.

Although the one-stage methods present their unique advantages to deal with zero-events studies, a very small proportion of meta-analyses utilized this group of methods in practice. As the most important methods under the one-stage framework, the GLMM<sup>23,24,29–33</sup> and BB<sup>5,25,26</sup> for meta-analysis have been widely discussed. The generalized estimating equation (GEE) is another one-stage framework method,<sup>34,35</sup> but it is seldom described in the meta-analysis literature. The lack of global recognition of one-stage methods results in unfamiliarity among most systematic reviewers with these methods. Kuss<sup>4</sup> has investigated the statistical properties of these methods and confirmed the superiority of them for meta-analysis with zero-events studies, while there are two additional issues to be addressed: (1) whether the total event count of a meta-analysis impacts the performance of one-stage methods? (2) whether the sample size ratio impacts the performance of one-stage methods? In this article, we give an overview of the one-stage frequentist methods for meta-analysis with zero-events studies, and compare these methods with continuity correction for the two-stage method.

## 2 | METHODS

### 2.1 | One-stage versus two-stage methods for meta-analysis

To begin with, we must distinguish the concepts of “one-stage” methods and “two-stage” methods. The “two-stage” methods refer to the standard meta-analysis

methods that contain a two-stage procedure. In the first stage, the study-specific effect sizes (and standard errors) are obtained or estimated from the included studies, and then these effect sizes are combined in the second stage through a certain weighting scheme (e.g., inverse variance).<sup>36,37</sup> For the “one-stage” methods, there is no need to obtain the study-specific effect sizes; instead, each study is treated as a stratum or cluster, and the overall effect size is directly obtained by either a generalized linear model or population-averaged method.<sup>38,39</sup> This article focuses on the frequentist framework and does not consider the one-stage Bayesian model. The “one-stage” methods include the “treat-as-one-trial” methods, the GLMMs, the stratified exact (logistic or Poisson) regression model (SERM), the GEE methods, and the BB model. The first three belong to a class of generalized linear models, while the GEE is a population-averaged method. The “treat-as-one-trial” method has been criticized as it may lead to Simpson’s paradox as weights are not assigned to the studies, and it is considered not an appropriate solution.<sup>40</sup> Of note, both “one-stage” and “two-stage” methods are feasible for aggregated and individual participant data. We focus on the application of these methods to aggregated data because such data are easily accessible and commonly seen in practice.

### 2.2 | The one-stage framework: Generalized linear mixed model

GLMMs are the most commonly used one-stage method in meta-analysis and are feasible to a wide range such as meta-analysis of prevalence,<sup>41</sup> network meta-analysis,<sup>33</sup> diagnostic meta-analysis,<sup>5</sup> dose–response meta-analysis,<sup>42</sup> and so on. GLMMs provide a good solution to modeling individual-specific correlation and between-study variance by setting various random effects. The standard random effects GLMMs for meta-analysis include the random slope GLMM and random intercept and slope GLMMs. We use  $i$  (1, 2, ...,  $k$ ) to denote the included studies and  $j$  to denote treatment status ( $j = 0$  for control and  $j = 1$  for treatment). Suppose the study-specific true event risk is  $\pi_{ij}$ . A random intercept and slope GLMM has the following form as described as model 3 by Jackson et al.<sup>30</sup>

$$\text{logit}\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \gamma_i + j\theta_i, \quad (1)$$

where  $\theta_i \sim N(\theta, \tau^2)$ ,  $\gamma_i \sim (\gamma, \sigma^2)$ ,  $\tau^2$  and  $\sigma^2$  refer to between-study variance on treatment effects and study effects. The model assumes that the true study effects (i.e., baseline risks) and true study-specific treatment effects differ across studies (i.e., random-effect model). If

we assume that true study effects are the same, then the above model degrades to a random slope GLMM:

$$\text{logit}\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \gamma + j\theta_i. \quad (2)$$

If we assume treatment effects are the same (i.e., fixed-effect model) but the study effects differ, that is, replacing  $\theta_i$  with  $\theta$ , then the model degrades to a random intercept GLMM. By centering the treatment status for Equation (1), say, replacing  $j$  with  $j-0.5$ , we can establish a modified random intercept and slope GLMM (i.e., model 5 in Jackson et al); similarly, by replacing  $j$  with  $j-0.5$  for Equation (2), we can establish a modified random slope GLMM, say, model 4 in Jackson et al.<sup>30,33</sup>

The joint likelihood function could be established when no random terms are involved, and we can estimate the parameters by maximizing the log-likelihood. Whenever there are random terms, the likelihood function has no closed-form, and some optimization methods should be used for the estimation. These include the Laplace approximation, the penalized quasi-likelihood, the (adaptive) Gauss-Hermite quadrature, iteratively weighted least squares, and the expectation-maximization algorithm.<sup>43</sup> A large-scale simulation study has found that Laplace approximation performs better than the penalized quasi-likelihood and adaptive Gauss-Hermite quadrature for GLMMs.<sup>32</sup>

GLMMs have been shown to have good properties for meta-analysis of rare events.<sup>30-33</sup> The major advantage of GLMMs is their ability to incorporate information from double-arm-zero-events studies without any prior information or post hoc correction. However, GLMMs cannot deal with the situation when the total event count of the meta-analysis is zero in either of the arm or both arms. In addition, when the total event in either arm is insufficient, the GLMMs may have poor performance.<sup>30</sup> Ju et al<sup>32</sup> suggested that at least 10 total events per arm should be contained to ensure the robustness of the results when utilizing GLMMs for meta-analysis.

### 2.3 | The one-stage framework: stratified exact regression model

The stratified exact regression model (SERM) provides a solution to analyzing single-arm-zero-events of binary outcomes as it does not rely on an asymptotic distribution; instead, it produces estimates through the exact distribution.<sup>4</sup> This method is similar to the random intercept GLMM; the difference is that the latter uses the likelihood estimation by an asymptotic distribution. There are two major SERMs for meta-analysis in terms of

the link functions, that is, logit or log, which are referred to as stratified exact logistic regression and stratified exact Poisson regression, respectively. A stratified logistic regression model can be written as:

$$\text{logit}\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \gamma_i + j\theta. \quad (3)$$

Here,  $\gamma_i$ 's are study-specific study effects, and  $\theta$  is the regression coefficient indicating the treatment effect, that is,  $\exp(\theta)$  represents the OR for the logistic model or the RR for the Poisson model. All studies are assumed to share a common effect. The SERM uses the permutation distribution of the sufficient statistics of  $\gamma_i$  and  $\theta$  to obtain the exact probability of each permutation for parameter estimation.<sup>44</sup> The detailed algorithm of the exact regression is provided by Tritchler.<sup>45</sup>

The advantage of SERM is that it can deal with single-arm zero events based on exact estimation and thus does not suffer from the problem of separation.<sup>46</sup> It should be noted that there would be no finite estimates when a conditional likelihood method is employed for strata with zero events, so the conditional likelihood method is unsuitable here. Another advantage is that it considers the variation of the study effects across studies that allows more “freedom” for the variance.<sup>33</sup> The major limitation is that this method is not sensible when dealing with double-arm-zero-events studies.<sup>4</sup> This can be easily illustrated by considering the models that include and exclude double-arm-zero-events studies with the same data set, which will give exactly the same results. Another limitation is that the exact estimation process is computationally expensive when the number of included studies is large.

### 2.4 | The one-stage framework: Generalized estimating equation

The GEE combined with robust variance estimators is an effective method to solve the problem of correlated data within studies. It was proposed by Liang and Zeger,<sup>34,35</sup> and is an extension of generalized linear model for the estimation equation of likelihood function, which establishes a relationship between response means and the linear combination of potential predictors.<sup>47</sup> The GEE focuses on the population mean estimates and is considered a marginal approach.<sup>48,49</sup> A working correlation matrix  $R(a)$  is used to define the GEE and obtain an efficient estimate of the variance.<sup>34</sup> The GEE is better described as a semiparametric estimation method, rather than a statistical model. We can specify the GEEs for meta-analysis as:

$$\text{logit}\left(E\left[\frac{\pi_{ij\lambda}}{1-\pi_{ij\lambda}}\right]\right) = j\theta. \quad (4)$$

The variance of subject  $\lambda$  has the form of  $V_\lambda = A_\lambda^{1/2}R(a)A_\lambda^{1/2}/\psi$ , where  $A_\lambda$  is the usual estimate of the variance based on the information matrix and is obtained as the negative expected value of the second derivative of the log-likelihood.<sup>47</sup> Moreover,  $\psi$  is the dispersion parameter and can be readily estimated.<sup>34,47</sup> The working correlation  $R(a)$  indicates the correlation of the probabilities of an event in treatment and control groups within each study (cluster). Liang et al<sup>34</sup> introduced seven types of working correlations, three of which are suitable for meta-analysis, that is, independent structure, exchangeable correlation, and unstructured correlation. The independent one assumes that there is no correlation between treatment and control arms; the exchangeable structure assumes an identical correlation between all effects within studies; and an unstructured one imposes no structure to the correlation matrix.

Similar to the GLMMs, the GEEs have a major advantage in dealing with double-arm-zero-events studies. Again, the GEEs cannot deal with the situation when the total event counts are zero in either of the arm or both arms across included studies. When the total event in either arm is insufficient, the estimation would be greatly biased.

## 2.5 | The one-stage framework: Beta-binomial model

The BB model was first introduced by Pearson<sup>50</sup> to deal with overdispersion for binomial data, and it has been widely discussed for its application to meta-analyses for zero-events studies in recent years.<sup>4,25,26,51–54</sup> It uses a mixture of binomial distribution of the number of events ( $r$ ) with a beta distribution for the true event risk  $\pi$ :  $r \sim \text{Bin}(n, \pi)$  and  $\pi \sim \text{Beta}(a, b)$ , where  $a > 0$ ,  $b > 0$ , and they are unknown parameters. Through the reparameterization for  $a$  and  $b$ , say,  $\mu = a/(a+b)$  and  $\phi = 1/(a+b+1)$ , we can obtain the expectation and variance:  $E(r) = n\mu$  and  $\text{Var}(r) = n\mu(1-\mu)[1+(n-1)\phi]$ . Here,  $\phi$  is the overdispersion parameter.<sup>53</sup> This reparameterization allows us to use the maximum likelihood method to obtain the unknown parameters  $a$  and  $b$ .<sup>55</sup> Some more sophisticated parameter estimation methods have been documented elsewhere.<sup>4,5,51</sup> A feasible method to establish a one-stage BB model and account for the variation of treatment effects (random-effect model) across studies is to treat study as a time-series variable  $t$ :

$$\text{logit}\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \gamma + j\theta_t. \quad (5)$$

The advantage of this procedure is that it is expected to have better convergence than the multilevel random effects models.

## 2.6 | The one-stage framework: Bivariate model

Several GLMMs can be applied to establish the bivariate model. A bivariate GLMM structure is<sup>24,30</sup>:

$$\begin{bmatrix} \text{logit}(\pi_{i0}/(1-\pi_{i0})) \\ \text{logit}(\pi_{i1}/(1-\pi_{i1})) \end{bmatrix} \sim N\left(\begin{pmatrix} \gamma \\ \gamma + \theta \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 \\ \rho\sigma_0\sigma_1 & \sigma_1^2 \end{pmatrix}\right). \quad (6)$$

Here,  $\sigma_0^2$  and  $\sigma_1^2$  are the variances according to the event risks on the logit scale,  $\text{logit}(\pi_{i0}/(1-\pi_{i0}))$  and  $\text{logit}(\pi_{i1}/(1-\pi_{i1}))$ , in the control and treatment arms, respectively. The  $\rho$  is the correlation coefficient between  $\text{logit}(\pi_{i0}/(1-\pi_{i0}))$  and  $\text{logit}(\pi_{i1}/(1-\pi_{i1}))$ . The advantage of the bivariate model is that it directly models the potential correlation of the event probabilities in two arms. When there are correlations, a bivariate model is expected to produce better point and variance estimates. However, due to the additional nuisance parameters, the estimation would be more complex and computationally expensive.

## 2.7 | The two-stage framework: Continuity correction

Within the two-stage framework, zero-events studies are generally dealt with Peto's OR method, Mantel-Haenszel method, or the continuity correction when the relative risk is utilized. When the absolute risk is utilized, zero-events studies are dealt with by the Mantel-Haenszel risk difference or the double-arcsine difference.<sup>1</sup> We focus on the former situation where the relative risk is used to measure the effect. When facing studies with no events in both arms, the only frequentist method is the continuity correction. Sweeting et al have documented several correction methods and discussed the performance of them.<sup>2</sup> The most widely used correction method is to add 0.5 to each cell of the  $2 \times 2$  table, which is primarily considered in this article.

We consider the increasingly popular tool, the inverse variance heterogeneous model,<sup>56,57</sup> to synthesis the study-

specific effects. This model discards the classic random effects assumption; instead, it treats the between-study variance as an overdispersion correction.<sup>56,57</sup> Simulation studies have demonstrated that the inverse variance heterogeneous (IVhet) model outperformed other two-stage methods.<sup>56,57</sup> This model can be implemented using the *admetan* module<sup>58</sup> in Stata. The pooled estimate and variance by this model are given by:

$$\hat{\theta}_{\text{IVhet}} = \sum_{i=1}^k w_i \theta_i, \quad \text{where } w_i = \frac{1}{s_i^2} / \sum_{i=1}^k \frac{1}{s_i^2}; \quad (7)$$

$$\text{Var}(\hat{\theta}_{\text{IVhet}}) = \sum_{i=1}^k w_i^2 s_i^2 \psi_i = \sum_{i=1}^k w_i^2 (s_i^2 + \tau^2) \quad (8)$$

Here,  $s_i^2$  is the study-specific variance of study  $i$  in the two-stage framework,  $w_i$  is the study-specific weight, and  $\psi_i$  is the study-specific overdispersion correction which can be calculated as  $\frac{s_i^2 + \tau^2}{s_i^2}$ . The advantage of this two-stage method compared with the classic random effects methods is that it is expected to have a better error estimation and keep a better coverage.<sup>56,57</sup>

### 3 | REAL-LIFE EXAMPLE

The COVID-19 pandemic in 2020 hugely impacted our society and threatened human lives. Physical prevention (e.g., physical distance, face mask) is currently the most important way to prevent person-to-person transmission of COVID-19. Chu et al.<sup>59</sup> recently conducted a meta-analysis on the effect of physical prevention for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and COVID-19. In this meta-analysis, they collected 44 comparative studies, including 29 that dealt with the preventive effect of face masks, which were used as our example here (Table S2). Among the 29 studies, 6 were single-arm-zero-events studies, and 6 were double-arm-zero-events studies, while no arms' total counts were zero. Due to the involvement of double-arm-zero-events studies, the SERM method could not be used. Therefore, we employed the BB, GLMMs, GEEs, and IVhet methods for the meta-analysis.

Table 1 presents the meta-analysis results. There was an obvious preventive effect of face masks on reducing the risk (ORs ranging from 0.16 to 0.53) of person-to-person transmission of COVID-19 and SARS-CoV-2, consistent with the original findings by Chu et al.<sup>59</sup> Except for the GEE with independent correlation and the bivariate GLMM with unstructured correlation, the effects were similar across the methods. When excluding six double-arm-zero-events studies, there were some changes in both ORs, CIs, and  $p$ -values.

## 4 | SIMULATION

### 4.1 | Data-generating mechanism

Table 2 presents the simulation settings. We conducted three sets of simulations to investigate the statistical properties of the above methods, and we chose the OR as the effect estimator. As mentioned earlier, the SERM can only be used to deal with single-arm-zero-events studies but is not feasible for double-arm-zero-events studies, so we did not fit the SERM method here. We generated grouped data for each meta-analysis based on the ‘‘PCRandom’’ method<sup>60</sup> and specified the event risk in the control group as a random variable. The event risk in the control group, the sample size in the control group, the (log) true treatment effect, and the between-study variance were first defined, and these parameters were used to estimate the event risk and sample size of the treatment group. Then the simulated meta-analysis with grouped data was generated.

In the first set of simulations, we considered an event risk in the control arm ranging from 0.05 to 0.1 with a uniform distribution, that is,  $\pi_0 \sim U(0.05, 0.1)$ ; in the second set of simulations, we considered  $\pi_0 \sim U(0.01, 0.05)$ ; and in the third set of simulations,  $\pi_0 \sim U(0, 0.01)$ . These settings were based on the empirical data (Supporting Information) from the meta-analyses with double-arm-zero-events studies in the *Cochrane Database of Systematic Reviews*. Among those meta-analyses, in the control arm, the sample sizes were fitted well by a log-normal distribution with the mean log value of 3.3537, the median event risk of 0 (interquartile range: 0–0.09), and a mean event risk of 0.07.<sup>13</sup> The sample size in the control arm ( $n_0$ ) was then set as  $\log n_0 \sim N(3.3537, 0.9992)$ . The sample size in the treatment arm was derived by  $\log n_1 = \log n_0 + \text{ratio}$  and the empirical sample size ratio:  $\log n_1 = \exp(\log n_0) \times \text{ratio}$ . In the empirical data, the ratio ranged from 0.12 ( $1/0.12 = 8.33$ ) to 6. We used the 5% to 95% quartiles of the empirical sample ratio to generate ratio  $\sim U(0.84, 2.04)$ . In addition, since the continuity correction has been criticized for its large bias when the sample ratio is strongly unbalanced,<sup>2</sup> we considered an additional scenario with ratio  $\sim U(2.04, 8.33)$  to examine if the one-stage framework might perform better in this situation.

In the first set of simulations, the majority of the meta-analyses had the total event count in each arm less than 10; the median event count in each study's control arm was  $\exp(3.3537) \times 0.01/2$ . In the second set of simulations, about 50% of the meta-analyses had the total event count in each arm more than 10; the median event count in each study's control arm was  $\exp(3.3537) \times (0.05 + 0.01)/2$ . In the third set of simulations, most of the meta-analyses had the total event count in each arm more than 10; the median event count in each study's

TABLE 1 Synthesized effects of face mask for preventing people-to-people transmission based on different meta-analysis methods

Method	Including double-arm-zero-events studies			Excluding double-arm-zero-events studies		
	OR	95% CI	p-Value	OR	95% CI	p-Value
IVhet	0.27	0.18–0.39	<0.001	0.27	0.18–0.41	<0.001
Beta-binomial	0.28	0.15–0.53	<0.001	0.27	0.15–0.52	<0.001
GEE with independent correlation	0.53	0.42–0.60	<0.001	0.54	0.45–0.64	<0.001
GEE with exchangeable correlation	0.32	0.29–0.36	<0.001	0.33	0.29–0.37	<0.001
GEE with unstructured correlation	0.34	0.30–0.38	<0.001	0.35	0.31–0.41	<0.001
GLMM with random intercept	0.21	0.17–0.27	<0.001	0.22	0.17–0.27	<0.001
GLMM with random slope	0.23	0.08–0.71	0.010	0.42	0.16–1.12	0.081
GLMM with random intercept and slope	0.16	0.10–0.26	<0.001	0.35	0.21–0.58	<0.001
Bivariate GLMM with exchangeable correlation	0.30	0.16–0.58	<0.001	0.29	0.15–0.56	<0.001
Bivariate GLMM with unstructured correlation	0.44	0.22–0.87	0.018	0.47	0.21–1.04	0.061

Note: All GEEs did not employ the robust variance estimator because it had poor coverage in our simulations (data not shown). All the p-values are estimated by the two-tailed z-test, where  $Z = \frac{\hat{\theta} - \theta}{\sqrt{V_{\hat{\theta}}}}$ ,  $\hat{\theta} = \log OR$ , and  $\alpha = 0.05$  was used as the criterion for statistical significance.

TABLE 2 Simulation settings

Parameter	Assigned values
Event risk of the control group	0–0.01; 0.01–0.05; or 0.05–0.1
Number of patients in control group ( $n_0$ )	mean(log) = 3.3537, SD(log) = 0.9992
Sample size ratio	Uniform (0.84, 2.04)
Number of patients in experimental group ( $n_1$ )	$n_1 = \exp(\log n_0) \times \text{ratio}$
Effect size (OR)	1, 2
Between-study standard deviation ( $\tau$ )	0.2; 0.4; 0.6; or 0.8
Number of studies included in each meta-analysis ( $k$ )	Uniform (4, 10)

control arm was  $\exp(3.3537) \times (0.1 + 0.05)/2$ . For the number of studies included  $k$ , we took the first to third quartiles of the empirical data; again, we assigned a discrete uniform distribution to  $k$  in all simulations, that is,  $k \sim U\{4, 10\}$  with a step width of 1. For the treatment effect  $\theta$ , two values were considered:  $\exp(\theta) = 1$  and 2, representing scenarios without true treatment effect ( $\log 1 = 0$ ) and with true treatment effect ( $\log 2$ ), respectively. For each of the three sets of simulations, four equally-spaced monotonically increasing values of  $\tau$  0.2 (small), 0.4 (moderate), 0.6 (substantial), and 0.8 (large) were set to generate small to large between-study variances.<sup>32</sup> Therefore, each set of simulations contained  $2 \times 4 = 8$  scenarios with different treatment effects and between-study variances.

The following eight methods were compared: (1) the two-stage method IVhet with continuity correction

(adding 0.5); (2) the BB model; (3) the GEE method with independent correlations; (4) the GEE method with an unstructured correlation matrix; (5) the random intercept GLMM; (6) the random slope GLMM; (7) the random intercept and slope GLMM; and (8) the bivariate GLMM model. We did not consider the reparameterization of GLMM models as in our previous study because we found that the standard GLMMs performed better than the former.<sup>32</sup> For each scenario, we first simulated 5000 meta-analyses, and then excluded the meta-analyses with zero total event counts in a single arm or in both arms. The percentage bias (PB, defined as  $(\hat{\theta} - \theta)/\theta$ ), mean squared error (MSE, defined as  $\text{Var}(\hat{\theta}) + (\hat{\theta} - \theta)^2$ ), coverage, and width of 95% CI were calculated; they were used to quantify the statistical properties of the candidate methods. Here,  $\hat{\theta}$  represents the estimated effects by these candidate methods. It should be noted that for the scenario  $\theta = 0$ , the PB is replaced by  $\hat{\theta} - \theta$ . The PB and MSE reflect the performance of point estimates and the width of CI reflects that of interval estimates. Therefore, a method was considered to have better performance if it had a lower PB, lower MSE, and narrower CI while the coverage was close to the nominal level 95%. All simulations were performed through the Stata/SE 14.0 (Stata, College Station, TX) software. The program for the simulations is available in the Supporting Information.

## 4.2 | Simulation results

Table S1 presents the proportions of meta-analyses that archived convergence. Generally, the BB model and the GEEs showed better convergence, while the GLMMs

showed poor convergence under the tolerance of 0.000001. Figures 1–4 present the simulation results. In the first set of simulations (first row in Figures 1–4), where most of the meta-analysis (>98.07%) had less than 10 total events in each arm, substantial biases were observed (about -24.14% to -39.78% for OR = 1 and -67.05% to -93.72% for OR = 2). In the second set of simulations (second row in Figures 1–4), about 62–70% of the meta-analyses had at least 10 total events in both arms, and we observed much smaller biases (-0.06 to 13.83% for OR = 1 and -22.59 to 20.38% for OR = 2). In the third set of simulations (third row in Figure 1–4),

about 96% of the meta-analyses had at least 10 total events in both arms, and the methods were almost unbiased in more homogeneous meta-analyses ( $\tau = 0.2$ ). As the number of total event counts increased, the estimation showed increased precisions (i.e., narrower CIs). This indicates that when the number of total events was less than 10 in each arm, none of these one-stage and two-stage methods showed good performance.

For the one-stage methods, when the event risk was low, say 0 to 0.01, where most of the meta-analyses had less than 10 total events in the arms, the performance was poor with large bias, large MSE, large width of CI,



FIGURE 1 Comparison of the 8 methods with respect to percentage bias. (1) IVhet; (2) BB model; (3) the GEE method with independent correlations (GEE); (4) the GEE method with unstructured working matrix (GEE\_B); (5) the random intercept GLMM (GLMM\_I); (6) the random slope GLMM (GLMM\_R); (7) the random intercept and slope GLMM (GLMM\_IR); (8) the bivariate GLMM with unstructured correlation matrix model (GLMM\_B) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]





**FIGURE 2** Comparison of the eight methods with respect to MSE. (1) IVhet; (2) BB model); (3) the GEE method with independent correlations (GEE); (4) the GEE method with unstructured working matrix (GEE\_B); (5) the random intercept GLMM (GLMM\_I); (6) the random slope GLMM (GLMM\_R); (7) the random intercept and slope GLMM (GLMM\_IR); (8) the bivariate GLMM with unstructured correlation matrix model (GLMM\_B) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

and extremely low coverage. As the event risk increased, the total events became sufficient ( $\geq 10$ ) in each arm, the performance of one-stage methods was good. The BB model, GEE with independent correlation, GEE with unstructured correlation, and random-intercept GLMM had better performance than the other methods in terms of bias, MSE, width of CI, and coverage. This finding is similar to that by Kuss.<sup>4</sup> Among these methods, the GEE with unstructured correlation had the best performance when the between-study variance was not obvious ( $\tau < 0.6$ ), while the other three also performed reasonably well. The BB model had the best performance when there was considerable between-study variance ( $\tau \geq 0.6$ ).

The IVhet model with the continuity correction performed well in terms of MSE, width of CI, and coverage. Generally, this method outperformed the one-stage methods when the event risk ranged from 0 to 0.01 and the sample size ratio was balanced. Under this setting, we recorded that when  $OR = 2$ , the IVhet showed moderately large bias but obviously better coverage than one-stage methods. As the event risk increased, the performance of one-stage methods became better, and the advantage of IVhet against these methods decreased. This was more obvious for the bias of all one-stage methods and for the MSE, width of CI, coverage of the BB and GEE with independent correlation. The IVhet model showed larger bias than the one-stage methods when

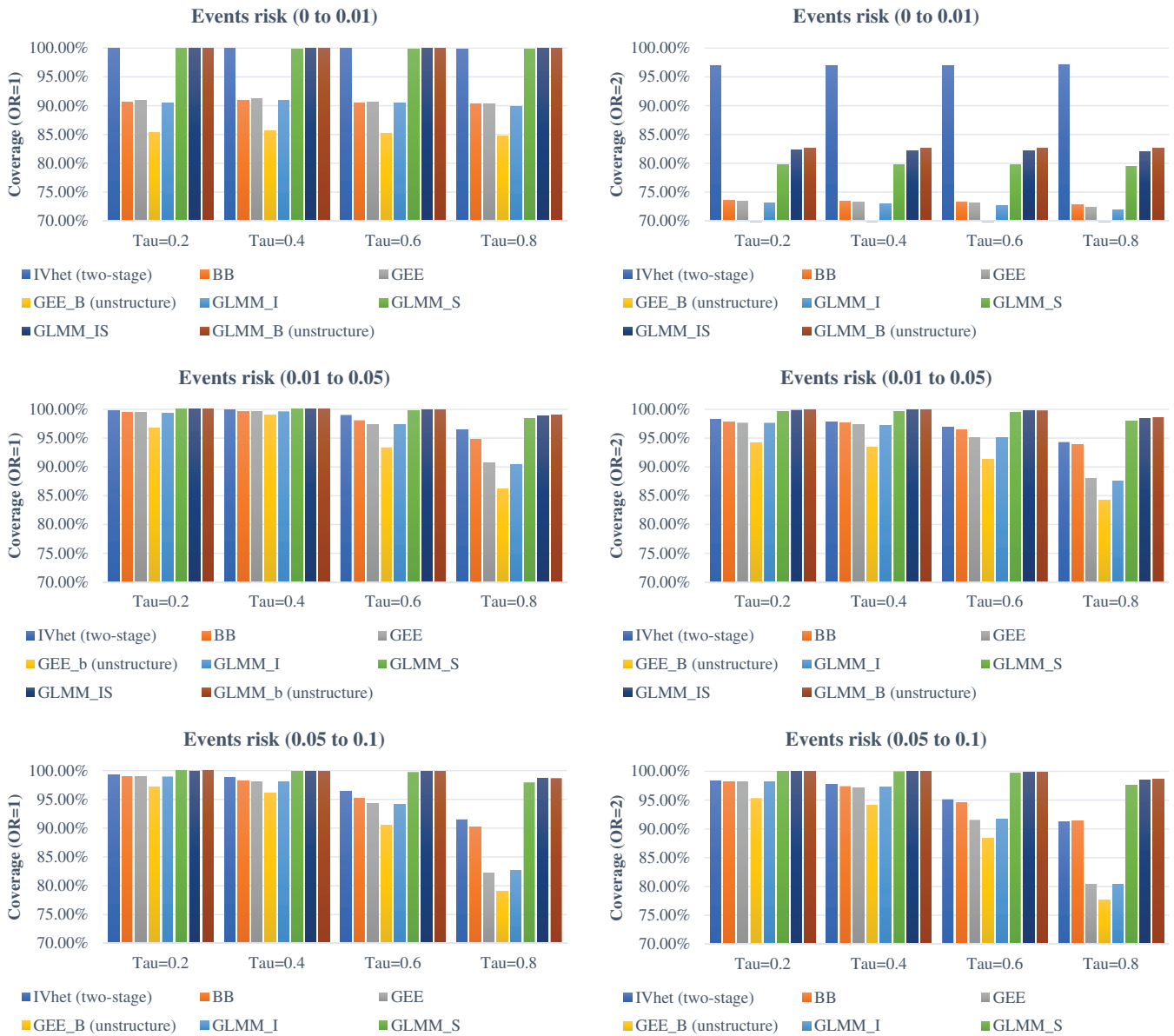


FIGURE 3 Comparison of the eight methods with respect to width of 95%CI. (1) IVhet; (2) BB model); (3) the GEE method with independent correlations (GEE); (4) the GEE method with unstructured working matrix (GEE\_B); (5) the random intercept GLMM (GLMM\_I); (6) the random slope GLMM (GLMM\_R); (7) the random intercept and slope GLMM (GLMM\_IR); (8) the bivariate GLMM with unstructured correlation matrix model (GLMM\_B) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

there was no obvious between-study variance. However, when the between-study variance was considerable ( $\tau \geq 0.6$ ), it showed similar or smaller bias than the one-stage methods. The forgoing was the situation when the group size of included studies was comparable. In our additional simulations, when the ratio of group size was not comparable, the continuity correction had poorer performance than the BB model, GEE with independent correlation, GEE with unstructured correlation, and random-intercept GLMM (Figure S2). The ratio of group size showed little impact on the performance of the one-stage methods.

## 5 | DISCUSSION

In this study, we summarized five one-stage frequentist methods for meta-analysis to deal with zero-events studies. In addition, we compared the three major one-stage methods to a two-stage method (IVhet) that uses the continuity correction to deal with studies with no events in both arms. Our simulation studies demonstrated that the one-stage methods performed poorly when the total events in each arm were insufficient. When the total events in each arm were sufficient, they generally performed well, and the GEE (in the case of low



**FIGURE 4** Comparison of the eight methods with respect to coverage. (1) IVhet; (2) BB model); (3) the GEE method with independent correlations (GEE); (4) the GEE method with unstructured working matrix (GEE\_B); (5) the random intercept GLMM (GLMM\_I); (6) the random slope GLMM (GLMM\_R); (7) the random intercept and slope GLMM (GLMM\_IR); (8) the bivariate GLMM with unstructured correlation matrix model (GLMM\_B) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

between-study variance) and BB model (in the case of obvious between-study variance) had better properties than other methods. The random intercepts GLMM also performed well when there was no obvious between-study variance. However, GLMMs seem to have serious convergence problems under the environment of the Stata program. We also found that, when the group size was balanced in a meta-analysis, the continuity correction based on the IVhet model performed well in the presence of obvious between-study variance. The continuity correction was susceptible to the group size ratio, but this was not for the one-stage methods.

Although the continuity correction based on the two-stage framework showed lower MSE and CI than the one-stage methods, we should not ignore the impact of the continuity correction on the error estimation. As we observed in the three sets of simulations, error estimation tended to be better when the total event count of each arm increased. For the continuity correction, 0.5 was added to each cell for zero-events studies, meaning the total event count was increased; as a result, the estimation is expected to be more precise, leading to smaller variances. This was more obvious in the scenario of no treatment effect, because studies with zero events and

those without zero events shared the same true effect and thus were more homogeneous, and precision substantially increased. In the simulations, there was a much smaller gap between the IVhet method (with the continuity correction) and the one-stage method (without the continuity correction) in terms of the MSE as the event risk increased. The increase in the event count by post hoc corrections could be traded-off by the increase in the total event count through the increase of the event risk.

The predominant advantage of the one-stage methods is their ability to synthesize zero-events studies without prior information or post hoc continuity corrections, preventing “additional noise” in estimation.<sup>61</sup> This feature makes them attractive for solving the problem of zero-events. However, we demonstrated several limitations that might restrict their application. In our simulations, the total event count had a large impact on the performance of both one-stage and two-stage methods. For a meta-analysis with less than 10 total events in either arm, the one-stage methods could have poor performance. In addition, as mentioned earlier, when the total event count was zero in either arm, the one-stage methods were no longer feasible for meta-analysis. This could be one of the major limitations of the one-stage framework methods.

Other limitations of the one-stage framework include that measuring the between-study variance is more complex than with the standard two-stage methods. We can use the intraclass correlation coefficients (ICC) to reflect the proportion of between-study variance to the total variance as a proxy for I-squared for GLMMs,<sup>62</sup> but this is difficult for GEEs. Second, likelihood-based methods generally face the problem of insufficient convergence, especially when random terms are added into the model. Many factors can influence the convergence, including the effect estimates (OR or RD), the setting of the tolerance for determining convergence, iteration times, the choice of model (e.g., adding random terms), optimization methods. For example, setting a less strict tolerance (e.g., 0.001) can increase the convergence rate. As shown in Table S1, the convergence rate was generally low, especially for GLMMs, while this was not the case when using the R program that we recorded previously.<sup>32</sup> This is likely due to the different default settings of parameters for model estimation between software. A comprehensive summary of the different settings in these programs in the future would be useful. Third, it is difficult for the one-stage methods to test for the potential publication bias<sup>63</sup>; the conventional methods for publication bias depend on study-specific effects and cannot be directly applied to the one-stage framework.

In this study, the one-stage Bayesian methods were not considered. Bayesian hierarchical models could be readily

established based on the GLMMs. As mentioned earlier, the setting of prior distribution largely impacts the final results. In the case of rare events, the prior information could predominate the weights, and the sample information might have little contribution to the meta-results. In addition, a random effects Bayesian model inevitably also has the same problem of frequentist GLMMs, due to the random effects assumption. Further investigation and improvement on the Bayesian methods are worthwhile in this area.

In summary, the one-stage framework has a unique advantage to deal with studies with zero events and could be considered in future meta-analyses whenever possible. The continuity correction based on the IVhet model is a good alternative when the group size is balanced. Based on our findings, researchers are advised to choose the GEE method with unstructured correlation to deal with zero-events problems when there is small between-study variance in a meta-analysis, and choose the BB method when there is considerable between-study variance.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## AUTHOR CONTRIBUTIONS

*Conception and design:* Chang Xu. *Manuscript drafting:* Chang Xu. *Data analysis and result interpretation:* Chang Xu, Lifeng Lin, Luis Furuya-Kanamori. *Statistical guidance:* Lifeng Lin. All authors have read and approved the manuscript.

## DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

## ORCID

Chang Xu  <https://orcid.org/0000-0002-2627-1250>

Lifeng Lin  <https://orcid.org/0000-0002-3562-9816>

## REFERENCES

1. Rucker G, Schwarzer G, Carpenter J, et al. Why add anything to nothing? The arcsine difference as a measure of treatment effect in meta-analysis with zero cells. *Stat Med.* 2009;28(5): 721-738.
2. 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.
3. Bradburn MJ, Deeks JJ, Berlin JA, et al. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Stat Med.* 2007;26(1):53-77.
4. 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):1097-1116.
5. Chu H, Nie L, Chen Y, et al. 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):621-633.
6. Yang G, Liu D, Wang J. Meta-analysis framework for exact inferences with application to the analysis of rare events. *Biometrics.* 2016;72(4):1378-1386.
  7. Bhaumik DK, Amatya A, Normand SL, et al. Meta-analysis of rare binary adverse event data. *J Am Stat Assoc.* 2012;107(498):555-567.
  8. Shuster JJ, Jones LS, Salmon DA. Fixed vs random effects meta-analysis in rare event studies: the rosiglitazone link with myocardial infarction and cardiac death. *Stat Med.* 2007;26(24):4375-4385.
  9. 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.
  10. Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomised trials. *Prog Cardiovasc Dis.* 1985;27(5):335-371.
  11. Cox DR. The continuity correction. *Biometrika.* 1970;57(1):217-219.
  12. Finkelstein MO, Levin B. Meta-analysis of “sparse” data: perspectives from the Avandia cases. *Jurimetrics J.* 2012;52:123-153.
  13. Xu C, Li L, Lin L, et al. Exclusion of studies with no events in both arms in meta-analysis impacted the conclusions. *J Clin Epidemiol.* 2020;123:91-99.
  14. JPT H, Thomas J, Chandler J, et al., eds. *Cochrane Handbook for Systematic Reviews of Interventions Version 6.2 (updated 2021)*. England: Cochrane; 2021 <https://www.training.cochrane.org/handbook>
  15. Bai O, Chen M, Wang X. Bayesian estimation and testing in random effects meta-analysis of rare binary adverse events. *Stat Biopharm Res.* 2016;8(1):49-59.
  16. Xu C, Zhou X, Zorzela L, et al. Utilization of the evidence from studies with no events in meta-analyses of adverse events: an empirical investigation. *BMC Med.* 2021;19(1):141.
  17. Liu D, Liu RY, Xie M. Exact meta-analysis approach for discrete data and its application to  $2 \times 2$  tables with rare events. *J Am Stat Assoc.* 2014;109(508):1450-1465.
  18. Tian L, Cai T, Pfeiffer MA, Piankov N, Cremieux PY, Wei LJ. Exact and efficient inference procedure for meta-analysis and its application to the analysis of independent  $2 \times 2$  tables with all available data but without artificial continuity correction. *Biostatistics.* 2009;10(2):275-281.
  19. Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. *Stat Meth Med Res.* 2001;10(4):277-303.
  20. Cai T, Parast L, Ryan L. Meta-analysis for rare events. *Stat Med.* 2010;29(20):2078-2089.
  21. Kuss O, Gromann C. An exact test for meta-analysis with binary endpoints. *Methods Inf Med.* 2007;46(6):662-668.
  22. Klingenberg B. A new and improved confidence interval for the Mantel-Haenszel risk difference. *Stat Med.* 2014;33(17):2968-2983.
  23. Platt RW, Leroux BG, Breslow N. Generalized linear mixed models for meta-analysis. *Stat Med.* 1999;18(6):643-654.
  24. Simmonds M, Higgins JPT. A general framework for the use of logistic regression models in meta-analysis. *Stat Methods Med Res.* 2016;35:2858-2877.
  25. Bakbergenuly I, Kulinskaya E. Beta-binomial model for meta-analysis of odds ratios. *Stat Med.* 2017;36(11):1715-1734.
  26. Chen Y, Hong C, Ning Y, et al. Meta-analysis of studies with bivariate binary outcomes: a marginal beta-binomial model approach. *Stat Med.* 2016;35(1):21-40.
  27. Warn DE, Thompson SG, Spiegelhalter DJ. Bayesian random effects meta-analysis of trials with binary outcomes: methods for the absolute risk difference and relative risk scales. *Stat Med.* 2002;21(11):1601-1623.
  28. Vázquez FJ, Moreno E, Negrín MA, et al. Bayesian robustness in meta-analysis for studies with zero responses. *Pharm Stat.* 2016;15(3):230-237.
  29. Stijnen T, Hamza TH, Ozdemir P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Stat Med.* 2010;29(29):3046-3067.
  30. Jackson D, Law M, Stijnen T, et al. A comparison of seven random-effects models for meta-analyses that estimate the summary odds ratio. *Stat Med.* 2018;37(7):1059-1085.
  31. Cheng LL, Ju K, Cai RL, et al. The use of one-stage meta-analytic method based on individual participant data for binary adverse events under the rule of three: a simulation study. *PeerJ.* 2019;7:e6295.
  32. Ju K, 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.
  33. Yu-Kang T. Use of generalized linear mixed models for network meta-analysis. *Med Decis Making.* 2014;34:911-918.
  34. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika.* 1986;73(1):13-22.
  35. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics.* 1988;44(4):1049-1060.
  36. Gurevitch J, Koricheva J, Nakagawa S, et al. Meta-analysis and the science of research synthesis. *Nature.* 2018;555(7695):175-182.
  37. Shadish WR, Lacy JD. The meta-analytic big bang. *Res Synth Methods.* 2015;6(3):246-264.
  38. Debray TPA, Moons KGM, Abozaid GMA, Koffijberg H, Riley RD. Individual participant data meta-analysis for a binary outcome: one-stage or two-stage? *PLOS One.* 2013;8(4):e60650.
  39. Burke DL, Ensor J, Riley RD. Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. *Stat Med.* 2017;36(5):855-875.
  40. Altman DG, Deeks JJ. Meta-analysis, Simpson's paradox, and the number needed to treat. *BMC Med Res Methodol.* 2002;2:3.
  41. Lin L, Chu H. Meta-analysis of proportions using generalized linear mixed models. *Epidemiology.* 2020;31(5):713-717. <https://doi.org/10.1097/EDE.0000000000001232>
  42. Crippa A, Discacciati A, Bottai M, Spiegelman D, Orsini N. One-stage dose-response meta-analysis for aggregated data. *Stat Methods Med Res.* 2019;28(5):1579-1596.
  43. Bolker BM, Brooks ME, Clark CJ, et al. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol Evol.* 2009;24(3):127-135.
  44. Mehta CR, Patel NR. Exact logistic regression: theory and examples. *Stat Med.* 1995;14(19):2143-2160.
  45. Tritchler D. An algorithm for exact logistic regression. *J Am Stat Assoc.* 1984;79(387):709-711.

46. Mansournia MA, Geroldinger A, Greenland S, et al. Separation in logistic regression: causes, consequences, and control. *Am J Epidemiol.* 2018;187(4):864-870.
47. Hardin JW, Hilbe JM, eds. *Generalized Estimating Equations, Second Edition.* New York: Chapman and Hall/CRC Press; 2012.
48. Spiess M, Fernández D, Nguyen T, Liu I. Generalized estimating equations to estimate the ordered stereotype logit model for panel data. *Stat Med.* 2020;39(14):1919-1940.
49. Pekár S, Brabec M. Generalized estimating equations: a pragmatic and flexible approach to the marginal GLM modelling of correlated data in the behavioural sciences. *Ethology.* 2018;124:86-93.
50. Pearson ES. Bayes' theorem, examined in the light of experimental sampling. *Biometrika.* 1925;17:388-442.
51. Kuss O, Hoyer A, Solms A. Meta-analysis for diagnostic accuracy studies: a new statistical model using beta-binomial distributions and bivariate copulas. *Stat Med.* 2014;33(1):17-30.
52. Efthimiou O, Mavridis D, Nikolakopoulou A, et al. A model for meta-analysis of correlated binary outcomes: the case of split-body interventions. *Stat Methods Med Res.* 2019;28(7):1998-2014.
53. Beisemann M, Doebler P, Holling H. Comparison of random-effects meta-analysis models for the relative risk in the case of rare events: a simulation study. *Biom J.* 2020;62(7):1597-1630.
54. Mathes T, Kuss O. Beta-binomial models for meta-analysis with binary outcomes: variations, extensions, and additional insights from econometrics. *Res Meth Med Health Sci.* 2021;2(2):82-89.
55. Hardin JW, Hilbe JM. Estimation and testing of binomial and beta-binomial regression models with and without zero inflation. *Stata J.* 2014;14(2):292-303.
56. Doi SAR, Furuya-Kanamori L. Selecting the best meta-analytic estimator for evidence-based practice: a simulation study. *Int J Evid Based Healthc.* 2020;18(1):86-94.
57. Doi SA, Barendregt JJ, Khan S, et al. Advances in the meta-analysis of heterogeneous clinical trials I: the inverse variance model. *Contemporary Clin Trials.* 2015;45:130-138.
58. David Fisher. ADMETAN: Stata module to provide comprehensive meta-analysis. Statistical Software Components S458561, Boston College Department of Economics; 2019. <https://ideas.repec.org/c/boc/bocode/s458561.html>
59. 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.
60. Pateras K, Nikolakopoulos S, Roes K. Data-generating models of dichotomous outcomes: heterogeneity in simulation studies for a random-effects meta-analysis. *Stat Med.* 2018;37(7):1115-1124.
61. Xu C, Furuya-Kanamori L, Zorzela L, et al. A proposed framework to guide evidence synthesis practice for meta-analysis with zero-events studies. *J Clin Epidemiol.* 2021;135:70-78.
62. Hemming K, Hughes JP, McKenzie JE, Forbes AB. Extending the I-squared statistic to describe treatment effect heterogeneity in cluster, multi-centre randomized trials and individual patient data meta-analysis. *Stat Methods Med Res.* 2021;30(2):376-395. <https://doi.org/10.1177/0962280220948550>
63. Shi L, Chu H, Lin L. A Bayesian approach to assessing small-study effects in meta-analysis of a binary outcome with controlled false positive rate. *Res Synth Methods.* 2020;11(4):535-552.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Xu C, Furuya-Kanamori L, Lin L. Synthesis of evidence from zero-events studies: A comparison of one-stage framework methods. *Res Syn Meth.* 2021;1-14. doi: 10.1002/jrsm.1521