Title

Bayesian meta-analysis of studies with rare events: Do the choice of prior distributions and the exclusion of studies without events in both arms matter?

Authors:

Soheila Aghlmandi[1,2], Peter Jüni[3], James Carpenter[4], Marcel Zwahlen[1]

[1] Institute of Social and Preventive Medicine (ISPM), University of Bern, Mittelstrasse 43, CH-3012 Bern, Switzerland

[2] Institute for Clinical Epidemiology and Biostatistics, University Hospital Basel, Spitalstrasse 12, CH- Basel, Switzerland

[3] Applied Health Research Centre (AHRC), St. Michael's Hospital, Department of Medicine, University of Toronto, 30 Bond Street, Toronto, Ontario M5B 1W8

[4] London School of Hygiene and Tropical Medicine (LSHTM), University of London, Keppel Street, London WC1E 7HT

Corresponding author:

Soheila Aghlmandi

soheila.aghlmandi@usb.ch

Short title: Bayesian meta-analysis of rare events

Word counts: 4700

Number of figures: 5

Number of tables: 6

Number of references: 30

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27732v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019

Abstract (max 250 words : currently 240)

Randomized controlled trials (RCTs) analyzing serious adverse events often observe low incidence and might even observe zero events in either or both of the treatment and control arms. In the meta-analysis of RCTs of adverse events, it is unclear whether trials with zero events in both arms provide any information for the summary risk ratio (RR) or odds ratio (OR). Studies with zero events in both arms are usually excluded in both frequentist and Bayesian meta-analysis**.** We used a fully probabilistic approach—a Bayesian framework—for the meta-analysis of studies with rare events, and systematically assessed whether exclusion of studies with no events in both arms produced different results compared to keeping all studies in the meta-analysis. We did this by conducting a simulation study in which we assessed the bias in the point estimate of the log(OR) and the coverage of the 95% posterior interval for the log(OR) for different analytical decisions and choices in fixed effect and random effects meta-analysis. We used simulated data generated from a known fixed effect or random effects data scenario (each scenario with a 1000 meta-analysis data-set). We found that the uniform and Jeffrey's prior on the baseline risk in the control group leads to biased results and a reduced coverage, and that setting the prior distribution on the log(odds) scale worked better. We also found nearly identical results regardless of whether studies with no events in both arms were excluded or not.

Keywords: meta-analysis, Bayesian approach, rare events, fixed effect, random effects

1 Introduction

Meta-analysis (MA) combines the results obtained from individual studies, usually randomized controlled trials (RCTs). The main outcome in such trials is often a clinical event, and the studies are powered for comparing the occurrence of that clinical event in the treatment arms. When an MA addresses treatment-associated adverse events, which are usually rare, no events might be observed in one or both arms of an individual trial, and effect measures such as the odds ratio (OR) or relative risk (RR), are undefined for all trials, or are biased [1]. In addition, when events are rare, but not all zero, the standard errors of the effect measures based on normal approximation theory are not robust, which can lead to unreliable statistical inferences.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27732v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019

The problem can be approached in several ways in meta-analysis. One approach is to exclude trials with zero events in one or both arms, which makes it more likely that the magnitude of the pooled treatment effect will be inflated [2]. Some research has pointed out that from an ethical point of view patients in double-zero studies deserve to be included in the analyses [3], while others have argued that such studies may carry information of relative treatment effects through their sample size [4]. Also, using a simulation study [5] showed that excluding studies with no events in both arms for meta-analyses introduced bias into the pooled estimates when there was no true treatment effect.

Another approach uses a continuity correction (CC) of 0.5 for each cell [6, 7]. Sweeting et al. [8] have proposed different CCs that perform better if the number of patients in the treatment and control groups are severely imbalanced. Based on simulation studies, [4] suggests that deleting trials with no events in either arm or adding CCs can introduce bias to the calculation of effect measure(s).

Various statistical methods have been proposed for using and combining information from trials with no events. A principled approach is to assume that the number of events given n (the number of patients in a treatment group) and the true risk follows a binomial distribution. Kuss [4] used beta-binomial regression methods to make inferences about OR, RR, and risk difference. Kuss's approach assumes that events in the treatment groups are binomially distributed, i.e. the likelihood for the observed events is the binomial distribution, and it can handle studies with no events. Cai et al. [9], proposed a method that uses the idea of conjugacy in the same way as the beta-binomial model. They used Poisson models for both fixed effect (FE) and random effects (RE) MA to make inferences about the RR between two treatment groups. Bohning et al. [10] proposed a Poisson model for RE and concluded that these techniques returned almost the same results as the Mantel-Haenszel (MH) method. Other methods along these lines can be found in serveral other publications [11-19].

Another approach to the MA of rare events is to take a fully probabilistic, Bayesian approach. Here, after the specification of prior distributions for all relevant parameters of the analysis model, the data and application of Bayes's theorem allows obtaining posterior distributions for all relevant parameters [20]. Smith (1995) and Warn [21, 22] showed how to implement a fully Bayesian FE and RE meta-analysis with exact binomial likelihood using WinBUGS. This of course needs a decision about the prior distributions to be used that could reflect expert opinion or be derived from external available information [23], or that could be set to reflect vague prior information. In an MA of rare events, the data contain limited information,

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27732v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019

and the information of the prior distributions is expected to contribute to the posterior distribution. Sweeting et al. [8] investigated, among other approaches, Bayesian inference in the FE meta-analysis in situations with rare events, and concluded that the method provided good coverage in all scenarios investigated. However, they excluded a priori trials with no events in both arms from the MA.

We used a Bayesian approach to conduct the MA of studies with rare events to estimate the odds ratio, more precisely the log of the odds ratio, and specifically assessed the importance of (1) excluding yes or no trials with zero events in both arms, and (2) the choice of priors for the true OR and τ for the heterogeneity in case of RE meta-analyses. We chose the OR as the target effect measure for ease of implementation because it is almost identical to the risk ratio in rare event situations and allows easier model implementation using the logit function. In **Section 2**, we define the statistical model and the different types of priors to be used both in FE and RE meta-analyses. In **Section 3**, we describe a simulation study and the range of scenarios in which we varied assumptions about true OR, the heterogeneity τ in RE standard deviation, the risk in the control group, the total number of patients in treatment and control groups, and the randomization ratio in the studies. In **Section 4**, we present the results of the simulation studies. In **Section 5**, we reanalyze studies on the cardiovascular risk of Rosiglitazone in the treatment of Type II diabetes.

2 Bayesian approach to meta-analysis of studies with rare events

Two approaches can be used to combine study findings:

1) The FE MA assumes that the treatment effect is the same in all of the studies. For FE, we consider that observed variation is caused by sampling variation.

2) The RE MA assumes that there is a variation of the true treatment effect across studies (heterogeneity). Therefore, one makes additional assumptions on how the study-specific treatment effects vary. In the binary case, one commonly assumes that the study-specific $\log(OR_i)$ follow a normal distribution, which then implies that one also estimates the standard deviation τ of this normal distribution [24]. No less than 16 methods have been identified to estimate τ or τ-squared [25]. In situations with rare events, it is particularly challenging to estimate τ and the choice of the prior distributions for τ is expected to be important.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27732v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019

2.1 Model structure for the meta-analysis

Throughout, we assume that data for each individual study $i = (1, \ldots, n)$ in the MA come from a two-arm randomized trial comparing a new treatment (received by the treatment group *t*) with a control treatment (received by the control group *c*) and that the outcome assessed in the MA is a binary adverse event. The numbers of events for *c* and *t* groups in each study i then follow a binomial distribution

$$x_{ic} \sim Binomial(p_{ic}, n_{ic}) \quad (1)$$

$$x_{it} \sim Binomial(p_{it}, n_{it}) \quad (2)$$

where n_{ic} and n_{it} are the total number of patients and p_{ic} and p_{it} the true risks in study i in the control and treatment groups. For the OR of each study *i,* we then have

$$OR_i = \left(\frac{p_{it}}{1-p_{it}} \Big/ \frac{p_{ic}}{1-p_{ic}}\right) \quad (3)$$

which can be rewritten as $logit(p_{it}) = log(OR_i) + logit(p_{ic})$, in which the logit function is $logit(p) = \log(\frac{p}{1-p})$. In the FE model, the true treatment effect is assumed to be identical in all studies to be meta-analyzed, i.e. $log(OR_i) = log(OR)$.

For the Bayesian approach, prior distributions must be specified for all unknown parameters, i.e. $log(OR_i)$ as well as the p_{ic} in FE MA. We chose an extremely vague prior distribution for log(OR) in the form of normal distribution with mean of zero and standard deviation (SD) of 10.

2.1.1 Prior distributions for a risk of the control group

For p_{ic}, we studied three different ways of defining the prior distributions: (a) use of a prior distribution that is conjugate to the binomial likelihood, (b) use of independent weakly informative distribution on the $\boldsymbol{logit(p_{ic})}$, and (c) allowing for hierarchical structure among the p_{ic}.

a. **Conjugate prior on $\boldsymbol{p_{ic}}$:** Due to mathematical convenience, one often chooses a prior distribution that is conjugate to the likelihood [26]. For the binomial likelihood, these are beta distributions with shape parameters a and b (defined in **Table 1**).
b. **Weakly informative prior on $\boldsymbol{logit(p_{ic})}$:** Using normal prior distributions on the logit scale has been proposed and used in previous studies [8, 21, 22, 27]. Therefore, we used a

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27732v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019

normal distribution with a mean of zero and SDs of 10 and 100 (precisions of 0.01 and 0.0001). To cover very small baseline risks, it seems reasonable to use these values for SDs. We also used uniform distribution with range of 20, which, when back transformed to the risk scale, has a substantial mass close to zero, but is bounded away from zero at 2×10^{-9}.

Table 1. List of prior distributions for $\boldsymbol{p_{ic}}$

Parameter		Prior distribution
a.	$\boldsymbol{p_{ic}}$	$beta(1,1)$
		$beta(0.5,0.5)$
b.	$\boldsymbol{logit(p_{ic})}$	$unif(-10,10)$
		$normal(0,10)$
		$normal(0,100)$
c.	$\boldsymbol{logit(p_{ic})}$*	$normal(\mu,\sigma)$ where $\mu \sim unif(-6,-3)$, $\sigma \sim unif(0,1)$

* hierarchical structure on $\boldsymbol{logit(p_{ic})}$, $\boldsymbol{i = 1, 2, \dots, n}$

c. **Hierarchical structure on prior for $\boldsymbol{logit(p_{ic})}$:** In the hierarchical model, we assume that multiple parameters of interest are drawn from the same common distribution. In this case, the $logit(p_{ic})$ come from a normal distribution with an unknown mean (μ) and standard deviation (σ). In addition to this structural assumption, one needs to specify prior distributions for both the mean (μ) and the standard deviation (σ). To reflect a rare events situation, we chose a uniform distribution U(-6 to -3) for μ and U(0 to 1) for σ. These specifications provide a 95% prior interval of 0.16% to 7.0% for the risk in the control group.

2.2 Additional model structure and assumptions for the RE MA

In the FE MA, we assumed a common true $\log(OR)$ for all studies. In the RE MA we assume that the true $\log(OR_i)$ from a normal distribution with mean $\log(OR)$ and a standard deviation (τ) which quantifies between-study heterogeneity [24]. We have

$$\log(OR_i) \sim normal(mean(\log(OR)),\ \tau), \quad i = 1, 2, \dots, n \qquad (5).$$

We specified a $normal(mean = 0,\ sd = 10)$ distribution as the prior distribution for $\log(OR)$ and investigated several prior distributions for τ as given in **Table 2**. Because it is

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27732v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019

175 particularly challenging to estimate τ in situations with rare events, we expected the
176 specification of the prior distributions for τ to be important. Working on the $\log(OR_i)$ implies

Table 2. List of prior distributions for τ

Parameter	Prior distribution	Mean
τ	$exp(2)$	0.5
	$unif(0, 2)$	1
	half-normal	0.5
	$lognormal(-4.07, 1.45^2)$	-4.07

177 that a τ of 0.5 to 1.0 already reflects large heterogeneity of the treatment effects across
178 studies, as discussed in Spiegelhalter [26]. Therefore, we set two prior distributions to have a
179 mean of 0.5,and a third, the uniform(0, 2), had a mean of 1. Finally, we used one of the prior
180 distributions suggested by Turner et al. [28], $lognormal(\text{-}4.06, 1.45^2)$, for τ^2.

181 In the RE MA, we investigated a subset of the prior distributions for logit(p_{ic}) we used in the
182 FE MA.
183

184 3 Simulations

185 3.1 Data generation scenarios

186 We conducted a simulation study to assess coverage of the 95% CIs and bias for $\log(OR)$
187 estimates. For the data simulations, we defined the following scenarios:

188 ▪ **Size of true $\boldsymbol{log(OR)}$:** For both FE and RE scenarios, we assessed scenarios with log(1)
189 and log(2).
190 ▪ **Statistical heterogeneity ($\boldsymbol{\tau}$):** For RE scenarios, we used τ of 0.2 and 0.5 for both sizes
191 of $\log(OR)$.
192 ▪ **The ratio of group sizes:** In MA of rare events, the imbalance between study groups can
193 make it hard to calculate effect measures. We therefore simulated data scenarios for 1:1
194 randomization of treatment vs. control groups, and for 1:2 and 1:4 randomizations. To
195 obtain higher proportions of zeros in both arms for ratios of 2 and 4, the values for p_{ic} and
196 n_{it} were set to smaller values than in the 1:1 randomization.
197 ▪ **Probability of control group ($\boldsymbol{p_{ic}}$):** We let the event risk in the control group vary
198 between 0.1% and 4%. Decreasing the probability of events in p_{ic} increases the
199 proportion of zeros in both arms. When more information is added to the control group
200 (e.g., ratio 1:2) the probability of events and the total number of patients in the control
201 group should be smaller to achieve trials with more zeros in both arms.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27732v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019

- **Percentage of trials with no events in both arms:** To assess the impact of the sparseness of data on coverage and bias for different specifications of the prior distributions, we varied the zeros in both arms from 5% to 65%. A high percentage of zeros in both arms indicates lower probability for p_{ic} and a smaller total number of patients in the treatment group.
- **Number of studies per MA:** We used a uniform distribution to vary the number of studies in each MA (**Table 3**).
- **Sample size of a single study:** We also used a uniform distribution to simulate the sample size of each study. Table 3 summarizes the values we used to simulate different scenarios of MA data sets.

Table 3. Parameter values used in the simulation of MA data sets

FE scenarios	
$\log(OR)$	0 or 0.69
Number of patients in treatment group (n_{it})	[20, 60]
Risk of control group (p_{ic})	[0.001, 0.04]
Number of trials in each MA	10, 20 or 50
RE scenarios	
$\log(OR_i)$*	
$\log(OR)$	0 or 0.69
Random effects standard deviation (τ)	0.2 or 0.5
Number of patients in treatment group (n_{it})	[10, 60]
Risk of control group (p_{ic})	[0.001, 0.035]
Number of trials in each MA	20 or 50
Both FE & REs scenarios	
Ratio of group sizes**	1:1, 1:2 or 1:4
Number of simulated MA data sets	1000

* follows a normal distribution with specified characteristics
** We assigned treatment vs. control group for the ratio of group sizes

When we combined all the above design factors, our simulation scenarios totaled 144 (Supplementary Tables S2-S7). The simulations were carried out with 1000 data sets for MA per scenario. Then we appended scenarios with less than or equal to 30% zeros in both arms vs. scenarios with more than 30% zeros based on the randomization ratio to calculate bias (bias = median of the 1000 estimated $\log(OR)$ – true $\log(OR)$). We obtained 95% CIs for the estimated $\log(OR)$ from the 2.5 and the 97.5 percentile of the posterior distribution and calculated the 95% coverage of true log(OR) by the proportion of times the 95% CI included the true log(OR). We summarize the results in detail from different perspectives in the figures and tables. We excluded MA data sets where all the generated studies had no events in either treatment or control group, i.e. no events across all studies. As a comparator to frequentist

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27732v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019

analysis methods, we used the MH method without any CC, which was identified in different publications as a robust method for sparse events MA [8] [10].

3.2 Software and implementation issues

There are no closed solutions for calculuating the posterior distributions for the analysis of the models we defined. We therefore used numerical simulation methods, in particular the Markov chain Monte Carlo (MCMC) method, to approximate the posterior distributions of the parameters of interest as implemented in the "Just Another Gibbs Sampler" (JAGS) software package, another variant of the BUGS language [29].

The necessary data simulations were implemented in R (R Core Team, http://www.R-project.org/) and called JAGS (http://mcmc-jags.sourceforge.net/) from within R using the jags function of the R2jags package. We used 4 chains and set 15,000 iterations with the first 5,000 simulated values as burn-in. We used Gelman and Rubin's diagnostic to check the convergence of multiple MCMC chains run in parallel. Details of the R and JAGS codes are provided in the supplementary documentation.

4 Results

We report simulation results separately for each effect measure and for the different standard deviations of RE. To avoid overloading this account, we present figures only of studies with no events in both arms that were included in the analyses. The results of FE for the conjugate family of priors and RE with $unif(0, 2)$, $exp(2)$ for τ, and, $log\text{-}normal(\text{-}4.06, 1.45^2)$ for τ^2 are in the supplementary documentation.

1. For FE scenarios

a. The family of conjugate priors showed increased bias and reduced coverage, but coverage improved when information in the control group increased (ratio 2,4), and estimates for true log(OR) were less biased (**Table S8**). Estimates of $beta(0.5, 0.5)$ in all the scenarios were less biased and had better coverage than $beta(1, 1)$. Excluding studies with zeros in both arms did not affect coverage or bias for true log(OR).

b. The weakly informative priors reached an average coverage of 94.6%, and bias showed a small negative change of the true log(OR). Almost all priors performed similarly for null effect and log(2). For different ratios, when we increased the proportion of zeros in both

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27732v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019

arms, bias increased slightly in a negative direction, but coverage was roughly the same. The uniform and normal distribution with SD of 100 behaved similarly with respect to both coverage and bias. Normal distribution with smaller SD (10) showed a small drop in coverage and an increase in bias (on average –0.05).

c. For the hierarchical structure, coverage slightly increased to 94.8%, but the bias was the same. When we compared performance of Bayesian methods for log(2) to performance for log(1), we found very similar coverage and estimates. The Bayesian method showed a slight increase in estimates of true effect measures when there was an imbalance in the ratios for log(2).

In general, for all the Bayesian methods, if studies with no events are excluded results for 95% coverage and bias are almost identical. Bayesian methods provide good coverage on average of 94.5%, slightly less than MH 96%. However, these methods are slightly biased from true log(OR). For log(1), null effect, in all the scenarios, the Bayesian machinery ran into difficulty calculating true log(OR), especially for scenarios that only included 10 studies in each MA data set. By increasing the information in the control group, although coverage improved slightly we observed an increase in bias (**Table 4**).

2. <u>For RE scenarios</u>

a. For τ ~ unif (0,2)

Average coverage for Bayesian methods was around 95% for both moderate and high heterogeneity, but for MH the coverage dropped for high heterogeneity to 92% on average. For the scenarios with under 30% zeros in both arms, the coverage decreased to 89.5% for MH, while for all the Bayesian methods it stayed around 94%. By increasing the information on the control group, the coverage dropped to 93% and the bias increased in the negative direction. For $\tau = 0.5$, the observed coverage was lower for 1:1 randomization than $\tau = 0.2$ but similar to the other randomization scenarios. In summary, uniform distribution is a poor choice to account for heterogeneity in RE MA due to high bias from true log(OR).

b. For $\tau^2 \sim lognormal(-4.06, 1.45^2)$

The mean coverage for log(OR) was similar for all the specified priors for p_{ic}, but different for scenarios with higher true heterogeneity $\tau = 0.5$, on average 93.5% and 85%, respectively. Bias was smaller for $\tau = 0.2$ than $\tau = 0.5$ for both true log(OR). Both mean coverage and bias were similar for low or high proportions of zeros in both arms irrespective of true log(OR). For different randomization scenarios (1:1,

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27732v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019

Table 4. 95% coverage and bias for different scenarios of FE MA for log(OR) = 0 and log(OR) = 0.69

Prior for $logit(p_{ic})$	Ratio [a]	Deletion [b]	Coverage	Bias	Coverage	Bias	Coverage	Bias	Coverage	Bias	Gel. & Rub. Statistic [c]
			log(OR) = 0				log(OR) = 0.69				
normal(0, 10)			≤ 30% [d]		>30%		≤ 30%		>30%		
	1:1	0	0.941	-0.033	0.945	-0.071	0.952	0	0.949	-0.045	1.0015
	1:2	0	0.947	-0.047	0.942	-0.083	0.946	-0.015	0.942	-0.045	1.0013
	1:4	0	0.942	-0.089	0.953	-0.093	0.951	-0.025	0.943	-0.044	1.0012
	1:1	1	0.942	-0.032	0.944	-0.071	0.953	-0.002	0.950	-0.044	1.0015
	1:2	1	0.942	-0.047	0.941	-0.082	0.946	-0.015	0.943	-0.045	1.0013
	1:4	1	0.942	-0.088	0.952	-0.094	0.951	-0.026	0.943	-0.044	1.0012
normal(0, 100)											
	1:1	0	0.939	0.002	0.939	0	0.942	0.057	0.940	0.062	1.0017
	1:2	0	0.948	-0.019	0.945	-0.034	0.943	0.028	0.939	0.023	1.0013
	1:4	0	0.945	-0.058	0.951	-0.063	0.951	0.006	0.939	0.015	1.0012
	1:1	1	0.939	0.003	0.939	0.001	0.942	0.056	0.939	0.059	1.0017
	1:2	1	0.948	-0.020	0.945	-0.033	0.944	0.027	0.940	0.023	1.0013
	1:4	1	0.944	-0.058	0.953	-0.063	0.952	0.006	0.940	0.012	1.0012
unif(-10, 10)											
	1:1	0	0.940	0.001	0.940	-0.004	0.945	0.056	0.942	0.051	1.0015
	1:2	0	0.946	-0.020	0.946	-0.036	0.944	0.027	0.940	0.019	1.0013
	1:4	0	0.945	-0.059	0.953	-0.064	0.952	0.006	0.940	0.010	1.0012
	1:1	1	0.939	0.002	0.940	-0.004	0.944	0.054	0.944	0.053	1.0015
	1:2	1	0.948	-0.020	0.946	-0.036	0.944	0.026	0.940	0.019	1.0013
	1:4	1	0.944	-0.057	0.951	-0.064	0.952	0.004	0.939	0.009	1.0012
Hierarchical											
	1:1	0	0.945	0.024	0.945	0.015	0.949	0.044	0.947	0.029	1.0128
	1:2	0	0.950	-0.016	0.947	-0.032	0.946	0.017	0.941	0.010	1.0075
	1:4	0	0.945	-0.057	0.954	-0.062	0.952	0.001	0.942	-0.008	1.0039
	1:1	1	0.945	0.023	0.944	0.016	0.947	0.042	0.946	0.031	1.0129
	1:2	1	0.949	-0.015	0.947	-0.032	0.946	0.019	0.942	0.009	1.0074
	1:4	1	0.946	-0.057	0.953	-0.063	0.953	0.001	0.943	-0.006	1.0039
Mantel-Haenszel											
	1:1	0	0.957	0.008	0.974	0.005	0.962	0.034	0.963	0.028	NA
	1:2	0	0.962	-0.003	0.970	-0.037	0.959	0.005	0.962	0.008	NA
	1:4	0	0.970	-0.063	0.963	-0.046	0.964	-0.017	0.961	-0.012	NA

[a] We assigned treatment vs. control group for the ratio of group sizes
[b] deletion is a logical argument; zero means trials with zero in both arms are excluded from the analyses.
[c] The Gelman and Rubin diagnostic is used to check the convergence of multiple mcmc chains run in parallel.
[d] Percentage of trials with no events in both arms.

288

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27732v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019

1:2, or 1:4), neither coverage nor bias changes for different priors for p_{ic}. When we excluded studies with zero events in both arms the results were almost identical. It is clear that lognormal as a prior for τ^2 returns better coverage and a less biased result than true log(OR), but the result can be further improved.

c. For τ ~ half-normal(mean = 0.5)

For all the scenarios with small to moderate heterogeneity for both true log(OR)s, coverage returned by the Bayesian methods was above 94% and there wass no specific pattern of increase or decrease when we had imbalanced randomization. In contrast, bias increased towards the negative by putting more information in the control group. The coverage was lower, 93% on average, for high heterogeneity (0.5) and the estimates were biased for true log(OR) with no specific direction. There was a clear pattern of increase in the coverage when we had more than 30% zeros in both arms for 0.5 heterogeneity scenarios (**Table 5** and **Table 6**).

Results for τ ~ exp(2) were very similar to τ ~ half-normal(mean = 0.5) in all the aspects (Figures S3 and S4, Tables S11 and S12).

In general, for all the RE Bayesian methods in the different data scenarios, the average coverage and bias were almost identical whether studies with no events were included or excluded. Bayesian methods provide good coverage of 94% on average, slightly higher than coverage when using the MH method, 92.6%, but both methods have a slight bias of the point estimate for the true log(OR). For log(1), null effect, bias was surprisingly large, especially for the scenarios in which there was high heterogeneity (0.5). By increasing the information in the control group, we observed an increase in bias, but coverage remained similar. As the proportion of zeros in the data increased, the hierarchical model with half-normal prior for τ showed better coverage and gave a less biased estimate compared to using a uniform distribution for τ. Estimates from the MH method displayed evidence of bias and poor coverage because the method was unable to account for heterogeneity when the standard deviation in the RE data generation scenario was high (0.5).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27732v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019

Table 5. 95% coverage and bias for different scenarios of REs MA log(OR) = 0 for τ ~ half-normal (mean = 0.5)

Prior for $logit(p_{ic})$	Ratio [a]	Deletion [b]	Coverage	Bias	Coverage	Bias	Coverage	Bias	Coverage	Bias	Gel. & Rub. Statistic [c]
			τ = 0.2				τ = 0.5				
			≤ 30% [d]		>30%		≤ 30%		>30%		
normal(0, 10)											
	1:1	0	0.949	-0.014	0.953	-0.043	0.927	0.082	0.946	0.074	1.0065
	1:2	0	0.945	-0.050	0.945	-0.111	0.935	0.038	0.935	-0.040	1.0074
	1:4	0	0.939	-0.135	0.957	-0.162	0.937	-0.007	0.955	-0.052	1.0095
	1:1	1	0.949	-0.010	0.954	-0.047	0.927	0.082	0.945	0.072	1.0066
	1:2	1	0.945	-0.053	0.944	-0.109	0.935	0.038	0.935	-0.040	1.0070
	1:4	1	0.942	-0.137	0.953	-0.160	0.938	-0.008	0.953	-0.051	1.0093
normal(0, 100)											
	1:1	0	0.947	0.029	0.948	0.019	0.917	0.127	0.932	0.137	1.0067
	1:2	0	0.949	-0.023	0.942	-0.057	0.933	0.065	0.934	0.020	1.0075
	1:4	0	0.943	-0.107	0.954	-0.122	0.937	0.015	0.951	-0.011	1.0092
	1:1	1	0.950	0.029	0.948	0.021	0.917	0.125	0.933	0.136	1.0066
	1:2	1	0.949	-0.026	0.943	-0.060	0.933	0.065	0.934	0.022	1.0074
	1:4	1	0.941	-0.103	0.956	-0.121	0.938	0.013	0.953	-0.012	1.0093
Hierarchical											
	1:1	0	0.943	-0.023	0.950	-0.032	0.928	0.067	0.938	0.073	1.0187
	1:2	0	0.942	-0.055	0.945	-0.093	0.932	0.025	0.936	-0.019	1.0140
	1:4	0	0.941	-0.127	0.953	-0.152	0.939	-0.016	0.953	-0.048	1.0121
	1:1	1	0.943	-0.021	0.948	-0.031	0.930	0.071	0.939	0.072	1.0187
	1:2	1	0.941	-0.053	0.941	-0.095	0.932	0.021	0.936	-0.019	1.0141
	1:4	1	0.940	-0.128	0.955	-0.147	0.941	-0.017	0.957	-0.046	1.0119
Mantel-Haenszel											
	1:1	0	0.955	0.027	0.959	0.031	0.902	0.125	0.946	0.129	NA
	1:2	0	0.944	0.007	0.957	0	0.894	0.107	0.948	0.068	NA
	1:4	0	0.955	-0.020	0.959	0.011	0.898	0.087	0.938	0.116	NA

[a] We assigned treatment vs. control group for the ratio of group sizes
[b] deletion is a logical argument; zero means trials with zero in both arms are excluded from the analyses.
[c] The Gelman and Rubin diagnostic is used to check the convergence of multiple mcmc chains run in parallel.
[d] Percentage of trials with no events in both arms.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27732v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019

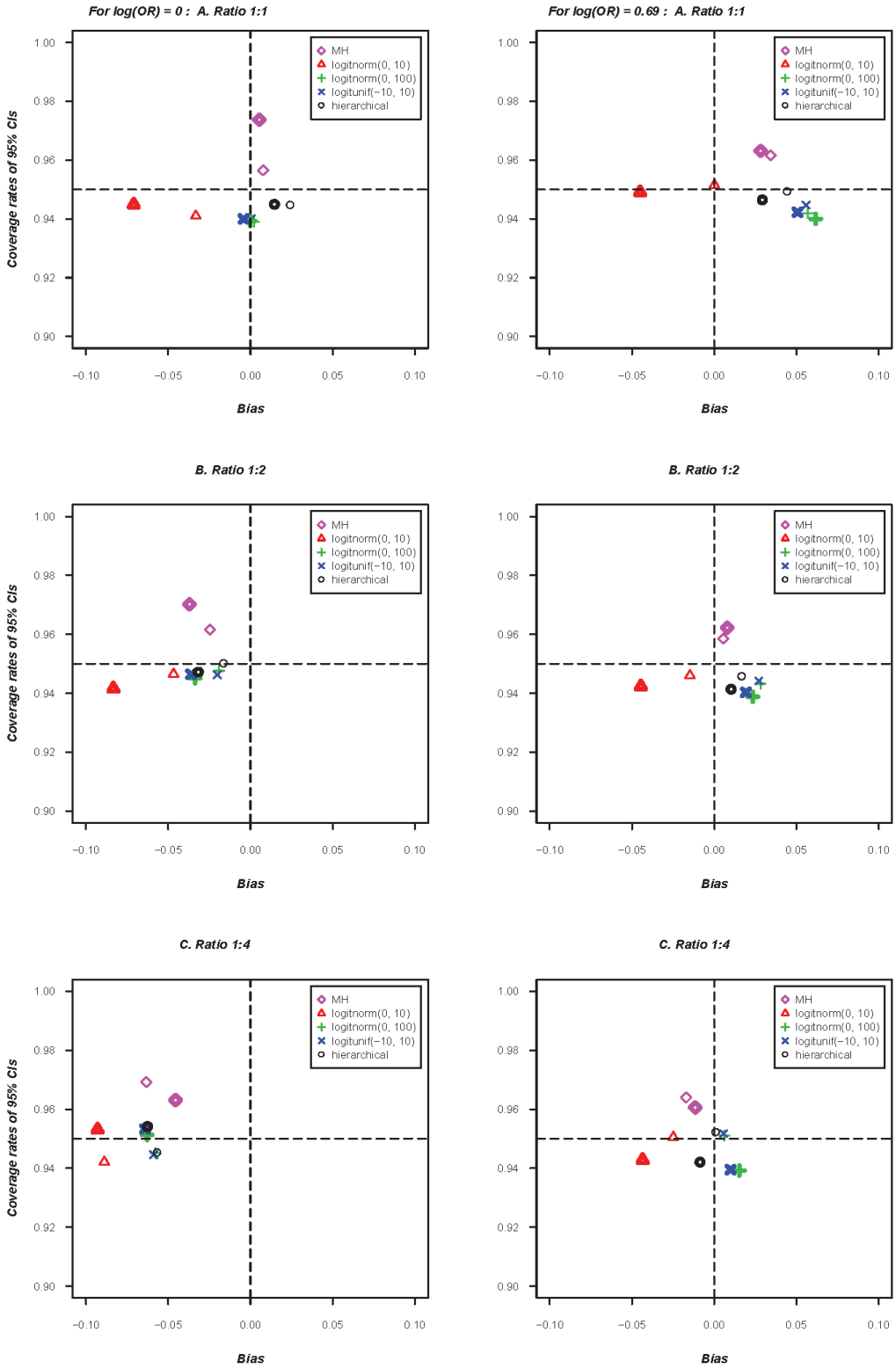
Table 6. 95% coverage and bias for different scenarios of REs MA log(OR) = 0.69 for $\tau \sim$ *half-normal (mean = 0.5)*

Prior for $logit(p_{ic})$	Ratio [a]	Deletion [b]	Coverage	Bias	Coverage	Bias	Coverage	Bias	Coverage	Bias	Gel. & Rub. Statistic [c]
			τ = 0.2				τ = 0.5				
			≤ 30% [d]		>30%		≤ 30%		>30%		
normal(0, 10)											
	1:1	0	0.956	0.023	0.960	0.020	0.918	0.128	0.951	0.098	1.0084
	1:2	0	0.947	0.002	0.948	-0.036	0.928	0.090	0.939	0.073	1.0071
	1:4	0	0.946	-0.037	0.944	-0.085	0.933	0.042	0.934	0.027	1.0071
	1:1	1	0.957	0.024	0.961	0.017	0.920	0.127	0.949	0.100	1.0087
	1:2	1	0.948	0.003	0.947	-0.036	0.929	0.091	0.937	0.072	1.0070
	1:4	1	0.944	-0.037	0.943	-0.085	0.933	0.043	0.939	0.028	1.0069
normal(0, 100)											
	1:1	0	0.940	0.083	0.946	0.128	0.895	0.189	0.912	0.209	1.0087
	1:2	0	0.943	0.038	0.944	0.039	0.912	0.132	0.920	0.150	1.0069
	1:4	0	0.945	-0.006	0.945	-0.036	0.983	0.074	0.927	0.081	1.0073
	1:1	1	0.940	0.086	0.941	0.130	0.896	0.192	0.914	0.216	1.0089
	1:2	1	0.944	0.039	0.941	0.041	0.914	0.132	0.920	0.148	1.0067
	1:4	1	0.944	-0.010	0.947	-0.034	0.928	0.073	0.925	0.080	1.0074
Hierarchical											
	1:1	0	0.945	0.005	0.949	0.027	0.923	0.083	0.934	0.111	1.0390
	1:2	0	0.939	-0.018	0.941	-0.025	0.929	0.059	0.930	0.079	1.0198
	1:4	0	0.939	-0.056	0.942	-0.087	0.932	0.018	0.936	0.026	1.0120
	1:1	1	0.948	0.004	0.949	0.026	0.922	0.087	0.933	0.111	1.0380
	1:2	1	0.943	-0.017	0.942	-0.026	0.927	0.057	0.931	0.074	1.0195
	1:4	1	0.939	-0.056	0.940	-0.086	0.933	0.019	0.935	0.020	1.0121
Mantel-Haenszel											
	1:1	0	0.951	0.037	0.966	0.060	0.909	0.136	0.954	0.153	NA
	1:2	0	0.934	0.024	0.959	0.027	0.895	0.121	0.937	0.131	NA
	1:4	0	0.944	0.008	0.963	0.007	0.895	0.104	0.934	0.106	NA

[a] We assigned treatment vs. control group for the ratio of group sizes
[b] deletion is a logical argument; zero means trials with zero in both arms are excluded from the analyses.
[c] The Gelman and Rubin diagnostic is used to check the convergence of multiple mcmc chains run in parallel.
[d] Percentage of trials with no events in both arms.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27732v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019

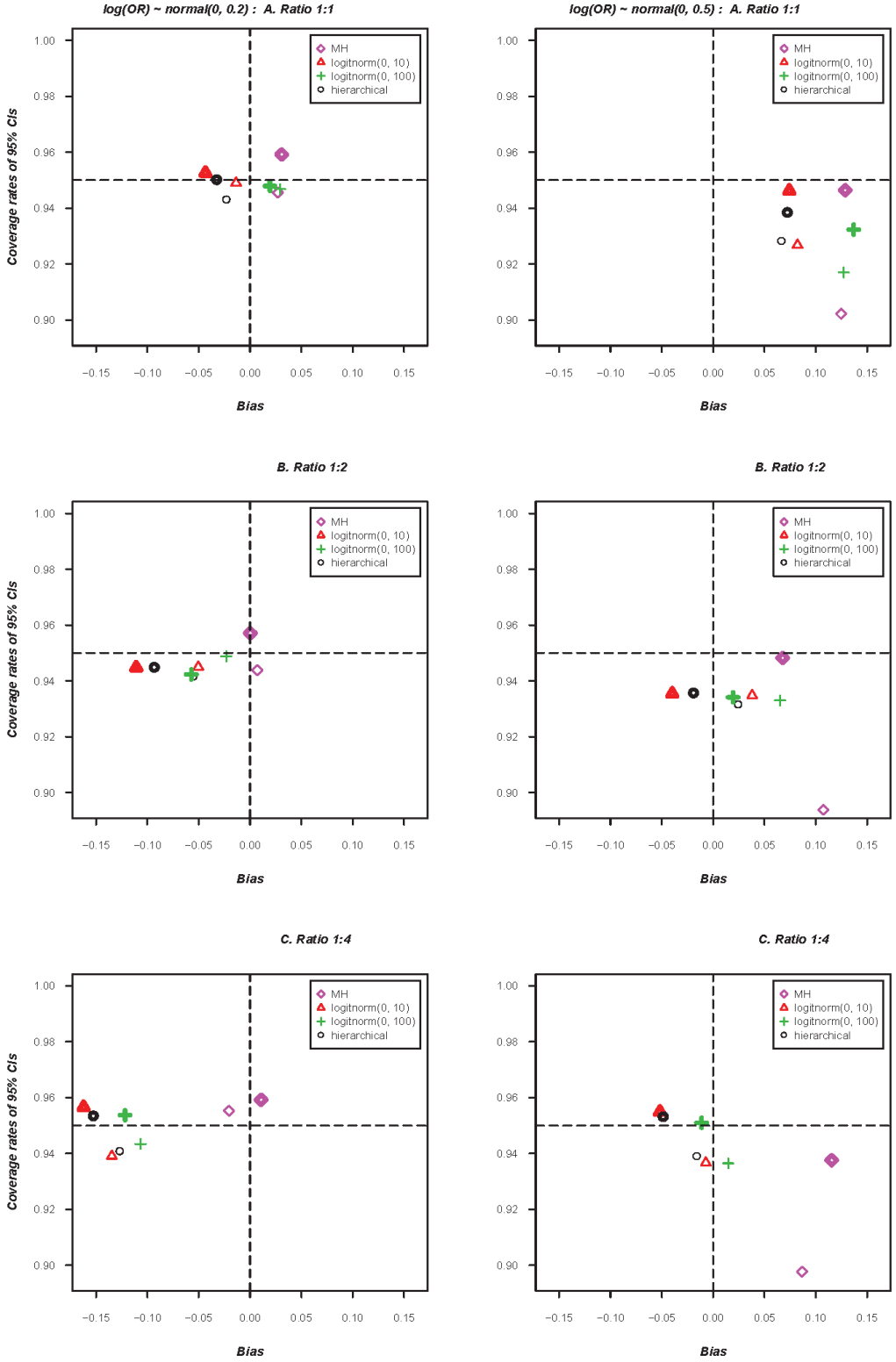
320 **Figure 1** Coverage probability of 95% CIs and bias for $\log(OR) = 0$ and $\log(OR) = 0.69$ estimate for
321 FE method when trials with no events in both arms were included (bold icons in the graph are
322 scenarios with more than 30% in both arms)



323

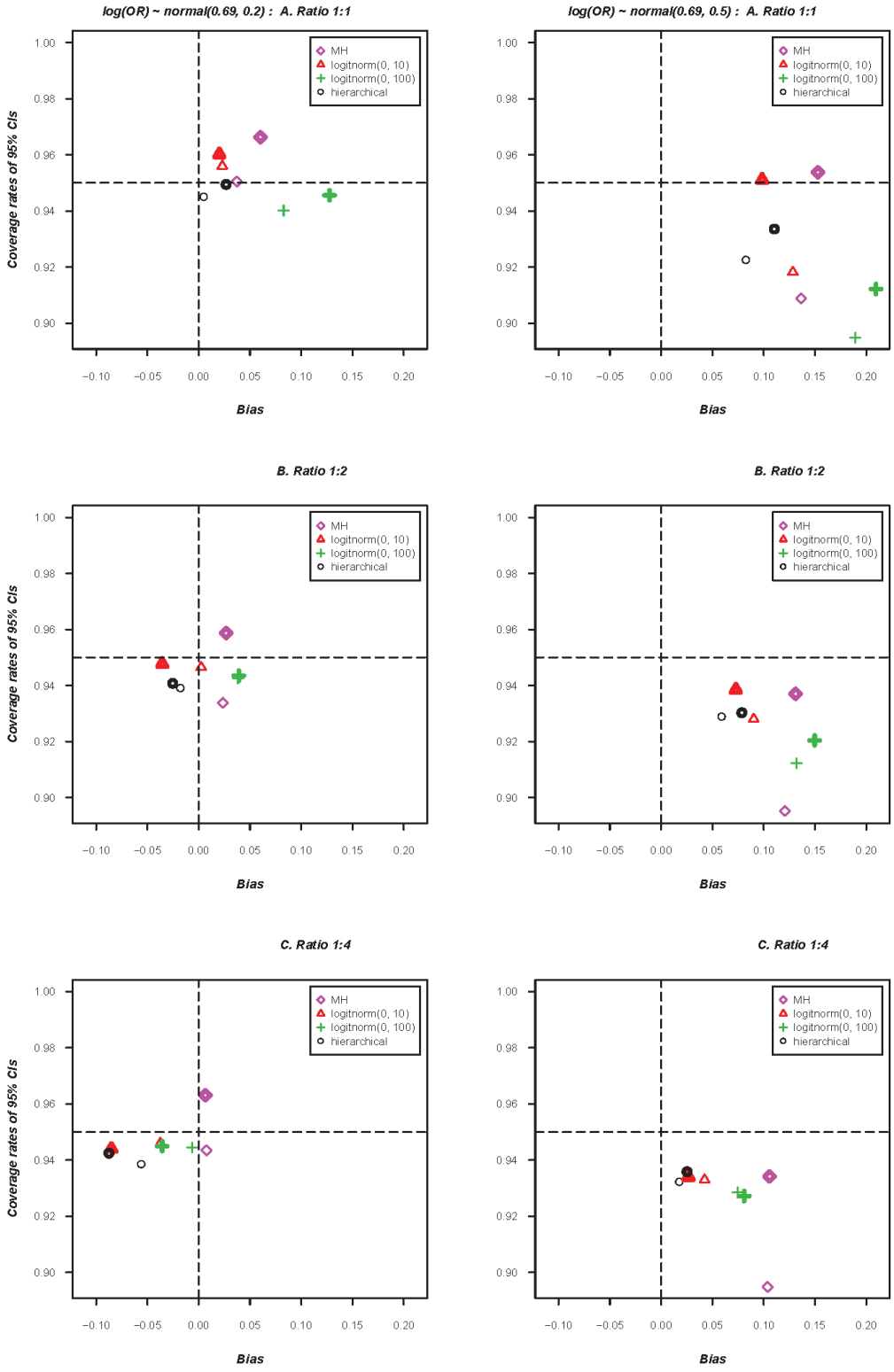
PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27732v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019

Figure 2 Coverage probability of 95% CIs and bias for $\log(OR_i)$ estimate for RE method with $\tau \sim$ *half-normal* (mean = 0.5) for different scenarios of $\log(OR_i) \sim \text{normal}(0, 0.2)$ & $\text{normal}(0, 0.5)$ (bold icons in the graph are scenarios with more than 30% in both arms)



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27732v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019

328 **Figure 3** Coverage probability of 95% CIs and bias for $\log(OR_i)$ estimate for RE method with τ~
329 *half-normal* (mean = 0.5) for different scenarios of $\log(OR_i) \sim \text{normal}(0.69, 0.2)$
330 & $\text{normal}(0.69, 0.5)$ (bold icons in the graph are scenarios with more than 30% in both arms)



331

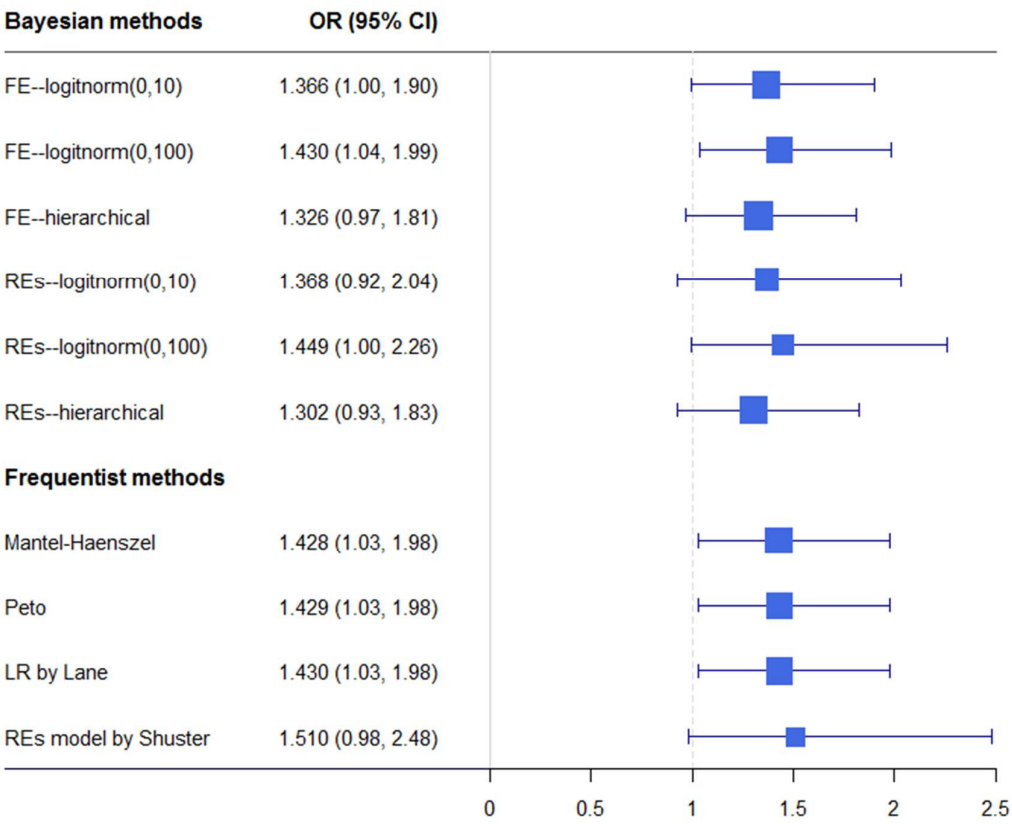
PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27732v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019

5 Illustration of the methods: example of Rosiglitazone

The Bayesian methods are illustrated with data from a meta-analysis of 48 comparative trials that examine the possible cardiac toxicity of Rosiglitazone in RCTs designed to study cardiovascular morbidity and mortality. Rosiglitazone, a Type II diabetes medicine, was introduced in 1999 and is known to reduce blood glucose and glycated hemoglobin levels. Adverse events of Rosiglitazone were studied and categorized as rare events. We used the MA data, which [27] also used. Events are rare for myocardial infarction (MI): 26 trials had zero in one arm, 10 trials had zero in both arms. The rare events problem is more pronounced for cardiovascular (CV) death since 25 studies had no events in both arms, and 17 had one arm with no event (the full data set is in supplemental Table S1). We illustrated the situation with this example using a selection of our Bayesian methods, and compared the results to the MH and Peto methods. We also compared our results with those reported by [11], and logistic regression (LR) by [27].

- For MI as a clinical outcome: Bayesian methods showed small sensitivity to the choice of priors (**Figure 4**).

Figure 4 Forest plot of an MA of Rosiglitazone for MI

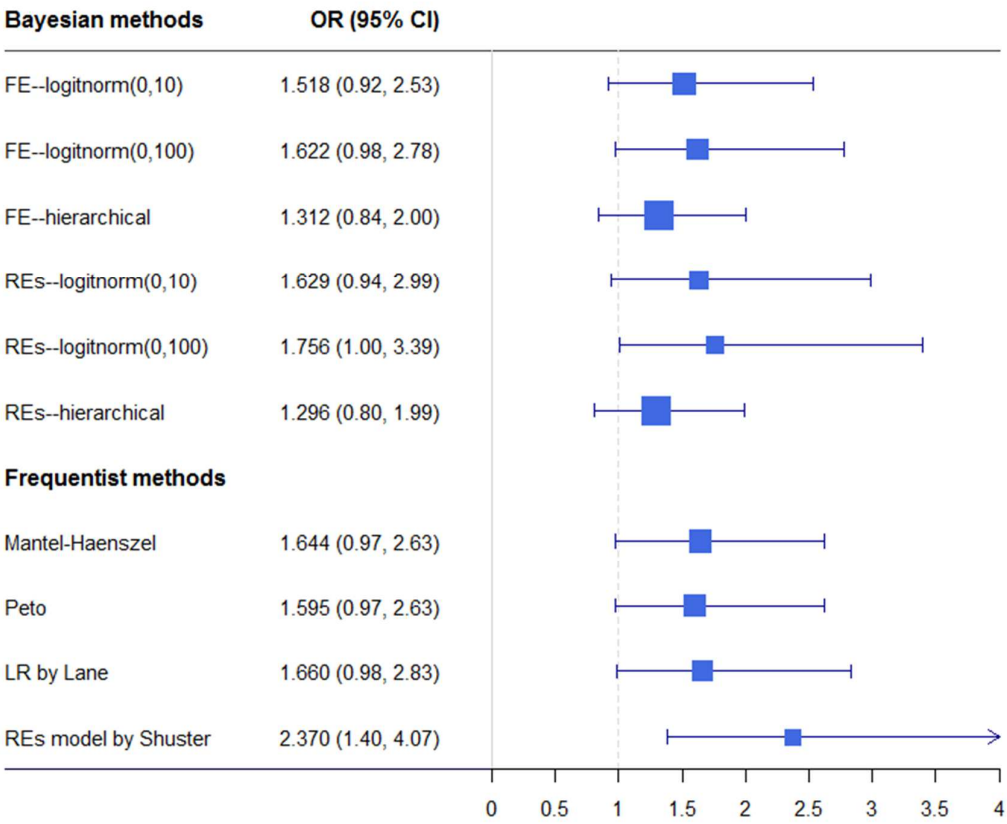


PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27732v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019

In FE, when we used a normal distribution with SD of 100 for the prior distribution of the logit of p_{ic}, the estimated OR was higher (OR = 1.43) than in all the other Bayesian approaches, and results were in line with both the MH and Peto methods (OR = 1.429 and 1.430) and logistic regression applied by [27]. For RE Bayesian, with the same prior for the logit of p_{ic} and a half-normal distribution (mean = 0.5) for the prior distribution of τ, we observed an OR of 1.45, which also was higher than the estimates from the other Bayesian methods. However, when implementing hierarchical prior distributions for the logit of p_{ic} for both FE and RE (τ ~ half-normal [mean = 0.5]) the estimated summary OR was clearly smaller (for FE, OR = 1.30; for RE, OR = 1.33) than in all the other methods. Shuster's RE model estimation is higher than our estimations with wider confidence interval than our CIs.

- Results of a forest plot (**Figure 5**) for CV death: Bayesian methods showed high sensitivity to the choice of priors.

Figure 5 Forrest plot of an MA of Rosiglitazone for CV death



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27732v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019

- In FE, Bayesian approaches' highest OR was 1.62, which is estimated by norm(0, 100) on logit of p_{ic}, and the 95% CI is slightly wider than other priors on baseline risk. We observed the same results for RE Bayesian approaches with the same prior on the risk of control group with half-normal (mean = 0.5) as τ, but the CI is even wider for RE than for FE.

  The MH and Peto effect measures were in line with the FE Bayesian method where we put the normal distribution of SD at 100. RE methods drew the same conclusion, but hierarchical Bayesian for both FE and RE (τ ~ half-normal(mean = 0.5)) seemed more robust for point estimate calculation, and showed more drastic change in the size of the effect measure than any other method. ORs of MH, and Peto and Lane's LR are very similar to norm(0, 100) on logit of p_{ic}. Shuster's RE model has the highest OR = 2.37 and also the widest 95% confidence interval.

  The high sensitivity to the choice of priors in CV death of Bayesian methods can be explained due to very low event rate, 0.5%, while for MI it is almost 2%.

6 Discussion

Conducting a meta-analysis of RCTs for rare but clinically relevant adverse events needs to be done with care. Different frequentist and fully probabilistic Bayesian approaches have been proposed and the results obtained seem to depend on the approach chosen [4, 8, 10, 14, 18, 30]. In addition some computational difficulties might occur, especially if one attempts to use a random-effects model because the available information is low when analyzing rare events. Here we focused on assessing the variability of the results, in terms of bias and coverage, for Bayesian approaches to implementing the MA. The fully probabilistic (Bayesian) analysis via MCMC methods has the advantage that exact binomial likelihoods can be used, and that studies with zero events in both arms do not need to be excluded from the analysis. However, in this approach prior distributions have to be defined for all relevant parameters in the chosen analysis model. In this simulation study implementing realistic, real-life situations, we found that point estimates for the log(OR) and coverage varied by the choice of the prior distributions for the baseline risk and the standard deviation of the random effect in RE meta-analysis. The results clearly showed that the uniform distribution and the Jeffrey's prior for the baseline risks in the control group lead to biased results and reduced coverage. Weakly informative distribution on the logit of the baseline risks in the control group and hierarchical

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27732v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019

structured prior distributions for the logit of the baseline risks provided similar results and coverage. Excluding studies with no events in both arms affected neither coverage nor bias compared to keeping all studies in the Bayesian analysis. This result is in clear contrast to the findings of [4] for frequentist methods, but we do not clearly understand the reasons for these differing conclusions.

For the simulated data scenarios with varying true log(OR) across the studies in the MA, the results of the Bayesian meta-analyses were also sensitive to the specification of the prior distributions for heterogeneity parameter τ. We found that using a uniform prior distribution from 0 to 2 resulted in high bias and lower coverage. Also, using lognormal distribution suggested by Turner et al. [28] for τ^2 resulted in slightly better results compared to uniform distribution but, using an informative prior exemplified by half-normal with mean = 0.5 for τ performed better.

In summary, in Bayesian MA of rare events the bias for the point estimate for the log(OR) and the coverage of the Bayesian CIs were similar whether studies with no events in both arms were excluded or not. However, bias and coverage were sensitive to the specification of the prior distributions for risk in the baseline groups and for the between-study heterogeneity. Therefore, in concrete situations, as in the case of the Rosiglitazone review, it is important to assess whether obtained results are robust to the specification of prior distributions, or, more generally, to the chosen analytical strategy.

Source of funding

SA is supported by the European Community's Seventh Framework Program FP7/2011: Marie Curie Initial Training Network MEDIASRES ("Novel Statistical Methodology for Diagnostic/Prognostic and Therapeutic Studies and Systematic Reviews", www.mediasres-itn.eu) with the Grant Agreement Number 290025.

Acknowledgments

SA would like to thank Alan Hayes and Georgia Salanti for their scientific inputs. We also thank Christopher Ritter and Kali Tal for language editing.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27732v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019

References:

1. Greenland, S., M.A. Mansournia, and D.G. Altman, *Sparse data bias: a problem hiding in plain sight.* BMJ, 2016. **352**: p. i1981.
2. Friedrich, J.O., J. Beyene, and N.K. Adhikari, *Rosiglitazone: can meta-analysis accurately estimate excess cardiovascular risk given the available data? Re-analysis of randomized trials using various methodologic approaches.* BMC Res Notes, 2009. **2**: p. 5.
3. Keus, F., et al., *Robustness Assessments Are Needed to Reduce Bias in Meta-Analyses That Include Zero-Event Randomized Trials.* American Journal of Gastroenterology, 2009. **104**(3): p. 546-551.
4. Kuss, O., *Statistical methods for meta-analyses including information from studies without any events-add nothing to nothing and succeed nevertheless.* Stat Med, 2015. **34**(7): p. 1097-116.
5. Cheng, J., et al., *Impact of including or excluding both-armed zero-event studies on using standard meta-analysis methods for rare event outcome: a simulation study.* Bmj Open, 2016. **6**(8).
6. Plackett, R.L., *The continuity correction in 2 x 2 tables.* Biometrika, 1964. **51**: p. 327-337.
7. Cox, D.R., *The continuity correction.* Biometrika, 1970. **57**.
8. Sweeting, M.J., A.J. Sutton, and P.C. Lambert, *What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data.* Statistics in Medicine, 2004. **26**: p. 53-77.
9. Cai, T., L. Parast, and L. Ryan, *Meta-analysis for rare events.* Stat Med, 2010. **29**(20): p. 2078-89.
10. Bohning, D., K. Mylona, and A. Kimber, *Meta-analysis of clinical trials with rare events.* Biom J, 2015. **57**(4): p. 633-48.
11. Shuster, J.J., L.S. Jones, and D.A. Salmon, *Fixed vs random effects meta-analysis in rare event studies: the rosiglitazone link with myocardial infarction and cardiac death.* Stat Med, 2007. **26**(24): p. 4375-85.
12. Bracken, M.B., *Rosiglitazone and cardiovascular risk.* N Engl J Med, 2007. **357**(9): p. 937-8; author reply 939-40.
13. Singh, S., Y.K. Loke, and C.D. Furberg, *Long-term Risk of Cardiovascular Events With Rosiglitazone: A Meta-analysis.* JAMA: The Journal of the American Medical Association, 2007. **298**(10): p. 1189-1195.
14. Bradburn, M.J., et al., *Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events.* Statistics In Medicine, 2007. **26**(1): p. 53-77.
15. Tian, L., et al., *Exact and efficient inference procedure for meta-analysis and its application to the analysis of independent 2 x 2 tables with all available data but without artificial continuity correction.* Biostatistics, 2009. **10**(2): p. 275-81.
16. Mannucci, E., et al., *Cardiac safety profile of rosiglitazone: a comprehensive meta-analysis of randomized clinical trials.* Int J Cardiol, 2010. **143**(2): p. 135-40.
17. Shuster, J.J., J.D. Guo, and J.S. Skyler, *Meta-analysis of safety for low event-rate binomial trials.* Res Synth Methods, 2012. **3**(1): p. 30-50.
18. Bhaumik, D.K., et al., *Meta-Analysis of Rare Binary Adverse Event Data.* Journal of the American Statistical Association, 2012. **107**(498): p. 555-567.
19. Li, L. and X. Wang, *Meta-analysis of rare binary events in treatment groups with unequal variability.* Stat Methods Med Res, 2017: p. 962280217721246.
20. Sutton, A.J. and K.R. Abrams, *Bayesian methods in meta-analysis and evidence synthesis.* Stat.Methods Med Res., 2001. **10**(4): p. 277-303.
21. Smith, T.C., D.J. Spiegelhalter, and A. Thomas, *Bayesian approaches to random-effects meta-analysis: a comparative study.* Stat Med, 1995. **14**(24): p. 2685-99.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27732v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019

22. Warn, D.E., T.S. G., and S.D. J., *Bayesian random effects of meta-analysis of trials with binary outcomes: methods for the absolute risk difference and relative risk scales.* Statistics in Medicine, 2002. **21**: p. 1601-1623.
23. Neuenschwander, B., et al., *Summarizing historical information on controls in clinical trials.* Clin Trials, 2010. **7**(1): p. 5-18.
24. Higgins, J.P.T., S.G. Thompson, and D.J. Spiegelhalter, *A re-evaluation of random-effects meta-analysis.* Journal of the Royal Statistical Society.Series A (Statistics in Society), 2009. **172**: p. 137-159.
25. Veroniki, A.A., et al., *Methods to estimate the between-study variance and its uncertainty in meta-analysis.* Res Synth Methods, 2016. **7**(1): p. 55-79.
26. Spiegelhalter, D., K. Abrams, and J. Myles, *Bayesian approaches to clinical trials and health-care evaluation.* John Wiley & Sons: London, 2004.
27. Lane, P.W., *Meta-analysis of incidence of rare events.* Stat Methods Med Res, 2013. **22**(2): p. 117-32.
28. Turner, R.M., et al., *Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis.* Stat Med, 2015. **34**(6): p. 984-98.
29. Lunn, D., et al., *The BUGS project: Evolution, critique and future directions.* Stat Med, 2009. **28**(25): p. 3049-3067.
30. Sutton, A.J., et al., *Meta-analysis of rare and adverse event data.* Expert Rev Pharmacoecon Outcomes Res, 2002. **2**(4): p. 367-79.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27732v1 | CC BY 4.0 Open Access | rec: 15 May 2019, publ: 15 May 2019