

# Meta-analysis of rare events under the assumption of a homogeneous treatment effect

Romain Piaget-Rossel<sup>1</sup> and Patrick Taffé<sup>1\*</sup>

<sup>1</sup> Institute of Social and Preventive Medicine, Lausanne University Hospital and Lausanne University, Route de la Corniche 10, 1010 Lausanne, Switzerland

We studied the performance of several meta-analysis methods in rare event settings, when the treatment effect is assumed to be homogeneous and baseline prevalences are either homogeneous or heterogeneous. We conducted extensive simulations that included the three most common effect sizes with count data: the odds ratio, the relative risk and the risk difference. We investigated several important scenarios by varying the level of rareness, the value of the trials' arms unbalance and the size of the treatment effect. We found that the Mantel-Haenszel method and the Binomial regression model provided the best results across all the scenarios investigated. The Peto method performed satisfactorily only when the true effect size was not too large and the degree of unbalance moderate. Inverse variance was the least reliable method. The use of a continuity correction factor slightly improved the performance of the inverse variance method but deteriorated that of the Peto and Mante-Haenszel methods. A method based on median unbiased estimators of the probabilities provided similar results to those obtained when using the inverse variance method with a continuity correction. Therefore, when the treatment effect can be assumed to be homogeneous and for either homogeneous or heterogeneous baseline prevalences, we highly recommend using the Mantel-Haenszel method without continuity correction (for all the effect sizes) or the Binomial regression model (for the odds ratio only) to meta-analyze the data.

*Key words:* Fixed-effect methods; Homogeneous treatment effect; Meta-analysis; Rare events; Simulation study

## 1 Introduction

The objective of a meta-analysis is to combine evidence from related but independent studies in order to improve the knowledge about a specific research question and to generalize the results (Normand, 1999). When the outcome is continuous, the investigator either uses a fixed-effect (FE) or a random-effects (RE) model according to his knowledge and the expected effect of the intervention. However, with count data, the investigator further has to distinguish between the setting of homogeneous and heterogeneous baseline prevalences, yielding a total of four different frameworks (Table 1).

- *Insert Table 1 here* -

While a clear distinction is made between the FE and RE models in the meta-analysis' literature, the question of homogeneous or heterogeneous baseline prevalences is seldomly discussed. Notably, in the three main published simulation studies assessing the performance of various meta-analysis methods in the context of rare events, the authors assumed a homogeneous baseline prevalence in all the simulated scenarios (Sweeting et al., 2004; Bradburn et al., 2007; Kuss 2015). The goal of this paper was thus to reassess the performance of FE meta-analysis methods in rare event settings, considering either homogeneous or heterogeneous baseline prevalences.

In practice, a typical setting under which the assumption of a homogeneous treatment effect is likely to hold is that of a multicenter randomized control trial, where each center has followed exactly the same research protocol. When patients between centers are comparable in terms of baseline characteristics, then the meta-analyst will likely assume a homogeneous baseline prevalence. On the contrary, when

---

\*Corresponding author: e-mail: patrick.taffe@unisante.ch

the primary studies have been conducted in different centers located in different countries or places in the world, the assumption of heterogeneous baseline prevalences might be more appropriate.

The classical method to conduct a meta-analysis in the framework of a homogeneous treatment effect is that of the inverse variance (IV) (Borenstein et al., 2009). This method is based on the generalized least squares technique and provides an estimator corresponding to a weighted average of the primary-studies' effect sizes (ESs). This estimator has good asymptotic properties (i.e. convergence, asymptotic normality). In finite samples however, and especially with rare events, asymptotic theory breaks down and the IV method yields biased estimates and invalid confidence intervals (CI) (Lane, 2013). Specific to binary data, two other methods are commonly used in FE meta-analyses: Mantel-Haenszel (MH) (Mantel and Haenszel, 1959) and Peto (Yusuf et al., 1985), the latter being applied for the estimation of the odds ratio (OR) only. While both methods were found to be more robust to rare events than the IV method, estimates provided by the Peto method were less reliable in settings with unbalanced trials' arms and/or large ES (Bradburn et al., 2007).

With very rare events, when some primary studies report zero event in one (single-zero studies; SZ studies) or both (double-zero studies; DZ studies) arms, the IV method may provide indefinite estimates. Under these circumstances, a straightforward way to compute the IV estimate is to exclude the problematic studies from the analysis. As for the MH and Peto methods, DZ studies are automatically excluded from the computation of the OR and the relative risk (RR; for MH), as can be checked by inspecting their mathematical formulae. However, excluding zero-event studies is suboptimal since they are likely to contain useful pieces of information, even DZ studies (e.g. no event in a sample of 50 patients is not the same as no event in a sample of 200 patients).

As discussed by Kuss et al. (2009), meta-analyses with zero-event studies are commonly encountered in practice. In a random sample of 500 Cochrane reviews, the authors found that 34% of these reviews contained at least one meta-analysis with a DZ study. To tackle this issue, the classical solution is to use a continuity correction factor (cc) (Sweeting et al., 2004). The problem with this solution is that the resulting estimates depend on the choice of the cc (Rücker et al., 2009). However, there is another way to include both SZ and DZ studies, without using a cc, which is based on the median unbiased estimator (MUE) method (Hirji et al., 1989; Parzen et al., 2002). This method provides estimates of the OR, RR and risk difference (RD) that always exist (i.e. even when no event is observed in both arms). Finally, the last alternative we have investigated to estimate the OR, which uses the information contained in DZ studies, is the Binomial regression model.

Through extensive simulations, we assessed the performance of these different methods (i.e. IV, MH, Peto, with or without cc, MUE and Binomial regression). Except for the Peto and Binomial regression methods that are specific to the OR, all the other methods were used to estimate the OR, the RR and the RD. Our simulations covered many important scenarios with different values of trials' arms unbalance (from strong unbalance in favor of either trials' arm to no unbalance), ES (from large reduction – or increase – in event prevalence to no effect), and baseline prevalences (from extremely rare to common events). In addition, all these scenarios were considered with either homogeneous or heterogeneous baseline prevalences.

## 2 Combining trials under the fixed-effect framework

Let  $\pi_t$  and  $\pi_c$  be the probability of the event in the treated and control populations respectively. In this paper, we focused on the three following effect sizes:

$$OR = (\pi_t * (1 - \pi_c)) / (\pi_c * (1 - \pi_t)) \quad (1)$$

$$RR = \pi_t / \pi_c \quad (2)$$

$$RD = \pi_t - \pi_c \quad (3)$$

Under the FE model, the ES is assumed homogeneous across the primary studies and the goal of the meta-analysis is to estimate a single population parameter. Inversely, under the RE model, one assumes

that each primary study seeks to estimate a different population parameter and the focus is on describing these parameters' distribution. Ideally, the selection of either framework should be grounded on contextual knowledge and not on statistical arguments. This selection will then determine the set of methods available to the meta-analyst. In the subsections below, we described six methods that can be used under the FE framework.

As already mentioned in the Introduction, when meta-analyzing count data in the FE framework, one should further distinguish between the settings of homogeneous and heterogeneous baseline prevalences (see Table 1). This subdivision has a direct implication regarding the notion of treatment effect homogeneity. With homogeneous baseline prevalences, the homogeneity of the ES holds whatever the metric adopted (OR, RR or RD). On the contrary, with heterogeneous baseline prevalences, homogeneity of the ES depends on the scale of measurement. Indeed, assume for instance that the baseline probability  $\pi_c$  is heterogeneous and that the RD is homogeneous. Then, if one were to use instead the RR, the ES estimate would turn out be heterogeneous (Table 2). Therefore, under the FE framework with baseline heterogeneity, the investigator has to specify on which scale the treatment effect is assumed homogeneous.

- Insert Table 2 here -

As this small example illustrates, the selection of the appropriate scale to measure the effect of the treatment is an important question. Actually, there is a broad debate in the literature on the advantages of absolute measures, such as the RD, versus relative measures, such as the RR and OR (Sinclair and Bracken, 1994; Papageorgiou et al., 2015).

In what follows, each primary study consists of one control group of size  $n_c$  and one treated group of size  $n_t$ . The number of events occurring in these two groups are denoted  $X_c$  and  $X_t$ , respectively.

## 2.1 The inverse variance method

For  $k = 1, \dots, K$ , the IV method is based on the following model:

$$\hat{\theta}_k = \theta + \epsilon_k, \quad \epsilon_k \sim N(0, \sigma_k^2) \quad (4)$$

where  $\hat{\theta}_k$  is the estimator of the parameter of interest  $\theta$  (i.e. RD, RR or OR) obtained from study  $k$  and  $\sigma_k^2$  its variance. The RR and OR's estimators are usually analyzed on the log scale, as their sampling distribution is more symmetrical on this scale.

The three ES estimators are computed as follows:

$$\widehat{OR}_k = \frac{\hat{\pi}_{kt} * (1 - \hat{\pi}_{kc})}{\hat{\pi}_{kc} * (1 - \hat{\pi}_{kt})} \quad (5)$$

$$\widehat{RR}_k = \frac{\hat{\pi}_{kt}}{\hat{\pi}_{kc}} \quad (6)$$

$$\widehat{RD}_k = \hat{\pi}_{kt} - \hat{\pi}_{kc} \quad (7)$$

where  $\hat{\pi}_{kt} = X_{kt}/n_{kt}$  and  $\hat{\pi}_{kc} = X_{kc}/n_{kc}$  are the maximum likelihood estimators.

Variances of these estimators, computed using the delta method for the OR and the RR, are estimated by

$$\hat{\sigma}_{\log \widehat{OR}_k}^2 = \frac{1}{n_{kt} \hat{\pi}_{kt} (1 - \hat{\pi}_{kt})} + \frac{1}{n_{kc} \hat{\pi}_{kc} (1 - \hat{\pi}_{kc})} \quad (8)$$

$$\hat{\sigma}_{\log RR_k}^2 = \frac{1 - \hat{\pi}_{kt}}{n_{kt}\hat{\pi}_{kt}} + \frac{1 - \hat{\pi}_{kc}}{n_{kc}\hat{\pi}_{kc}} \quad (9)$$

$$\hat{\sigma}_{RD_k}^2 = \frac{\hat{\pi}_{kt}(1 - \hat{\pi}_{kt})}{n_{kt}} + \frac{\hat{\pi}_{kc}(1 - \hat{\pi}_{kc})}{n_{kc}} \quad (10)$$

The IV estimator is obtained by applying the generalized least squares method, assuming independence between the  $K$  primary studies and the within-study variances  $\sigma_k^2$  known:

$$\hat{\theta}_{IV} = \frac{\sum_{k=1}^K W_k \hat{\theta}_k}{\sum_{k=1}^K W_k} \quad (11)$$

where  $W_k = 1/\sigma_k^2$ . This estimator is equivalent to the maximum likelihood estimator whenever the assumption of normality (4) holds. The variance of  $\hat{\theta}_{IV}$  is given by

$$Var(\hat{\theta}_{IV}) = \frac{1}{\sum_{k=1}^K W_k} \quad (12)$$

In practice,  $\sigma_k^2$  is not observed and  $\hat{W}_k = 1/\hat{\sigma}_k^2$  is used instead.

As discussed in the Introduction section, the good performances of the IV method hold asymptotically. However, in finite samples, estimates obtained with this method systematically deviate from the true parameter value and observed coverage probabilities of the CI depart from the nominal value. These undesirable properties, which are exacerbated with rare events, can be explained by several reasons. First, primary-study estimators of the RR and OR are biased (Firth, 1993; Nemes et al. 2009). Therefore, even when  $\sigma_k^2$  is known, one has for RR and OR:

$$E(\hat{\theta}_{IV} | \sigma_k^2) = \frac{\sum_{k=1}^K W_k E(\hat{\theta}_k)}{\sum_{k=1}^K W_k} = E(\hat{\theta}_k) \neq \theta \quad (13)$$

Second, the primary-study estimate  $\hat{\theta}_k$  is correlated with its estimated variance and this correlation is not taken into account by the IV method (Berkey et al., 1995). Third, the sampling distribution of the IV estimator is not well approximated by a normal distribution.

In very rare event settings, when some primary studies report zero event, the IV method can lead to indefinite estimates. For the RR and the OR, this happens whenever  $X_{kc} = 0$  or  $X_{kt} = 0$  (or both) for any  $k$ . For the RD,  $\hat{\sigma}_k^2 = 0$  and, thus,  $\hat{W}_k = \infty$  when  $X_{kc} = X_{kt} = 0$ . As a result, to obtain defined IV estimates, one has to exclude DZ studies from the computation of all ES and SZ studies from the computation of OR and RR.

## 2.2 The Mantel-Haenszel method

The MH method was first proposed in 1959 to estimate the risk of an exposure by means of an Odds Ratio while adjusting for confounding factors (Mantel and Haenszel, 1959). This method has, then, been extended to other ES (Rothman et al., 2008) and can be applied to conduct meta-analyses of OR, RR and RD using the following formulae:

$$\widehat{OR}_{MH} = \frac{\sum_{k=1}^K X_{kt}(n_{kc} - X_{kc})/n_k}{\sum_{k=1}^K X_{kc}(n_{kt} - X_{kt})/n_k} \quad (14)$$

$$Var(\log(\widehat{OR}_{MH})) = \frac{1}{2} \left( \frac{\sum_{k=1}^K R_k P_k}{R^2} + \frac{\sum_{k=1}^K (P_k S_k + Q_k R_k)}{RS} + \frac{\sum_{k=1}^K S_k Q_k}{S^2} \right) \quad (15)$$

$$\widehat{RR}_{MH} = \frac{\sum_{k=1}^K X_{kt} n_{kc} / n_k}{\sum_{k=1}^K X_{kc} n_{kt} / n_k} \quad (16)$$

$$Var(\log(\widehat{RR}_{MH})) = \frac{\sum_{k=1}^K \left[ \frac{(X_{kt} + X_{kc}) n_{kt} n_{kc}}{n_k^2} - \frac{X_{kt} X_{kc}}{n_k} \right]}{\left( \sum_{k=1}^K \frac{X_{kt} n_{kc}}{n_k} \right) \left( \sum_{k=1}^K \frac{X_{kc} n_{kt}}{n_k} \right)} \quad (17)$$

$$\widehat{RD}_{MH} = \frac{\sum_{k=1}^K (X_{kt} n_{kc} - X_{kc} n_{kt}) / n_k}{\sum_{k=1}^K n_{kt} n_{kc} / n_k} \quad (18)$$

$$Var(\widehat{RD}_{MH}) = \frac{\sum_{k=1}^K \left( \frac{n_{kt} n_{kc}}{n_k} \right)^2 \left[ \frac{X_{kt} (n_{kt} - X_{kt})}{n_{kt}^2 (n_{kt} - 1)} - \frac{X_{kc} (n_{kc} - X_{kc})}{n_{kc}^2 (n_{kc} - 1)} \right]}{\left( \sum_{k=1}^K \frac{n_{kt} n_{kc}}{n_k} \right)^2} \quad (19)$$

where  $n_k = n_{kt} + n_{kc}$ ,  $P_k = (X_{kt} + (n_{kc} - X_{kc})) / n_k$ ,  $Q_k = (X_{kc} + (n_{kt} - X_{kt})) / n_k$ ,  $R_k$

$= X_{kt} (n_{kc} - X_{kc}) / n_k$ ,  $S_k = X_{kc} (n_{kt} - X_{kt}) / n_k$ ,  $R = \sum_{k=1}^K R_k$ ,  $S = \sum_{k=1}^K S_k$ .

The MH formulae do not rely on the primary-study ES estimates (only on the counts). As a result, MH estimators are more robust to zero-event issues than IV's. For RR and OR it provides indefinite estimate only when all control groups report zero event (i.e.  $X_{kc} = 0 \forall k$ ). The variance estimates of the log(OR) and log(RR) are indefinite when either  $X_{kt} = 0$  or  $X_{kc} = 0 \forall k$  (or both). Although quite rare, such extreme scenarios are sometimes encountered in practice. For instance, a systematic review on the occurrence of lactic acidosis with metformin use in type 2 diabetes mellitus gathered 148 studies that were all DZ studies (Salpeter et al., 2010). From the above formulae, one can see that SZ studies contribute to the computation of all ESs (i.e. to either the numerator of the denominator for the OR and RR and to both for the RD). Contrariwise, DZ studies do not contribute to the computation of the OR and RR estimates, whereas these studies do contribute to the RD estimate (i.e. to the denominator).

Silcocks (2005) showed that the MH estimator for the OR corresponded to the maximum likelihood estimator (based on the Binomial distribution for the number of events), whenever the probability of event in the control group is homogeneous and the ratio of sample sizes is constant across the primary studies. Similarly, one can show that the MH estimators of the RR and RD correspond to their likelihood counterpart, whenever the ratio of sample sizes is constant (whatever the prevalences). For the RD, the MH estimator is given by (18). With homogeneous ratios of sample sizes (i.e.  $n_{kc} / n_{kt} = R \forall k$ ), we have  $n_{kc} = n_{kt} * R$  and  $n_k = n_{kt} * (1 + R)$ . Substituting these results into the above formula yields:

$$\begin{aligned} \widehat{RD}_{MH} &= \frac{\sum_{k=1}^K (X_{kt} n_{kt} R - X_{kc} n_{kt}) / (n_{kt} * (1 + R))}{\sum_{k=1}^K n_{kt}^2 R / (n_{kt} * (1 + R))} \\ &= \frac{\sum_{k=1}^K X_{kt} R / (1 + R) - \sum_{k=1}^K X_{kc} / (1 + R)}{\sum_{k=1}^K n_{kt} R / (1 + R)} \\ &= \frac{\sum_{k=1}^K X_{kt}}{\sum_{k=1}^K n_{kt}} - \frac{\sum_{k=1}^K X_{kc}}{\sum_{k=1}^K n_{kt} R} \end{aligned} \quad (20)$$

Since  $n_{kc} = n_{kt} * R$ , the equivalence with the maximum likelihood estimator follows. Using similar arguments, one can show that  $\widehat{RR}_{MH} = \widehat{RR}_{ML}$  when  $n_{kc} / n_{kt} = R$ .

### 2.3 The Peto method

The Peto method was introduced as a user-friendly solution to estimate the OR in the setting of rare events (Yusuf et al., 1985). Peto proposed the following estimator for the log(OR):

$$\widehat{\log(OR)}_{Peto} = \frac{\sum_{k=1}^K (O_k - E_k)}{\sum_{k=1}^K V_k} \quad (21)$$

where  $O_k = X_{kt}$  is the observed number of events in the treatment group of study  $k$ ,  $E_k = X_k * \frac{n_{kt}}{n_k}$  is the expected number of events in the treatment group under the null hypothesis of no treatment effect,  $X_k = X_{kc} + X_{kt}$  is the total number of event, and  $V_k = E_k n_{kc} (n_k - X_k) / (n_k (n_k - 1))$  is the hypergeometric variance of  $O_i$  under the null. The variance of this estimator is given by:

$$Var(\widehat{\log(OR)}_{Peto}) = \frac{1}{\sum_{k=1}^K V_k} \quad (22)$$

Peto estimator is obtained using exact likelihood theory (Cox, 1977) and corresponds to the estimate of the common log(OR) obtained in the first step of a Newton-Raphson procedure to maximize the conditional log-likelihood when the starting value for the log(OR) is zero (McCullagh and Nelder, 1981). Hence its other name: the "one-step estimator".

DZ studies do not contribute to the Peto log(OR) estimate (i.e. the quantities  $O_k$ ,  $E_k$  and  $V_k$  are all null), whereas SZ studies do contribute. The only setting under which the Peto estimator and its variance are undefined is when all included studies are DZ studies.

#### 2.4 The continuity correction factor method

As already mentioned in the three previous subsections, the three classical FE methods have difficulties to deal with zero-event studies. The IV method yields indefinite estimates in the presence of either SZ or DZ studies when pooling ORs or RRs, and in the presence of DZ studies when pooling RDs. As for MH and Peto, although these two methods are robust to zero-event studies, they are based on formulae that discard DZ studies when pooling ORs or RRs (for MH).

A simple remedy to the issue of SZ and DZ studies, which dates back to 1934 and has been adopted by many researchers is to use a cc. It consists in adding a constant  $c$  to each cell of each contingency table containing one or more 0 frequency (Yates, 1934). Plackett (1964) provided a detailed account of this method. In this paper, we used a cc of 0.5, which can be justified by theoretical arguments (Bhaumik et al., 2012).

In this paper, the motivation of using a cc was to allow all studies to contribute to the combined ES estimate. Therefore, we additionally evaluated IV, MH and Peto methods with a cc. The table below summarizes the various strategies applied with these three methods to tackle the issue of zero-event studies:

- Insert Table 3 here -

#### 2.5 The median unbiased estimator method

The MUE method works in two steps. First, one computes in each primary study  $k$  the MUE of  $\pi_{kj}$ ,  $j \in \{c, t\}$  (Parzen et al., 2002):

$$\hat{\pi}_{kj} = \begin{cases} (1 - 0.5^{1/n_{kj}})/2 & \text{if } X_{kj} = 0 \\ (p_{kj}^L + p_{kj}^U)/2 & \text{if } 0 < X_{kj} < n_{kj} \\ (0.5^{1/n_{kj}} + 1)/2 & \text{if } X_{kj} = n_{kj} \end{cases} \quad (23)$$

with  $p_{kj}^L = F^{-1}(0.5 | \alpha = X_{kj}, \beta = n_{kj} - X_{kj} + 1)$  and  $p_{kj}^U = F^{-1}(0.5 | \alpha = X_{kj} + 1, \beta = n_{kj} - X_{kj})$ , where  $F^{-1}(Q | \alpha, \beta)$  is the  $Q$ th quantile of the beta-distribution with parameters  $\alpha$  and  $\beta$ . Note that the expression we used for  $p_{kj}^U$  differed from that provided in the paper of Parzen and colleagues, which, we believe, contains an error. Mathematical justifications can be found in the Appendix.

Second, from these two estimated probabilities, one can compute the primary-study ESs using equations (5)-(7) as well as their corresponding variance using equations (8)-(10), and combined them using a weighted average as in (11). Variance of the MUE combined estimate is then given by (12).

Clearly, the MUE method provides estimators that are always well defined, whatever the degree of sparseness of the events. Moreover, this method includes the information from DZ studies.

## 2.6 The Binomial regression model

Another option to deal with zero-event studies is to use a Binomial logistic regression model:

$$X_{kc} \sim \text{Binomial}(n_{kc}, \pi_{kc}) \quad (24)$$

$$X_{kt} \sim \text{Binomial}(n_{kt}, \pi_{kt}) \quad (25)$$

$$\text{logit}(\pi_{kc}) = \alpha_k \quad (26)$$

$$\text{logit}(\pi_{kt}) = \alpha_k + \delta \quad (27)$$

where  $\alpha_k$  represents the logit of the control group probability and  $\delta$  the log(OR).

Under the assumption of homogeneous baseline prevalences,  $\alpha_k = \alpha, \forall k$ , parameters are estimated by maximizing the following likelihood function:

$$L(\alpha, \delta) = \prod_{k=1}^K \binom{n_{kc}}{X_{kc}} \pi_c^{X_{kc}} (1 - \pi_c)^{n_{kc} - X_{kc}} \binom{n_{kt}}{X_{kt}} \pi_t^{X_{kt}} (1 - \pi_t)^{n_{kt} - X_{kt}} \quad (28)$$

Under the assumption of baseline heterogeneity,  $\alpha_k$  is treated as a nuisance parameter. One can deal with this nuisance parameter by adopting either a fixed effects or a random effects approach. The main advantage of a fixed effects approach is that no distributional assumption have to be made, whereas the random effects approach implies the choice of a distribution for  $\alpha_k$ . However, particularly with rare events, it is advantageous to treat  $\alpha_k$  as a random variable to limit as much as possible the number of parameters to be estimated and allow DZ studies to contribute to the estimation.

We assumed  $\alpha_k \sim N(\alpha, \sigma_\alpha^2)$  and estimated the parameters by maximizing the following marginal likelihood function:

$$L(\alpha, \delta) = \prod_{k=1}^K \int_{-\infty}^{+\infty} \binom{n_{kc}}{X_{kc}} \pi_c^{X_{kc}} (1 - \pi_c)^{n_{kc} - X_{kc}} \binom{n_{kt}}{X_{kt}} \pi_t^{X_{kt}} (1 - \pi_t)^{n_{kt} - X_{kt}} f(\alpha_k | \alpha, \sigma_\alpha^2) d\alpha_k \quad (29)$$

where  $f(\cdot)$  is the Normal density.

## 3 Illustrative example

### 3.1 Perinatal death in post-term pregnancy

To motivate and illustrate the use of the six methods described in the previous section, we considered the systematic review conducted by Crowley (2000). In this review, the author compared the number of deaths induced by routine and selective induction of pregnancies that go beyond term. Data of the 19 randomized control trials included in Crowley's review are shown in Table 4.

- Insert Table 4 here -

All trials reported at least one arm with zero event and 11 trials were DZ studies. Most of the trials' arms included between 50 and 150 women. Except for one study where an imbalance of 2:1 in favor of the treatment group was observed, studies' arms were mostly balanced.

### 3.2 Results from the methods when fitted to the illustrative dataset

Results obtained by the different methods applied to the illustrative dataset are displayed in Table 5. All methods found a decrease of the number of perinatal deaths in the treated women (i.e. those in the group of routine induction). However, the methods differed markedly in terms of the magnitude of the effect and confidence interval obtained, especially for the OR and RR. Since this review included only zero-event studies, the IV method without cc was unable to provide finite estimates for the OR and RR. As expected in such a rare event setting, estimates obtained for these two ES were quite similar.

- Insert Table 5 here -

## 4 Simulation study

### 4.1 Model

We considered various scenarios, which varied according to (i) the level of rareness (extremely rare, very rare, moderately rare, common), (ii) the assumption regarding baseline prevalences (homogeneous or heterogeneous), (iii) the level of unbalance between the trials' arms (strong unbalance in favor of either arms, moderate unbalance in favor of either arms, no unbalance), (iv) the size of the treatment effect (large reduction, moderate reduction, no effect, moderate increase, large increase). Extremely rare events had a median baseline probability of  $M_{\pi_c} = 0.005$ . The three other rareness levels corresponded to  $M_{\pi_c} = 0.01, 0.05, 0.1$ , respectively. With homogeneous baseline prevalences we set  $\pi_{kc} = M_{\pi_c} \forall k$ . With heterogeneous baseline prevalences, we used  $\pi_{kc} \sim \text{Logit-normal}(\text{logit}(M_{\pi_c}), \sigma_{\text{logit}}^2)$ , with  $\sigma_{\text{logit}}^2 = 0.5$ . As regards unbalance, the five scenarios considered were obtained with mean levels  $r = 0.25, 0.5, 1, 2, 4$  ( $r = 0.25$  corresponding to setting with control groups four times smaller than treatment groups, in average). For the OR, and the RR, the sizes of treatment effect – measured on the log(OR) and log(RR) scales, respectively – were  $-1.5, -0.5, 0, 0.5, 1.5$ . For the RD, we did not consider reduction in event prevalence to avoid cases with negative treatment probability. We used the three following values: 0, 0.05, 0.1.

Combinations of these four characteristics resulted in 200 simulation scenarios for the OR and RR, and 120 for the RD. For each scenario, 10,000 meta-analyses were generated, each of them consisting of  $K = 20$  primary studies with treatment arms' sample sizes ranging from 50 to 150 (i.e.  $n_{kt} \sim \text{Uniform}(50; 150)$ ). Sample sizes for the control arms were then obtained as  $n_{kc} = n_{kt} * R$ , where  $R \sim \text{Uniform}(r - 0.1; r + 0.1)$  and  $r$  controls the degree of arms unbalance. Treatment probabilities  $\pi_{kt}$  were derived from the control probabilities and the ES considered (e.g. for RD:  $\pi_{kt} = \pi_{kc} + RD$ ). Finally, the number of events in both arms were generated by two binomial draws with respective sample sizes and event probabilities.

As pointed out by a reviewer, meta-analyses of  $K = 20$  studies might seem an optimistic scenario, as many published meta-analyses include a smaller number of studies. Nevertheless, for our goal to study the impact of rare events, it was better to have a large enough  $K$  to avoid fluctuation issues related to scarcity of primary-studies. Therefore, we set the number of studies at a somewhat ideal level, but still realistic. For instance, Moher et al. (2007) found a median number of 23 studies out of 88 systematic reviews analyzed.

For each of the generated meta-analyses, we estimated the ES of interest, its standard error and the 95% Wald CI. Performances of the different methods were assessed in terms of bias, and coverage rate. We additionally computed CIs' width and reported them, along with the coverage rate, in the file containing the detailed simulations results (see Supplementary Material). We decided to compute median instead of mean values for the bias and CIs' width to avoid the influence of exceedingly large or small values obtained in some simulations. Both OR and RR were analyzed on the log scale.



## 4.2 Results

In this section, we presented abridged results focusing on the most interesting findings for each ES in the setting of heterogeneous baseline prevalences, since results and conclusions obtained with homogeneous and heterogeneous baseline prevalences were quite similar. Furthermore, we only discussed results for  $r = 1, 2, 4$  and  $\log(ES) = 0, -0.5, -1.5$  ( $ES = RR$  or  $OR$ ), which illustrate well the issues of unbalanced sample sizes and large treatment effects. Detailed results for all the ES, simulated scenarios and outcomes can be found in the Supplementary Material.

### 4.2.1 Odds ratio

In terms of bias (Figure 1), the MH and Binomial regression methods provided the best estimates whereas the IV method was the least robust across almost all settings considered. The use of a cc reduced the bias of the IV estimator but increased that of MH and Peto. The MUE method can be seen as an improved version of the IV method with performances comparable to those obtained when using the IV method with cc. The Peto method obtained similar results to those of the MH and Binomial regression methods for moderate and no treatment effects, particularly under balanced settings. However, it was clearly not reliable with large treatment effects and even failed to converge for increasing baseline probabilities (whatever the degree of unbalance). Under the scenario of no treatment effect and balanced trials, all the methods provided unbiased estimates, except the MH + cc method.

- Insert Figure 1 here -

In terms of coverage rates (Figure 2), ranking of the different meta-analysis methods was similar to what was obtained in terms of bias. The following additional observations can be made: (i) MH and Binomial regression were slightly conservative in extremely rare event settings, (ii) coverage rates obtained by the IV (with or without cc) and MUE methods were below 90% in many scenarios considered, and (iii) ranking between these three methods varied depending on the setting.

- Insert Figure 2 here -

### 4.2.2 Relative risk

In terms of bias (Figure 3), comparisons between the different methods yielded to the same conclusions as those drawn for the OR: (i) MH estimator was the least biased, (ii) the use of a cc increased the bias for MH estimator and reduced it for the IV estimator, and (iii) performances of the MUE method were comparable to those of the IV method with cc.

- Insert Figure 3 here -

In terms of coverage rates (Figure 4), results obtained by the six methods investigated to estimate the RR were comparable to those obtained for the OR.

- Insert Figure 4 here -

### 4.2.3 Risk difference

In terms of bias (Figure 5), the MH estimator of the RD was the least biased across the different settings. Ranking of IV, IV + cc and MUE depended on the setting and none of the methods strictly dominated – or was strictly dominated by – the others. For all the methods, bias values for the RD were much smaller than those obtained for the RR and OR.

- Insert Figure 5 here -

In terms of coverage rates (Figure 6), the MH method again obtained the best results. Contrarily to CIs for the OR and RR, those obtained for the RD tended to have slightly below 95% coverage rates in

extremely rare event settings. Coverage rates provided by IV, IV + cc and MUE methods did not systematically converge towards 95% as the event prevalence increased.

- Insert Figure 6 here -

## 5 Discussion

Traditionally, in meta-analysis, a clear distinction is made between the FE and the RE models. In the FE model, one assumes that the treatment effect is homogeneous across the primary studies and the goal of the meta-analysis is to estimate a single population parameter. In the RE model, one assumes that the treatment effect is heterogeneous across the primary studies (there is a population of parameters) and focus is on characterizing the distribution of these parameters. The selection of either model must be based on contextual knowledge, as this choice has implications in terms of target of inference and appropriate statistical methods to conduct the analysis.

When considering count data, one must further distinguish between settings where baseline prevalences are expected to be homogeneous or heterogeneous. Indeed, we have shown that this distinction has a direct implication regarding the notion of treatment effect homogeneity. Again, the choice between these two settings should be grounded on contextual knowledge. In the literature, the distinction between homogeneous and heterogeneous baseline prevalences is seldomly discussed. Actually, we found only one paper where simulations were carried out under the FE framework with heterogeneous baseline prevalences (Böhning and Sarol, 2000). However, the authors focused on the RD only and did not consider the issue of rare events. In the three most comprehensive simulation studies that tackled the issue of rare event meta-analysis (Sweeting et al., 2004; Bradburn et al., 2007; Kuss 2015), the authors did not consider heterogeneous baseline prevalences. Therefore, the objective of this paper was to investigate the impact of homogeneous versus heterogeneous baseline prevalences on the performance of FE meta-analysis methods in the context of rare events and homogeneous treatment effect. Through extensive simulations, we assessed the ability of the IV, MH, Peto, with or without cc, MUE and Binomial regression methods to estimate the three most commonly used effect sizes with count data (RD, RR and OR) under various settings.

We found that whatever the baseline prevalences (i.e. either homogeneous or heterogeneous), under all the scenarios considered, and for all the ESs, the most reliable methods were the MH method without cc and the Binomial regression model (for estimating the OR only). Interestingly, the fact that the MH method discards DZ studies did not seem to introduce a bias. On the contrary, using a cc to include the information contained in DZ studies deteriorated dramatically this method's performances.

Under the setting of homogeneity in baseline prevalences, MH estimates for the OR, RR and RD correspond to those of the simple pooling method. In other words, when assuming a FE model with homogeneous baseline prevalences, one does not need to apply meta-analysis techniques to compute a combined estimate. This is because, under this particular framework, the three ES are collapsible (Greenland et al., 1999; Guo and Geng, 2005; Hernan et al., 2011). However, when introducing heterogeneity in baseline prevalences, the OR is no more collapsible and simple pooling leads to biased estimates (this was confirmed by additional simulations; results not shown). As for the RR and RD, they remain collapsible and, thus, one can obtain unbiased estimates by simply pooling the studies, even in the presence of heterogeneous baseline prevalences. Nonetheless, our simulations have shown that simple pooling yielded CIs with coverage rates below nominal. The reason was linked to a failure of the simple pooling method to account for a larger sampling distribution of the ES.

Consistently with the statistical literature, our simulations ranked IV as the worst method in almost all rare events settings. Moreover, we found that this method failed to provide valid CIs and unbiased estimates in scenarios with common events (i.e. baseline prevalences around 10%). When estimating the RR and OR, the use of a cc improved the performance of the IV method in terms of bias. However, this was not the case for the RD and coverage rates obtained for the three ESs were not nominal. As a solution to improve the coverage rates, we additionally considered the modified Hartung-Knapp-Sidik-Jonkman (HKSJ) method (Röver et al., 2015), which involves the use of the Student-t distribution and a multiplicative correction term for the variance (additional simulations; results not shown). By increasing the size of the CI, this method improved the coverage rates in scenarios where IV's CIs were

below nominal. Most of the time, however, coverage rate of HKSJ's CIs remained below 95%. In addition, when coverage rates were nominal, this method provided wider CIs than those obtained using the MH method. Likewise, HKSJ's CIs were too conservative in scenarios where IV's CIs were valid.

The Peto method, which is often recommended in cases of rare events (Higgins and Green, 2011), yielded contrasted results. Whereas the amount of bias and coverage rates were similar to those obtained by the MH and Binomial regression method in scenarios with no effect or medium treatment effect, results provided by this method under settings with large treatment effect critically deteriorated, especially with large unbalance between the trial's arms. These findings seem to corroborate the conclusions made by some authors stating that the Peto method should only be applied with great caution (Brockhaus et al., 2016). Moreover, we found that Peto's estimates did not converge towards the true OR when the baseline probability increased, which gives additional credits to the argument that Peto's OR should be viewed as a different ES, and not as an OR's estimator (Brockhaus et al., 2014). Finally, the use of a cc deteriorated the performance of the Peto method.

The MUE method is an interesting – since less subjective – alternative to the use of a cc. It tackles the issue of zero-event studies elegantly and does not involve complex computations (i.e. it only requires the computation of Beta quantiles). While meta-analyses techniques based on MUE of probability have already been discussed by Li and Wang (2017), we innovated here by combining these estimates using a weighted average. The amount of bias found for the OR and RR were similar to that obtained with the IV + cc method. On the other hand, results obtained for the RD were more contrasted and the MUE method sometimes provided the most biased estimates.

## **6 Conclusion**

Based on our findings, we make the following recommendations to applied researchers conducting meta-analyses of count data, under the framework of a homogeneous treatment effect. First, it is important to clarify the question of heterogeneous vs homogeneous baseline prevalences. When prevalences are expected to be heterogeneous, the researcher has to decide on which scale the treatment effect is assumed homogeneous (i.e. OR, RR or RD). Second, we highly recommend using the MH method without cc in all circumstances (i.e. whatever the ES of interest, the assumption regarding the heterogeneity in baseline prevalences, and the scenario considered). To estimate the OR, the Binomial regression method is a sound alternative, which allows one to adjust for covariates. Third, the use of a cc should be definitively abandoned.

The main limitation to these recommendations is that they only apply to the framework of a homogeneous treatment effect. The reader must keep in mind that the MH method is not valid when the treatment effect is heterogeneous, which is often the case in practice (Kontopantelis et al., 2013). Moreover, we did not investigate Bayesian methods. However, this choice was deliberate and motivated by the fact that these methods require the use of subjective priors, which can have – even non-informative ones – substantial effects on the estimates, especially with rare events (Lambert et al., 2005; Senn, 2007). Future researches should focus on extending the present work to the setting of heterogeneous treatment effects, with or without baseline heterogeneity.

### **Acknowledgements**

The work of RPR was supported by the Swiss National Science Foundation (grant number: 200021\_169504/1). The authors would like to thank one associate editor and two anonymous reviewers for their helpful comments and suggestions.

### **Conflict of Interest**

*The authors have declared no conflict of interest.*

### **Supplementary Material**

A word document (.docx) containing detailed simulations results for all the ES (OR, RR, RD), all the statistical outcomes (bias, coverage rate, CI width), and all the simulated scenarios (200 for the OR and RR and 120 for the RD) may be found online in the Supporting Information tab for this article. Source code to reproduce the results is also available as Supporting Information.

## Appendix: derivation of $p_{kj}^U$ for the median unbiased estimator method

Parzen et al. (2002) showed that  $p_{kj}^U$  could be obtained by solving

$$0.5 = \sum_{i=0}^{X_{kj}} (p_{kj}^U)^i (1 - p_{kj}^U)^{n_{kj}-i} \quad (\text{A1})$$

Then, one can use the following relationship between the cumulative Beta distribution function and the cumulative Binomial distribution function (Johnson and Kotz 1969; Daly, 1992):

$$F(p|\alpha, \beta) = \sum_{i=\alpha}^n \binom{n}{i} p^i (1-p)^{n-i} \quad (\text{A2})$$

where  $F(\cdot|\alpha, \beta)$  is the cumulative Beta distribution with integer parameters  $\alpha$  and  $\beta = n - \alpha + 1$ . Plugging (A2) into (A1) yields

$$0.5 = 1 - F(p_{kj}^U|\alpha = X_{kj} + 1, \beta = n_{kj} - \alpha + 1) \quad (\text{A3})$$

where  $\beta = n_{kj} - (X_{kj} + 1) + 1 = n_{kj} - X_{kj}$ , and not  $n_{kj} - X_{kj} + 2$  as Parzen and colleagues wrote in their paper on page 425.

Finally, an expression for  $p_{kj}^U$  is given by

$$p_{kj}^U = F^{-1}(0.5|\alpha = X_{kj} + 1, \beta = n_{kj} - X_{kj}) \quad (\text{A4})$$

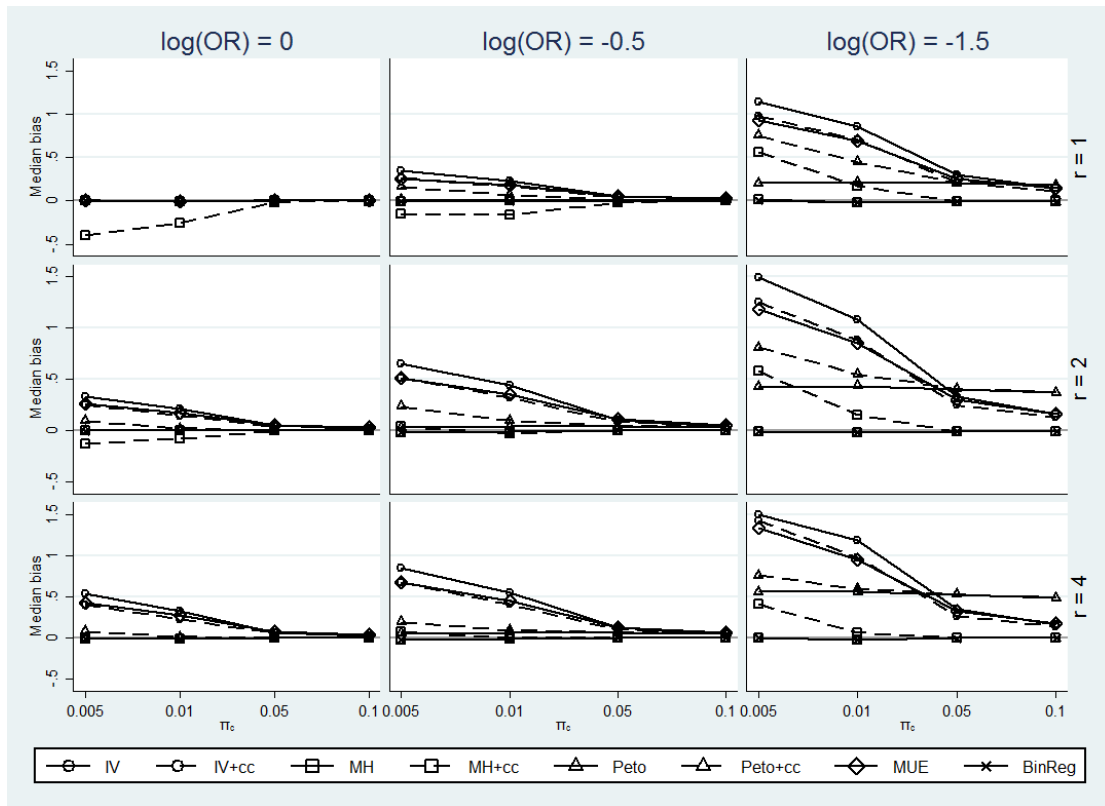
which corresponds to the expression we used in this paper.

## References

- Berkey, C. S., Hoaglin, D. C., Mosteller, F., et al. (1995). A random-effects regression model for meta-analysis. *Statistics in Medicine* **14**, 395-411.
- Bhaumik, D. K., Amaty, A., Normand, S. L., et al. (2012). Meta-analysis of rare binary adverse event data. *Journal of American Statistical Association* **107**, 555-567.
- Böhning, D., and Sarol, J. (2000). Estimating risk difference in multicenter studies under baseline-risk heterogeneity. *Biometrics* **56**, 304-308.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., et al. (2009). *Introduction to Meta-Analysis*. John Wiley & Sons, Ltd, Chichester.
- Bradburn, M. J., Deeks, J. J., Berlin, J. A., et al. (2007). Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Statistics in Medicine* **26**, 53-77.
- Brockhaus, A. C., Grouven, U., and Bender, R. (2016). Performance of the Peto odds ratio compared to the usual odds ratio estimator in the case of rare events. *Biometrical Journal* **58**, 1428-1444.
- Brockhaus, A. C., Bender, R., and Skipka, G. (2014). The Peto odds ratio viewed as a new effect measure. *Statistics in Medicine* **33**, 4861-4874.
- Cox, D. R. (1977). *The Analysis of Binary Data*. Chapman and Hall, New York.
- Crowley, P. (2000). Interventions for preventing or improving the outcome of delivery at or beyond term. *Cochrane Database Systematic Reviews* **2**, CD000170.
- Daly, L. (1992). Simple SAS Macros for the calculation of exact binomial and poisson confidence intervals. *Computer in Biology and Medicine* **22**, 351-361.
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 27-38.
- Greenland, S., Robins, J. M., and Pearl, J. (1999). Confounding and collapsibility in causal inference. *Statistical Science* **14**, 29-46.
- Guo, J., and Geng, Z. (2005). Collapsibility of logistic regression coefficients. *Journal of the Royal Statistical Society Series B (Methodological)* **57**, 263-267.
- Hernan, M. A., Clayton, D., and Keiding, N. (2011). The Simpson's paradox unraveled. *International Journal of Epidemiology* **40**, 780-785.
- Higgins, J. P. T., and Green, S. (eds). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]*. The Cochrane Collaboration, 2011. Available from: [www.training.cochrane.org/handbook](http://www.training.cochrane.org/handbook) [Accessed on 10 December 2018]
- Hirji, K. F., Tsiatis, A. A., and Mehta, C. R. (1989). Median unbiased estimation for binary data. *The American Statistician* **43**, 7-11.
- Johnson, N. L., and Kotz, S. (1969). *Distribution in Statistics: Discrete Distributions*. Wiley, New York.
- Kontopantelis, E., Springate, D. A., Reeves, D. (2013). A re-analysis of the Cochrane library data: the dangers of unobserved heterogeneity in meta-analyses. *Plos One* **8**, e69930.
- Kuss, O. (2015). Statistical methods for meta-analyses including information from studies without any events – add nothing to nothing and succeed nevertheless. *Statistics in Medicine* **34**, 1097-1116.
- Kuss, O., Wandrey, M., Kunze, M. (2009). How frequent are meta-analyses with “double-zero” studies in systematic reviews? Available from: <http://www.egms.de/static/de/meetings/gmds2009/09gmds155.shtml> [Accessed 7 February 2019].
- Lambert, P. C., Sutton, A. J., Burton, P. R., et al. (2005). How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine* **24**, 2401-2428.
- Lane, P. W. (2013). Meta-analysis of incidence of rare events. *Statistical Methods in Medical Research* **22**, 117-132.
- Li, L., and Wang, X. (2017). Meta-analysis of rare binary events in treatment groups with unequal variability. *Statistical Methods in Medical Research* **0**, 1-12.
- Mantel, N., and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* **22**, 719-748.
- McCullagh, P., and Nelder, J. A. (1981). *Generalized Linear Models*. Chapman and Hall, New York.

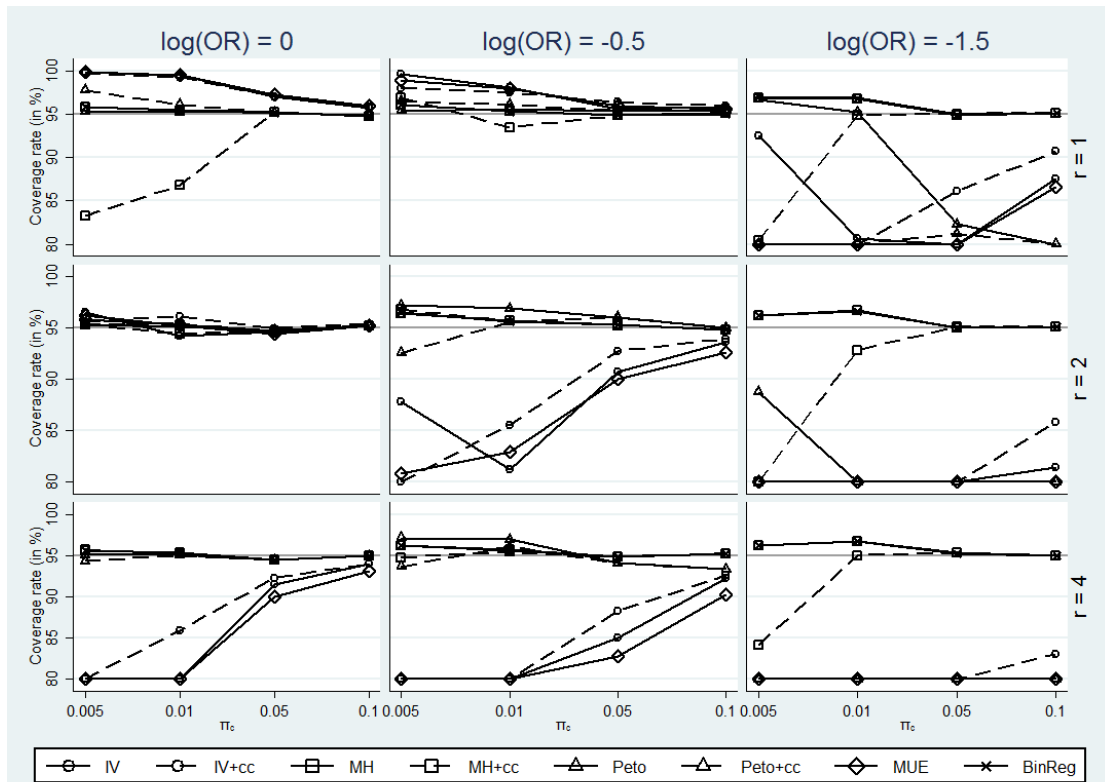
- Moher, D., Tetzlaff, J., Tricco, A. A., et al. (2007). Epidemiology and reporting characteristics of systematic reviews. *PLoS Medicine* **4**, e78.
- Nemes, S., Jonasson, J. M., Genell, A., et al. (2009). Bias in odds ratios by logistic regression modelling and sample size. *BMC Medical Research Methodology* **9**, 56.
- Normand, S. L. T. (1999). Meta-analysis: formulating, evaluating, combining, and reporting. *Statistics in Medicine* **18**, 321-359.
- Papageorgiou, S. N., Tsiranidou, E., Antonoglou, G. N., et al. (2015). Choice of effect measure for meta-analysis of dichotomous outcomes influenced the identified heterogeneity and direction of small-study effects. *Journal of Clinical Epidemiology* **68**, 534-541.
- Parzen, M., Lipsitz, S., Ibrahim, J., et al. (2002). An estimate of the odds ratio that always exists. *Journal of Computational and Graphical Statistics* **11**, 420-436.
- Plackett, R. L. (1964). The continuity correction in 2 x 2 tables. *Biometrika* **51**, 327-337.
- Rothman, K. J., Greenland, S., and Lash, T. L. (2008). *Modern Epidemiology*. Lippincott Williams & Wilkins, Philadelphia.
- Röver, C., Knapp, F., Friede, T. (2015). Hartung-Knapp-Sidik-Jonkman approach and its modification for random-effects meta-analysis with few studies. *BMC Medical Research Methodology* **15**, 99.
- Rücker, G., Schwarzer, G., Carpenter, J., et al. (2009). Why add anything to nothing? The arcsine difference as a measure of treatment effect in meta-analysis with zero cells. *Statistics in Medicine* **28**, 721-738.
- Salpeter, S. R., Greyber, E., Pasternak, G. A., et al. (2010). Risk of fatal and nonfatal lactic acidosis with metformin use in type 2 diabetes mellitus. *Cochrane Database System Reviews* **4**, CD002967.
- Senn, S. (2007). Trying to be precise about vagueness. *Statistics in Medicine* **26**, 1417-1430.
- Silcocks, P. (2005). An easy approach to the Robins-Breslow-Greenland variance estimator. *Epidemiologic Perspectives & Innovations* **2**, 9.
- Sinclair, J. C., Bracken, M. B. (1994). Clinically useful measures of effect in binary analyses of randomized trials. *Journal of Clinical Epidemiology* **47**, 881-889.
- Sweeting, M. J., Sutton, A. J., and Lambert, P.C. (2004). What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Statistics in Medicine* **23**, 1351-1375.
- Yates, F. (1934). Contingency tables involving small numbers and the chi<sup>2</sup> test. *Supplement of the Journal of the Royal Statistical Society* **1**, 217-235.
- Yusuf, S., Peto, R., Lewis, J., et al. (1985). Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Progress in Cardiovascular Diseases* **27**, 335-371

## Figures and tables

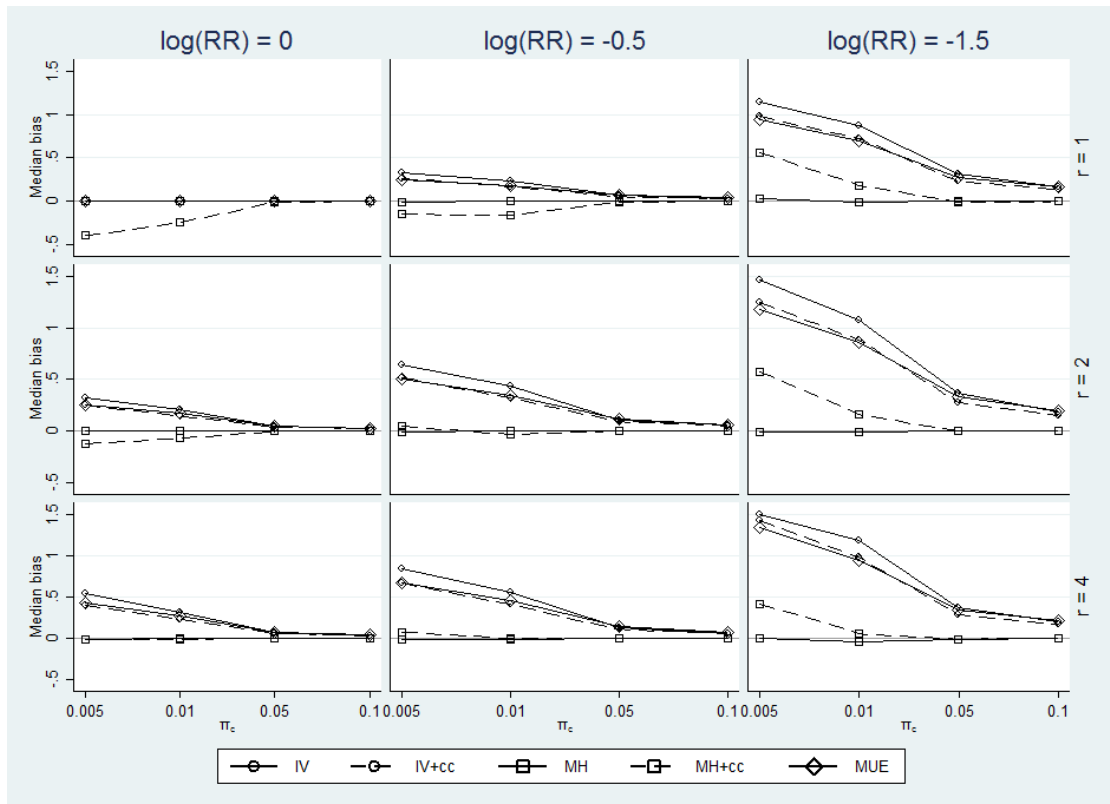


**Figure 1** Biases obtained by the IV, IV+cc, MH, MH+cc, Peto, Peto+cc, MUE and Binomial regression methods when estimating the  $\log(\text{OR})$ . Values were right-truncated at 1.5 to allow a better visual inspection.

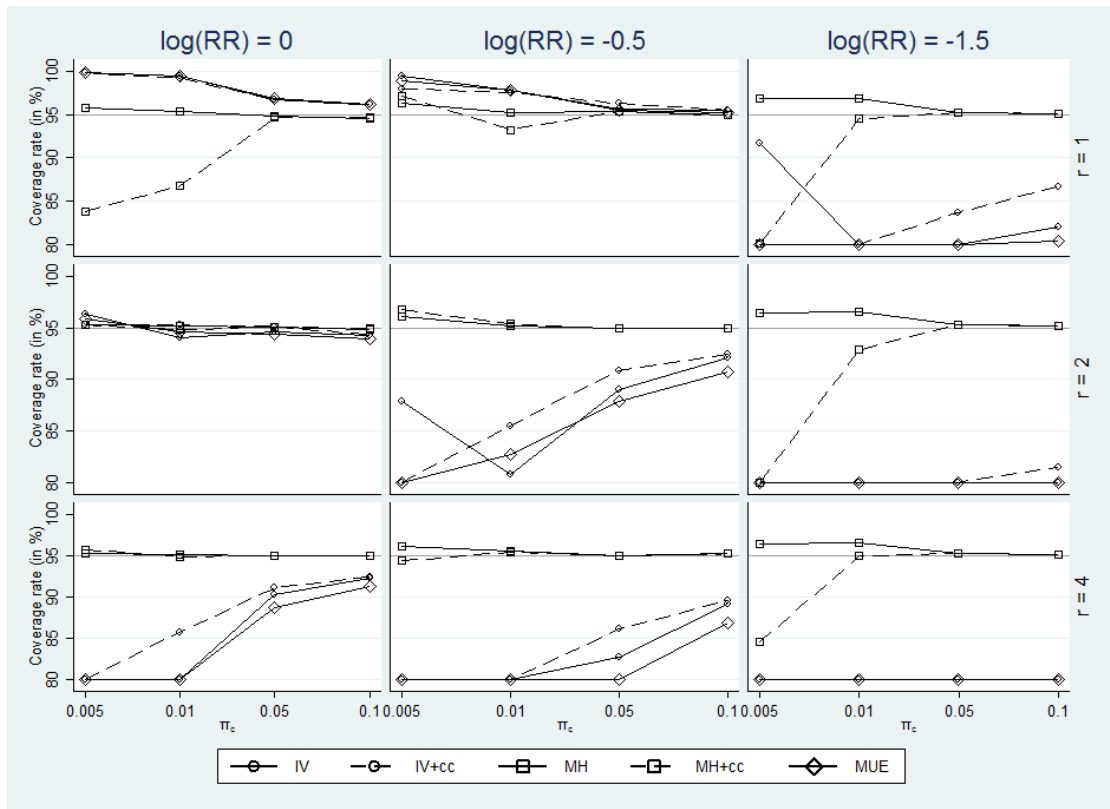




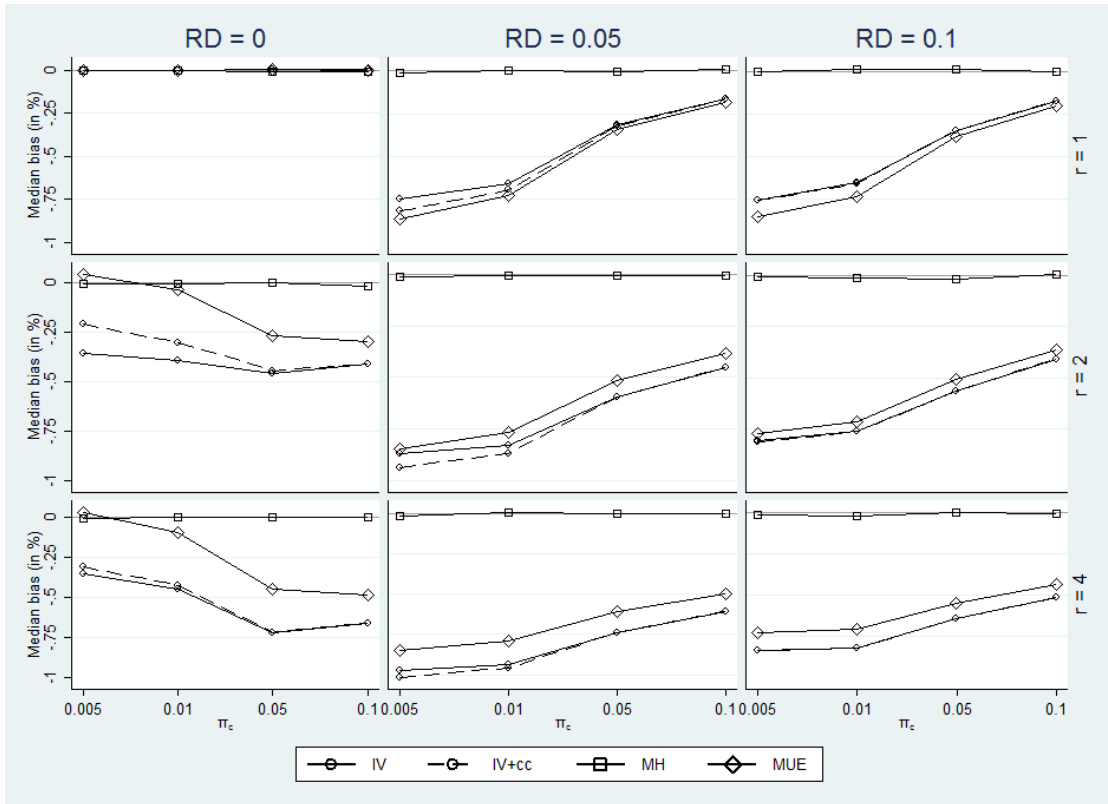
**Figure 2** Coverage rates (in %) obtained by the IV, IV+cc, MH, MH+cc, Peto, Peto+cc, MUE and Binomial regression methods for the  $\log(\text{OR})$ . Values were left-truncated at 80 to allow a better visual inspection.



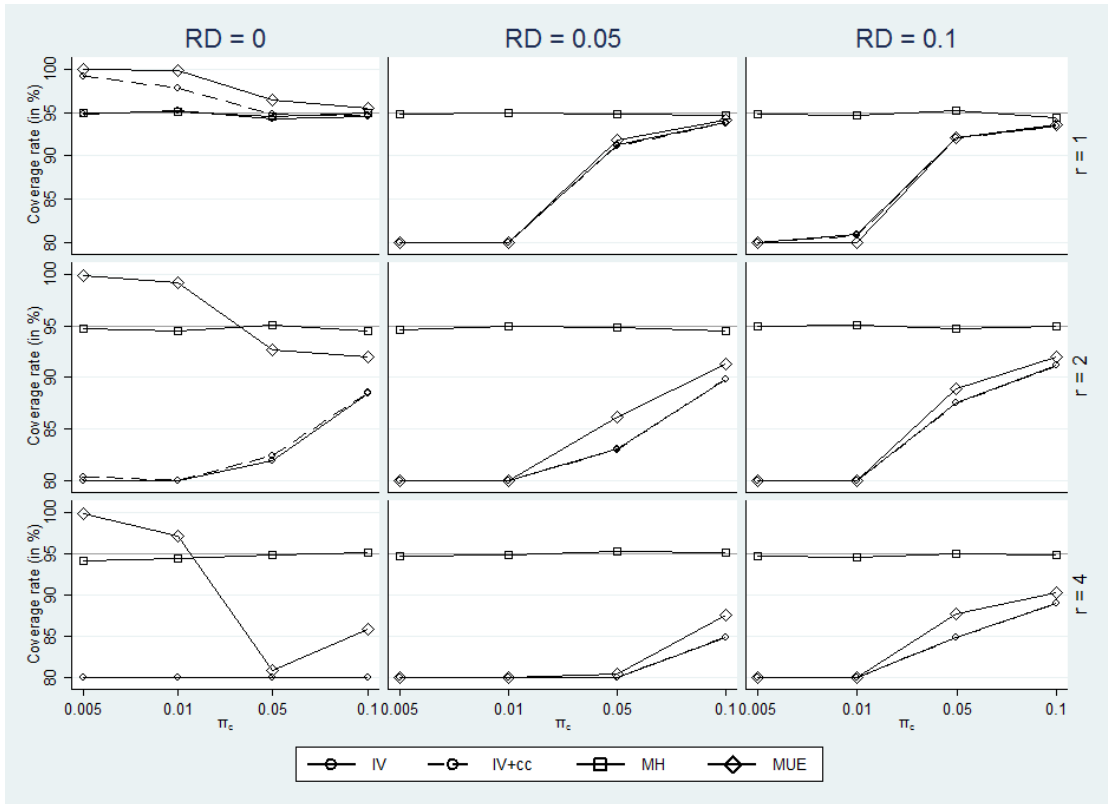
**Figure 3** Biases obtained by the IV, IV+cc, MH, MH+cc and MUE methods when estimating the  $\log(RR)$ . Values were right-truncated at 1.5 to allow a better visual inspection.



**Figure 4** Coverage rates (in %) obtained by the IV, IV+cc, MH, MH+cc and MUE methods for the  $\log(RR)$ . Values were left-truncated at 80 to allow a better visual inspection.



**Figure 5** Biases (in %) obtained by the IV, IV+cc, MH and MUE methods when estimating the RD.



**Figure 6** Coverage rates (in %) obtained by the IV, IV+cc, MH and MUE methods for the RD. Values were left-truncated at 80 to allow a better visual inspection.

**Table 1** Frameworks for the meta-analysis of count data

	Homogeneous treatment effect	Heterogeneous treatment effects
Homogeneous baseline prevalence	Fixed effect (FE) with baseline heterogeneity	Random effects (RE) with baseline homogeneity
Heterogeneous baseline prevalences	FE with baseline heterogeneity	RE with baseline heterogeneity

**Table 2** Consequence of baseline heterogeneity under the FE framework

$\pi_c$	$\pi_t$	RD	RR
0.16	0.26	0.1	1.63
0.33	0.43	0.1	1.30
0.03	0.13	0.1	4.45
0.17	0.27	0.1	1.59
0.07	0.17	0.1	2.39

**Table 3** Strategy applied to deal with zero-event studies

	IV method		MH method		Peto method	
	without cc	with cc	without cc	with cc	without cc	with cc
OR	Discard DZ and SZ studies	Add 0.5 to DZ and SZ studies	No action required <sup>a</sup>	Add 0.5 to DZ studies	No action required <sup>a</sup>	Add 0.5 to DZ studies
RR	Discard DZ and SZ studies	Add 0.5 to DZ and SZ studies	No action required <sup>a</sup>	Add 0.5 to DZ studies	-	-
RD	Discard DZ studies	Add 0.5 to DZ studies	No action required <sup>a</sup>	No cc required <sup>b</sup>	-	-

Note: <sup>a</sup>MH and Peto methods provide estimates that are well defined even in the presence of SZ or DZ studies. <sup>b</sup>The MH method provides RD estimates that include the information contained in both SZ and DZ studies.

**Table 4.** Illustrative dataset (perinatal death in post-term pregnancy)

Trial	Routine induction		Selective induction	
	$n_c$	$X_c$ (Deaths)	$n_t$	$X_t$ (Deaths)
Henry (1969)	57	2	55	0
Cole (1975)	119	0	118	0
Martin (1978)	134	1	131	0
Tylleskar (1979)	55	0	57	0
Breart (1982)	235	0	481	0
Katz (1983)	78	0	78	1
Suikkari (1983)	53	0	66	0
Sande (1983)	90	0	76	0
Cardozo (1986)	207	1	195	0
Augensen (1987)	195	0	214	0
Dyson (1987)	150	1	152	0
Witter (1987)	97	0	103	0
Bergsjo (1989)	94	1	94	0
Egarter (1989)	168	1	188	0
Martin (1989)	10	0	12	0
Heden (1991)	129	0	109	0
Hannah (1992)	1706	2	1701	0
Herabuyta (1992)	51	0	57	0
NICH (1994)	175	0	235	0

**Table 5.** Results of meta-analysis of the illustrative dataset

	Effect size estimate	95% Wald CI
<i>Odds ratio</i>		
IV	.	.
IV+cc	0.56	(0.25; 1.27)
MH	0.11	(0.01; 0.88)
MH+cc	0.35	(0.14; 0.88)
Peto	0.20	(0.06; 0.70)
Peto+cc	0.44	(0.18; 1.03)
MUE	0.52	(0.20; 1.38)
BinReg (with fixed intercept)	0.10	(0.01; 0.81)
BinReg (with random intercept)	0.10	(0.01; 0.81)
<i>Relative risk</i>		
IV	.	.
IV+cc	0.57	(0.25; 1.27)
MH	0.11	(0.01; 0.88)
MH+cc	0.35	(0.14; 0.89)
MUE	0.52	(0.20; 1.38)
<i>Risk difference (values provided in %)</i>		
IV	-0.15	(-0.31; 0.00)
IV+cc	-0.14	(-0.28; 0.00)
MH	-0.20	(-0.36; -0.05)
MUE	-0.11	(-0.27; 0.04)