

# Pitfalls of Using the Risk Ratio in Meta-Analysis

Ilyas Bakbergenuly\*, David C. Hoaglin<sup>†</sup> and Elena Kulinskaya<sup>‡</sup>

November 26, 2018

## Abstract

For meta-analysis of studies that report outcomes as binomial proportions, the most popular measure of effect is the odds ratio (OR), usually analyzed as  $\log(\text{OR})$ . Many meta-analyses use the risk ratio (RR) and its logarithm, because of its simpler interpretation. Although  $\log(\text{OR})$  and  $\log(\text{RR})$  are both unbounded, use of  $\log(\text{RR})$  must ensure that estimates are compatible with study-level event rates in the interval  $(0, 1)$ . These complications pose a particular challenge for random-effects models, both in applications and in generating data for simulations. As background we review the conventional random-effects model and then binomial generalized linear mixed models (GLMMs) with the logit link function, which do not have these complications. We then focus on log-binomial models and explore implications of using them; theoretical calculations and simulation show evidence of biases. The main competitors to the binomial GLMMs use the beta-binomial (BB) distribution, either in BB regression or by maximizing a BB likelihood; a simulation produces mixed results. Two examples and an examination of Cochrane meta-analyses that used RR suggest bias in the results from the conventional inverse-variance-weighted approach. Finally, we comment on other measures of effect that have range restrictions, including risk difference, and outline further research.

Keywords: Relative risk, risk difference, response ratio, log-binomial model, beta-binomial model

## 1 Introduction

For meta-analysis of studies that report binary outcomes (usually summarized as the number of subjects who had an event and the number who had no event, in a treatment group and a control group), the most popular measure of effect is the odds ratio (OR), usually analyzed in the log scale, as the difference in log-odds between the two groups. Many meta-analyses, however, use the risk ratio (RR), the ratio of the probability of an event in the treatment group ( $\pi_T$ ) to that in the control group

---

\* *Corresponding author:* School of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, United Kingdom; i.bakbergenuly@uea.ac.uk

<sup>†</sup>University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA.

<sup>‡</sup>School of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, United Kingdom; e.kulinskaya@uea.ac.uk

( $\pi_C$ ). Importantly, the benefits of analyzing  $\log(\text{RR})$  are offset by the restrictions  $\pi_C < 1$  and  $\pi_T < 1$ , which need to be explicitly applied to their estimates, unlike in the analysis of  $\log(\text{OR})$ . Thus, one must balance the mathematical convenience of the odds ratio against the simpler interpretation of the risk ratio.

When  $\pi_C$  and  $\pi_T$  are small (e.g.,  $< .1$ ),  $\text{OR} \approx \text{RR}$ . If  $\pi_C$  or  $\pi_T$  is not small, however, RR (also called the relative risk) is often considered a better measure of effect than OR, despite the latter's mathematical convenience. In applications the RR and its complement, the percentage reduction in risk, have a direct interpretation. Fleiss et al.<sup>1</sup> point out that Cornfield<sup>2</sup> proposed the odds ratio, in 1951, only because it provided a good approximation to the relative risk. (Interestingly, Cornfield<sup>2</sup> did not use the term "odds ratio.") In general, when  $\pi_T < \pi_C$ ,  $\text{OR} < \text{RR} < 1$ ; and when  $\pi_T > \pi_C$ ,  $\text{OR} > \text{RR} > 1$ <sup>3</sup>. That is, OR is always farther from 1 than RR. Various authors have discussed reasons for choosing RR instead of OR and the ease with which OR can be misinterpreted (e.g., Sinclair and Bracken<sup>4</sup>, Sackett et al.<sup>5</sup>, Altman et al.<sup>6</sup>, Deeks<sup>7</sup>, Newcombe<sup>8</sup>). The OR is necessary in case-control studies (where the risk ratio cannot be estimated directly), and it readily allows adjustment for covariates via logistic regression, but those applications are usually separate from meta-analysis.

When a population consists of strata, it may be possible to summarize an effect more simply if the measure for the entire sample adequately represents the stratum-specific measures, that is, if the measure of effect can be collapsed over the strata. The risk ratio for the entire sample must lie within the range of the stratum-specific risk ratios, but the odds ratio for the entire sample can be closer to 1 than the odds ratio for any of the strata<sup>9</sup>. Even in ideal cases in which the risk ratio or odds ratio is the same in all strata, however, the corresponding measure for the entire population may not equal that common value. Certain conditions must hold for collapsibility of the risk ratio or the odds ratio<sup>10</sup>.

Methods for meta-analysis of risk ratios have received much less attention than methods for odds ratios, in part because analysis of their performance involves complications arising mainly from the restrictions on the ranges of  $\hat{\pi}_C$  and  $\hat{\pi}_T$ . The impact of those complications on actual meta-analyses is not widely understood and may not be apparent to users. We discuss the role of the restrictions in models for fixed-effect and, especially, random-effects meta-analysis, examine their impact on generation of data for simulation studies and on the results, and deduce their likely contribution to bias in examples and in a sizable number of Cochrane reviews. As background, Section 2 reviews the conventional random-effects model (REM), which uses the sample log-odds-ratio or log-risk-ratio as the measure of effect.

To avoid the assumptions and approximations in the conventional methods, an alternative approach bases the analysis on the likelihood for pairs of independent binomial distributions. Section 3 discusses this approach, an application of generalized linear mixed models (GLMMs) with the logit and log transformations as the usual link functions and normal as the usual distribution of random effects. In random-effects models for the log-risk-ratio, the constraint on  $\pi_T$  imposes a truncation on the distribution of the random effects; we use simulation to explore the impact on

estimation of the between-study variance ( $\tau^2$ ) and the overall log-risk-ratio.

The main alternative to the binomial GLMMs is beta-binomial regression. The beta-binomial (BB) distribution arises as a mixture of binomial distributions in which the probability of an event,  $p$ , follows a beta distribution. Section 4 reviews the beta-binomial distribution and BB regression and discusses its application in meta-analysis of log-risk-ratios. In Section 5 we analyze two examples to compare conventional procedures and the procedures based on beta-binomial distributions. Using a collection of 1286 meta-analyses of RR, in a 2004 Cochrane Library issue, we explore (in Section 6) several practical implications of the restriction on the range of the binomial rates (represented by truncation of the distribution of random effects). Finally, the discussion in Section 7 puts our investigation and results in perspective. We have focused on the risk ratio, but other measures of effect also have range restrictions, including risk difference, response ratio (i.e., the log of the ratio of means), and  $\arcsin(\sqrt{p})$  for binomial proportions; methods for these need further research.

## 2 Conventional random-effects model

Random-effects meta-analysis aims to estimate an overall effect,  $\theta$ , defined as the mean of a distribution of study-level effects whose variance is  $\tau^2$ . When  $\tau^2 = 0$ , the random-effects model reduces to the fixed-effect model. For the usual choice of a normal distribution, the effects in the individual studies are  $\theta_j \sim N(\theta, \tau^2)$ ,  $j = 1, \dots, K$ . Study  $j$  yields the estimate  $y_j$  of  $\theta_j$ , along with an estimate,  $s_j^2$ , of its within-study variance,  $\sigma_j^2$ . For some measures of effect, such as mean difference (i.e., the difference between the mean outcome in the treatment group and the mean outcome in the control group),  $y_j$  and  $\theta_j$  are in the same scale as the data. For other measures,  $y_j$  comes from applying a transformation to the data of Study  $j$  or to a summary measure based on those data. In the most common example  $y$  is the log of the sample odds ratio for the occurrence of an event.

The theory associated with the conventional random-effects analysis assumes that the distribution of  $y_j$  can be adequately approximated by  $N(\theta_j, \sigma_j^2)$ . In the resulting normal-normal model, the marginal distribution of  $y_j$  is  $N(\theta_j, \sigma_j^2 + \tau^2)$ . The conventional approach then estimates  $\theta$  by a weighted mean of the  $y_j$  with inverse-variance weights. Theory yields the optimal weights,  $1/(\sigma_j^2 + \tau^2)$ , but both  $\sigma_j^2$  and  $\tau^2$  are unknown. Thus, applications use  $s_j^2$  instead of  $\sigma_j^2$  and estimate  $\tau^2$ , producing the weights  $w_j = 1/(s_j^2 + \hat{\tau}^2)$ . A key assumption is that one can substitute  $s_j^2$  for  $\sigma_j^2$  without allowing for its variability. Despite its documented shortcomings<sup>11;12</sup>, this approach has remained an acceptable part of research on random-effects meta-analysis, and it serves as the basis for most applications. For the log-odds-ratio, for example, the logit transformation yields the logit-normal-normal model: if  $n_{jC}$  and  $y_{jC}$  denote the sample size and number of events in the control group of Study  $j$  and  $n_{jT}$  and  $y_{jT}$  are the corresponding data in the treatment group, the log of the

sample odds ratio is

$$y_j = \log(p_{jT}/(1 - p_{jT})) - \log(p_{jC}/(1 - p_{jC})) = \text{logit}(p_{jT}) - \text{logit}(p_{jC})$$

where  $p_{jT} = y_{jT}/n_{jT}$  and  $p_{jC} = y_{jC}/n_{jC}$ , which results in the estimate of the log-odds-ratio

$$y_j = \log\left(\frac{y_{jT}/(n_{jT} - y_{jT})}{y_{jC}/(n_{jC} - y_{jC})}\right),$$

and the customary estimate of its variance is

$$s_j^2 = \frac{1}{y_{jT}} + \frac{1}{n_{jT} - y_{jT}} + \frac{1}{y_{jC}} + \frac{1}{n_{jC} - y_{jC}}$$

(though in common use, these estimates are biased<sup>13</sup>); if the  $2 \times 2$  table contains one zero cell, 0.5 is usually added to all four cells; if the  $2 \times 2$  table contains two zero cells, Study  $j$  is omitted from the analysis. The assumption of a normal distribution for  $y_j$  in typical finite samples, however, has little empirical support, and correlation between  $y_j$  and  $s_j^2$  is a potential source of bias.

Similarly, the log of the sample risk ratio,

$$y_j = \log(p_{jT}) - \log(p_{jC}),$$

results in the estimate

$$y_j = \log\left(\frac{y_{jT}/n_{jT}}{y_{jC}/n_{jC}}\right),$$

and the customary estimate of its variance is

$$s_j^2 = \frac{1}{y_{jT}} - \frac{1}{n_{jT}} + \frac{1}{y_{jC}} - \frac{1}{n_{jC}}$$

(as above, these estimates are biased<sup>14</sup>).

### 3 Binomial GLMMs

For the important class of applications in which the individual outcome is binary, the available data from each study usually include the sample size and number of events in each group. Then a likelihood-based analysis can avoid the assumptions and approximations of using a normal distribution for  $y_j$ . For simplicity, we consider only generalized linear mixed models based on the summary data available from  $K$   $2 \times 2$  tables (i.e., the numbers of events  $Y_{jC}$  and  $Y_{jT}$  out of  $n_{jC}$  and  $n_{jT}$  binomial trials, with probability of an event  $\pi_{jC}$  and  $\pi_{jT}$ , respectively). In their discussion of these multilevel models, Turner et al.<sup>15</sup> use the term “individual data methods” when the individual subjects’ data are binary (and “summary data methods” when the measure of effect is the sample log-odds-ratio), but their analyses use the data from  $2 \times 2$  tables.

This section reviews logistic linear mixed models, discusses log-binomial models and their complications, and examines the consequences of using the log-binomial-normal model to generate data.

We assume that, given the probabilities  $\pi_{ji}$ ,

$$Y_{ji}|\pi_{ji} \sim \text{Binomial}(n_{ji}, \pi_{ji}) \quad \text{for } i = C, T \quad \text{and } j = 1, \dots, K. \quad (1)$$

For link function  $g$  the basic GLMM for random-effects meta-analysis of treatment versus control is

$$g(\pi_{ji}) = \alpha_j + (\theta + b_j)x_i, \quad (2)$$

where, for Study  $j$ ,  $\alpha_j$  is the control group effect;  $\theta$  is the overall treatment effect;  $b_j$  is the random treatment effect, representing the departure of Study  $j$ 's true treatment effect ( $\theta_j$ ) from  $\theta$ ; and  $x_i$  is an indicator variable for the treatment group ( $x_C = 0$ ,  $x_T = 1$ ); the  $b_j$  are independent, and usually  $b_j \sim N(0, \tau^2)$ .

The most common link function is the logit transformation  $g(\pi) = \text{logit}(\pi) = \log(\pi/(1-\pi))$ . The resulting mixed-effects logistic regression, with log-odds-ratio as the effect measure, belongs to the class of generalized linear mixed models, discussed in meta-analysis by Turner et al.<sup>15</sup> and Stijnen et al.<sup>16</sup>. We also consider the log link, which corresponds to the log-risk-ratio. Meta-regression models expand Equation (2) to include study-level covariates.

For actual analyses and for simulation studies of log-odds-ratios, the two-level logit-binomial-normal model is attractive, for several reasons: log-odds is compatible with binomial likelihoods, the values of  $\theta$  are not bounded, and it is not necessary to rely on asymptotic normality of the sample  $\log(\text{OR})$ . This model is logistic regression with a random effect. Conveniently, it can be fitted by all modern software for GLMMs. Alternatively, one can use a conditional hypergeometric-normal model, implemented in SAS NLMIXED<sup>16</sup> and in R in *metafor*<sup>17</sup>.

### 3.1 Logistic linear mixed models

As background for examining the log link function and the log-binomial models for the log-risk-ratio, we review the more-familiar logistic linear mixed models (LMMs).

#### 3.1.1 Fixed-effects logistic model

The standard fixed-effects logistic model does not account for heterogeneity of the odds-ratios between studies. Assuming a binomial distribution in the two arms, the model is ( $j = 1, \dots, K$ )

$$\log\left(\frac{\pi_{ji}}{1-\pi_{ji}}\right) = \alpha_j + \theta x_i \quad (3)$$

where the  $\alpha_j$  are fixed control group effects (usually regarded as nuisance parameters) and  $\theta$  is the overall log-odds-ratio. The  $K + 1$  parameters of this model can be estimated using maximum likelihood (ML).

### 3.1.2 Logistic linear mixed models

A basic mixed-effects logistic regression model fits fixed effects for the studies' control groups and accounts for heterogeneity in odds-ratios among studies. Given the binomial distributions in the two arms (1), the model is ( $j = 1, \dots, K$ )

$$\log\left(\frac{\pi_{ji}}{1 - \pi_{ji}}\right) = \alpha_j + (\theta + b_j)x_i, \quad (4)$$

where the  $\alpha_j$  are fixed control group effects (usually regarded as nuisance parameters),  $\theta$  is the overall log-odds-ratio,  $b_j \sim N(0, \tau^2)$  are random effects, and  $\tau^2$  is the between-study variance. The fixed study-specific intercepts  $\alpha_j$  have to be estimated, along with  $\theta$  and  $\tau^2$ . These  $K + 2$  parameters are estimated iteratively, using marginal quasi-likelihood, penalized quasi-likelihood, or a first- or second-order Taylor-expansion approximation. A fixed-effect meta-analysis corresponds to  $\tau^2 = 0$ .

As  $K$  becomes large, it may be inconvenient, even problematic, to have a separate  $\alpha_j$  for each study. One can replace those fixed effects with random effects  $\alpha + u_j$ , centered at  $\alpha$ :

$$\log\left(\frac{\pi_{ji}}{1 - \pi_{ji}}\right) = \alpha + u_j + (\theta + b_j)x_i. \quad (5)$$

As before,  $\theta$  is the overall log-odds-ratio, and  $b_j \sim N(0, \tau^2)$ . Now  $u_j \sim N(0, \sigma^2)$ , and  $u_j$  and  $b_j$  can be correlated:  $\text{Cov}(u_j, b_j) = \rho\sigma\tau$ . Heterogeneity of log-odds in the control groups is represented by the variance  $\sigma^2$ , and in the treatment groups, by  $\sigma^2 + 2\rho\sigma\tau + \tau^2$ . In contrast, the conventional random-effects model, which works with the sample log-odds-ratios, involves only a single between-study variance,  $\tau^2$ . Turner et al.<sup>15</sup> point out that  $\rho$  should be estimated. Assuming that  $\rho = 0$  would impose the potentially inappropriate restriction that the variation among trials for control groups ( $\sigma^2$ ) must be less than or equal to the variation among trials for treatment groups ( $\sigma^2 + \tau^2$ ).

Estimation of  $\alpha$ ,  $\mu$ ,  $\sigma^2$ ,  $\tau^2$  and  $\rho$  is similar to estimation of the parameters in Model (4) (Turner et al.<sup>15</sup>). The related bivariate logistic-normal model discussed by Stijnen et al.<sup>16</sup> assumes a bivariate normal distribution for log-odds in the two arms of each study.

## 3.2 Log-binomial models

This section examines the use of the log link function in binomial GLMs and GLMMs.

### 3.2.1 Fixed-effects log-binomial model

The log-binomial model is a constrained GLM with the log link function:

$$\begin{aligned} \log(\pi_{jC}) &= \alpha_j \leq 0, \\ \log(\pi_{jT}) &= \alpha_j + \theta \leq 0. \end{aligned} \quad (6)$$

As in the logistic model, the  $\alpha_j$  are nuisance parameters; here  $\theta$  is the overall log-risk-ratio. The linear constraints in (6) guarantee that  $\pi_{jC} < 1$  and  $\pi_{jT} < 1$ . Likelihood-based methods must ensure that the estimates satisfy these restrictions. They may cause convergence problems, but neglecting them may lead to wrong estimates. Luo et al.<sup>18</sup> provide a brief review of the existing methods and the requisite R code. They propose an adaptive-barrier approach to ML estimation that is easily implemented in R, and they compare several methods by simulation. An approach by Donoghoe and Marschner<sup>19</sup> based on the EM algorithm is implemented in the R package *logbin*<sup>20</sup>. Marschner<sup>21</sup> gives a comprehensive review of contemporary maximum-likelihood and alternative methods, mostly based on unconstrained quasi-likelihood estimation procedures.

### 3.2.2 Log-binomial linear mixed models

To the authors' knowledge, no theoretical developments so far have produced log-binomial mixed models. The main reason, in our opinion, is the restricted parameter space. We now examine this in more detail. In Model (4) mechanically replacing the logit link function by the log link produces the following model for the log-risk-ratio:

$$\log(\pi_{ji}) = \alpha_j + (\theta + b_j)x_i, \quad b_j \sim N(0, \tau^2); \quad j = 1, \dots, K; \quad i = C, T. \quad (7)$$

Here the  $\alpha_j = \log(\pi_{jC}) < 0$ , but the restriction  $\pi_{jT} < 1$  implies that

$$b_j < -\log(\pi_{jC}) - \theta, \quad j = 1, \dots, K. \quad (8)$$

The probability that  $b_j$  satisfies this restriction is  $\Phi((-\log(\pi_{jC}) - \theta)/\tau)$ , where  $\Phi$  is the cumulative distribution function of the standard normal distribution. Thus, as written, the model in Equation (7) is improper. If  $\pi_{jC}$  and  $\theta$  are very small, this probability may be almost 1, so that the restriction has little impact; but for moderate  $\pi_{jC}$  and/or larger values of  $\theta$ , it becomes a serious issue. As an example, for  $\tau^2 = 1$  and  $\theta = 0$  the probability is 0.989 when  $\pi_{jC} = 0.1$  and 0.886 when  $\pi_{jC} = 0.3$ , decreasing to 0.904 and 0.581 when  $\theta = 1$ . These probabilities apply to an individual  $b_j$ . For  $\tau^2 = 1, \theta = 0$ , and  $\pi_{jC} = 0.1$ , for example, the probability that all  $K$  of the  $b_j$  satisfy the restriction is  $(0.989)^K$ , which equals 0.948 when  $K = 5$ , 0.898 when  $K = 10$ , and 0.807 when  $K = 20$ . To summarise, restriction (8) is not compatible with the model (7), which needs to be replaced by an appropriate model. A simple modification by Warn et al.<sup>22</sup>, in the context of Bayesian modelling of RR and RD, replaces Equation (7) by

$$\begin{aligned} \log(\pi_{ji}) &= \alpha_j + (\theta + b_j^U)x_i, \quad b_j \sim N(0, \tau^2); \\ b_j^U &= \min(-\log(\pi_{jC}) - \theta, b_j); \quad j = 1, \dots, K; \quad i = C, T. \end{aligned} \quad (9)$$

This introduces a point mass of probability  $1 - \Phi(\beta)$  at  $c_j^*$ , equal to or just below  $c_j = -\log(\pi_{jC}) - \theta$ , for  $\beta = (c_j^* - \theta)/\tau$  and, equivalently, imputes the values of  $\pi_{jT}$  at or just below 1. Rhodes et al.<sup>23</sup> use this model for Bayesian analysis of inconsistency in the Cochrane database.

An alternative truncates the normal distribution of random effects ( $b_j$ ) from the right at  $A_j = -\log(\pi_{jC}) - \theta$ . We denote the Normal distribution  $N(\mu, \sigma^2)$  truncated from above at  $A$  by  $TN(\mu, \sigma^2, A)$ . Then the model is

$$\log(\pi_{ji}) = \alpha_j + (\theta + b_j)x_i, \quad b_j \sim TN(0, \tau^2, -\log(\pi_{jC}) - \theta); \quad j = 1, \dots, K; \quad i = C, T. \quad (10)$$

Instead of implementing the restriction in Equation (8), an impossible task, both models distort the distribution of  $\theta_j$ . Restrictions depending on the values of  $\pi_{jC}$  and  $\theta$  make both models very artificial.

In both models, the expected value of the log-risk-ratio,  $E(\theta_j)$  no longer equals  $\theta$ , the overall log-risk-ratio in Equation (7). The expected value of  $\theta_j$  in Model (9) is

$$E(\theta_j) = \theta - \frac{1}{\sqrt{2\pi\tau^2}} \int_{c^*}^{\infty} x \exp(-((x - \theta)/\tau)^2) dx + c^*(1 - \Phi(\beta)) < \theta,$$

where  $\phi$  is the probability density function of the standard normal distribution and  $c^* = -\log(\pi_{jC})$ . A parallel calculation yields  $E(\theta_j^2)$  and hence  $\text{Var}(\theta_j)$ .

Next we determine the corresponding mean and variance of  $\theta_j$  in Model (10). For  $X$  having a  $TN(\mu, \tau^2, A)$  distribution, let  $\beta = (A - \mu)/\tau$ . Then (see, e.g., Barr and Sherrill<sup>24</sup>)

$$E(X) = \mu - \tau \frac{\phi(\beta)}{\Phi(\beta)} \text{ and } \text{Var}(X) = \tau^2 \left[ 1 - \beta \frac{\phi(\beta)}{\Phi(\beta)} - \left( \frac{\phi(\beta)}{\Phi(\beta)} \right)^2 \right].$$

In our context  $A = -\log(\pi_{Cj})$ ,  $\mu = \theta$  and  $\beta = (-\log(\pi_{Cj}) - \theta)/\tau$ . Therefore, the mean of  $\theta_j$  is less than  $\theta$ , and it decreases with increasing  $\pi_{jC}$ . The variance of  $\theta_j$  is noticeably smaller than  $\tau^2$ , decreasing as  $\pi_{jC}$  increases. This model is also clearly not satisfactory.

Importantly, in both models, the expected values of the log-risk-ratios  $\theta_j$  depend on the individual values of  $\pi_{jC}$ , making the meta-analysis of the  $\theta_j$  rather pointless.

In Section 3.3 we consider in more detail what happens when Model (7) is used and the restrictions are neglected. As we shall see, this mistake results in considerable biases. Overall, we find the log-binomial LMMs with fixed  $\alpha_j$  not suitable for modeling the risk ratio.

The analog of Model (5), with random effects for the control groups, is

$$\log(\pi_{ji}) = \alpha + u_j + (\theta + b_j)x_{ji}, \quad b_j \sim N(0, \tau^2), \quad u_j \sim N(0, \sigma^2) \text{ and } \text{Cov}(u_j, b_j) = \rho\sigma\tau. \quad (11)$$

This model involves even more restrictions:

$$u_j < -\log(\pi_{jC}) - \alpha, \quad b_j + u_j < -\log(\pi_{jC}) - \alpha - \theta, \quad j = 1, \dots, K, \quad (12)$$

so it also is not suitable.

To summarize, we do not think that a GLMM with the log link is a feasible option for modeling relative risk.



### 3.3 Generating data from the log-binomial LMM

In this section we discuss the consequences of using the log-binomial-normal mixed model, Equation (7), to generate data, and we use a small simulation study for illustration.

#### 3.3.1 Practicalities

In practice the studies in a meta-analysis come from a systematic review, bringing with them the underlying pairs of event probabilities,  $(\pi_{jC}, \pi_{jT})$ . For random-effects models it is convenient to regard the  $(\pi_{jC}, \pi_{jT})$ , and hence the  $(\log\pi_{jC}, \log\pi_{jT})$ , as a sample from some bivariate distribution. We can also approach the joint distribution of  $(\pi_{jC}, \pi_{jT})$  via the marginal distribution of  $\pi_{jC}$  and the conditional distribution of  $\pi_{jT}$  given  $\pi_{jC}$  and a value of  $\theta$ . Thus, to obtain data from the log-binomial LMM for meta-analysis of log-risk-ratio, we can choose values of  $\pi_{jC}$ , generate study effects  $\theta_j$  from  $N(\theta, \tau^2)$ , calculate the  $\pi_{jT} = \pi_{jC} \exp(\theta_j)$ , generate observations  $Y_{jC}$  from the Binomial( $n_{jC}, \pi_{jC}$ ) distributions, and generate observations  $Y_{jT}$  from the Binomial( $n_{jT}, \pi_{jT}$ ) distributions. (This approach parallels a common method of generating data for meta-analyses of odds ratios.)

However, this process may produce values of  $\pi_{jT} > 1$ . As a remedy one has two practical options: either impute values of  $\pi_{jT}$  at or slightly below 1, or reject values of  $\theta_j$  that are too large and generate replacement values of  $\theta_j$ . The first option is equivalent to using Model (9), and the second option (rejection sampling) is equivalent to truncating the normal distribution of random effects ( $b_j$ ) as in Model (10). Both options introduce bias; that is,  $E(\theta_j)$  no longer equals  $\theta$ , the overall log-risk-ratio in Equation (7). The first option appears to be more popular in meta-analysis. IntHout et al.<sup>25</sup> use it in their simulations. The second option seems uncommon, but many authors who use simulation in meta-analysis do not report details of implementation. Some authors use “truncate” but create a point mass (e.g., Panityakul et al.<sup>26</sup>). Pedroza and Truong<sup>27</sup> use truncation in simulating risk difference in multicenter trials. Both options aim to approximate the actual situation, in which  $\pi_{jC} < 1$  and  $\pi_{jT} < 1$ . The basic difficulty lies in using a normal distribution for the random effects. A different approach is required to obtain unbiased inference, or the bias needs to be estimated and eliminated.

#### 3.3.2 Simulation study

To evaluate the size of these biases in conventional random-effects meta-analysis (i.e., the log-normal-normal model, Section 2), we conducted a small simulation study of the two options for generating data from a log-binomial LMM. We used equal sample sizes  $n_{jC} = n_{jT} = n/2$  and the same value of  $\pi_C$  for all studies. We set  $K = 5, 10, 20$ ;  $n = 100, 200, 500$ ;  $\theta = -1.5, -1, -0.5, 0, 0.5, 1, 1.5$ ;  $\pi_{jC} = 0.1$  or  $0.3$ ; and  $\tau^2 = 0.1$  and  $1$ . If a ‘study’ had  $Y_T + Y_C = 0$  or  $n$ , we followed customary practice by discarding it and reducing  $K$  accordingly. We estimated  $\tau^2$  by three methods: DerSimonian-Laird<sup>28</sup>, Mandel-Paule<sup>29</sup>, and restricted maximum

likelihood (REML). From 1000 replications we studied estimation of  $\tau^2$  and  $\theta$  and the coverage of confidence intervals for  $\theta$  based on the normal approximation.

Figures 1 and 2 show the results for estimation of  $\tau^2$  (i.e.,  $\text{var}(\theta_j)$ ) when  $\tau^2 = 1$ . Similar results for  $\tau^2 = 0.1$  appear in Figures 10 and 11 in the Supplementary material (Appendix B.1). The substantial biases (usually negative) in estimation of  $\tau^2$  for both options have two sources. For the larger values of  $\theta$  they arise from the restriction  $\pi_{jT} < 1$ , which has greater impact for  $\pi_{jC} = 0.3$  than for  $\pi_{jC} = 0.1$ . As the traces for the theoretical value of  $\tau^2$  show, this source of bias plays a steadily decreasing role as  $\theta$  decreases from 1.5 to  $-1.5$ . For the smaller (i.e., more-negative) values of  $\theta$ , the source of the bias is progressively small values of  $\pi_{jT}$ , as  $\theta$  becomes more negative. For example, when  $\pi_{jC} = 0.1$ , and  $\theta = -0.5$ , the median value of  $\pi_{jT}$  is 0.0607, for  $\theta = -1$ , the median value is 0.0368, and for  $\theta = -1.5$  it is 0.0223. For sparse data, the distribution of  $\log(\text{risk})$  and hence  $\log(\text{RR})$  is not well approximated by a normal distribution. It is well known that the standard REM does not perform well in these circumstances<sup>30</sup>.

In relation to  $\theta$  the point-mass option (Figure 1) has similar patterns of bias in  $\hat{\tau}^2$  for the three methods as  $K$  increases and as  $n$  increases. The MP method consistently has the smallest bias, followed by REML and then DL. As  $K$  increases, the patterns for each  $n$  change little. As  $n$  increases, the traces for each  $K$  move closer to 1, and the trace for REML moves closer to that for MP. In contrast, the traces for DL generally move farther away from the other estimators.

For the truncation option the plots of bias in  $\hat{\tau}^2$  versus  $\theta$  (Figure 2) are qualitatively similar to those for the point-mass option (Figure 1), with several main differences. For  $\pi_{jC} = 0.1$  and each combination of  $n$  and  $K$ , the biases are larger than those in Figure 1, especially for  $\theta \geq 0$ . For  $\pi_{jC} = 0.3$  and  $\theta > 0$  the slopes are not as steep, and the biases at  $\theta = 1.5$  are not as large, as in Figure 1.

Biases in estimating  $\theta$  are almost the same for the three methods of estimating  $\tau^2$ . Therefore Figure 3 shows the results for the Mandel-Paule method and, for comparison, the theoretical expectations. For both options and both values of  $\pi_{jC}$  the bias in  $\hat{\theta}$  is strongly related to  $\theta$ . When  $\pi_{jC} = 0.1$ , the two options produce the same bias for  $\theta \leq 0$ : positive at  $\theta = 0$  and roughly linear in  $\theta$ , with negative slope, for  $\theta < 0$ . (We expect the restriction  $\pi_{jT} < 1$  to have little impact.) For  $\theta > 0$  the traces for the two options diverge; the point-mass option has bias of relatively small magnitude, and the truncation option has increasingly negative bias as  $\theta$  increases. When  $\pi_{jC} = 0.3$ , both traces show substantial curvature. For  $\theta \leq -0.5$ , truncation often produces smaller (and positive) bias than the point-mass option, but for  $\theta \geq 0$  its bias is negative and considerably larger in magnitude. These patterns change little with  $K$  and only slightly with  $n$ .

In summary, neither the point-mass option nor the truncation option responds satisfactorily to  $b_j$  sampled from a random-effects distribution that produces  $\pi_{jT} > 1$ . The resulting biases in estimating  $\theta$  and  $\tau^2$  are often unacceptably large. Our choice of  $\tau^2 = 1$  as the true value may have magnified the biases, but it serves to illustrate the difficulties, **and the same general patterns in estimating  $\tau^2$  are present when  $\tau^2 = 0.1$** . The biases seen in this small simulation raise questions

about the results of numerous meta-analyses that have employed the REM for risk-ratios. We explore this further in Section 6.

## 4 Beta-binomial model

In this section we explore the main alternative to the binomial GLMMs, beta-binomial regression, and its application to meta-analysis of risk ratios.

### 4.1 The beta-binomial distribution

The beta-binomial (BB) distribution arises as a mixture of binomial distributions,  $\text{Binom}(n, p)$ , according to a beta distribution for  $p$ . If  $Y \sim \text{Binom}(n, p)$  and  $p \sim \text{Beta}(\alpha, \beta)$ , then, unconditionally,  $Y$  follows a beta-binomial distribution with parameters  $n$ ,  $\alpha$ , and  $\beta$  (Johnson et al.<sup>31</sup>, p.270). It is convenient to parametrize this distribution as  $\text{BetaBinom}(n, \pi, \rho)$ , where  $\pi = \alpha/(\alpha + \beta)$ ,  $\rho = 1/(\alpha + \beta + 1)$ ,  $\alpha > 0$ , and  $\beta > 0$ . Then the beta distribution has mean  $\pi$  and variance  $\pi(1 - \pi)\rho$ , and

$$E(Y) = n\pi, \quad \text{Var}(Y) = n\pi(1 - \pi)(1 + (n - 1)\rho), \quad (13)$$

which shows overdispersion relative to  $\text{Binom}(n, \pi)$ . The distribution of the sum of  $n$  Bernoulli( $\pi$ ) random variables with intra-cluster correlation (ICC)  $\rho$  has the same mean and variance, but its actual shape may be very different<sup>32</sup>.

### 4.2 Beta-binomial regression

Like the conventional REM and the binomial GLMMs, beta-binomial regression is a two-stage random-effects model. Assume that, as in a randomized controlled trial, the treatment and control groups of each study are independent, and that within the two groups, conditional on the probabilities, the numbers of events follow binomial distributions. Allowing beta-distributed variation of the probabilities among the studies, the resulting marginal distributions are BB distributions. If (for simplicity) they have the same  $\rho$  for both groups and all studies, then

$$Y_{jC} \sim \text{BetaBinom}(n_{jC}, \pi_C, \rho) \quad \text{and} \quad Y_{jT} \sim \text{BetaBinom}(n_{jT}, \pi_T, \rho). \quad (14)$$

That is, as in the binomial model, the two groups differ only on  $\pi$  (and  $n$ ). The corresponding beta distributions are

$$p_{ji} \sim \text{Beta}(\pi_i(\rho^{-1} - 1), (1 - \pi_i)(\rho^{-1} - 1)). \quad (15)$$

Importantly, because the GLMMs and BB regression make very different distributional assumptions about the random effects, their results may differ. Bakbergenuly and Kulinskaya<sup>33</sup> considered meta-analysis of odds ratios under a beta-binomial model. A more general model uses a bivariate beta distribution, and therefore may have different values of  $\rho$  in the two groups, and also correlation between the groups<sup>34</sup>.

Like GLMMs, the beta-binomial model can incorporate a matrix of covariates  $\mathbf{X}$  by using an appropriate link function  $g$  for the effect measure of interest, so that  $g(\pi) = \mathbf{X}\theta$  for a vector of unknown parameters  $\theta$ .

For beta-binomial meta-analysis,

$$g(\pi_i) = \alpha + x_i\theta;$$

as before,  $x_i$  is a treatment indicator, and  $\theta$  is a treatment effect. Thus, beta-binomial regression yields a random-effects model in which  $\theta$  determines the association between  $p_{jT}$  and  $p_{jC}$  (through the link function).

Most, if not all, beta-binomial regression programs use the logit link<sup>35;36</sup>, so the probabilities  $\pi = \text{expit}(\mathbf{X}\theta)$ . The logit link function guarantees that the probabilities  $\pi$  lie in the interval  $(0, 1)$ . The log link encounters the same complications as in the log-binomial model, because the estimation process needs to incorporate the constraint  $\mathbf{X}\theta < 0$ . We are not aware of any theoretical work for this log-beta-binomial model.

### 4.3 Using standard beta-binomial regression for risk ratio

In a single study the risk ratio can be estimated as the ratio  $\hat{\pi}_T/\hat{\pi}_C$  of the maximum-likelihood estimators of  $\pi_T$  and  $\pi_C$  or from the estimated logits obtained by using a beta-binomial regression with the logit link.

The likelihood for the beta-binomial model (14) is

$$\prod_{j=1}^K \prod_{i=1}^2 \binom{n_{ji}}{Y_{ji}} \frac{\text{Beta}(\pi_i(\rho^{-1} - 1) + Y_{ji}, (1 - \pi_i)(\rho^{-1} - 1) + n_{ji} - Y_{ji})}{\text{Beta}(\pi_i(\rho^{-1} - 1), (1 - \pi_i)(\rho^{-1} - 1))}, \quad (16)$$

where  $\text{Beta}(u, v)$  is the beta function. The parameters  $\pi_T$ ,  $\pi_C$ , and  $\rho$  can be estimated by maximizing the log-likelihood. This process may encounter computational problems because the beta-binomial distribution does not belong to an exponential family<sup>37</sup>. Maximum-likelihood estimation requires numerical methods such as the Newton-Raphson method. The approximate covariance matrix of the parameter estimates is obtained by evaluating the inverse of the Hessian matrix at those estimates.

In R, the package *bbmle* provides a program for maximizing the beta-binomial likelihood<sup>38</sup>. General beta-binomial regression with the logit link function is implemented in a number of R packages, including *gamlss*<sup>39</sup> and *hglm*<sup>36</sup>. The use of the SAS procedure NLMIXED is explained in Martinez et al.<sup>37</sup>.

From the MLEs of  $\pi_T$  and  $\pi_C$  and estimates of their variances  $\text{Var}(\hat{\pi}_T)$  and  $\text{Var}(\hat{\pi}_C)$  (say, from *bbmle*), the delta method yields an approximation for the variance of  $\log(\text{RR})$ :

$$\text{Var}(\log(\text{RR})) \approx \left[ \frac{1}{\hat{\pi}_T} \right]^2 \text{Var}(\hat{\pi}_T) + \left[ \frac{1}{\hat{\pi}_C} \right]^2 \text{Var}(\hat{\pi}_C). \quad (17)$$

Similarly, when using the logit link, output from beta-binomial regression (say, *gamlss*) provides estimates of the log-odds  $\hat{\eta}_C = \hat{\alpha}$  and  $\hat{\eta}_T = \hat{\alpha} + \hat{\theta}$  and their standard errors. To obtain the estimate of the RR, the expit transformation yields the estimated probabilities

$$\hat{\pi}_C = \frac{\exp(\hat{\alpha})}{1 + \exp(\hat{\alpha})} \quad \text{and} \quad \hat{\pi}_T = \frac{\exp(\hat{\alpha} + \hat{\theta})}{1 + \exp(\hat{\alpha} + \hat{\theta})}.$$

Then  $\log(\text{RR})$  is given by  $\log(\hat{\pi}_T/\hat{\pi}_C)$  with variance approximated by the delta method:

$$\begin{aligned} \text{Var}(\log(\text{RR})) &\approx \text{Var}\left(\log\left(\frac{\exp(\hat{\eta}_T)}{1 + \exp(\hat{\eta}_T)}\right) - \log\left(\frac{\exp(\hat{\eta}_C)}{1 + \exp(\hat{\eta}_C)}\right)\right) \\ &= \left[\frac{1}{1 + \exp(\hat{\eta}_T)}\right]^2 \text{Var}(\hat{\eta}_T) + \left[\frac{1}{1 + \exp(\hat{\eta}_C)}\right]^2 \text{Var}(\hat{\eta}_C). \end{aligned} \quad (18)$$

The overdispersion in the BB model may be parametrized in various ways: *bbmle* estimates  $\gamma = (1 - \rho)/\rho$ , and *gamlss* estimates  $\gamma = \log(\rho)$ . In Appendix A we provide R functions for using *bbmle* and *gamlss* for meta-analysis of RR.

#### 4.4 Conventional meta-analysis of risk ratios under the beta-binomial model

Conventional meta-analysis calculates the sample risk ratios,  $\hat{\psi}_j = \hat{\pi}_{jT}/\hat{\pi}_{jC}$ , and their logarithms,  $\hat{\theta}_j$ , and uses inverse-variance weights based on estimated variances of the  $\hat{\theta}_j$ . To obtain an unbiased (to  $O(n^{-2})$ ) estimate of  $\theta_j$  and an unbiased (to  $O(n^{-3})$ ) estimate of  $\text{Var}(\hat{\theta}_j)$  under a binomial distribution, Pettigrew et al.<sup>14</sup> add 1/2 to the number of events and the total in each group:

$$\hat{\theta}_j = \log\left(\frac{Y_{jT} + 1/2}{n_{jT} + 1/2}\right) - \log\left(\frac{Y_{jC} + 1/2}{n_{jC} + 1/2}\right); \quad (19)$$

we retain this estimate. When  $Y_{jC}$  and  $Y_{jT}$  have BB distributions, Equation (14), the approximate variance of  $\hat{\theta}_j$ , obtained via the delta method, is

$$\text{Var}(\hat{\theta}_j) \approx \frac{1 - \pi_T}{n_{jT}\pi_T}(1 + (n_{jT} - 1)\rho) + \frac{1 - \pi_C}{n_{jC}\pi_C}(1 + (n_{jC} - 1)\rho). \quad (20)$$

Substituting the variances of the  $\hat{\pi}_i$  from the line above (13) yields the same variance as in Equation (17). Setting  $\rho = 0$  in (20) yields the within-study variance of  $\hat{\theta}_j$  for binomially distributed data, so under the BB model ( $\rho > 0$ ) this variance is inflated. In a direct parallel, in the conventional REM, compared with the fixed-effect model, the variances of the  $\hat{\theta}_j$  are inflated by an additive variance component,  $\tau^2$ . Thus, the BB model is similar to the conventional REM, but the variance-inflation is multiplicative instead of additive; it is of order  $O(1)$  and increases with  $\rho$ ; the variance also may be large when  $\pi_C$  or  $\pi_T$  is close to 0.

The conventional REM uses inverse-variance weights to obtain an estimate of the overall effect. Estimating  $\pi_{jT}$  and  $\pi_{jC}$  as in (19) yields the estimated variance

$$\begin{aligned} \widehat{\text{Var}}(\hat{\theta}_j) &= \left(\frac{1}{y_{jT}+1/2} - \frac{1}{n_{jT}+1/2}\right)(1 + (n_{jT} - 1)\hat{\rho}) + \\ &\quad \left(\frac{1}{y_{jC}+1/2} - \frac{1}{n_{jC}+1/2}\right)(1 + (n_{jC} - 1)\hat{\rho}). \end{aligned} \quad (21)$$

To use these estimated variances, however, we must estimate  $\rho$ .

## 4.5 Estimation of $\rho$

To estimate  $\rho$ , we modify two established methods: a method-of-moments estimator based on Cochran’s  $Q$  statistic, similar to the DerSimonian-Laird<sup>28</sup> estimator of  $\tau^2$ , and a restricted-maximum-likelihood estimator. According to Viechtbauer<sup>40</sup>, these two approaches perform best for estimation of the between-studies variance  $\tau^2$  in the additive REM. We also use a version of the method of Mandel and Paule<sup>29</sup> to estimate  $\rho$  from the large-sample approximation of  $Q$  by a chi-squared distribution. All three methods were proposed by Kulinskaya and Olkin<sup>41</sup>, but they have not previously been explored by simulation. We refer to these estimators as  $\hat{\rho}_{MoM}$ ,  $\hat{\rho}_{REML}$  and  $\hat{\rho}_{MP}$ , respectively. However, in finite samples the chi-squared distribution is a poor approximation to the distribution of the  $Q$  statistic<sup>42</sup>, and we propose a new method for point and interval estimation of  $\rho$  based on inverting the modified Breslow-Day (BD) test<sup>43</sup>,  $\hat{\rho}_{BD}$ . Bakbergenuly and Kulinskaya<sup>33</sup> proposed a similar method in meta-analysis of odds ratios. The detailed derivations for these four estimators of  $\rho$  are given in Appendix A1.

## 4.6 Simulation study

To explore the performance of beta-binomial methods for meta-analysis of risk ratios, we conducted a small simulation study. We used equal sample sizes  $n_{jC} = n_{jT} = n/2$ , the same value of  $\theta$  for all studies, and  $\pi_T = \pi_C \exp(\theta)$  for  $\theta < -\log(\pi_C)$ . The data in the Treatment and Control groups were generated from independent beta-binomial distributions. Parallel to Section 3.3.2, we set  $K = 5, 10, 20$ ;  $n = 100, 200, 500$ ;  $\theta = -1.5, -1.0, -0.5, 0, 0.5, 1, 1.5$ ;  $\pi_C = 0.1$  or  $0.3$ ; and  $\rho = 0.1$ . As in our simulation study for the log-binomial LMM (Section 3.3.2), when a ‘study’ had  $Y_T + Y_C = 0$  or  $n$ , we discarded it and reduced  $K$  accordingly.

We estimated  $\rho$  by the modified MoM and MP methods, REML, and the BD-based method. The estimates of  $\theta$  used inverse-variance weights, and their CIs used the normal approximation. We also included *bbmle* and *gamlss*, the latter with logit link. For all methods, we estimated bias in estimation of  $\theta$  and  $\rho$ , and coverage of confidence intervals for  $\theta$ . For the nine combinations of  $n$  and  $K$ , Figures 4, 5, and 6 plot (versus  $\theta$ ) the estimated bias for  $\hat{\rho}$  and  $\hat{\theta}$ , and the coverage of  $\theta$ , respectively. Because the results are almost the same for the MP, MoM, and REML methods, the figures show only the MP results.

For *bbmle* and *gamlss*, the bias in  $\hat{\rho}$  does not vary appreciably with  $\theta$ ,  $\pi_C$ , or  $n$ . It is roughly  $-0.02$  for  $K = 5$ , and it approaches 0 as  $K$  increases. The bias of the other methods is generally negative, unrelated to  $n$ , and only weakly related to  $K$ . The relation of those biases to  $\theta$  is roughly linear, with similar positive slopes (except for high positive values with BD when  $\pi_C = 0.3$ ). When  $\pi_C = 0.1$ , the bias of MP rises from  $-0.0855$  at  $\theta = -1.5$  to  $-0.0386$  at  $\theta = +1.5$ ; it is closer to 0 by roughly 0.05 when  $\pi_C = 0.3$ . The trace for BD when  $\pi_C = 0.1$  closely resembles that for MP when  $\pi_C = 0.3$ , and it shifts upward by roughly 0.035 when  $\pi_C = 0.3$ . Since the true value of  $\rho$  is 0.1, these biases are substantial.

The bias in  $\hat{\theta}$  follows two patterns, both of which change little with  $N$  or  $\pi_C$ .

One pattern (RR estimated with *bbmle* or *gamlss*) goes linearly from small negative bias at  $\theta = -1.5$  to small positive bias at  $\theta = 1.5$ , and its magnitude decreases as  $K$  increases. The other pattern (MP, BD and other inverse-variance methods) goes, roughly linearly, from around 0.3 at  $\theta = -1.5$  to around  $-0.3$  at  $\theta = 1.5$ , when  $K = 5$ ; and the traces become steeper as  $K$  increases. All of the methods have essentially no bias at  $\theta = 0$ , but otherwise MP and BD underestimate the magnitude of  $\theta$  by about 20%. This is due to the effect of transformation bias, which is almost linear in  $\rho$ . Section 6.2.3 in Bakbergenuly and Kulinskaya<sup>33</sup> gives a detailed explanation of the similar bias in BB meta-analysis of log-odds-ratio.

The confidence intervals for *bbmle* have the best coverage of  $\theta$ , slightly below the nominal 95% when  $K = 5$  (particularly when  $\theta = -1.5$ ) but differing little from 95% when  $K = 10$  and  $K = 20$ . Coverage of the *gamlss* intervals is lower, especially when  $\pi_C = 0.3$ , and it declines steadily as  $\theta$  increases from  $-1.5$  to  $+1.5$ ; the pattern changes little with  $n$  or  $K$ . The inverse-variance-based methods give close to nominal coverage when  $\theta = 0$ , but they deteriorate rapidly as  $\theta$  departs from 0 in either direction, and that pattern becomes worse as  $K$  increases (for example, when  $K = 20$ , coverage of the MP interval is near or below 50% at  $\theta = -1.5$  and  $\theta = 1.5$ , for all three values of  $n$ ). Patterns of bias such as those in Figure 5 would lead us to expect the patterns of coverage in Figure 6.

Overall, the inverse-variance methods do not help with estimation of RR in the beta-binomial model, as it requires the same constraints as the log-binomial model. Beta-binomial regression performs much better in estimation of  $\rho$  and RR, especially for  $K \geq 10$ . However, the use of the logit link, as in *gamlss*, does not provide sufficient coverage of RR, and only *bbmle* provides a viable option when the data follow a BB model. We discuss model misspecification issues in Sections 6 and 7.

## 5 Examples

In this section, we re-analyze the data from two medical meta-analyses, comparing two conventional random-effects methods (DL and REML) with the six methods developed for the beta-binomial model that we discussed in Section 4. The first meta-analysis, on the effect of diuretics on pre-eclampsia<sup>44</sup> considered the beneficial effects of treatment (i.e., the RR of benefit), whereas the second meta-analysis focused on side-effects of low-dosage tricyclic antidepressants in acute depression<sup>45</sup> (i.e., the RR of harm).

### 5.1 Example 1: Effect of diuretics on pre-eclampsia

A meta-analysis of nine trials, with a total of 6942 patients, evaluated the effect of diuretics on pre-eclampsia, a serious complication in pregnancy<sup>44</sup>. These data (Table 1) have also been analyzed (as odds ratios) by Thompson and Pocock<sup>46</sup>, Hardy and Thompson<sup>47</sup>, Biggerstaff and Tweedie<sup>48</sup>, Whitehead<sup>49</sup>, Viechtbauer<sup>50</sup>, Makambi and Lu<sup>51</sup>, and Kulinskaya and Olkin<sup>41</sup>. The incidence of pre-eclampsia

Table 1: Data from nine trials of diuretics for treatment of pre-eclampsia in pregnancy. The study-level estimate of  $\log(\text{RR})$ ,  $\hat{\theta}_j$ , comes from Equation (19).

Study	$Y_{jC}$	$n_{jC}$	$Y_{jC}/n_{jC}$	$Y_{jT}$	$n_{jT}$	$\hat{\theta}_j$
1	14	136	0.103	14	131	0.0373
2	17	134	0.127	21	385	-0.8471
3	24	48	0.500	14	57	-0.6947
4	18	40	0.450	6	38	-0.9953
5	35	760	0.046	12	1011	-1.3290
6	175	1336	0.131	138	1370	-0.2619
7	20	524	0.038	15	506	-0.2447
8	2	103	0.019	6	108	0.9083
9	40	102	0.392	65	153	0.0969

in the control group shows considerable heterogeneity: from 1.9% in Study 8 to 50.0% in Study 3. The incidence in over half of the studies is large enough that the odds ratio does not provide a satisfactory approximation for the risk ratio. The study-level estimates of  $\log(\text{RR})$  ( $\hat{\theta}_j$  from (19)) range from  $-1.33$  to  $+0.91$ .

For various methods Table 2 shows the estimated values of  $\tau^2$  for the conventional REM, and of  $\rho$  in the BB model. In the conventional REM the DerSimonian-Laird estimate of  $\tau^2$  is  $\hat{\tau}_{DL}^2 = 0.156$ , and  $\hat{\tau}_{REML}^2 = 0.199$ . Viechtbauer<sup>50</sup> gives Q-profile confidence intervals for DL, and Hardy and Thompson<sup>47</sup> give profile-likelihood confidence intervals for the REML method.

For the beta-binomial model, six methods provide estimates of  $\rho$ : 0.138 for *bbmle* and 0.139 for *gamlss*, and 0.008 to 0.019 for the method-of-moments, REML, MP, and BD estimators. The separation between results from BB regression methods (*bbmle/gamlss*) and the inverse-variance BB methods is in the direction that we would expect from the simulation results on bias in Figure 4, but the magnitude of  $\hat{\rho}$  from BB regression methods is greater (perhaps because of the particular mixture of values of the incidence of pre-eclampsia in the control group).

In *bbmle*, the maximum-likelihood-based estimates of the means of the two beta-binomial distributions are  $\hat{\pi}_T = 0.143$  and  $\hat{\pi}_C = 0.185$ , which result in  $RR = 0.774$ . The confidence intervals for estimates of probabilities and the overdispersion parameter  $\gamma$  ( $= (1 - \rho)/\rho = \alpha + \beta$ ) are based on the standard errors obtained from the inverse of the observed information matrix. The standard error for the log-risk-ratio is obtained by the delta method as a function of  $\pi_T$  and  $\pi_C$ , Equation (17).

In *gamlss* the estimates for probabilities are obtained from the beta-binomial regression model with logit link function. The estimates for probabilities obtained by inverting the logit link function,  $\hat{\pi}_T = 0.143$  and  $\hat{\pi}_C = 0.185$ , yield  $RR = 0.773$ . In *gamlss* the intra-cluster correlation is defined as  $\rho = \sigma/(\sigma+1)$ , where  $\sigma = 1/(\alpha+\beta)$  is the overdispersion parameter. The parameter  $\sigma$  has a log link function. Standard



Table 2: Point estimates and confidence intervals for  $\tau^2$ ,  $\rho$ , log-risk-ratio (LRR), and risk ratio (RR) in the example of diuretics in pre-eclampsia; FEM is the fixed-effect model, REM is the random-effects model, and BB is the beta-binomial model. *bbmle* and *gamlss* are beta-binomial maximum-likelihood-based and generalized-additive-regression models. The heterogeneity parameter is  $\tau^2$  in REM and  $\rho$  in the BB model.  $L$  and  $U$  denote the lower and upper limits of the 95% confidence intervals (CIs).

Model	Method	Overdispersion parameter	$L$	$U$	LRR	$L$	$U$	Length of CI	RR	$L$	$U$
		$\tau^2$									
FEM					-0.305	-0.449	-0.161	0.288	0.737	0.638	0.851
REM	DL&IV	0.156	0.049	1.582	-0.437	-0.768	-0.107	0.661	0.646	0.464	0.899
REM	REML&IV	0.199	0.032	0.989	-0.439	-0.799	-0.079	0.720	0.645	0.450	0.924
		$\rho$									
BB	MoM&IV	0.008	0.002	0.093	-0.297	-0.563	-0.032	0.530	0.743	0.570	0.969
BB	REML&IV	0.010	0.001	0.061	-0.305	-0.595	-0.014	0.581	0.737	0.551	0.986
BB	MP&IV	0.016	0.002	0.093	-0.316	-0.644	0.012	0.632	0.729	0.525	1.011
BB	BD&IV	0.019	0.003	0.106	-0.321	-0.668	0.025	0.693	0.725	0.513	1.026
BB	bbmle	0.138	0.077	0.258	-0.257	-1.008	0.495	1.504	0.774	0.365	1.640
BB	gamlss	0.139	0.057	0.300	-0.257	-0.948	0.433	1.381	0.773	0.388	1.542

errors and confidence intervals for  $\sigma$  are obtained in the log scale. The relation between  $\rho$  and  $\sigma$  yields a confidence interval for  $\rho$ .

The estimate of the overall risk ratio is highest (0.774) in the *bbmle* model, and its confidence interval is the longest (1.275). The estimate of the risk ratio is lowest (0.645) in the conventional REM with  $\hat{\tau}_{REML}^2 = 0.199$ .

The results of this example can be compared with simulation results for  $K = 10$  (in Figure 5), since it has nine studies. Thus, for  $\hat{\theta}_{bbmle} = -0.257$ , the bias of *bbmle* and *gamlss* is about 0.20, which leads to the estimate  $\hat{\theta}_{True} = -0.257 - 0.20 = -0.457$ . For  $\hat{\theta}_{BD} = -0.321$ , the bias of the Breslow-Day method for estimating  $\rho$  is 0.10, which leads to the estimate  $\hat{\theta}_{True} = -0.321 - 0.10 = -0.421$ .

## 5.2 Example 2: Side-effects of low-dosage tricyclic antidepressants in acute-phase depression

Systematic review CD003197<sup>45</sup> in the Cochrane Library compared the effects and side effects of low-dosage tricyclic antidepressants (TCA) with placebo and with standard-dosage tricyclics in acute-phase treatment of depression. Comparison 2 Outcome 6 is the meta-analysis of the rate of side effects in the low-dosage TCA group vs placebo. Table 3 gives the data on numbers of patients experiencing at least one side effect.

The incidence of side effects in the placebo group is substantial; only 2 of the 16 studies have  $Y_{jC}/n_{jC} < 10\%$ , and the highest is 75%. Again, the odds ratio does not provide a satisfactory approximation for the risk ratio. One would expect more patients in the treatment group to report side effects ( $RR > 1, \theta > 0$ ). The values of

Table 3: Data on numbers of patients experiencing at least one side effect in studies of low-dosage tricyclic antidepressants vs. placebo. The study-level estimate of  $\log(\text{RR})$ ,  $\hat{\theta}_j$ , comes from Equation (19).

Study	$Y_{jC}$	$n_{jC}$	$Y_{jC}/n_{jC}$	$Y_{jT}$	$n_{jT}$	$Y_{jT}/n_{jT}$	$\hat{\theta}_j$
1	17	28	0.607	16	24	0.667	0.092
2	7	10	0.700	12	12	1.000	0.336
3	3	12	0.250	8	13	0.615	0.810
4	30	62	0.484	29	60	0.483	-0.001
5	14	53	0.264	34	60	0.567	0.744
6	5	21	0.238	14	20	0.700	1.017
7	13	46	0.283	37	45	0.822	1.043
8	45	60	0.750	56	60	0.933	0.217
9	31	82	0.378	52	95	0.547	0.364
10	0	10	0.000	3	16	0.188	1.494
11	9	47	0.191	51	110	0.464	0.846
12	5	20	0.250	8	20	0.400	0.435
13	3	16	0.188	7	15	0.467	0.825
14	43	72	0.597	63	72	0.875	0.378
15	1	29	0.034	8	28	0.286	1.769
16	5	23	0.217	5	17	0.294	0.295

$\hat{\theta}_j$  show substantial heterogeneity, and they suggest a mixture of four groups: eight values from  $-0.001$  to  $0.435$ , six values from  $0.744$  to  $1.043$ , one at  $1.494$ , and one at  $1.769$ .

Table 4 shows the estimated values of  $\tau^2$  in the conventional REM and of  $\rho$  in the BB model. The DerSimonian-Laird estimate is  $\hat{\tau}_{DL}^2 = 0.047$ , and  $\hat{\tau}_{REML}^2 = 0.068$ . For the beta-binomial model, the six estimates of  $\rho$  again separate into two clumps: the method-of-moments, REML, MP, and BD estimates range from  $0.006$  to  $0.031$ , and *bbmle* and *gamlss* both produce  $0.175$ . From the simulation results in Figure 4 (especially those for  $n = 100$  and  $K = 20$ ), we might expect such a separation, but the values of the other parameters in the simulations are not close to the estimated values in this example.

Both *bbmle* and *gamlss* yield  $\hat{\pi}_T = 0.593$ ,  $\hat{\pi}_C = 0.347$ , and hence  $RR = 1.708$ . This is likely to be a reasonable estimate of RR. From the simulation results in Figure 5, we would expect the other BB estimates to have negative bias. The estimates from the REM are also likely to be low. Of *bbmle* and *gamlss*, the latter has a substantially wider confidence interval.

## 6 Risk ratio in Cochrane reviews

To explore the practical implications of the restricted range in meta-analyses of risk ratio, we reviewed random-effects meta-analyses that used RR in Cochrane Library Issue 4, 2004. As in the Cochrane Collaboration's Review Manager<sup>52</sup>, we used

Table 4: Point estimates and confidence intervals for  $\tau^2$ ,  $\rho$ , log-risk-ratio (LRR), and risk ratio (RR) in the example of side effects of low-dosage tricyclic antidepressants vs. placebo; FEM is the fixed-effect model, REM is the random-effects model, and BB is the beta-binomial model. *bbmle* and *gamlss* are beta-binomial maximum-likelihood-based and generalized-additive-regression models. The heterogeneity parameter is  $\tau^2$  in REM and  $\rho$  in the BB model.  $L$  and  $U$  denote the lower and upper limits of the 95% confidence intervals (CIs).

Model	Method	Overdispersion parameter	L	U	LRR	L	U	Length of CI	RR	L	U
		$\tau^2$									
FEM					0.355	0.258	0.452	0.194	1.426	1.294	1.571
REM	DL&IV	0.047	0.005	0.329	0.461	0.286	0.636	0.350	1.586	1.331	1.889
REM	REML&IV	0.068	0.005	0.275	0.480	0.286	0.674	0.388	1.616	1.331	1.961
		$\rho$									
BB	MoM&IV	0.028	0.002	0.107	0.368	0.217	0.520	0.303	1.445	1.242	1.682
BB	REML&IV	0.031	0.004	0.109	0.369	0.214	0.525	0.311	1.447	1.238	1.690
BB	MP&IV	0.026	0.002	0.107	0.368	0.219	0.517	0.298	1.445	1.245	1.676
BB	BD&IV	0.006	-0.008	0.073	0.359	0.246	0.473	0.227	1.432	1.279	1.604
BB	bbmle	0.175	0.107	0.276	0.535	0.384	0.686	0.302	1.708	1.469	1.987
BB	gamlss	0.175	0.102	0.284	0.535	0.326	0.745	0.419	1.708	1.384	2.107

inverse-variance-weighted meta-analysis and estimated the between-study variance  $\tau^2$  by the DerSimonian-Laird method. We also included the BB-based analysis using *bbmle*.

We considered only the 2115 meta-analyses with  $K \geq 3$  studies. Among those, 1286 MAs had  $\hat{\tau}^2 > 0$  (by our calculations, using *metabin* from the R package *meta*). Those 1286 MAs included 8940 studies with  $n_C \geq 5$  and  $n_T \geq 5$ .

For Study  $j$  in MA  $m$  we calculated the estimated log(RR),  $\hat{\theta}_{mj}$ , from (19) and its within-study variance  $v_{mj}$  from (21) with  $\rho = 0$ . (These calculations aim to minimize bias in  $\hat{\theta}_{mj}$  and  $v_{mj}$ . They add 1/2 to each cell of the  $2 \times 2$  table for each study, whereas the conventional ones add 1/2 only when the  $2 \times 2$  table contains a zero cell.) The FE weights are  $w_{mj}^F = 1/v_{mj}$ , and the RE weights are  $w_{mj}^R = (v_{mj} + \hat{\tau}_m^2)^{-1}$ . For the FEM and the REM, we set  $w_{mj}$  equal to  $w_{mj}^F$  and  $w_{mj}^R$ , respectively, to obtain the combined effects  $\hat{\theta}_m$  and their estimated variances  $1/\sum_j w_{mj}$ .

For the FEM and the REM we calculated studentized residuals  $r_{mj} = (\hat{\theta}_{mj} - \hat{\theta}_m)/s_{mj}$ , now defining  $s_{mj}^2 = \text{Var}(\hat{\theta}_{mj} - \hat{\theta}_m) = 1/w_{mj} - 1/\sum_j w_{mj}$  for the respective weights. If, in the model for the log-risk-ratio, the assumption  $b_{mj} \sim N(0, \tau_m^2)$  holds for these Cochrane reviews, these  $r_{mj}$  should have approximately the standard normal distribution. Because no single meta-analysis involves enough studies to assess the reasonableness of that assumption, we combined the  $r_{mj}$  from all studies in MAs with  $\hat{\theta} \geq 0$  and  $\hat{\tau}^2 > 0$ . The Q-Q plot in Figure 7 shows that the distribution of the  $r_{mj}$  is not well approximated by a normal distribution. Although less striking, the corresponding plot of the  $r_{mj}$  from the studies in MAs with  $\hat{\theta} < 0$  and  $\hat{\tau}^2 > 0$  (not

Table 5: Studies in 1286 meta-analyses in the Cochrane Database that used REM for RR and had  $\hat{\tau}^2 > 0$ , cross-classified by the estimated probability of truncation for  $\pi_T$  and whether  $\hat{\theta} \geq 0$

	Estimated probability of truncation of $\pi_T$				
	< 0.05	[0.05 – 0.15)	[0.15 – 0.25)	[0.25 – 0.35)	[0.35 – 0.45)
$\hat{\theta} \geq 0$	1725	192	68	59	50
$\hat{\theta} < 0$	6249	362	88	56	18
Total	7974	554	156	115	68

	Estimated probability of truncation of $\pi_T$				
	[0.45 – 0.55)	[0.55 – 0.65)	[0.65 – 0.75)	[0.75 – 0.85)	$\geq 0.85$
$\hat{\theta} \geq 0$	16	22	16	14	5
$\hat{\theta} < 0$	0	0	0	0	0
Total	16	22	16	14	5

shown) reinforces that message. The example in Section 5.2 suggests an additional departure: the studies' effects may come from a mixture of distributions. This could help to account for the appearance, in Figure 7, of a distribution whose tails are lighter than normal.

Building on the analysis in Section 3.2, we also inquire into the impact of the restriction in Equation (8). Because conventional meta-analysis starts with the log of the sample RR, and the range of the log function is unbounded, it might seem that the restriction would have no impact. However, the basic data for each study include  $y_{jC}$  and  $n_{jC}$ , and under the usual binomial model  $\hat{\pi}_{jC} = y_{jC}/n_{jC}$  is an unbiased estimate of  $\pi_{jC}$ . Thus, when  $\hat{\pi}_{jT} = y_{jT}/n_{jT} > \hat{\pi}_{jC}$  and hence  $\hat{\theta}_j > 0$ , larger  $\pi_{jC}$  (and hence  $-\log\pi_{jC}$  closer to 0) will increase the impact of the restriction. We investigate the impact by estimating the probability of violating Equation (8), which we refer to as the truncation probability. For a normally distributed random effect  $b_{mj}$  this is approximated by  $1 - \Phi((-\log(\hat{\pi}_{mjC}) - \hat{\theta}_m)/\hat{\tau}_m)$ . We grouped these estimated probabilities into 10 bins: < 0.05, [0.05 – 0.15), [0.15 – 0.25), ...,  $\geq 0.85$ . Table 5 shows the numbers of studies in each bin.

In total, 966 studies had truncation probability  $\geq 0.05$ : 442 studies from 188 MAs with  $\hat{\theta} \geq 0$  and 524 studies from 241 MAs with  $\hat{\theta} < 0$ . These 429 MAs out of 1286 (exactly 1/3 of the MAs using REM for RR) are likely to have reported biased conclusions. For the MAs with  $\hat{\theta} \geq 0$  and  $\hat{\tau}^2 > 0$ , Figure 8 shows boxplots of the  $r_{mj}$  in each of the ten intervals of probability of truncation for  $\pi_{jT}$ . The distance between  $\hat{\theta}_{mj}$  and  $\hat{\theta}_m$  is strongly related to the truncation probability. When the truncation probability is  $\geq 0.35$ , the median of the  $r_{mj}$  is at or below  $-1$ . In these 123 studies the violation of the assumptions of the conventional REM is likely to be greatest. Separate Q-Q plots of the  $r_{mj}$  for the studies with truncation probability  $< 0.35$  and the studies with truncation probability  $\geq 0.35$  (not shown) support this conclusion.

The three panels of Figure 9 plot the estimates of  $\log(\text{RR})$  from the REM,

the estimates of  $\log(\text{RR})$  from *bbmle*, and the difference between them versus the estimates from the FEM. For the 353 MAs with  $\hat{\theta}_{REM} \geq 0$  and  $\hat{\tau}^2 > 0$  in panel (a), the majority of REM estimates are above their FEM counterparts, sometimes very substantially. This pattern supports the impression that the positive risk ratios and their significance reported from the conventional REM are overestimates. We conclude that the positive values of  $\log(\text{RR})$  estimated from REM are likely to be overestimates. Simulations performed previously to ascertain the quality of those estimates are likely to have provided downward-biased results, compensating for this overestimation.

The *bbmle* estimates from the 713 MAs with  $\hat{\theta}_{bbmle} \geq 0$  and  $\hat{\rho} > 0$  in panel (b) follow the same pattern, perhaps even more so. There are more positive values of  $\hat{\theta}$  from *bbmle* than from REM, and they appear to have higher values. Relative to the estimates from the FEM, the differences between the estimates from the REM and those from *bbmle*, in panel (c), are more often negative, and they are often not small. A difference of 0.223 in  $\log(\text{RR})$  corresponds to a factor of 1.25 in RR. These differences may be due to the differences in the underlying assumptions about the distribution of random effects in the two models, and hence to model misspecification in one of them, and/or to the biases discussed in Sections 3 and 4.

At the suggestion of a referee, we repeated the analyses of this section, restricting the participating meta-analyses to only the 128 with  $K \geq 10$  studies, to guard against imprecise estimation of  $\tau^2$  for small  $K$ . The results, in Appendix B.2 (Figures 12-14), are qualitatively similar.

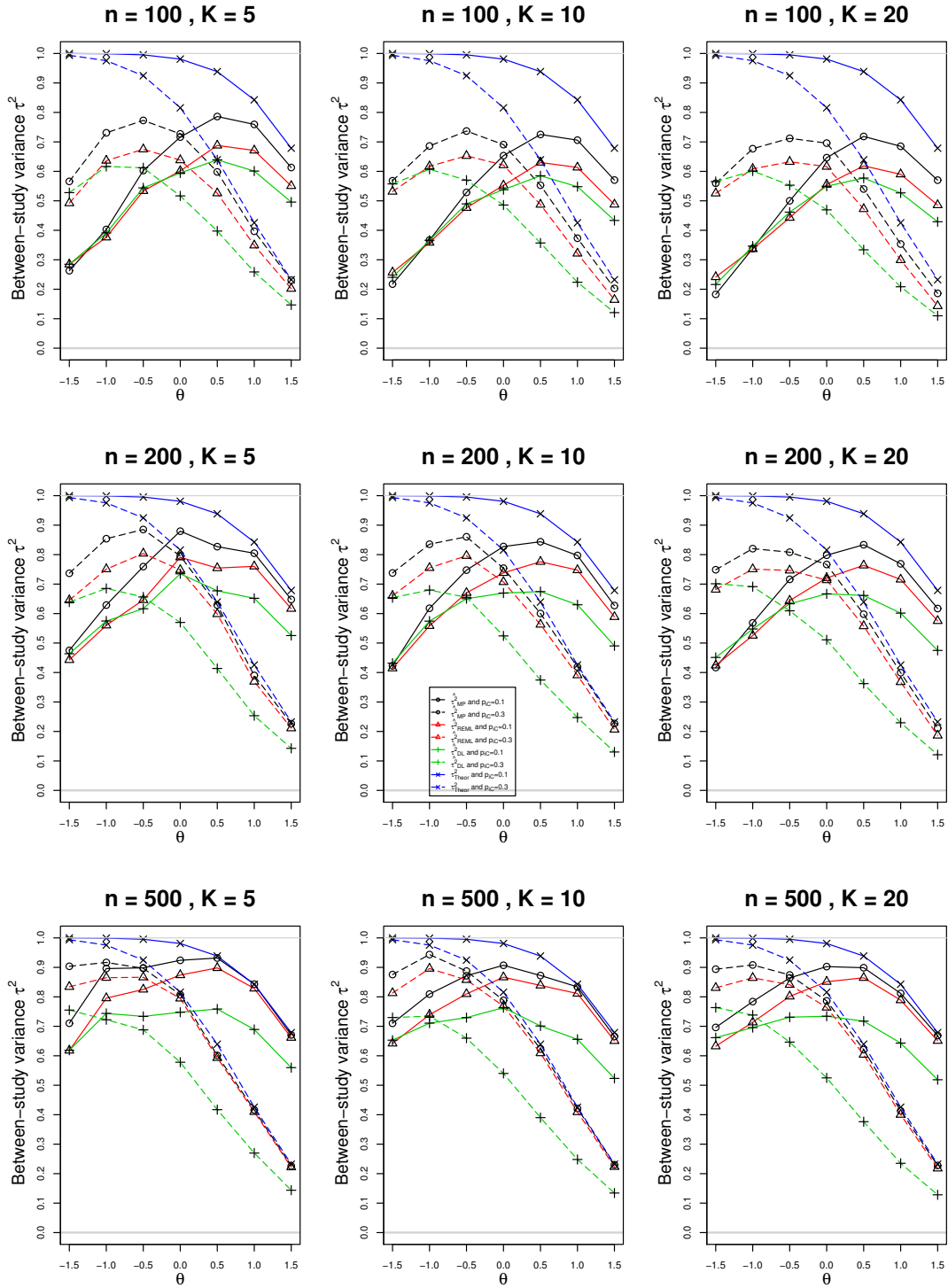


Figure 1: Relation of estimates of the between-studies variance ( $\tau^2$ ) to the overall log-risk-ratio ( $\theta$ ) in  $K$  studies, each of total sample size  $n$ , when data come from the binomial-normal model with point mass for  $\tau^2 = 1$  and  $\pi_{jC} = 0.1$  (solid lines) and  $0.3$  (dashed). The Mandel-Paule (circle), REML (triangle), and DerSimonian-Laird (plus) estimators are compared with the true variance (cross). Light grey line at 1.

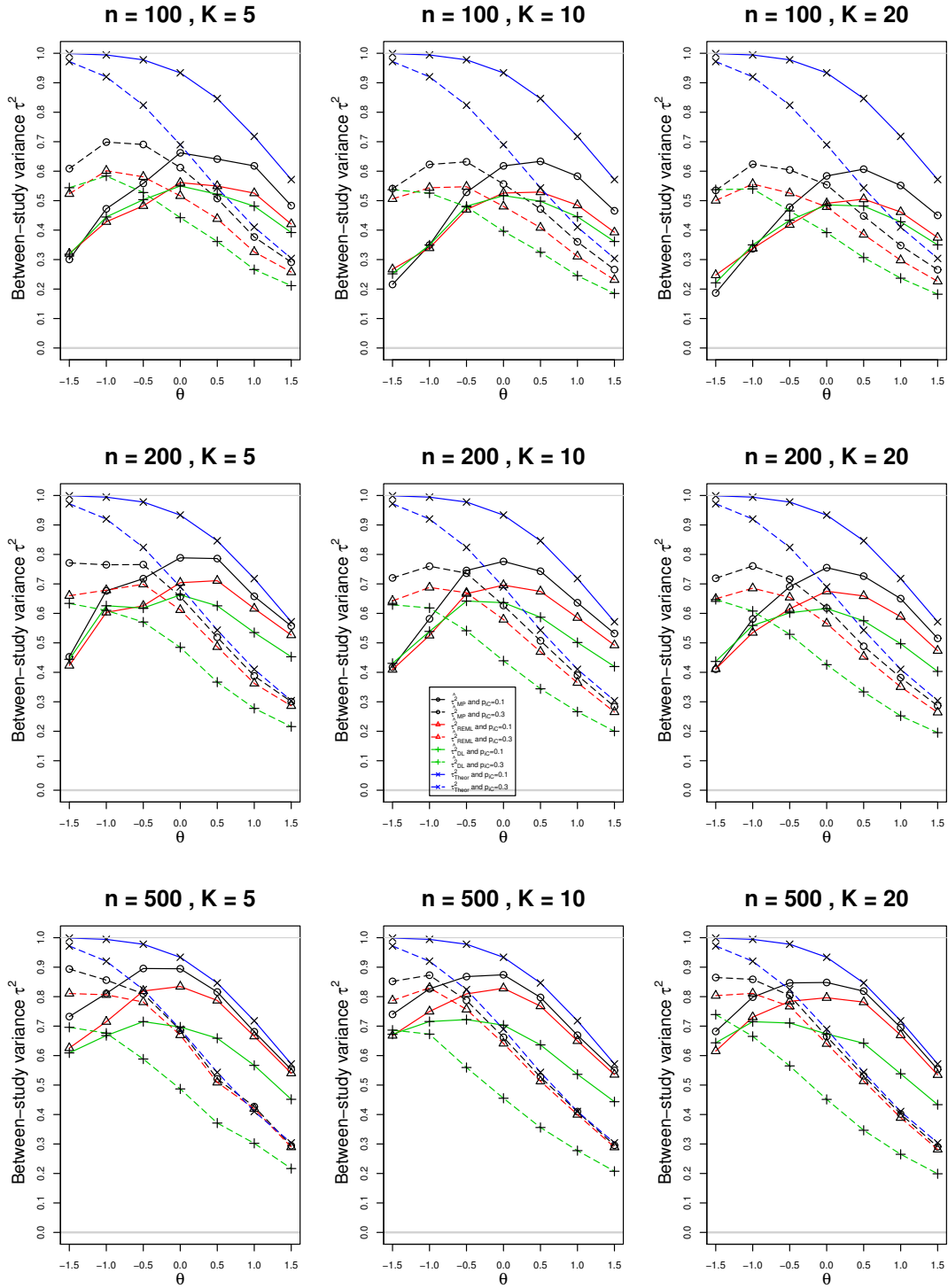


Figure 2: Relation of estimates of the between-studies variance  $\tau^2$  to the overall log-risk-ratio ( $\theta$ ) in  $K$  studies, each of total sample size  $n$ , when data come from the binomial-normal model with truncation for  $\tau^2 = 1$  and  $\pi_{jC} = 0.1$  (solid lines) and 0.3 (dashed). The Mandel-Paule (circle), REML (triangle), and DerSimonian-Laird (plus) estimation methods are compared with the true variance (cross). Light grey line at 1.

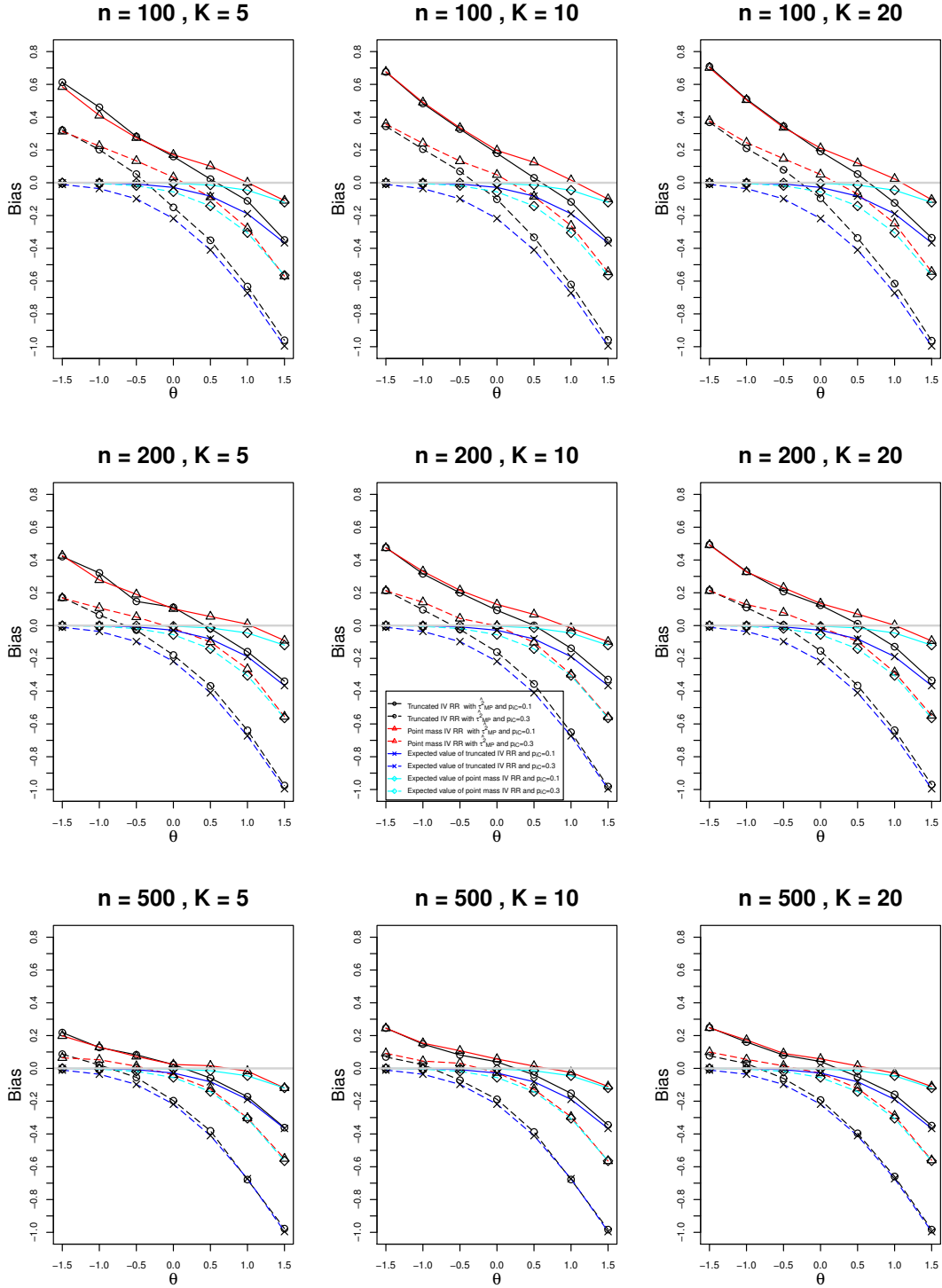


Figure 3: Relation (to the overall log-risk-ratio,  $\theta$ ) of bias in the conventional method of estimating the log-relative-risk,  $\theta$ , in the binomial-normal model from  $K$  studies, each of total sample size  $n$ , with truncation (circle) or point-mass (triangle) option, when  $\tau^2$  (true value,  $\tau^2 = 1$ ) is estimated by the Mandel-Paule method, compared with true bias from truncation (cross) and point mass (diamond).  $\pi_{jC} = 0.1$  (solid lines) and 0.3 (dashed). Light grey line at 0.



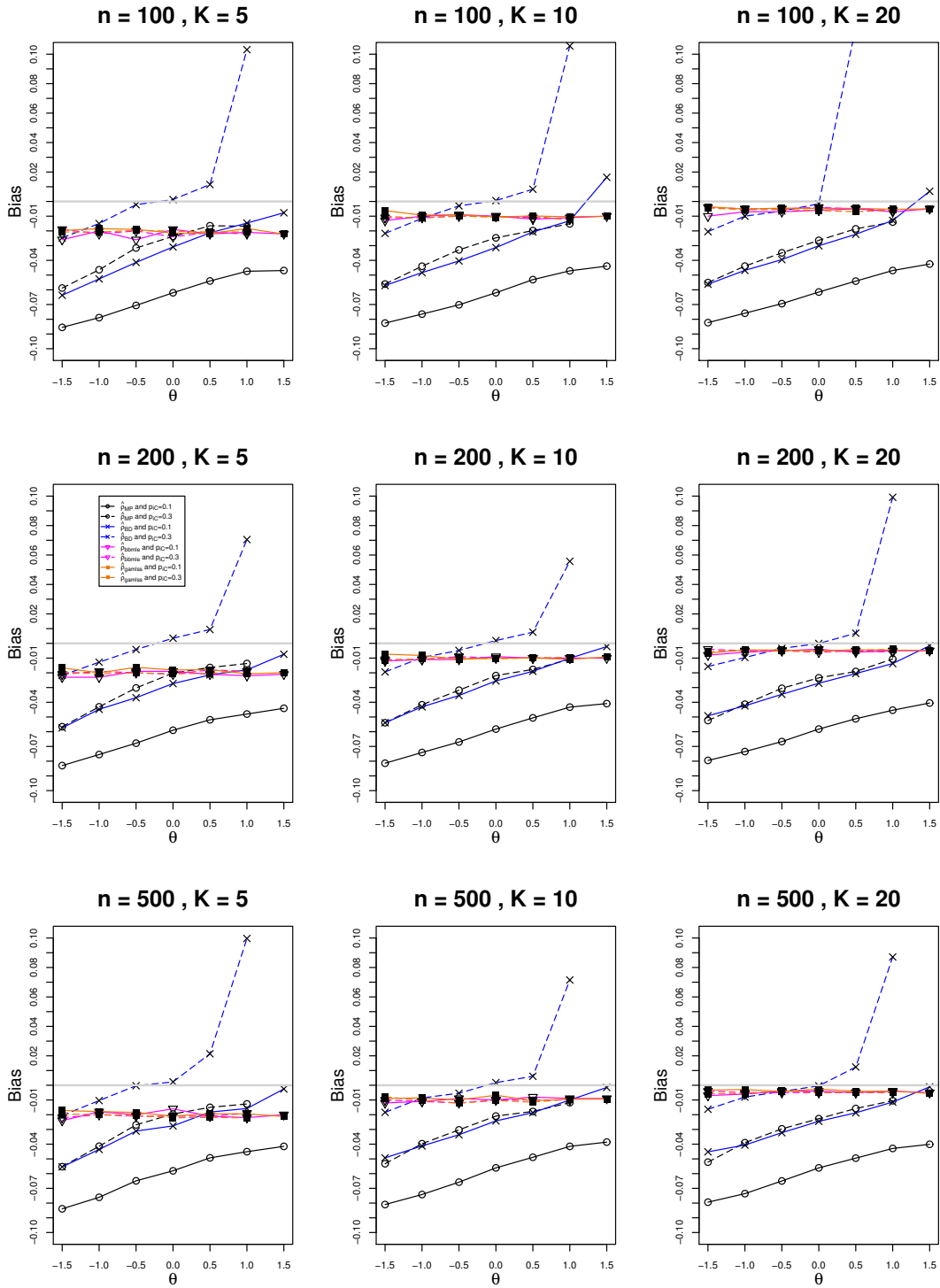


Figure 4: Relation (to the overall log-risk-ratio,  $\theta$ ) of bias in estimating  $\rho$  from  $K$  studies, each of total size  $n$ , in the beta-binomial model for  $\rho = 0.1$  and  $\pi_C = 0.1$  (solid lines) and 0.3 (dashed). The methods are Mandel-Paule (circle), Breslow-Day (cross), *bbmle* (reverse triangle), and *gamlss* (filled square). Light grey line at 0.

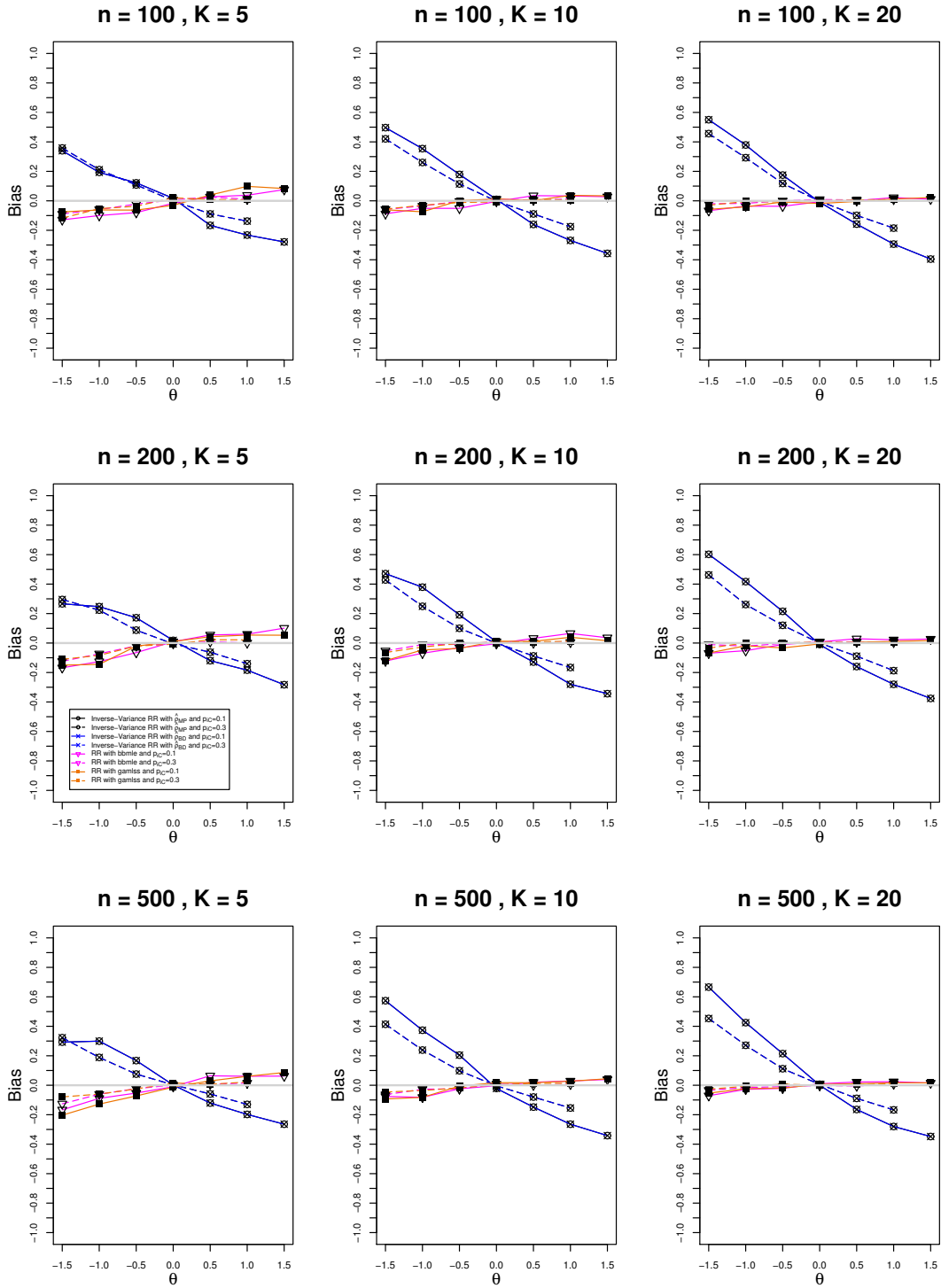


Figure 5: Bias in estimating the overall log-risk-ratio,  $\theta$ , from  $K$  studies, each of total size  $n$ , in the beta-binomial model for  $\rho = 0.1$  and  $\pi_C = 0.1$  (solid lines) and 0.3 (dashed). The log-relative-risk is estimated by using inverse-variance weights. The methods for estimation of  $\rho$  are Mandel-Paule (circle), Breslow-Day (cross), *bbmls* (reverse triangle), and *gamlss* (filled square). Light grey line at 0.

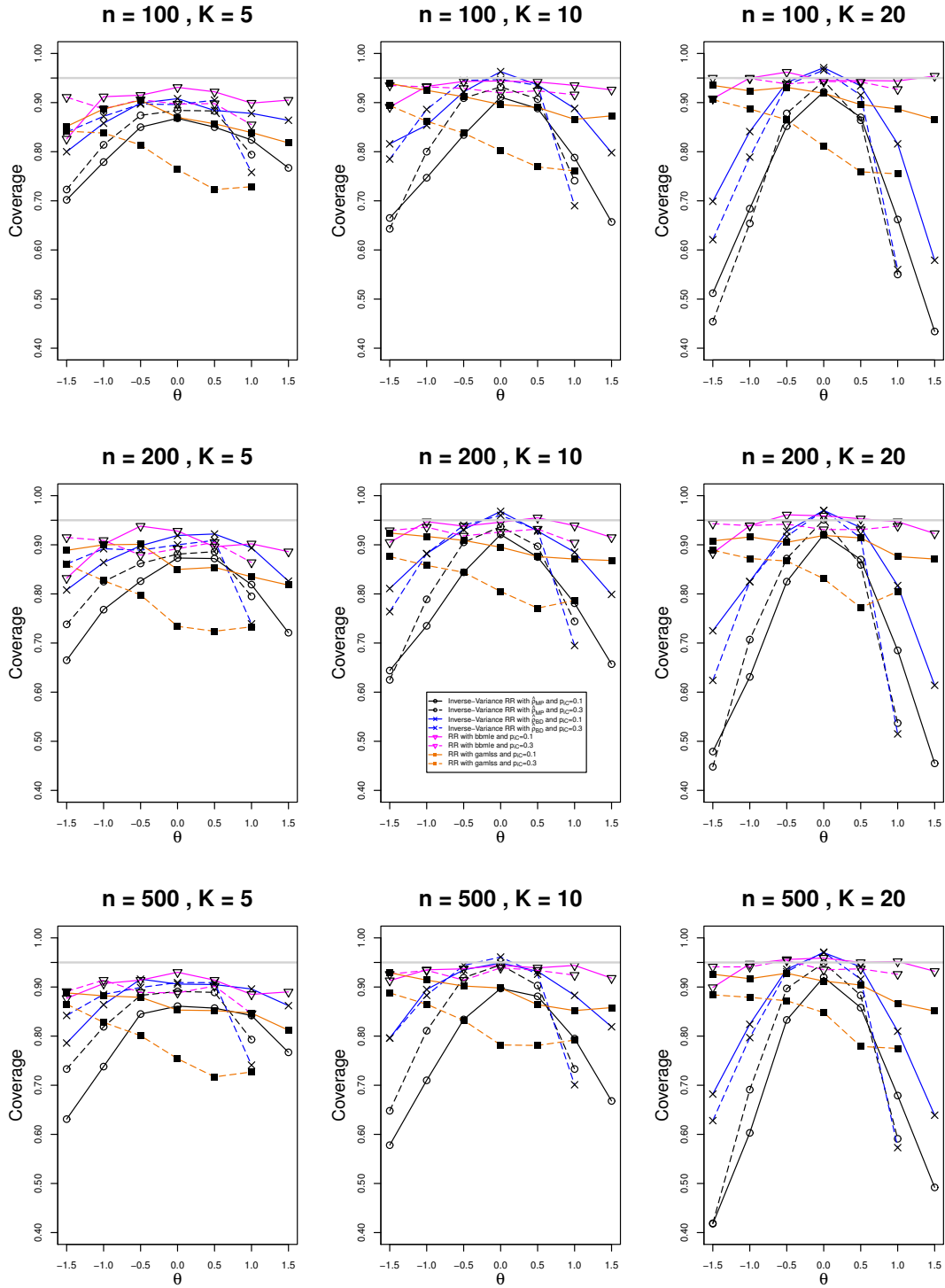


Figure 6: Coverage of the overall log-risk-ratio,  $\theta$ , from  $K$  studies, each of total size  $n$ , in the beta-binomial model for  $\rho = 0.1$  and  $\pi_C = 0.1$  (solid lines) and  $0.3$  (dashed). The log-relative-risk is estimated by using inverse-variance weights. The methods for estimation of  $\rho$  are Mandel-Paule (circle), Breslow-Day (cross), *bbmls* (reverse triangle), and *gamlss* (filled square). Light grey line at 0.95.

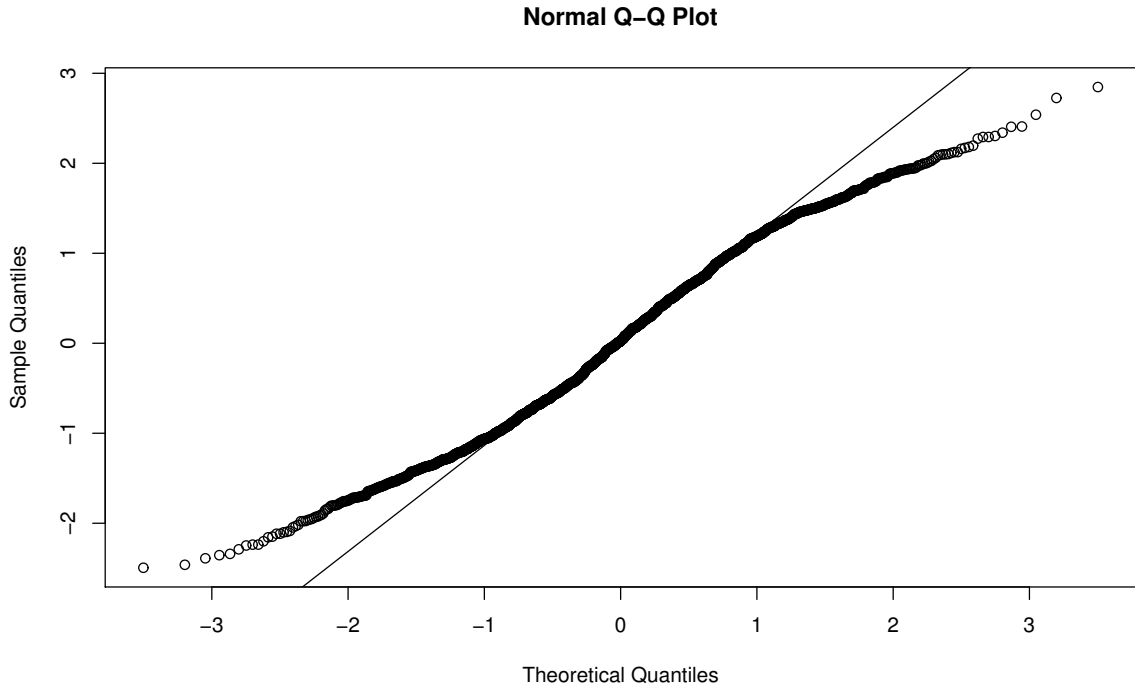


Figure 7: Normal Q-Q plot of the studentized residuals for the studies from REM meta-analyses of log-risk-ratio with  $\hat{\theta} \geq 0$ ,  $\hat{\tau}^2 > 0$  in Cochrane Library Issue 4.

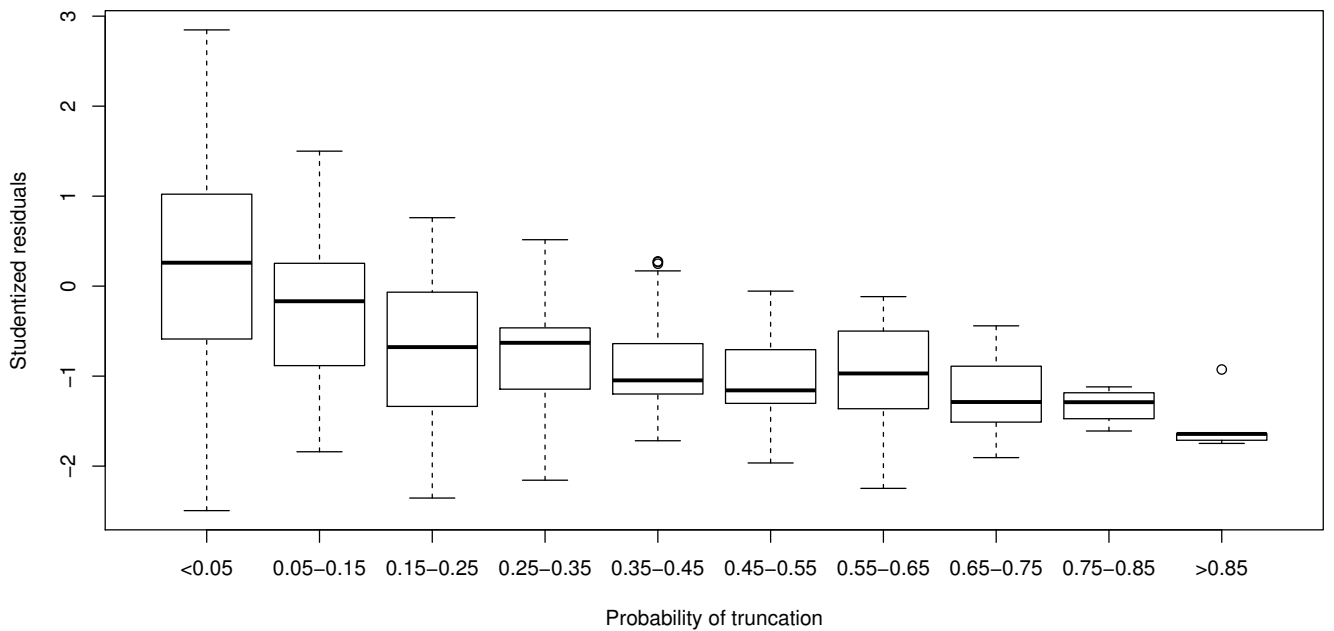


Figure 8: Boxplots of studentized residuals by truncation probability, for the studies from REM meta-analyses of log-risk-ratio with  $\hat{\theta} \geq 0$ ,  $\hat{\tau}^2 > 0$  in Cochrane Library Issue 4.

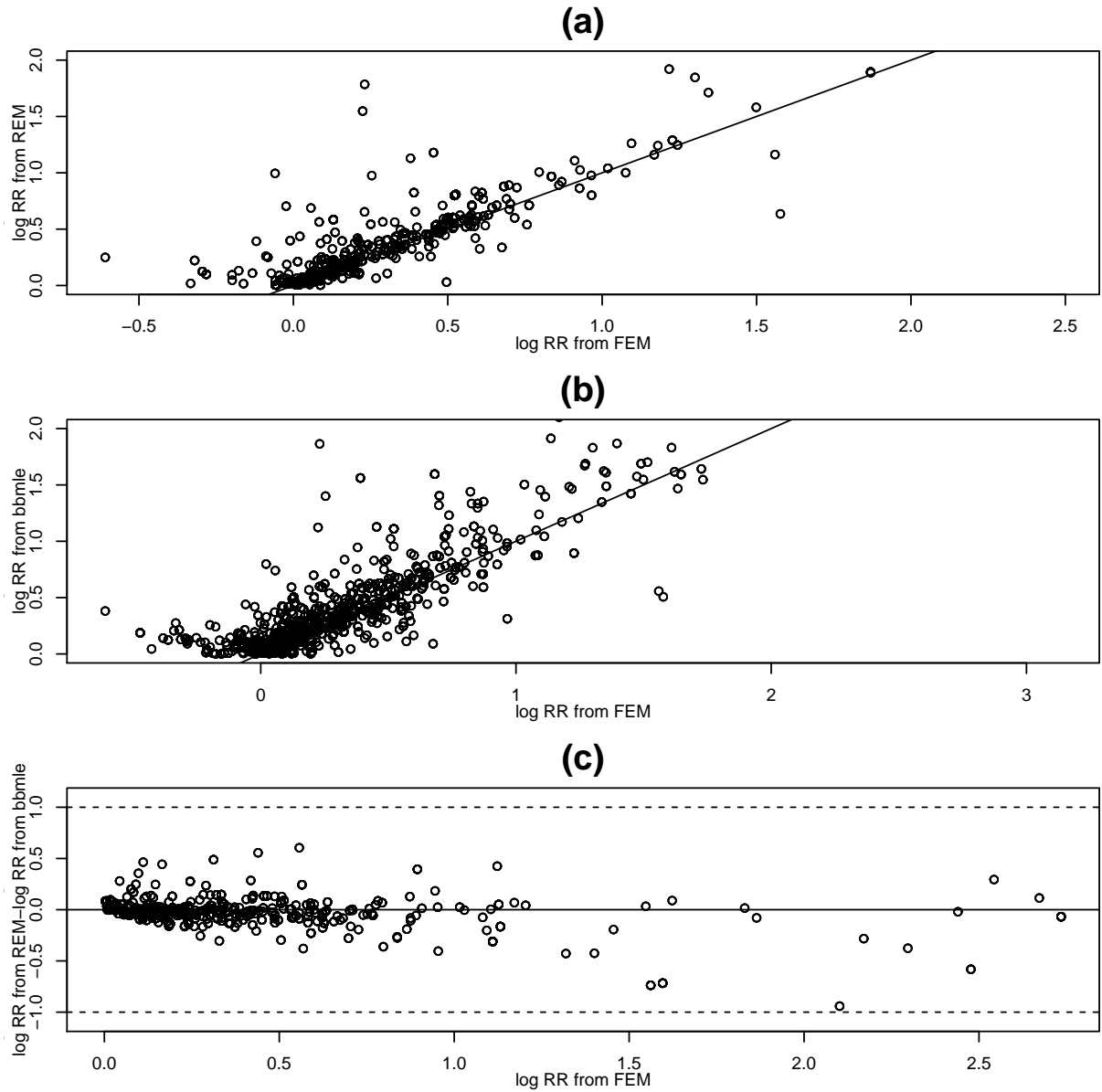


Figure 9: Scatterplot (vs. log-risk-ratio from FEM) of the meta-analytic estimates of log-risk-ratio obtained by: (a) REM, for the 353 REM meta-analyses of RR with  $\hat{\theta}_{REM} \geq 0$ ,  $\hat{\tau}^2 > 0$ ; (b) *bbmle*, for the 713 meta-analyses of RR with  $\hat{\theta}_{bbmle} \geq 0$ ,  $\hat{\rho} > 0$ ; (c) Difference between log(RR) from REM and *bbmle* for the 353 meta-analyses with  $\hat{\rho} > 0$  and  $\hat{\theta}_{bbmle} \geq 0$  and  $\tau^2 > 0$  and  $\hat{\theta}_{REM} \geq 0$ .

## 7 Discussion

With models for log-odds-ratio as background, we have focused on models for meta-analysis of log-risk-ratio, for two main reasons. First, when the event probabilities are not small, the risk ratio is often more appropriate than the odds ratio. Second, in generating binomial data for study-level  $2 \times 2$  tables under a random-effects model for log-risk-ratio, one must impose a restriction to ensure that  $\pi_{jT} < 1$  (in addition to having  $0 < \pi_{jC} < 1$ ). Thus, in the conventional random-effects model, we explored the consequences of the restrictions on the parameter space. A small simulation study showed that they lead to bias in estimates of  $\tau^2$  and in the estimate of the overall log-risk-ratio.

The alternative of obtaining the data in the  $2 \times 2$  tables from beta-binomial distributions, and using the log link function, has the same complications as in the log-binomial model. In order to use the conventional meta-analysis models for  $\log(\text{RR})$ , one can estimate the  $\pi_{jT}$  and  $\pi_{jC}$  in each study by maximum likelihood or from the estimated logits in a beta-binomial regression with logit link. The variances of the resulting estimates of  $\log(\text{RR})$  involve  $\rho$  (intra-cluster correlation). We considered several ways of estimating  $\rho$ , but another small simulation study showed bias in estimates of  $\rho$  that was often unacceptable, though *bbmle* provided reasonable point and interval estimates of the overall  $\log(\text{RR})$  when the data were generated from BB distributions.

Thus, neither the log-binomial model nor the beta-binomial model is satisfactory for meta-analysis of (log of) risk ratio. Because the range of the log function is unbounded, it might seem that conventional meta-analysis of  $\log(\text{RR})$  would avoid the complications associated with restrictions on the parameter space, but it does not. Many meta-analyses use (log of) risk ratio as the effect measure (e.g., in the Cochrane reviews summarized in Section 6), however, so reliable methods are needed.

Importantly, the standard log-binomial-normal model and the beta-binomial model are based on different assumptions about the unobservable mixing distributions. The results from these and similar two-stage models are not always robust against violations of distributional assumptions. For instance, misspecification of the random-effects distribution in GLMMs can induce bias in the estimates of the linear predictor parameters and severe bias in estimates of the variance components. Alonso et al.<sup>53</sup> give a comprehensive discussion. For meta-analysis of odds ratios using BB and REM, these misspecification biases are demonstrated in Bakbergenuly and Kulinskaya<sup>33</sup> (Supplementary Material D). Unfortunately, it is very difficult to determine the true data-generating mechanism for the random effects, especially when dealing with sparse data; Drikvandi et al.<sup>54</sup> discuss some developments.

For the log-risk-ratio, the complications in the log-binomial model, Equation (6), arise from the restriction on  $\theta$  introduced by the relation between  $\theta$  and the nuisance parameters,  $\alpha_j = \log(\pi_{jC})$ . More technically, the joint range of  $\theta$  and  $\log(\pi_C)$  is a proper subset of the set  $\{(\log(\pi_C), \theta) : \log(\pi_C) \leq 0, -\infty < \theta < +\infty\}$ . By contrast, for the odds ratio,  $-\infty < \text{logit}(\pi_C) < +\infty$ ,  $-\infty < \log(\text{OR}) < +\infty$ , and the joint range of  $\log(\text{OR})$  and  $\text{logit}(\pi_C)$  is the entire real plane. To circumvent this difficulty

with the log-risk-ratio, Richardson et al.<sup>55</sup> propose a new nuisance parameter, the log of the odds-product (OP). The log-OP ranges from  $-\infty$  to  $+\infty$ , and this choice of nuisance parameter has the advantage that both  $\log(\text{RR})$  and the transformed risk difference,  $\text{arctanh}(\text{RD}) = \log((1+\text{RD})/(1-\text{RD}))$ , can be modeled independently of  $\log(\text{OP})$ . The introduction of the log of the odds-product as the nuisance parameter in models of  $\log(\text{RR})$  opens up a promising approach that should be the focus of substantial further research. However, it is plausible that this approach will suffer from transformation bias and other biases, which are the bane of the existing models for RR. In the interim, the use of odds ratio instead of relative risk appears to be a safer option.

For the two examples in Section 5, the choice of method seems to matter. Although the confidence intervals overlap, the estimates of the overall  $\log(\text{RR})$  separate into three groupings: conventional REM with inverse-variance weights, BB with inverse-variance weights, and BB maximum likelihood. In interpreting the estimates, however, some caution is appropriate. The study-level estimates of  $\log(\text{RR})$  in both examples suggest a mixture. Thus, a single distribution (as in the random-effects models) may not be an adequate description of the heterogeneity. Lin et al.<sup>56</sup> argue that, if heterogeneity is present, it should permeate the entire collection of studies, instead of being limited to a small number of outlying studies. We add that presence of distinct groupings also represents a departure from regular heterogeneity. In such situations it may be appropriate to model a cluster structure by a finite-mixture distribution<sup>57</sup> or a product-partition model<sup>58</sup> or to consider a fixed-effects analysis<sup>59</sup>.

For a sizable number of meta-analyses of risk ratio in Cochrane reviews, we derived studentized residuals from the difference between the study-level estimate and the overall estimate of the log-risk-ratio. When combined across meta-analyses, the studentized residuals had a distribution that departed from a normal distribution, by having lighter tails. And when categorized into intervals of the estimated truncation probability, the studentized residuals showed a strong association between larger truncation probability and more-negative difference between the study-level estimate and the overall estimate. Because the vast majority of studentized residuals in meta-analyses with  $\hat{\theta} \geq 0$  belonged to the bins where the truncation probability was  $< 0.35$ , we suspect upward bias in the overall estimates.

Other effect measures have restrictions on their parameter spaces. We would expect similar results for risk difference, response ratio, and  $\arcsin(\sqrt{p})$  for binomial proportions. The development of appropriate methodology for this important problem is an urgent task.

## Declaration of Conflicting Interests

The authors declare that they have no conflicts of interest.

## Funding

The work by E. Kulinskaya was supported by the Economic and Social Research Council [grant number ES/L011859/1].

## References

- [1] J. L. Fleiss, B. Levin, and M. C. Paik, *Statistical Methods for Rates and Proportions*. Hoboken, NJ: John Wiley & Sons, 3rd ed., 2003.
- [2] J. Cornfield, “A method of estimating comparative rates from clinical data: Application to cancer of the lung, breast and cervix,” *Journal of the National Cancer Institute*, vol. 11, pp. 1269–1275, 1951.
- [3] N. P. Jewell, *Statistics for Epidemiology*. Chapman & Hall/CRC, 2004.
- [4] J. C. Sinclair and M. B. Bracken, “Clinically useful measures of effect in binary analyses of randomized trials,” *Journal of Clinical Epidemiology*, vol. 47, no. 8, pp. 881–889, 1994.
- [5] D. L. Sackett, J. J. Deeks, and D. G. Altman, “Down with odds ratios!” *Evidence-Based Medicine*, vol. 1, no. 6, pp. 164–166, 1996.
- [6] D. G. Altman, J. J. Deeks, and D. L. Sackett, “Odds ratios should be avoided when events are common,” *BMJ*, vol. 317, no. 7168, p. 1318, 1998.
- [7] J. J. Deeks, “Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes,” *Statistics in Medicine*, vol. 21, no. 11, pp. 1575–1600, 2002.
- [8] R. G. Newcombe, “A deficiency of the odds ratio as a measure of effect size,” *Statistics in Medicine*, vol. 25, no. 24, pp. 4235–4240, 2006.
- [9] S. Greenland, K. J. Rothman, and T. L. Lash, “Measures of effect and measures of association,” in *Modern Epidemiology* (K. J. Rothman, S. Greenland, and T. L. Lash, eds.), ch. 4, pp. 51–70, Lippincott Williams & Wilkins, third edition ed., 2008.
- [10] N. Wermuth, “Parametric collapsibility and the lack of moderating effects in contingency tables with a dichotomous response variable,” *Journal of the Royal Statistical Society, Series B*, vol. 49, no. 3, pp. 353–364, 1987.
- [11] Y. Li, L. Shi, and H. Daniel Roth, “The bias of the commonly-used estimate of variance in meta-analysis,” *Communications in statistics-theory and methods*, vol. 23, no. 4, pp. 1063–1085, 1994.
- [12] A. L. Rukhin, “Weighted means statistics in interlaboratory studies,” *Metrologia*, vol. 46, no. 3, p. 323, 2009.



- [13] J. J. Gart, H. M. Pettigrew, and D. G. Thomas, “The effect of bias, variance estimation, skewness and kurtosis of the empirical logit on weighted least squares analyses,” *Biometrika*, vol. 72, no. 1, pp. 179–190, 1985.
- [14] H. M. Pettigrew, J. J. Gart, and D. G. Thomas, “The bias and higher cumulants of the logarithm of a binomial variate,” *Biometrika*, vol. 73, no. 2, pp. 425–435, 1986.
- [15] R. M. Turner, R. Z. Omar, M. Yang, H. Goldstein, and S. G. Thompson, “A multilevel model framework for meta-analysis of clinical trials with binary outcomes,” *Statistics in Medicine*, vol. 19, no. 24, pp. 3417–3432, 2000.
- [16] T. Stijnen, T. H. Hamza, and P. Özdemir, “Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data,” *Statistics in Medicine*, vol. 29, no. 29, pp. 3046–3067, 2010.
- [17] W. Viechtbauer, “Conducting meta-analyses in r with the `metafor` package,” *Journal of Statistical Software*, vol. 36, no. 3, pp. 1–48, 2010.
- [18] J. Luo, J. Zhang, and H. Sun, “Estimation of relative risk using a log-binomial model with constraints,” *Computational Statistics*, vol. 29, no. 5, pp. 981–1003, 2014.
- [19] M. W. Donoghoe and I. C. Marschner, “Flexible regression models for rate differences, risk differences and relative risks,” *The International Journal of Biostatistics*, vol. 11, no. 1, pp. 91–108, 2015.
- [20] M. W. Donoghoe, “logbin: Relative risk regression using the log-binomial model.” <https://cran.r-project.org/package=logbin>, 2015. (accessed 02 June 2016).
- [21] I. C. Marschner, “Relative risk regression for binary outcomes: Methods and recommendations,” *Australian & New Zealand Journal of Statistics*, vol. 57, no. 4, pp. 437–462, 2015.
- [22] D. Warn, S. Thompson, and D. Spiegelhalter, “Bayesian random effects meta-analysis of trials with binary outcomes: methods for the absolute risk difference and relative risk scales,” *Statistics in Medicine*, vol. 21, no. 11, pp. 1601–1623, 2002.
- [23] K. M. Rhodes, R. M. Turner, and J. P. T. Higgins, “Empirical evidence about inconsistency among studies in a pair-wise meta-analysis,” *Research Synthesis Methods*, vol. 7, no. 4, pp. 346–370, 2015.
- [24] D. R. Barr and E. T. Sherrill, “Mean and variance of truncated normal distributions,” *The American Statistician*, vol. 53, no. 4, pp. 357–361, 1999.

- [25] J. IntHout, J. P. Ioannidis, and G. F. Borm, “The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard dersimonian-laird method,” *BMC Medical Research Methodology*, vol. 14, no. 25, 2014.
- [26] T. Panityakul, C. Bumrungrsup, and G. Knapp, “On estimating residual heterogeneity in random-effects meta-regression: a comparative study,” *Journal of Statistical Theory and Applications*, vol. 12, no. 3, pp. 253–265, 2013.
- [27] C. Pedroza and V. T. T. Truong, “Performance of models for estimating absolute risk difference in multicenter trials with binary outcome,” *BMC Medical Research Methodology*, vol. 16, no. 113, 2016.
- [28] R. DerSimonian and N. Laird, “Meta-analysis in clinical trials,” *Controlled Clinical Trials*, vol. 7, no. 3, pp. 177–188, 1986.
- [29] J. Mandel and R. C. Paule, “Interlaboratory evaluation of a material with unequal numbers of replicates,” *Analytical Chemistry*, vol. 42, no. 11, pp. 1194–1197, 1970. Correction 43: 1287 (1971).
- [30] O. Kuss, “Statistical methods for meta-analyses including information from studies without any events – add nothing to nothing and succeed nevertheless,” *Statistics in Medicine*, vol. 34, no. 7, pp. 1097–1116, 2015.
- [31] N. L. Johnson, A. W. Kemp, and S. Kotz, *Univariate Discrete Distributions*. John Wiley & Sons, 1992.
- [32] I. Bakbergenuly, E. Kulinskaya, and S. Morgenthaler, “Inference for binomial probability based on dependent Bernoulli random variables with applications to meta-analysis and group level studies,” *Biometrical Journal*, vol. 58, no. 4, pp. 896–914, 2016.
- [33] I. Bakbergenuly and E. Kulinskaya, “Beta-binomial model for meta-analysis of odds ratios,” *Statistics in Medicine*, vol. 36, no. 11, pp. 1715–1734, 2017.
- [34] H. Chu, L. Nie, Y. Chen, Y. Huang, and W. Sun, “Bivariate random effects models for meta-analysis of comparative studies with binary outcomes: Methods for the absolute risk difference and relative risk,” *Statistical Methods in Medical Research*, vol. 21, no. 6, pp. 621–633, 2012.
- [35] D. Stasinopoulos and R. Rigby, “Generalized additive models for location, scale and shape (gamlss) in R,” *Journal of Statistical Software*, vol. 23, no. 7, 2008.
- [36] L. Rönnegård, X. Shen, and M. Alam, “hglm: A package for fitting hierarchical generalized linear models,” *The R Journal*, vol. 2, no. 2, pp. 20–28, 2010.
- [37] E. Z. Martinez, J. A. Achcar, and D. C. Aragon, “Parameter estimation of the beta-binomial distribution: an application using the sas software,” *Ciência e Natura*, vol. 37, no. 3, pp. 12–19, 2015.

- [38] B. Bolker, “Maximum likelihood estimation and analysis with the bbmle package.” <https://cran.r-project.org/web/packages/bbmle/vignettes/mle2.pdf>, 2016.
- [39] M. Stasinopoulos, B. Rigby, C. Akantziliotou, G. Heller, R. Ospina, N. Motpan, F. McElduff, V. Voudouris, M. Djennad, M. Enea, A. Ghalanos, and C. Argyropoulos, “Package gamlss, version 5.0-2.” <https://cran.r-project.org/web/packages/gamlss/gamlss.pdf>, 2017.
- [40] W. Viechtbauer, “Bias and efficiency of meta-analytic variance estimators in the random-effects model,” *Journal of Educational and Behavioral Statistics*, vol. 30, no. 3, pp. 261–293, 2005.
- [41] E. Kulinskaya and I. Olkin, “An overdispersion model in meta-analysis,” *Statistical Modelling*, vol. 14, no. 1, pp. 49–76, 2014.
- [42] D. C. Hoaglin, “Misunderstandings about Q and Cochran’s Q test’ in meta-analysis,” *Statistics in Medicine*, vol. 35, no. 4, pp. 485–495, 2016.
- [43] N. Breslow and N. Day, *Statistical methods in cancer research. Volume 1 The analysis of case-control studies*. Lyon: International Agency for Research on Cancer, 1980.
- [44] R. Collins, S. Yusuf, and R. Peto, “Overview of randomised trials of diuretics in pregnancy,” *British Medical Journal*, vol. 290, no. 6461, pp. 17–23, 1985.
- [45] T. A. Furukawa, H. McGuire, and C. Barbui, “Low dosage tricyclic antidepressants for depression,” *The Cochrane Library*, 2003.
- [46] S. G. Thompson and S. J. Pocock, “Can meta-analysis be trusted?,” *Lancet*, vol. 338, pp. 1127–1130, 1991.
- [47] R. Hardy and S. Thompson, “A likelihood approach to meta-analysis with random effects,” *Statistics in Medicine*, vol. 15, pp. 619–629, 1996.
- [48] B. Biggerstaff, R. Tweedie, *et al.*, “Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis,” *Statistics in Medicine*, vol. 16, no. 7, pp. 753–768, 1997.
- [49] A. Whitehead, *Meta-Analysis of Controlled Clinical Trials*. Chichester, England: John Wiley & Sons, Ltd, 2002.
- [50] W. Viechtbauer, “Confidence intervals for the amount of heterogeneity in meta-analysis,” *Statistics in Medicine*, vol. 26, no. 1, pp. 37–52, 2007.
- [51] K. H. Makambi and W. Lu, “Combining study outcome measures using dominance adjusted weights,” *Research Synthesis Methods*, vol. 4, no. 2, pp. 188–197, 2013.

- [52] J. Deeks and J. Higgins, “Statistical algorithms in Review Manager 5,” tech. rep., Cochrane Collaboration, August 2010. on Behalf of the Statistical Methods Group of The Cochrane Collaboration.
- [53] A. Alonso, S. Litière, and A. Laenen, “A note on the indeterminacy of the random-effects distribution in hierarchical models,” *The American Statistician*, vol. 64, no. 4, pp. 318–324, 2010.
- [54] R. Drikvandi, G. Verbeke, and G. Molenberghs, “Diagnosing misspecification of the random-effects distribution in mixed models,” *Biometrics*, vol. 73, pp. 63–71, 2017.
- [55] T. S. Richardson, J. M. Robins, and L. Wang, “On modeling and estimation for the relative risk and risk difference,” *Journal of the American Statistical Association*, vol. 112, no. 519, pp. 1121–1130, 2017.
- [56] L. Lin, H. Chu, and J. S. Hodges, “Alternative measures of between-study heterogeneity in meta-analysis: reducing the impact of outlying studies,” *Biometrics*, vol. 73, no. 1, pp. 156–166, 2017.
- [57] R. Kuhnert and D. Böhning, “A comparison of three different models for estimating relative risk in meta-analysis of clinical trials under unobserved heterogeneity,” *Statistics in Medicine*, vol. 26, no. 11, pp. 2277–2296, 2007.
- [58] E. Moreno, F.-J. Vázquez-Polo, and M. A. Negrín, “Bayesian meta-analysis: The role of the between-sample heterogeneity,” *Statistical Methods in Medical Research*, vol. 27, no. 12, pp. 3643–3657, 2018.
- [59] K. Rice, J. Higgins, and T. Lumley, “A re-evaluation of fixed effect(s) meta-analysis,” *Journal of the Royal Statistical Society, Series A*, vol. 181, no. 1, pp. 205–227, 2018.
- [60] N. Mantel and W. Haenszel, “Statistical aspects of the analysis of data from retrospective studies of disease,” *Journal of the National Cancer Institute*, vol. 22, no. 4, pp. 719–748, 1959.
- [61] G. A. Darlington and A. Donner, “Meta-analysis of community-based cluster randomization trials with binary outcomes,” *Clinical Trials*, vol. 4, no. 5, pp. 491–498, 2007.

## A Appendix

### A.1 Estimation of $\rho$ in the BB model

In this section, we give detailed derivations of the four estimators of  $\rho$  introduced in Section 4.5.

#### Method-of-moments estimator

Denote by  $\sigma_j^2(\rho) = \text{Var}(\hat{\theta}_j)$  the variance of  $\log(\text{RR})$  given in Equation (20). Then  $\sigma_j^2(\rho) = \sigma_j^2(0) + a_j\rho$ , where  $\sigma_j^2(0)$  is the variance under the fixed-effect model and  $a_j = (\pi_T^{-1} - 1)(1 - n_{jT}^{-1}) + (\pi_C^{-1} - 1)(1 - n_{jC}^{-1})$ . The inverse-variance weights are  $w_j(\rho) = \sigma_j^{-2}(\rho)$ . Cochran's statistic is  $Q = \sum \hat{w}_j(\hat{\theta}_j - \bar{\theta}_w)^2$ , where  $\hat{w}_j = \hat{\sigma}_j^{-2}(0)$  and  $\bar{\theta}_w = \sum \hat{w}_j \hat{\theta}_j / \sum \hat{w}_j$ . Under the null hypothesis of no over- or under-dispersion ( $\rho = 0$ ), the  $Q$ -statistic has approximately a chi-squared distribution with  $K - 1$  degrees of freedom, so  $E(Q) = K - 1$ . Using the theoretical weights  $w_j = w_j(0)$  instead of  $\hat{w}_j$ , the derivation of the expected value of  $Q$  yields

$$E(Q) = \sum w_j \sigma_j^2(\rho) - \frac{\sum w_j^2 \sigma_j^2(\rho)}{\sum w_j}.$$

After substitution of  $\sigma_j^2(\rho) = \sigma_j^2(0) + a_j\rho$ , this becomes

$$E(Q) = K - 1 + \left[ \sum w_j a_j - \frac{\sum w_j^2 a_j}{\sum w_j} \right] \rho.$$

The DerSimonian-Laird-inspired estimator of  $\rho$ , denoted by  $\hat{\rho}_{MoM}$ , is obtained by substituting  $Q$  for  $E(Q)$  and the estimated weights  $\hat{w}_j$  for the theoretical weights  $w_j$  and solving the resulting equation for  $\rho$ .

#### Modified Mandel-Paule estimator

Similarly, the modified Mandel-Paule estimator,  $\hat{\rho}_{MP}$ , is obtained by considering  $Q(\rho) = \sum \hat{w}_j(\rho)(\hat{\theta}_j - \bar{\theta}_{w(\rho)})^2$  and solving the equation  $Q(\rho) = K - 1$ . The weighted mean is  $\bar{\theta}_{w(\rho)} = \sum \hat{w}_j(\rho)\hat{\theta}_j / W(\rho)$ , where  $W(\rho) = \sum \hat{w}_j(\rho)$ .

#### Restricted-maximum-likelihood estimator

Following Kulinskaya and Olkin<sup>41</sup>, the REML equation for  $\rho$  is

$$(W(\rho))^{-1} \sum \hat{w}_j(\rho) \frac{a_j}{1 + a_j\rho} + \sum \hat{w}_j(\rho)(\theta_j - \theta)^2 \frac{a_j}{1 + a_j\rho} = \sum \frac{a_j}{1 + a_j\rho}, \quad (\text{A1})$$

and an iterative procedure readily yields a solution, denoted by  $\hat{\rho}_{REML}$ .

### Inverting the modified Breslow-Day test

The Breslow-Day test for homogeneity of odds ratios is based on the statistic

$$\sum_{j=1}^K \frac{(Y_{jT} - Y_{jT}(\hat{\psi}))^2}{\text{Var}(Y_{jT}|\hat{\psi})},$$

where  $Y_{jT}(\hat{\psi})$  and  $\text{Var}(Y_{jT}|\hat{\psi})$ , respectively, denote the expected value and the asymptotic variance of the number of events in the treatment group of Study  $j$  based on the Mantel-Haenszel<sup>60</sup> (MH) estimate of the odds ratio,  $\hat{\psi}$ . We apply the same approach to the risk ratio.

The expected number of events, given RR  $\hat{\psi}$  and the total number of events  $Y_j$ , is

$$Y_{jT}(\hat{\psi}) = \frac{\hat{\psi}n_{jT}Y_j}{n_j - n_{jT} + \hat{\psi}n_{jT}}.$$

The MH estimate for the relative risk is

$$\hat{\psi} = \frac{\sum y_{jT}n_{jC}/n_j}{\sum y_{jC}n_{jT}/n_j},$$

and the asymptotic variance<sup>61</sup> of  $Y_{jT}$  is

$$\text{Var}(Y_{jT}|\hat{\psi}) = \left[ \frac{1}{Y_{jT}(\hat{\psi})C_{jT}} + \frac{1}{(Y_j - Y_{jT}(\hat{\psi}))C_{jC}} + \frac{1}{(n_{jT} - Y_{jT}(\hat{\psi}))C_{jT}} + \frac{1}{(n_j - Y_j - n_{jT} + Y_{jT}(\hat{\psi}))C_{jC}} \right]^{-1}, \quad (\text{A2})$$

where  $C_{ji} = 1 + (n_{ji} - 1)\hat{\rho}$  are correction factors;  $i = C, T$ ;  $j = 1, \dots, K$ . The Breslow-Day statistic has a  $\chi_{K-1}^2$  distribution, and this modification of it can be used for testing the homogeneity of RRs. Equating the BD statistic to  $K - 1$  yields the Mandel-Paule-type estimate for  $\rho$ , denoted by  $\hat{\rho}_{BD}$ .

## A.2 Function for meta-analysis of relative risk using bbmle

```
#####grp-level outcome data#####
###          | Event | No event |
### Treatment | ai    | bi        |
### Control   | ci    | di        |
### K is the number of studies
### ai and ci are number of events in T and C arms
###required packages : bbmle and emdbook
###initial parameters
my_bbmle<-function(ai , bi , ci , di) {
  require(emdbook)
  require(bbmle)
  ##initial parameters
  pT0<-sum(( ai+0.5)/(ni+1))/k
  pC0<-sum(( ci+0.5)/(ni+1))/k
  rho0<-0.01
  #####
  ### negative log-likelihood functions for
  ### a pair of beta-binomial distribution
  mtmp <- function(probT , sizeT , theta , probC , sizeC) {
    -sum(dbetabinom( ai , probT , sizeT , theta , log=TRUE) )
    -sum(dbetabinom( ci , probC , sizeC , theta , log=TRUE) )
  }
  m0 <- suppressWarnings( try( mle2( mtmp , start=list( probT=pT0 ,
    theta=rho0^(-1)-1 , probC=pC0) , data=list( sizeT=ni , sizeC=ni) ,
    silent=TRUE) )
  if( (!( inherits( m0 , "try-error" ) ) ) ) {
    if( (!( is.na( stdEr( m0) [[ 1 ] ] ) ) ) & (!( is.na( stdEr( m0) [[ 3 ] ] ) ) ) ) {
      probT1<-coef( m0) [[ 1 ] ]
      probC1<-coef( m0) [[ 3 ] ]
      varT1<-(stdEr( m0) [[ 1 ] ] ) ^2
      varC1<-(stdEr( m0) [[ 3 ] ] ) ^2
      varlogRR<-(1/probT1)^2*varT1+(1/probC1)^2*varC1
      RR1<-(probT1)/(probC1)
      log_RR<-log( RR1)
      logRRL<-log_RR-qnorm(.975)*sqrt( varlogRR)
      logRRU<-log_RR+qnorm(.975)*sqrt( varlogRR)
      rhoEst<-1/( coef( m0) [[ 2 ] ] +1)
      return( list( rhoEst , log_RR , logRRL , logRRU) )
    }
  }
}
```

### A.3 Function for meta-analysis of relative risk using gamlss

```

##R version 3.5.0
#####grp-level outcome data#####
###          | Event | No event |
### Treatment | ai    | bi      |
### Control   | ci    | di      |
### K is the number of studies
### ai and ci are number of events in T and C arms
###required packages: gamlss
  my_gamlss<-function(ai, bi, ci, di) {
require(emdbook)
require(gamlss)
r<-c(ai, ci)
N<-c(ni, ni)
y<-cbind(r, N-r)
arm<-c(rep("T", times=k), rep("C", times=k))
arm<-as.factor(arm)
sink(tempfile())
modelGamlss <- suppressWarnings(try(gamlss(y~arm-1, sigma.
  formula=~1, family=BB), silent=TRUE))
sink()
if(!(inherits(modelGamlss, "try-error"))){
#####fitting the model using gamlss#####
sink(tempfile())
res<-suppressWarnings(summary(modelGamlss, save=TRUE))
sink()
pC3<-exp(modelGamlss$mu. coefficients [[1]])/(1+exp(
  modelGamlss$mu. coefficients [[1]]) )
pT3<-exp(modelGamlss$mu. coefficients [[2]])/(1+exp(
  modelGamlss$mu. coefficients [[2]]) )
st.error1<-(res$coef.table [,2] [[1]])
st.error2<-(res$coef.table [,2] [[2]])
varC3<-(1/(1+exp(modelGamlss$mu. coefficients [[1]]))^2)^2*(st
  .error1)^2
varT3<-(1/(1+exp(modelGamlss$mu. coefficients [[2]]))^2)^2*(st
  .error2)^2
RR3<-pT3/pC3
logRRgamlss<-log(RR3)
varRR3<-varT3+varC3
logRRL_gamlss=0
logRRU_gamlss=0
if(!is.null(st.error2)){
logRRL_gamlss<-logRRgamlss-qnorm(.975)*sqrt(varRR3)

```



```
logRRU_gamlss<-logRRgamlss+qnorm(.975)*sqrt(varRR3)
}
rhoEst<-exp(modelGamlss$sigma.coefficients[[1]])/(1+exp(
  modelGamlss$sigma.coefficients[[1]]) )
}
return(list(rhoEst ,logRRgamlss ,logRRL_gamlss ,logRRU_gamlss))
}
```

#### A.4 Function for meta-analysis of relative risk using $\rho_{DL}$

```

##R version 3.5.0
#####grp-level outcome data#####
###      | Event | No event |
### Treatment | ai      | bi      |
### Control   | ci      | di      |
### K is the number of studies
### ai and ci are number of events in T and C arms
###required packages: None
my_DL_function_generic<-function(ai, ci, bi, di){
  p1i<-(ai+0.5)/(ai+bi+1)
  p2i<-(ci+0.5)/(ci+di+1)
  psi_i<-(ai+0.5)*((ci+di)+0.5)/((ci+0.5)*(ai+bi+0.5))
  yi      <- log(psi_i)
  vi      <- 1/(ai+0.5)-1/(ai+bi+0.5)+1/(ci+0.5)-1/(ci+di+0.5)
  n1i<-ai+bi
  n2i<-ci+di
  n_i<-n1i+n2i
  Ri<-n1i/n2i
  a_i<-(Ri*n_i/(Ri+1))*((1+psi_i-2*p1i)/(Ri*(1-p1i)+psi_i-p1i)
    )-1
  wi<-1/vi
  W<-sum(wi)
  thetabar<-sum(yi*wi)/W
  Q<-sum(wi*(yi-thetabar)^2)
  barn_w<-sum(a_i*wi)/W
  barn<-sum(a_i)/k
  rho_M<-max((Q-k+1)/(k*barn-barn_w),-1/max(ni-1))
  ###estimated weights
  w_M<-wi/(1+a_i*rho_M)
  thetabar_M<-sum(yi*w_M)/sum(w_M)
  Ltheta_M<-thetabar_M-qnorm(.975)/sqrt(sum(w_M))
  Utheta_M<-thetabar_M+qnorm(.975)/sqrt(sum(w_M))
  return(list(rho_M, thetabar_M, Ltheta_M, Utheta_M))
}

```

## A.5 Function for meta-analysis of relative risk using $\rho_{MP}$

```

##R version 3.5.0
#####grp-level outcome data#####
###      | Event | No event |
### Treatment | ai      | bi      |
### Control   | ci      | di      |
### K is the number of studies
### ai and ci are number of events in T and C arms
my_MP_function_generic<-function(ai, ci, bi, di){
k<-length(ai)
fMP<-function(g, sigma2, theta, a_i, k){
sum((theta-sum(theta/(sigma2*(1+a_i*g)))/
sum(1/(sigma2*(1+a_i*g)))^2/(sigma2*(1+a_i*g)))-(k-1)}
pli<-(ai+0.5)/(ai+bi+1)
p2i<-(ci+0.5)/(ci+di+1)
psi_i<-(ai+0.5)*((ci+di)+0.5)/((ci+0.5)*(ai+bi+0.5))
yi      <- log(psi_i)
vi      <- 1/(ai+0.5) - 1/(ai+bi+0.5) + 1/(ci+0.5) - 1/(ci+di
+0.5)
wi<-1/vi
n1i<-ai+bi
n2i<-ci+di
n_i<-n1i+n2i
Ri<-n1i/n2i
a_i<-(Ri*n_i/(Ri+1))*((1+psi_i-2*pli)/(Ri*(1-pli)+psi_i-pli)
)-1
ll<-max(-1/max(a_i), -1/max(n1i-1), -1/max(n2i-1))+0.00000001
uu<-1000
if (fMP(ll, sigma2=vi, theta=yi, a_i=a_i, k=k)*
fMP(uu, sigma2=vi, theta=yi, a_i=a_i, k=k)<0)
{
rho_MP<-as.numeric(uniroot(fMP, c(ll, uu), tol = 0.0001,
sigma2=vi, theta=yi, a_i=a_i, k=k)[1])
w_MP<-wi/(1+a_i*rho_MP)
thetabar_MP<-sum(yi*w_MP)/sum(w_MP)
if (rho_MP==(-1/max(ni-1)) | is.na(thetabar_MP) | (rho_MP==(-1/
max(ni-1))&is.na(thetabar_MP))){
yi_maxni<-yi[ni==max(ni)]
wi_maxni<-wi[ni==max(ni)]
thetabar_MP<-sum(yi_maxni*wi_maxni)/sum(wi_maxni)
w_MP<-wi[ni==max(ni)]
}
Ltheta_MP<-thetabar_MP-qnorm(.975)/sqrt(sum(w_MP))

```

```
Utheta_MP<-thetabar_MP+qnorm(.975)/sqrt(sum(w_MP))  
return(list(rho_MP, thetabar_MP, Ltheta_MP, Utheta_MP))  
}
```

## A.6 Function for meta-analysis of relative risk using $\rho_{BD}$

```

##R version 3.5.0
#####grp-level outcome data#####
###          | Event | No event |
### Treatment | ai    | bi      |
### Control   | ci    | di      |
### K is the number of studies
### ai and ci are number of events in T and C arms
###required packages: None
my_BD_function_generic<-function(ai, ci, bi, di) {
nT<-ni
nC<-ni
XT<-ai
XC<-ci
k<-length(ai)
pli<-(ai+0.5)/(ai+bi+1)
p2i<-(ci+0.5)/(ci+di+1)
psi_i<-(ai+0.5)*((ci+di)+0.5)/((ci+0.5)*(ai+bi+0.5))
yi<-log(psi_i)
vi<-1/(ai+0.5)-1/(ai+bi+0.5)+1/(ci+0.5)-1/(ci+di+0.5)
wi<-1/vi
n1i<-ai+bi
n2i<-ci+di
n_i<-n1i+n2i
Ri<-n1i/n2i
a_i<-((Ri*n_i/(Ri+1))*((1+psi_i-2*pli)/(Ri*(1-pli)+psi_i-pli)
)-1)
#####
#####Breslow-day method#####
rho_BD<-numeric(1)
###Breslow-Day test conf. inter.
rho_BDU<-numeric(1)
rho_BDL<-numeric(1)
ll<-max(-1/max(a_i), -1/max(n1i-1), -1/max(n2i-1))+0.00000001
uu<-1000
##Mantel-Haenzsel estimator of relative risk
psiMH<-sum((nC*XT)/(nC+nT))/sum((nT*XC)/(nC+nT))
if (psiMH==1) {X1j=((ai+ci)*n1i)/(n1i+n2i)} else {
X1j<-psiMH*n1i*(ai+ci)/((n1i+n2i)-n1i+psiMH*n1i)
}
#function for the Standard Breslow day test
f<-function(rho, X_1j, X1j, xj, n1j, n2j, nj, k) {
varj<-((1/(X1j*(1+(n1j-1)*rho)))+1/((xj-X1j)*(1+(n2j-1)*rho)))

```

```

+1/((n1j-X1j)*(1+(n1j-1)*rho))+1/((nj-xj-n1j+X1j)*(1+(n2j-1)*rho)))^-1)
varj0<-((1/((X1j+.5)*(1+(n1j-1)*rho))+1/((xj-X1j+.5)*(1+(n2j-1)*rho))+1/((n1j-X1j+.5)*(1+(n1j-1)*rho))+1/((nj-xj-n1j+X1j+.5)*(1+(n2j-1)*rho))))^-1)
var<-ifelse(is.na(varj),varj0,varj)
var<-ifelse(var==0,varj0,var)
sum((X_1j-X1j)^2/var)-k+1
}

if(f(uu,ai,X1j,ai+ci,n1i,n2i,n1i+n2i,k=k)>0){rho_BD<-1}
if((f(ll,ai,X1j,ai+ci,n1i,n2i,n1i+n2i,k=k)>0)&(f(uu,ai,X1j,ai+ci,n1i,n2i,n1i+n2i,k=k)<0)){
rho_BD<-as.numeric(uniroot(f,c(ll,uu),tol=0.00001,X_1j=ai,X1j=X1j,xj=ai+ci,n1j=n1i,n2j=n2i,nj=(n1i+n2i),k=k)[1])
}
###Upper Limit
f1<-function(rho,X_1j,X1j,xj,n1j,n2j,nj,k){
varj<-((1/(X1j*(1+(n1j-1)*rho))+1/((xj-X1j)*(1+(n2j-1)*rho))+1/((n1j-X1j)*(1+(n1j-1)*rho))+1/((nj-xj-n1j+X1j)*(1+(n2j-1)*rho))))^-1)
varj0<-((1/((X1j+.5)*(1+(n1j-1)*rho))+1/((xj-X1j+.5)*(1+(n2j-1)*rho))+1/((n1j-X1j+.5)*(1+(n1j-1)*rho))+1/((nj-xj-n1j+X1j+.5)*(1+(n2j-1)*rho))))^-1)
var<-ifelse(is.na(varj),varj0,varj)
var<-ifelse(var==0,varj0,var)
sum((X_1j-X1j)^2/var)-qchisq(0.025,k-1)
}

if(f1(uu,ai,X1j,ai+ci,n1i,n2i,n1i+n2i,k=k)>0){rho_BDU<-1}
if((f1(ll,ai,X1j,ai+ci,n1i,n2i,n1i+n2i,k=k)>0)&(f1(uu,ai,X1j,ai+ci,n1i,n2i,n1i+n2i,k=k)<0)){
rho_BDU<-as.numeric(uniroot(f1,c(ll,uu),tol=0.00001,X_1j=ai,X1j=X1j,xj=ai+ci,n1j=n1i,n2j=n2i,nj=(n1i+n2i),k=k)[1])
}
###Lower Limit
f2<-function(rho,X_1j,X1j,xj,n1j,n2j,nj,k){
varj<-((1/(X1j*(1+(n1j-1)*rho))+1/((xj-X1j)*(1+(n2j-1)*rho))+1/((n1j-X1j)*(1+(n1j-1)*rho))+1/((nj-xj-n1j+X1j)*(1+(n2j-1)*rho))))^-1)
varj0<-((1/((X1j+.5)*(1+(n1j-1)*rho))+1/((xj-X1j+.5)*(1+(n2j-1)*rho))+1/((n1j-X1j+.5)*(1+(n1j-1)*rho))+1/((nj-xj-n1j+X1j+.5)*(1+(n2j-1)*rho))))^-1)
var<-ifelse(is.na(varj),varj0,varj)

```

```

var<-ifelse(var==0, varj0, var)
sum((X_1j-X1j)^2/var)-qchisq(0.975, k-1)
}

if(f2(uu, ai, X1j, ai+ci, n1i, n2i, n1i+n2i, k=k)>0){rho_BDL<-11}
if((f2(ll, ai, X1j, ai+ci, n1i, n2i, n1i+n2i, k=k)>0)&(f2(uu, ai, X1j,
, ai+ci, n1i, n2i, n1i+n2i, k=k)<0)){
rho_BDL<-as.numeric(uniroot(f2, c(ll, uu), tol=0.00001, X_1j=ai,
X1j=X1j, xj=ai+ci, n1j=n1i, n2j=n2i, nj=(n1i+n2i), k=k)[1])
}
#####Inverse variance estimation of theta using rho_BD
#####
w_BD<-wi/(1+a_i*rho_BD)
thetabar_BD<-sum(yi*w_BD)/sum(w_BD)
Ltheta_BD<-thetabar_BD-qnorm(.975)/sqrt(sum(w_BD))
Utheta_BD<-thetabar_BD+qnorm(.975)/sqrt(sum(w_BD))
return(list(rho_BD, rho_BDL, rho_BDU, thetabar_BD, Ltheta_BD,
Utheta_BD))
}

```

## B Additional figures

- B.1 Estimation of the between-studies variance when  $\tau^2 = 0.1$  and the data are generated under the point-mass option and the truncation option



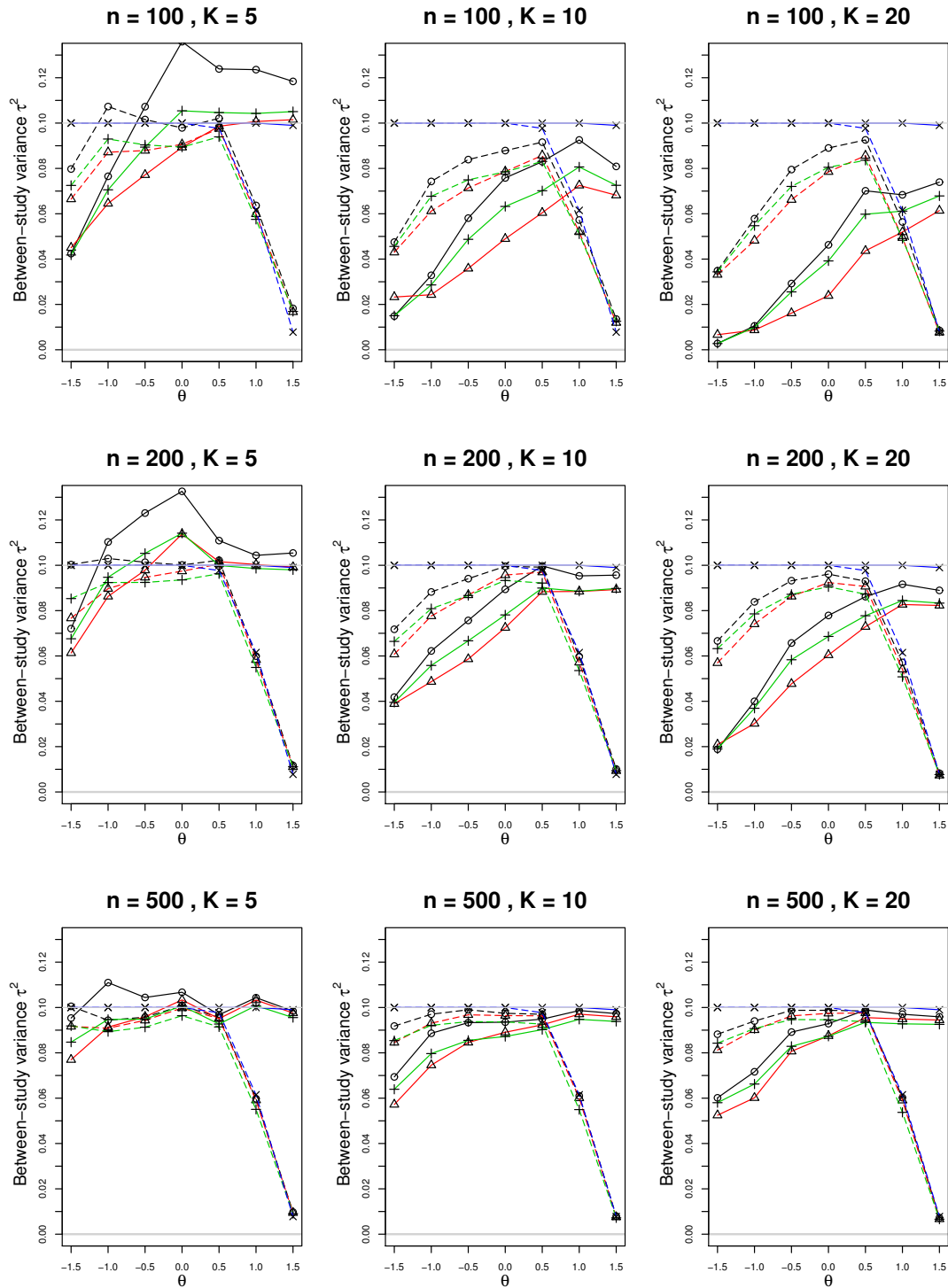


Figure A.1: Relation of estimates of the between-studies variance  $\tau^2$  to the overall log-risk-ratio ( $\theta$ ) in  $K$  studies, each of total sample size  $n$ , when data come from the binomial-normal model with point mass for  $\tau^2 = 0.1$  and  $\pi_{jC} = 0.1$  (solid lines) and 0.3 (dashed). The Mandel-Paule (circle), REML (triangle), and DerSimonian-Laird (plus) estimation methods are compared with the true variance (cross). Light grey line at 0.1.

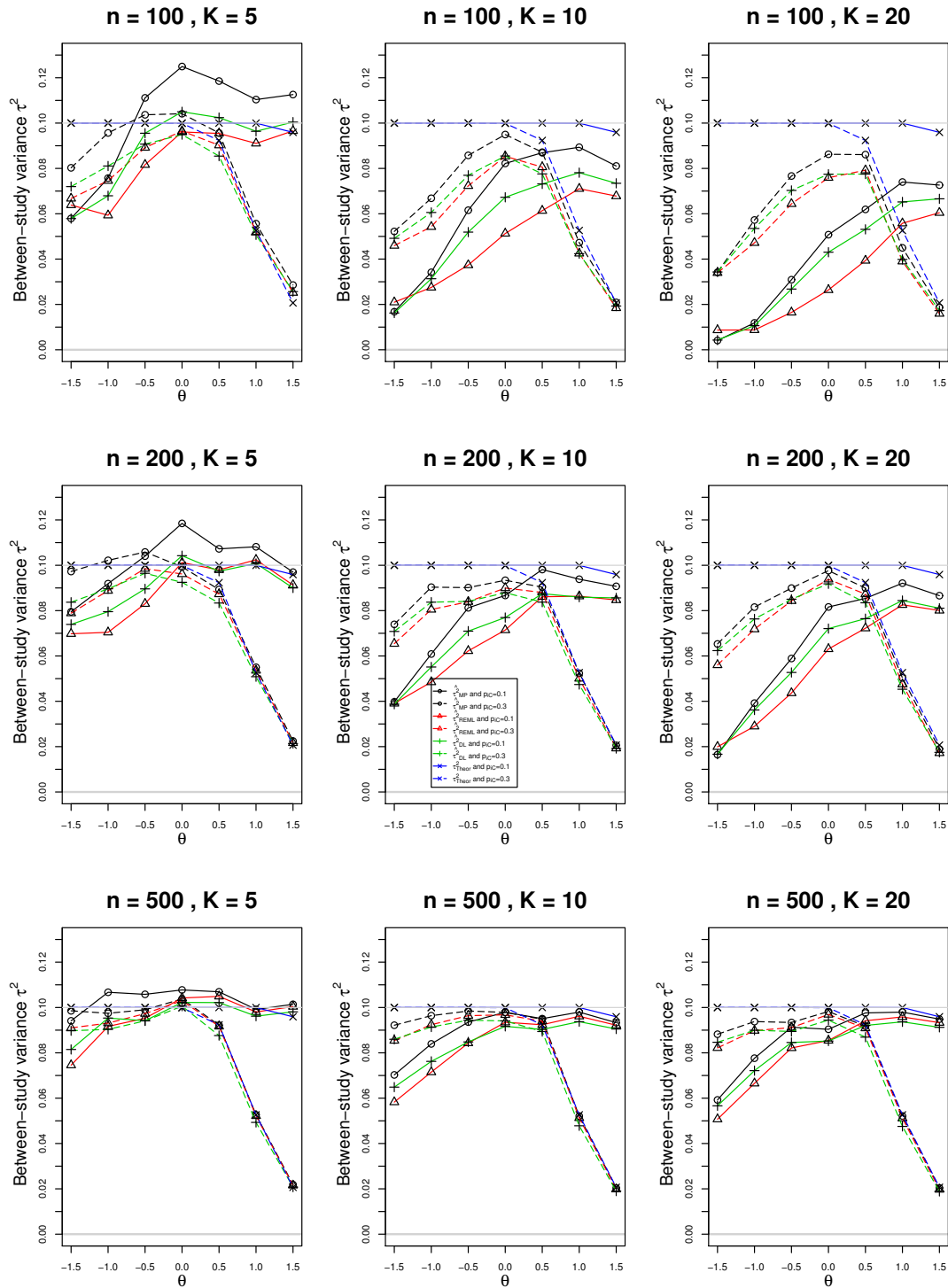


Figure A.2: Relation of estimates of the between-studies variance  $\tau^2$  to the overall log-risk-ratio ( $\theta$ ) in  $K$  studies, each of total sample size  $n$ , when data come from the binomial-normal model with truncation for  $\tau^2 = 0.1$  and  $\pi_{JC} = 0.1$  (solid lines) and 0.3 (dashed). The Mandel-Paule (circle), REML (triangle), and DerSimonian-Laird (plus) estimation methods are compared with the true variance (cross). Light grey line at 0.1.

## B.2 Figures for studentized residuals when $K \geq 10$

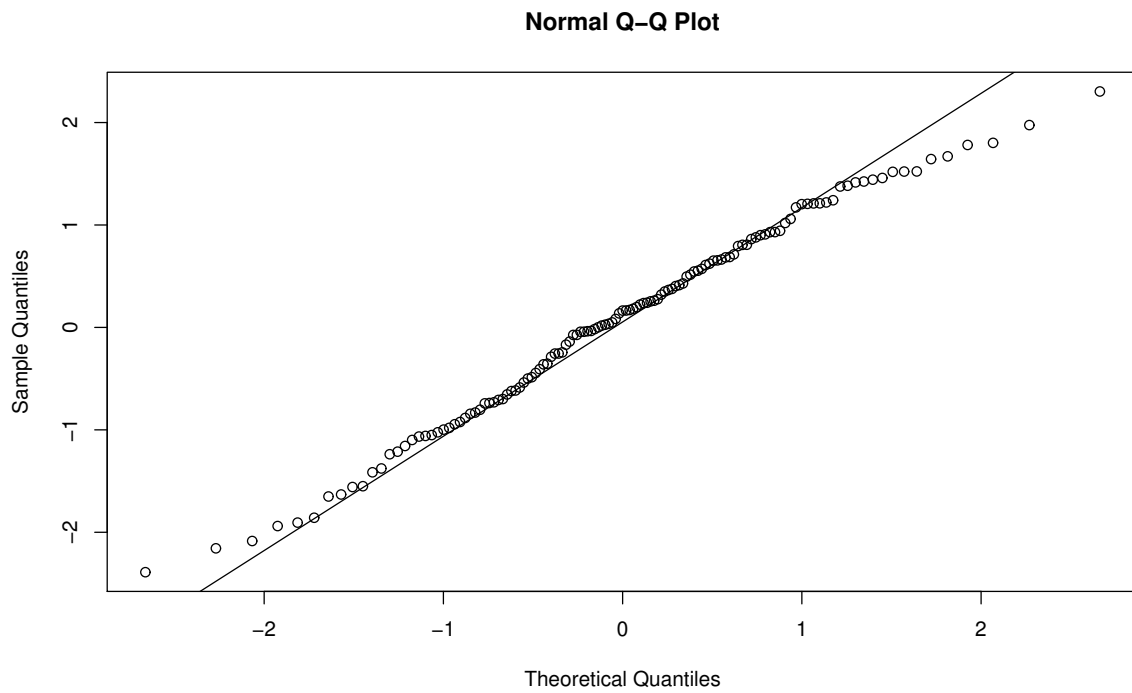


Figure A.3: Normal Q-Q plot of the 129 studentized residuals for the studies from 11 REM meta-analyses of RR with  $\hat{\theta} \geq 0$ ,  $\hat{\tau}^2 > 0$  and  $K \geq 10$  in Cochrane Library Issue 4.

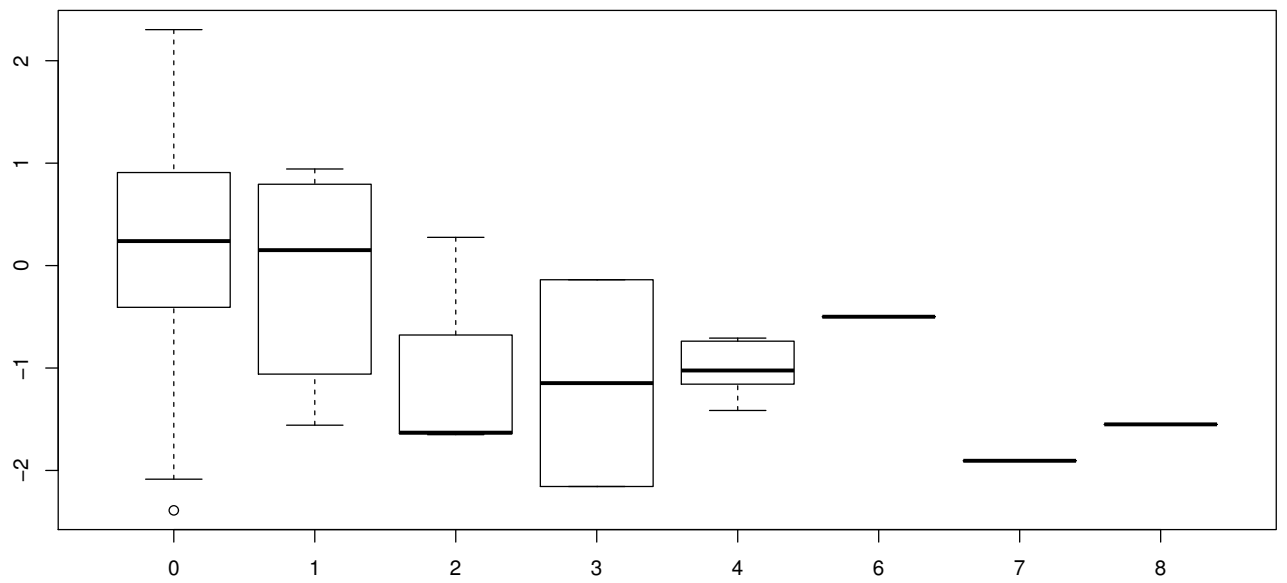


Figure A.4: Boxplots of 129 studentized residuals by truncation probability, for the studies from 11 REM meta-analyses of RR with  $\hat{\theta} \geq 0$ ,  $\hat{\tau}^2 > 0$  and  $K \geq 10$  in Cochrane Library Issue 4.

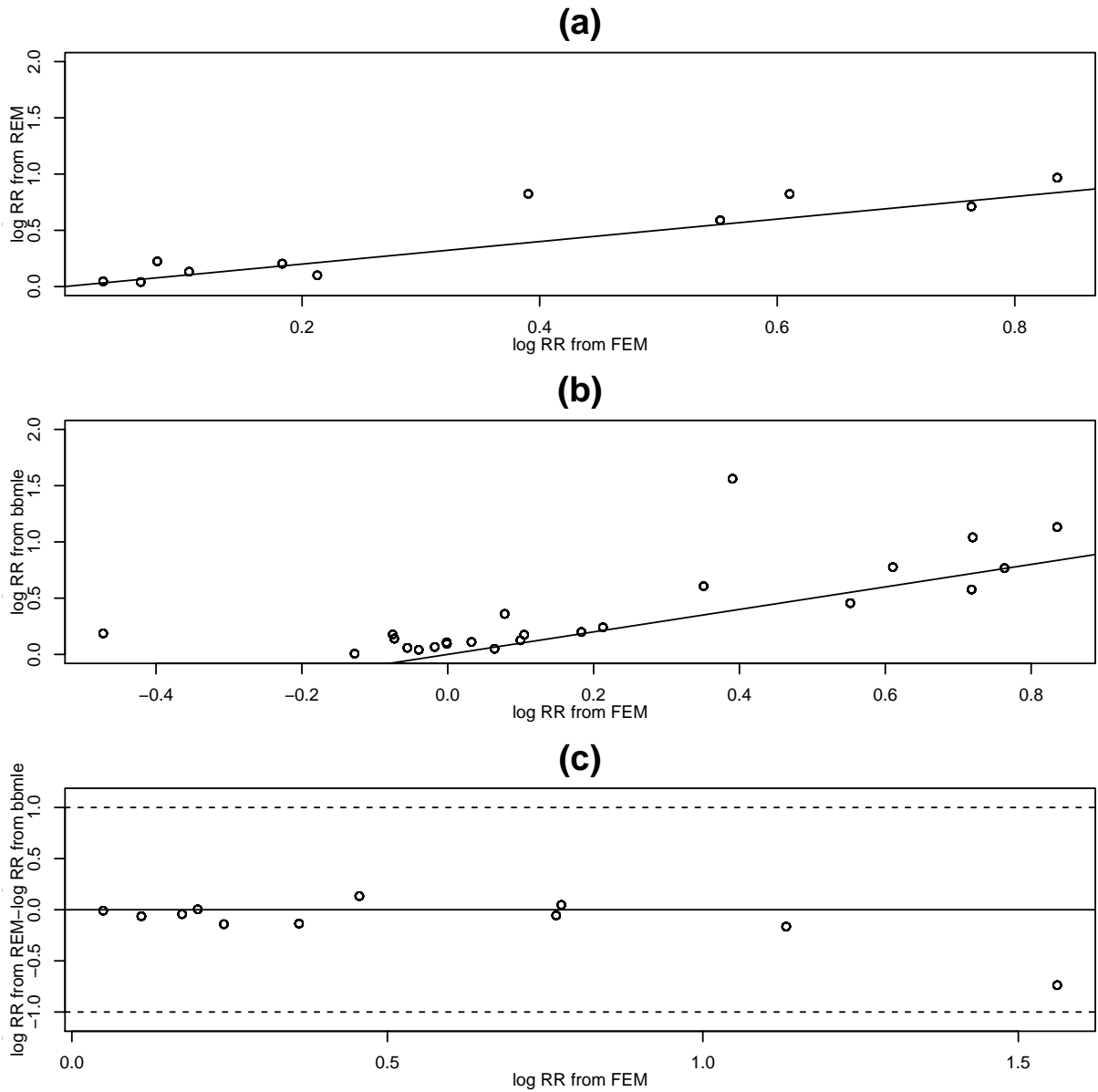


Figure A.5: Scatterplot (vs. log-risk-ratio from FEM) of the meta-analytic estimates of log-risk-ratio obtained by:(a) REM, for the 11 REM meta-analyses of RR with  $K \geq 10$ ,  $\hat{\theta}_{REM} \geq 0$  and  $\hat{\tau}^2 > 0$ ; (b) *bbmle*, for the 24 meta-analyses of RR with  $K \geq 10$ ,  $\hat{\theta}_{bbmle} \geq 0$  and  $\hat{\rho} > 0$ ; (c) Difference between log(RR) from REM and *bbmle* for the 11 meta-analyses with  $K \geq 10$ ,  $\hat{\rho} > 0$  and  $\hat{\theta}_{bbmle} \geq 0$  and  $\tau^2 > 0$  and  $\hat{\theta}_{REM} \geq 0$ .