



Article

# Bayesian Methods for Meta-Analyses of Binary Outcomes: Implementations, Examples, and Impact of Priors

Fahad M. Al Amer<sup>1,2</sup>, Christopher G. Thompson<sup>3</sup> and Lifeng Lin<sup>2,\*</sup> 

<sup>1</sup> Department of Mathematics, College of Science and Arts, Najran University, Najran 55461, Saudi Arabia; fma16@my.fsu.edu

<sup>2</sup> Department of Statistics, Florida State University, Tallahassee, FL 32306, USA

<sup>3</sup> Department of Educational Psychology, Texas A&M University, College Station, TX 77843, USA; cgthompson@tamu.edu

\* Correspondence: linl@stat.fsu.edu

**Abstract:** Bayesian methods are an important set of tools for performing meta-analyses. They avoid some potentially unrealistic assumptions that are required by conventional frequentist methods. More importantly, meta-analysts can incorporate prior information from many sources, including experts' opinions and prior meta-analyses. Nevertheless, Bayesian methods are used less frequently than conventional frequentist methods, primarily because of the need for nontrivial statistical coding, while frequentist approaches can be implemented via many user-friendly software packages. This article aims at providing a practical review of implementations for Bayesian meta-analyses with various prior distributions. We present Bayesian methods for meta-analyses with the focus on odds ratio for binary outcomes. We summarize various commonly used prior distribution choices for the between-studies heterogeneity variance, a critical parameter in meta-analyses. They include the inverse-gamma, uniform, and half-normal distributions, as well as evidence-based informative log-normal priors. Five real-world examples are presented to illustrate their performance. We provide all of the statistical code for future use by practitioners. Under certain circumstances, Bayesian methods can produce markedly different results from those by frequentist methods, including a change in decision on statistical significance. When data information is limited, the choice of priors may have a large impact on meta-analytic results, in which case sensitivity analyses are recommended. Moreover, the algorithm for implementing Bayesian analyses may not converge for extremely sparse data; caution is needed in interpreting respective results. As such, convergence should be routinely examined. When select statistical assumptions that are made by conventional frequentist methods are violated, Bayesian methods provide a reliable alternative to perform a meta-analysis.

**Keywords:** Bayesian analysis; Markov chain Monte Carlo; meta-analysis; odds ratio; prior distribution



**Citation:** Al Amer, F.M.; Thompson, C.G.; Lin, L. Bayesian Methods for Meta-Analyses of Binary Outcomes: Implementations, Examples, and Impact of Priors. *Int. J. Environ. Res. Public Health* **2021**, *18*, 3492. <https://doi.org/10.3390/ijerph18073492>

Academic Editor: Paul B. Tchounwou

Received: 25 February 2021

Accepted: 22 March 2021

Published: 27 March 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Systematic reviews and meta-analyses have been a popular tool for synthesizing evidence in many fields, including evidence-based medicine, public health, and environmental research [1–3]. Through systematic reviews and meta-analyses, researchers can combine findings from independent studies on common research topics, with one goal being to produce overall results that may be more precise than individual studies.

Coming along with the mass production of meta-analyses in recent years, the quality of many meta-analyses may be low, being partly related to certain methodological limitations [4]. For example, the method that was proposed by DerSimonian and Laird [5] is the most widely-used meta-analysis approach, with over 30,000 Google Scholar citations as of March 21, 2021. However, this approach has been shown to be outperformed by several alternative estimation methods, such as the restricted maximum-likelihood (REML) [6,7]. In addition, conventional meta-analysis methods often assume that within-study sample

variances are known, fixed values. Thus, sampling errors from within-study variances are ignored, which may be unrealistic and they can lead to substantial biases in some situations (e.g., small sample sizes) [8,9]. For example, the sample variance of the log odds ratio (OR) is typically calculated as  $1/a + 1/b + 1/c + 1/d$ , where  $a$ ,  $b$ ,  $c$ , and  $d$  represent the four data cells in a  $2 \times 2$  table for a binary outcome. This sample variance is conventionally treated as a fixed variable in a meta-analysis, although the four data cells are actually random variables. More advanced frequentist approaches, such as generalized mixed-effects models, can avoid this unrealistic assumption, particularly for meta-analyses of binary outcomes [10–13]. Many simulation studies have compared various frequentist methods for meta-analyzing binary outcomes and they have offered useful recommendations under different scenarios [14–17]. However, these approaches may require maximizing certain likelihood functions that involve complicated integrals and are generally less familiar among meta-analysis practitioners.

When comparing frequentist approaches to meta-analysis, Bayesian methods are widely used in medical research among other fields, and they provide a flexible way for implementing complicated models [18–24]. Another major benefit of Bayesian methods is that prior information can be explicitly incorporated in meta-analytic models and, thus, have an impact on the results [25]. In the literature of multivariate meta-analysis of multiple outcomes and/or multiple treatments, Bayesian methods are a standard approach for effectively modeling complicated variance-covariance structures [26,27]. However, frequentist methods still dominate conventional univariate meta-analyses that compare each pair of treatments for each outcome separately [28]. The widespread use of frequentist methods in univariate meta-analyses is partly due to the relatively easy implementation in several popular software packages [29]. Bayesian methods, on the other hand, typically require nontrivial statistical coding, which may limit their promotion in applications. Fortunately, with the advance of meta-analysis methodology, more software programs are being developed to provide user-friendly functions for implementing Bayesian meta-analyses. For example, the R package “bayesmeta” contains a collection of functions for deriving and evaluating posterior distributions of parameters in a random-effects meta-analysis [30]. Another R package “brms” fits Bayesian multilevel models and supports a wide range of distributions and link functions [31]. The R packages “gemtc” and “pcnetmeta” can be used to implement Bayesian network meta-analysis of multiple treatments [32,33]. The relatively new bayesmh command in Stata and BGLIMM procedure in SAS can be also used to fit Bayesian meta-analysis models.

The objectives of this article are three-fold, with a focus on univariate meta-analyses of ORs for binary outcomes. First, we review Bayesian methods, as well as frequentist methods, for conducting meta-analyses. As the heterogeneity of effect sizes plays a critical role when combining primary study results, we introduce three commonly used estimators of the between-studies variance under the frequentist framework. Under a Bayesian framework, we present four prior distributions with different sets of hyper-parameters. It is of note that this manuscript does not aim to provide recommendations for prior selection, because preferable priors may differ case-by-case, particularly when experts’ opinions are available to derive informative priors. Second, we present detailed steps to implement a Bayesian meta-analysis via the Markov chain Monte Carlo (MCMC), with an emphasis on the importance of validating the posterior samples (i.e., MCMC convergence). Third, we apply the various methods to data from five meta-analyses that were published in *The BMJ*. Based on these examples, we empirically examine the impact of different priors on the overall effect estimates, including point and interval estimates, and explore potential problems that are caused by sparse data.

## 2. Materials and Methods

### 2.1. Bayesian Meta-Analysis of Odds Ratios

The Supplementary Materials (Data A) review the frequentist meta-analysis methods. The frequentist framework assumes unknown parameters of interest to be fixed; by

contrast, the Bayesian framework treats unknown parameters as random variables with assigned prior distributions. Bayesian methods have several advantages over frequentist ones. For example, they can be easily applied to complicated models [34,35], improving convergence issues, and improving estimates using informative priors [36–38]. Nevertheless, the required computation is usually more extensive than frequentist methods [39]. Under the Bayesian framework, the estimation and inference of parameters are based on their respective posterior distributions. In some instances, such posterior distributions may not have explicit forms. To overcome this difficulty, MCMC algorithms are widely used to numerically draw samples from posterior distributions [40].

In order to implement Bayesian random-effects meta-analysis of ORs, we consider a meta-analysis containing  $k$  studies with binary outcomes; study  $i$  has  $r_{Ti}$  and  $r_{Ci}$  events in the treatment and control groups, respectively. The event counts are assumed to independently follow binomial distributions with sample sizes  $n_{Ti}$  and  $n_{Ci}$  within each study. The Bayesian hierarchical model is [41–44]:

$$\begin{aligned} r_{Ci} &\sim \text{Bin}(n_{Ci}, \pi_{Ci}), \quad r_{Ti} \sim \text{Bin}(n_{Ti}, \pi_{Ti}); \\ \text{logit}(\pi_{Ci}) &= \mu_i, \quad \text{logit}(\pi_{Ti}) = \mu_i + \delta_i; \\ \delta_i &\sim N(\theta, \tau^2); \\ \mu_i, \theta, \text{ and } \tau &\sim \text{priors}, \end{aligned}$$

where  $\pi_{Ti}$  and  $\pi_{Ci}$  are the underlying true event rates in study  $i$ 's treatment and control groups. Additionally,  $\mu_i$  denotes the baseline risk in each study. Because baseline characteristics may differ greatly across studies, researchers often treat it as a nuisance. Of note, one may also treat these baseline risks as random effects; we do not make this assumption in this article because there are ongoing debates about this assumption for randomized controlled trials [10,45–47].

Using the logit link function  $\text{logit}(t) = \log[t/(1-t)]$  for the true event rates,  $\delta_i$ 's represent the underlying true log ORs within studies. To account for potential heterogeneity, they are assumed to be random effects, following a normal distribution with mean  $\theta$  and variance  $\tau^2$ , where  $\theta$  is interpreted as the overall log OR, and  $\tau^2$  is the between-studies variance.

The vague normal prior  $N(0, 100^2)$  is frequently assigned for  $\mu_i$  and  $\theta$ ; meta-analysis results are usually robust to the choice of priors for these parameters, as long as the priors are sufficiently vague. If available, more precise priors may be used for them based on experts' opinions. Similar to the frequentist framework, where multiple estimators are available for  $\tau^2$  [7], choosing the prior for the heterogeneity parameter is critical in Bayesian meta-analyses, especially when  $k$  is small or events are rare, as this prior may greatly impact the credible interval (CrI) length [48].

## 2.2. Priors for Heterogeneity

We consider four types of priors for the heterogeneity parameter: inverse-gamma, uniform, half-normal, and log-normal priors. Different hyper-parameters are used for each prior distribution. We choose these priors, because they have been used in many Bayesian meta-analyses. One may also tailor a specific prior using experts' opinions on a case-by-case basis; such practice is out of the scope of this article.

The inverse-gamma prior, which is denoted by  $IG(\alpha, \beta)$ , is widely used for  $\tau^2$ . It is conjugate (producing a posterior within the same distribution family and, thus, yielding a closed-form expression) and facilitates computation [40]. Small parameter values are often assigned to the hyper-parameters  $\alpha$  and  $\beta$ , which determine the distribution shape and scale, respectively. As both hyper-parameters approach zero, this prior corresponds to a flat prior for  $\log(\tau^2)$ . We consider setting both hyper-parameters as equivalent and to values 0.001, 0.01, or 0.1; these are common choices in practice [49]. The inverse-gamma prior may be advantageous for dealing with sparse data, as it may improve stability and convergence [50]. Nevertheless, because the inverse-gamma prior is not truly "non-informative", the choice of hyper-parameters may have a substantial impact on meta-analytic results.

The uniform prior, denoted by  $U(0, c)$ , is also widely used for the heterogeneity standard deviation  $\tau$  [51], where  $c$  determines the prior’s upper bound. The lower bound is fixed to zero, as  $\tau$  must be non-negative. We consider  $c = 2, 10,$  and  $100$  [52]. The justification for choosing an appropriate hyper-parameter  $c$  should be based on specific cases. For example, it might be reasonable to set  $c = 2$  for log ORs, as they usually range between  $-2$  and  $2$ . However, when the effect measure is the mean difference for continuous outcomes, the choice of  $c$  should be based on the outcomes’ scales.

The half-normal prior, which is denoted by  $HN(0, \sigma^2)$ , is another candidate prior for  $\tau$  [38,53]. This prior is a special case of the folded normal distribution, i.e., the absolute value of a random variable following  $N(0, \sigma^2)$ , where  $\sigma^2$  controls the extent of heterogeneity. We consider  $\sigma^2 = 0.1, 1,$  or  $2$  [38].

In addition, the evidence-based informative log-normal prior has been suggested for  $\tau^2$  [54]. We denote this prior by  $LN(\mu, \sigma^2)$ ; equivalently,  $\log(\tau^2)$  has the normal prior  $N(\mu, \sigma^2)$ . The hyper-parameters  $\mu$  and  $\sigma^2$  are derived from over 10,000 meta-analyses from the Cochrane Library, which may accurately predict the extent of heterogeneity of an external meta-analysis. As a result, the prior is viewed as “informative”, and it may be useful for meta-analyses with a few studies. The values of hyper-parameters depend on the type of treatment comparison and outcome in a meta-analysis, as detailed in Table 1. Treatment comparisons are classified as pharmacological treatment vs. placebo/control, pharmacological treatment vs. pharmacological treatment, or comparisons involved with non-pharmacological treatments (e.g., medical deviance, surgery), according to Turner et al. [54]. Outcomes are classified as all-cause mortality, semi-objective outcomes (e.g., cause-specific mortality), and subjective outcomes (e.g., mental health condition).

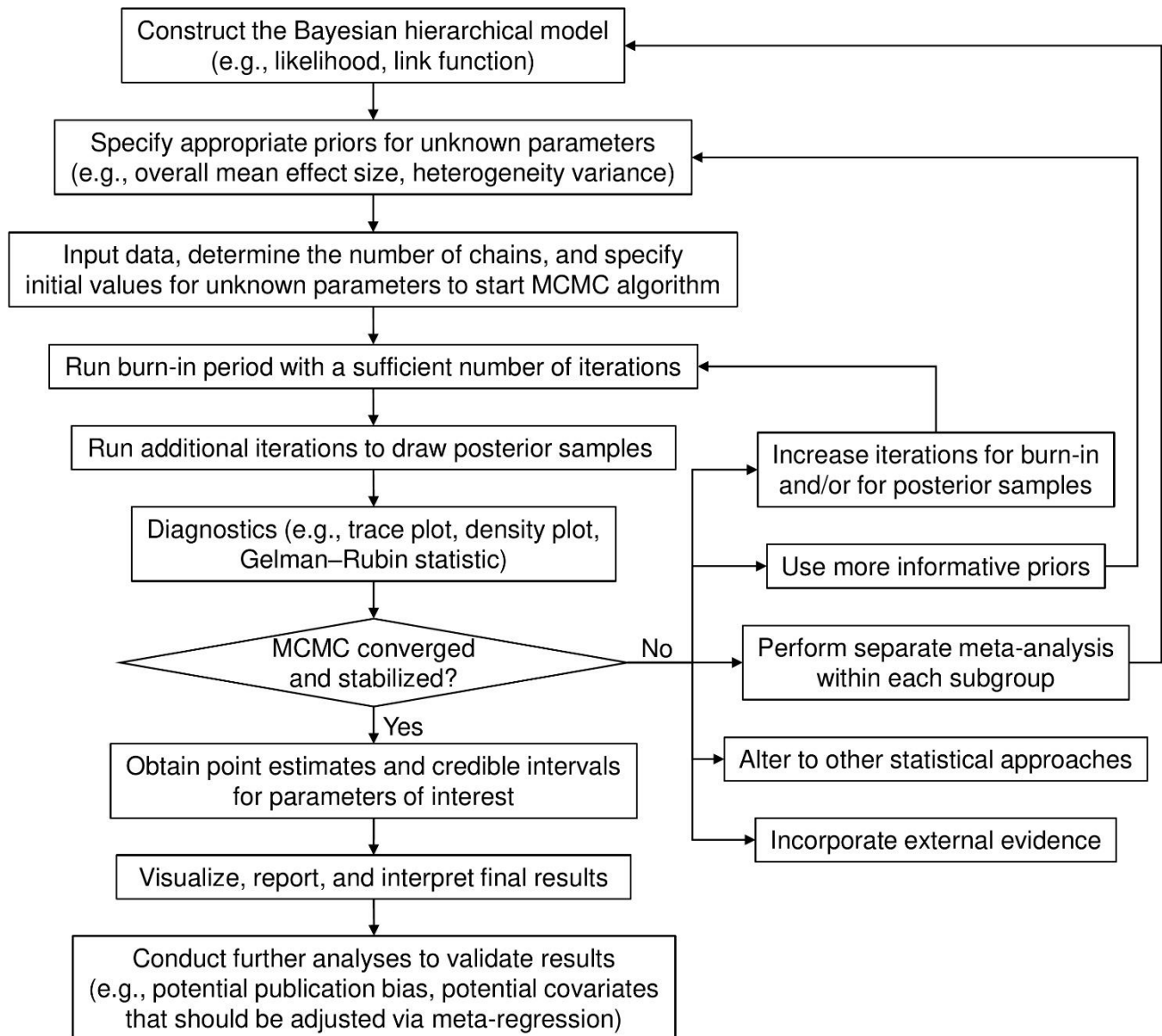
**Table 1.** Summary of prior distributions for the heterogeneity component (variance  $\tau^2$  or standard deviation  $\tau$ ) in a meta-analysis of odds ratios.

| Prior Distribution                    | Used for | Hyper-Parameter  |
|---------------------------------------|----------|--|
| Inverse-gamma,<br>$IG(\alpha, \beta)$ | $\tau^2$ | $\alpha = \beta = 0.1;$ or<br>$\alpha = \beta = 0.01;$ or<br>$\alpha = \beta = 0.001.$   |
| Uniform, $U(0, c)$                    | $\tau$   | $c = 2;$ or<br>$c = 10;$ or<br>$c = 100.$  |
| Half-normal,<br>$HN(0, \sigma^2)$     | $\tau$   | $\sigma^2 = 0.1;$ or<br>$\sigma^2 = 1;$ or<br>$\sigma^2 = 2.$  |
| Log-normal,<br>$LN(\mu, \sigma^2)$    | $\tau^2$ | Pharmacological vs. placebo/control comparison:<br>$\mu = -4.06, \sigma = 1.45$ (all-cause mortality);<br>$\mu = -3.02, \sigma = 1.85$ (semi-objective outcome);<br>$\mu = -2.13, \sigma = 1.58$ (subjective outcome).<br>Pharmacological vs. pharmacological comparison:<br>$\mu = -4.27, \sigma = 1.48$ (all-cause mortality);<br>$\mu = -3.23, \sigma = 1.88$ (semi-objective outcome);<br>$\mu = -2.34, \sigma = 1.62$ (subjective outcome).<br>Non-pharmacological comparison:<br>$\mu = -3.93, \sigma = 1.51$ (all-cause mortality);<br>$\mu = -2.89, \sigma = 1.91$ (semi-objective outcome);<br>$\mu = -2.01, \sigma = 1.64$ (subjective outcome). |

### 2.3. Implementation

Figure 1 shows the flowchart of a general process for implementing a Bayesian meta-analysis via MCMC algorithms. The burn-in period draws posterior samples early in the iteration process to achieve the convergence and stabilization of the Markov chains. Posterior draws during the burn-in period are discarded before making statistical inference. The diagnostic procedures are often overlooked in applications of Bayesian meta-analyses. If the Markov chains have not converged and stabilized, estimates that are based on the

posterior samples may be invalid and/or misleading. Several factors may contribute to this problem. For example, if the number of iterations is insufficient, drawing more posterior samples may solve the problem. The convergence issue may also arise from a multi-modal posterior distribution, possibly because of subgroup effects. In this case, separate meta-analyses may be performed for each subgroup.



**Figure 1.** Flow chart for implementing a Bayesian meta-analysis.

Various approaches, including trace and density plots, are available for diagnosing MCMC samples and their possible convergence. A trace plot presents the posterior samples after the burn-in period in each Markov chain iteration for each parameter. If it shows well-mixed samples, this provides evidence that the Markov chains have stabilized and converged. A density plot depicts the posterior sample density for each parameter. If the density shows multiple modes (e.g., “peaks”), then the Markov chains may not have converged. When Markov chains’ convergence and stabilization have been justified, point estimates (usually posterior medians), and their 95% CrIs, often formed by 2.5% and 97.5% posterior quantiles, can be computed for the Bayesian meta-analysis.

#### 2.4. Application to Real Data

We used five real-world datasets with different sizes (number of studies and within-study sample sizes), event rates, and heterogeneity extents to illustrate the implementation of Bayesian meta-analyses, the impact of prior distributions on meta-analytic results, and the potential problems of MCMC convergence. All of the datasets had binary outcomes; regardless of original analyses, (log) ORs were used as effect measures in our re-analyses.

Example 1. The meta-analysis by Lamont et al. [55] investigated the risk of recurrent stillbirth from 13 cohort studies, all with large sample sizes. The comparison type was non-pharmacological (previous stillbirth vs. previous live birth), and the outcome (recurrent stillbirth) might be classified as semi-objective.

Example 2. The meta-analysis by Crocker et al. [56] combined eight studies to investigate the impact of patient and public involvement (PPI) on patient enrollment in clinical trials. The intervention PPI was non-pharmacological, and the outcome (enrollment in clinical trials) was likely semi-objective.

Example 3. The meta-analysis, as reported by Baxi et al. [57], included 13 studies to synthesize the association between anti-programmed cell death 1 (anti-PD-1) drugs and immune-related adverse events. This meta-analysis compared pharmacological treatments with control. In this example, the outcome was the adverse event colitis, which may be considered subjective according to Turner et al. [54] The event rate in each primary study was fairly low; many zero event counts existed.

Example 4. This meta-analysis is also from Baxi et al. [57]; it combined 15 studies with the adverse event hepatitis, which was also considered as a subjective outcome. The treatment comparison was also pharmacological treatments (anti-PD-1 drugs) vs. control. All of the studies had zero events in the control group.

Example 5. The fifth meta-analysis by Martineau et al. [58] collected 25 randomized controlled trials to assess the effect of vitamin D supplementation on the risk of acute respiratory tract infection (ARTI). The comparison was the pharmacological treatment (vitamin D supplementation) vs. control. The outcome was the experience with at least one ARTI, which might be considered to be subjective.

#### 2.5. Statistical Analyses

For each dataset, we implemented the random-effects meta-analysis using the DerSimonian–Laird (DL) [5], maximum-likelihood (ML), and REML estimators [59] for  $\tau^2$  under the frequentist framework and using the Bayesian models with 12 different priors for  $\tau^2$  or  $\tau$ . Specifically, we considered all four prior distributions that are reviewed in Section 2.2. For each of the inverse-gamma, uniform, and half-normal priors, we used three sets of hyper-parameters, as shown in Table 1. For the informative log-normal priors proposed by Turner et al. [54], we identified the treatment comparison type of each meta-analysis. Recall that the outcomes were classified as all-cause mortality, semi-objective outcomes, and subjective outcomes (Section 2.2). Such classifications might be ambiguous in some cases, particularly when distinguishing semi-objective and subjective outcomes. Therefore, according to the treatment comparison type of each meta-analysis, we used all three possible sets of hyper-parameters for three outcome types.

In the presence of zero event counts, when implementing the conventional frequentist methods, studies with zero counts in both the treatment and control groups had to be excluded (see Data A.3 in the Supplementary Materials for details). The continuity correction of 0.5 was applied to studies with zero counts in only a single treatment arm. Such ad hoc corrections were not needed in Bayesian analyses.

We used the R package “rjags” (version 4–7) to implement the Bayesian models via the MCMC algorithm. For each Bayesian model, we used three Markov chains, each having a burn-in period of 50,000 iterations, followed by 200,000 iterations for drawing posterior samples with thinning rate 2. Trace plots were used to assess the chains’ convergence. We obtained the posterior medians of the overall OR, the heterogeneity standard deviation  $\hat{\tau}$ , and 95% CrIs for both parameter estimates. The frequentist methods were implemented

via the R package “metafor” [60]. We used the  $Q$ -profile method to obtain 95% confidence intervals (CIs) of  $\hat{\tau}$  [61]. The Supplementary Materials (Data B) provide all of the statistical code.

This study did not require ethical approval and patient consent, because it focused on statistical methods for meta-analyses, and all the analyses were performed using publicly available data.

### 3. Results

Table 2 provides the estimated ORs and heterogeneity standard deviations of the five real-world meta-analyses using all priors under the Bayesian framework, as well as three different estimators under the frequentist framework. All of the ORs were computed for comparing treatment against control. Figure 2 shows the forest plot for the meta-analysis on stillbirth; Figures S1–S4 in the Supplementary Materials (Data C) present those for the remaining four meta-analyses. In addition, Figures S5–S64 in the Supplementary Materials (Data D) provide trace plots related to MCMC processes, and Figures S65–S69 in the Supplementary Materials (Data E) present posterior density plots.

#### 3.1. Example 1: Meta-Analysis on Stillbirth

The estimated overall ORs by all methods ranged from 4.30 to 4.59, and their CIs/CrIs were all above 1, which suggested a significantly increased risk of recurrent stillbirth among women with a previous stillbirth compared with those with a previous live birth. The results were relatively robust to the different meta-analytic methods, likely because this meta-analysis contained a moderate number of studies ( $k = 13$ ), with each study having many patients. These two combined qualities likely provided enough information for similar estimation results. The trace plots (Figures S5–S16) and posterior density plot (Figure S65) indicate that the Markov chains converged and stabilized well.

Nevertheless, different methods indicated some noticeable dissimilar point and interval estimates. For example, the frequentist DL, ML, and REML methods estimated the overall OR as 4.59, 4.52, and 4.47, respectively, and  $\hat{\tau}$  was, accordingly, 0.38, 0.43, and 0.46. Among the Bayesian methods, the inverse-gamma and uniform priors produced fairly similar results. Those by the half-normal and log-normal priors led to some small differences. Specifically, as the hyper-parameter of the half-normal prior increased from 0.1 to 2, the estimated overall OR changed from 4.35 to 4.26, with  $\hat{\tau}$  changing from 0.45 to 0.53 and yielding wider CrIs. The estimated overall ORs ranged from 4.31 to 4.39 for the log-normal priors.

In addition, the 95% CrIs that were produced by the Bayesian methods were generally wider than the 95% CIs produced by the frequentist methods. This was possibly because the frequentist methods failed to account for sampling errors within sample variances and the uncertainty in  $\hat{\tau}$ .

#### 3.2. Example 2: Meta-Analysis on Patient Enrollment in Clinical Trials

All three frequentist methods produced  $\hat{\tau} = 0$ , which then corresponded to the same estimated overall OR 1.16 with 95% CI [1.03, 1.30]. This suggested a statistically significant effect of PPI on patient enrollment in clinical trials. However, all of the Bayesian methods produced 95% CrIs containing 1, which indicated that the effect of PPI was not significant. The trace plots (Figures S17–S28) and the posterior density plot (Figure S66) indicate that the Markov chains converged and stabilized well.

The number of studies ( $k = 8$ ) was relatively small. The uniform and half-normal priors with different hyper-parameters led to fairly similar results. Nevertheless, the inverse-gamma prior had an influence on the results. As its hyper-parameters changed from 0.001 to 0.1, although the OR estimates remained nearly unchanged, the associated 95% CrI became much wider, changing from [0.99, 1.39] to [0.87, 1.57], with  $\hat{\tau}$  changing from 0.09 to 0.29. The informative log-normal prior also had a noticeable impact on the 95% CrIs, and  $\hat{\tau}$  was 0.10, 0.12, and 0.16 while using different hyper-parameters.

**Table 2.** The estimated odds ratios (OR) and heterogeneity standard deviations (Tau) with their 95% credible/confidence intervals using different methods in the five examples.

| Estimate | Bayesian Method   |                       |                      |                            |                      |                                   |                                |                       |                       |                               |                       |                       | Frequentist Method <sup>e</sup> |                       |                       |
|----------|---|-----------------------|----------------------|----------------------------|----------------------|-----------------------------------|--------------------------------|-----------------------|-----------------------|-------------------------------|-----------------------|-----------------------|---------------------------------|-----------------------|-----------------------|
|          | Inverse-Gamma Prior <sup>a</sup>                                  |                       |                      | Uniform Prior <sup>b</sup> |                      |                                   | Half-Normal Prior <sup>c</sup> |                       |                       | Log-Normal Prior <sup>d</sup> |                       |                       | DL                              | ML                    | REML                  |
|          | IG1   | IG2                   | IG3                  | U1                         | U2                   | U3                                | HN1                            | HN2                   | HN3                   | LN1                           | LN2                   | LN3                   |                                 |                       |                       |
|          | Example 1: meta-analysis on stillbirth                            |                       |                      |                            |                      |                                   |                                |                       |                       |                               |                       |                       |                                 |                       |                       |
| OR       | 4.30<br>(2.95, 5.86)  | 4.30<br>(2.95, 5.87)  | 4.26<br>(2.90, 5.90) | 4.26<br>(2.85, 5.95)       | 4.25<br>(2.85, 5.94) | 4.25<br>(2.85, 5.95)              | 4.35<br>(3.13, 5.78)           | 4.26<br>(2.89, 5.91)  | 4.26<br>(2.86, 5.92)  | 4.39<br>(3.16, 5.76)          | 4.33<br>(3.04, 5.82)  | 4.31<br>(3.01, 5.83)  | 4.59<br>(3.56, 5.93)            | 4.52<br>(3.42, 5.96)  | 4.47<br>(3.34, 5.99)  |
| Tau      | 0.49<br>(0.29, 0.89)  | 0.50<br>(0.30, 0.89)  | 0.52<br>(0.32, 0.91) | 0.53<br>(0.31, 0.98)       | 0.54<br>(0.31, 0.98) | 0.54<br>(0.31, 0.98)              | 0.45<br>(0.28, 0.71)           | 0.52<br>(0.31, 0.93)  | 0.43<br>(0.31, 0.95)  | 0.43<br>(0.26, 0.72)          | 0.47<br>(0.28, 0.80)  | 0.48<br>(0.29, 0.82)  | 0.38<br>(0.28, 0.91)            | 0.43<br>(0.28, 0.91)  | 0.46<br>(0.28, 0.91)  |
|          | Example 2: meta-analysis on patient enrollment in clinical trials |                       |                      |                            |                      |                                   |                                |                       |                       |                               |                       |                       |                                 |                       |                       |
| OR       | 1.17<br>(0.99, 1.39)  | 1.17<br>(0.95, 1.44)  | 1.17<br>(0.87, 1.57) | 1.17<br>(0.96, 1.43)       | 1.17<br>(0.96, 1.43) | 1.17<br>(0.96, 1.43)              | 1.17<br>(0.98, 1.41)           | 1.17<br>(0.96, 1.42)  | 1.17<br>(0.98, 1.41)  | 1.17<br>(0.99, 1.39)          | 1.17<br>(0.97, 1.41)  | 1.17<br>(0.95, 1.45)  | 1.16<br>(1.03, 1.30)            | 1.16<br>(1.03, 1.30)  | 1.16<br>(1.03, 1.30)  |
| Tau      | 0.09<br>(0.02, 0.34)  | 0.15<br>(0.06, 0.42)  | 0.29<br>(0.16, 0.63) | 0.11<br>(0.01, 0.46)       | 0.12<br>(0.01, 0.47) | 0.11<br>(0.01, 0.47)              | 0.10<br>(0.01, 0.36)           | 0.10<br>(0.01, 0.44)  | 0.11<br>(0.01, 0.45)  | 0.10<br>(0.03, 0.28)          | 0.12<br>(0.03, 0.36)  | 0.16<br>(0.05, 0.43)  | 0.00<br>(0.00, 0.46)            | 0.00<br>(0.00, 0.46)  | 0.00<br>(0.00, 0.46)  |
|          | Example 3: meta-analysis on colitis <sup>f</sup>                  |                       |                      |                            |                      |                                   |                                |                       |                       |                               |                       |                       |                                 |                       |                       |
| OR       | 6.90<br>(2.28, 102)   | 7.15<br>(2.25, 84.30) | 7.99<br>(2.25, 141)  | 7.59<br>(2.24, 47.85)      | 9.83<br>(2.24, 2008) | 11.37<br>(2.25, 10 <sup>5</sup> ) | 6.09<br>(2.23, 22.46)          | 6.88<br>(2.23, 35.32) | 7.62<br>(2.28, 56.53) | 5.89<br>(2.25, 21.04)         | 5.95<br>(2.18, 23.08) | 6.34<br>(2.20, 26.20) | 3.39<br>(1.45, 7.95)            | 3.39<br>(1.45, 7.95)  | 3.39<br>(1.45, 7.95)  |
| Tau      | 0.25<br>(0.03, 3.89)  | 0.44<br>(0.08, 3.37)  | 0.73<br>(0.22, 3.98) | 0.74<br>(0.03, 1.90)       | 1.13<br>(0.05, 7.32) | 1.33<br>(0.05, 12.84)             | 0.20<br>(0.01, 0.68)           | 0.51<br>(0.02, 1.78)  | 0.66<br>(0.03, 2.45)  | 0.13<br>(0.03, 0.49)          | 0.21<br>(0.04, 1.02)  | 0.31<br>(0.07, 1.26)  | 0.00<br>(0.00, 0.00)            | 0.00<br>(0.00, 0.00)  | 0.00<br>(0.00, 0.00)  |
|          | Example 4: meta-analysis on hepatitis <sup>g</sup>                |                       |                      |                            |                      |                                   |                                |                       |                       |                               |                       |                       |                                 |                       |                       |
| OR       | NA  | NA                    | NA                   | NA                         | NA                   | NA                                | NA                             | NA                    | NA                    | NA                            | NA                    | NA                    | 3.14<br>(0.76, 12.98)           | 3.14<br>(0.76, 12.98) | 3.14<br>(0.76, 12.98) |
| Tau      | NA  | NA                    | NA                   | NA                         | NA                   | NA                                | NA                             | NA                    | NA                    | NA                            | NA                    | NA                    | 0.00<br>(0.00, 0.00)            | 0.00<br>(0.00, 0.00)  | 0.00<br>(0.00, 0.00)  |
|          | Example 5: meta-analysis on acute respiratory tract infection     |                       |                      |                            |                      |                                   |                                |                       |                       |                               |                       |                       |                                 |                       |                       |
| OR       | 0.83<br>(0.70, 0.95)  | 0.82<br>(0.69, 0.95)  | 0.81<br>(0.67, 0.95) | 0.82<br>(0.68, 0.95)       | 0.82<br>(0.68, 0.95) | 0.82<br>(0.68, 0.95)              | 0.82<br>(0.69, 0.95)           | 0.82<br>(0.69, 0.95)  | 0.82<br>(0.69, 0.95)  | 0.83<br>(0.71, 0.95)          | 0.83<br>(0.70, 0.95)  | 0.82<br>(0.70, 0.95)  | 0.83<br>(0.72, 0.95)            | 0.83<br>(0.72, 0.95)  | 0.83<br>(0.71, 0.95)  |
| Tau      | 0.21<br>(0.05, 0.41)  | 0.23<br>(0.10, 0.43)  | 0.29<br>(0.18, 0.48) | 0.25<br>(0.08, 0.46)       | 0.25<br>(0.08, 0.45) | 0.25<br>(0.08, 0.46)              | 0.23<br>(0.07, 0.41)           | 0.25<br>(0.08, 0.45)  | 0.25<br>(0.08, 0.45)  | 0.19<br>(0.06, 0.36)          | 0.22<br>(0.08, 0.40)  | 0.24<br>(0.10, 0.42)  | 0.21<br>(0.08, 0.47)            | 0.21<br>(0.08, 0.47)  | 0.23<br>(0.08, 0.47)  |

<sup>a</sup> Inverse-gamma priors for  $\tau^2$  with three sets of hyper-parameters: IG1,  $IG(0.001, 0.001)$ ; IG2,  $IG(0.01, 0.01)$ ; IG3,  $IG(0.1, 0.1)$ . <sup>b</sup> Uniform priors for  $\tau$  with three sets of hyper-parameters: U1,  $U(0, 2)$ ; U2,  $U(0, 10)$ ; U3,  $U(0, 100)$ . <sup>c</sup> Half-normal priors for  $\tau$  with three sets of hyper-parameters: HN1,  $HN(0, 0.1)$ ; HN2,  $HN(0, 1)$ ; HN3,  $HN(0, 2)$ . <sup>d</sup> Log-normal priors for  $\tau^2$  with three sets of hyper-parameters. For the meta-analyses on stillbirth and on patient enrollment in clinical trials (non-pharmacological treatment comparisons), they are: LN1,  $LN(-3.93, 1.51^2)$ ; LN2,  $LN(-2.89, 1.91^2)$ ; LN3,  $LN(-2.01, 1.64^2)$ . For the meta-analyses on colitis, on hepatitis, and on acute respiratory tract infection (pharmacological treatments vs. placebo/control), they are: LN1,  $LN(-4.06, 1.45^2)$ ; LN2,  $LN(-3.02, 1.85^2)$ ; LN3,  $LN(-2.13, 1.58^2)$ . The results in italic are based on the set of hyper-parameters that best matches the outcome categorization (all-cause mortality, semi-objective, and subjective outcomes) according to Turner et al. [54] <sup>e</sup> The frequentist methods include the DerSimonian–Laird (DL), maximum likelihood (ML), and restricted maximum likelihood (REML) estimators. <sup>f</sup> Markov chains might have convergence issues when using some priors, so the results may be interpreted with cautions. <sup>g</sup> Markov chains had poor convergence when using all priors, so the results by the Bayesian methods may not be reliable and are reported as not available (NA). The results by the frequentist methods should also be interpreted with great caution.



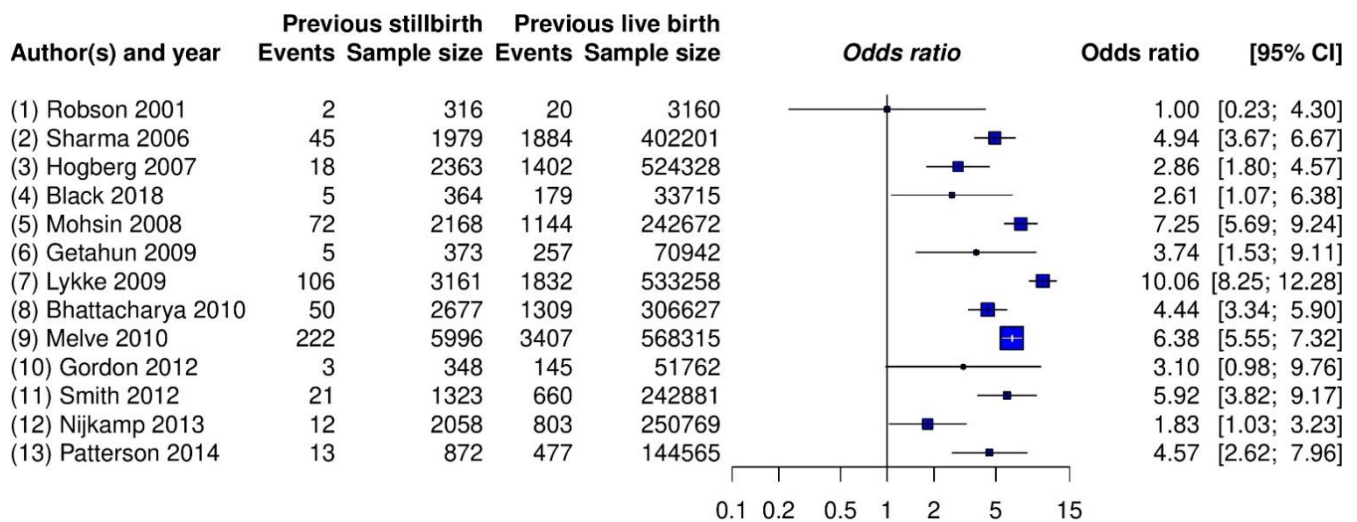


Figure 2. Forest plot of the meta-analysis on stillbirth (Example 1).

### 3.3. Example 3: Meta-Analysis on Colitis

This meta-analysis contained  $k = 13$  studies; four studies had zero event counts in both arms, thus their ORs were not estimable, and they were subsequently removed when using the frequentist methods. All three frequentist estimators gave  $\hat{\tau} = 0$  and estimated the overall OR as 3.39 with 95% CI [1.45, 7.95], indicating that patients that were treated with anti-PD-1 drugs were at significantly higher risk of developing colitis as compared with those in the control group.

The trace plots (Figures S29–S40) and the posterior density plot (Figure S67) indicate that some Markov chains might have convergence issues when using some priors. The posterior samples of the overall OR might take extreme values in a few MCMC iterations. Caution was advised when interpreting the results to assess their reliability. If more informative priors were available for colitis than the 12 priors considered here, one may further examine whether such priors could improve the MCMC convergence and, thus, the validity of the results.

The Bayesian OR estimates, ranging from 5.89 to 11.37, were much larger than those that were obtained from the frequentist methods. The large difference was likely because the Bayesian methods effectively accounted for the double-zero-event studies, while the conventional frequentist methods could not use such studies.

Besides the four double-zero-event studies, six studies contained zero counts in the control group. Because of the sparse data, the results from the Bayesian methods were fairly sensitive to the choice of hyper-parameters. The 95% CrIs were also much wider than the 95% CIs produced by the frequentist methods, although they still suggested a significant effect of anti-PD-1 drugs on colitis. The frequentist 95% CIs were narrow, likely yielding low coverage probabilities, because the frequentist methods used large-sample properties to derive within-study sample variances and treated them as known, fixed values.

### 3.4. Example 4: Meta-Analysis on Hepatitis

This meta-analysis was from the same systematic review as the previous meta-analysis on colitis [57]; it investigated another adverse event, hepatitis. The data were even sparser in this meta-analysis, which may serve as an excellent example to illustrate the importance of checking MCMC convergence in Bayesian meta-analysis. All  $k = 15$  studies had zero counts in the control group, and only five studies had non-zero counts in the treatment group. In total, this meta-analysis only had six events among all 7156 patients.

Again, all three frequentist methods yielded  $\hat{\tau} = 0$  and produced the overall OR estimate 3.14 with 95% CI [0.76, 12.98], indicating no significant effect of anti-PD-1 drugs on hepatitis. The results that were produced by the Bayesian methods with the considered 12 priors were not reliable, because the trace plots (Figures S41–S52) showed that nearly all Markov chains did not converge. The posterior density plot (Figure S68) indicated multiple modes of the posterior distributions. These issues were likely caused by the extremely sparse data with few events. The estimated heterogeneity standard deviation  $\hat{\tau}$  was very sensitive to the choice of priors; it was primarily influenced by the priors, because the sparse data contained little information about heterogeneity. If the researchers had better prior beliefs than the priors used in this article, they may try such alternatives to examine if the inference for this extremely sparse dataset could be improved.

Of note, although the frequentist methods successfully produced the results, as reported in Table 2, they may also be unreliable. For example, the likelihood function for the sparse data may be fairly flat, making the ML algorithm used for deriving the estimates highly unstable [62,63]. The event rate was close/equal to the boundary value 0, and technical issues existed when making inference on the ML estimates.

### 3.5. Example 5: Meta-Analysis on Acute Respiratory Tract Infection

The number of studies ( $k = 25$ ) was relatively large in this meta-analysis, and no zero-count appeared in any arm. All three frequentist methods produced nearly the same point and interval estimates for the overall OR, showing that vitamin D supplementation significantly reduced the risk of ARTI. Because of sufficient data information, the results were fairly robust to prior distribution specifications. Across all 12 sets of priors, the overall OR estimate was about 0.82, and its 95% CrI lower and upper bounds were about 0.70 and 0.95, respectively. The trace plots (Figures S53–S64) and posterior density plot (Figure S69) indicate that the Markov chains converged and stabilized well.

## 4. Conclusions

This article has provided a practical review of Bayesian meta-analyses of binary outcomes and summarized several commonly used priors for the heterogeneity variance. We have shown five worked examples to highlight the implementations of incorporating different prior distributions and the importance of checking MCMC convergence. The Bayesian methods are advantageous over the conventional frequentist methods, primarily because they make more practical assumptions and they can incorporate informative priors.

Generally, when the data provided by a meta-analysis are limited (e.g., few studies, small sample sizes, or low event rates), the choice of prior may have a noteworthy impact on meta-analysis results. Ideally, meta-analyses may use priors that represent experts' opinions. If such priors are unavailable or disputed, sensitivity analyses that use different candidate priors are recommended. In addition, meta-analysts should routinely examine the convergence and stabilization of MCMC processes when performing Bayesian meta-analyses; this is often overlooked in current practice, as illustrated in Examples 3 and 4. When event rates are low and zero event counts appear in many arms in a meta-analysis, the MCMC algorithm may not converge; as such, the inference that is based on posterior samples is unreliable. We have summarized several possible solutions to handle this problem in Figure 1. If the Bayesian meta-analysis still fails to produce valid results after considering these solutions, researchers may refer to several papers that especially focused on the cases of few studies [38,64–67] and rare events [68–75] for alternative approaches. Moreover, if individual participant data are available, incorporating them into the meta-analysis might help to improve the estimation of treatment effects [13,76–78].

Bayesian methods for meta-analysis are not problem-free, and their results must be carefully validated before being applied to decision-making, as discussed above. This article did not intend to promote Bayesian meta-analyses to all situations and involve in the debates on frequentist vs. Bayesian inference. Instead, we aimed at providing some practical guidelines for researchers who are interested in implementing Bayesian

meta-analyses. One may refer to other papers that especially aimed at comparing these two types of methods to better understand the pros and cons of frequentist and Bayesian meta-analyses [74,79–84].

This article limited the review of Bayesian methods to the case of ORs. Bayesian methods are broadly available for meta-analyses of other effect measures (e.g., relative risks) with binary outcomes and for meta-analyses with other outcomes (e.g., continuous or count data) [42,85]. Similar implementations of Bayesian analyses may be applied to these cases with proper adjustments (e.g., different specifications of likelihood, link functions, and priors).

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/ijerph18073492/s1>, In the Supplementary Materials, Supplementary Materials data A presents conventional frequentist meta-analysis of odds ratios, Supplementary Materials data B gives R code for analyzing the five examples, Supplementary Materials data C presents forest plots, Supplementary Materials data D presents trace plots, and Supplementary Materials data E presents posterior density plots.

**Author Contributions:** F.M.A.A. collected data, performed statistical analyses, and drafted the manuscript; C.G.T. helped interpret results and substantially revised the manuscript; L.L. designed this study, supervised the project, and substantially revised the manuscript. All authors have approved the final version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The forest plots in Figure 2 and Figures S1–S4 (Supplementary Materials data C) include the data used in this article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Niforatos, J.D.; Weaver, M.; Johansen, M.E. Assessment of publication trends of systematic reviews and randomized clinical trials, 1995 to 2017. *JAMA Intern. Med.* **2019**, *179*, 1593–1594. [[CrossRef](#)]
2. Requia, W.J.; Adams, M.D.; Arain, A.; Papatheodorou, S.; Koutrakis, P.; Mahmoud, M. Global association of air pollution and cardiorespiratory diseases: A systematic review, meta-analysis, and investigation of modifier variables. *Am. J. Public Health* **2018**, *108*, S123–S130. [[CrossRef](#)]
3. Hoffmann, R.; Dimitrova, A.; Muttarak, R.; Crespo Cuaresma, J.; Peisker, J. A meta-analysis of country-level studies on environmental change and migration. *Nat. Clim. Chang.* **2020**, *10*, 904–912. [[CrossRef](#)]
4. Ioannidis, J.P.A. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *Milbank Q.* **2016**, *94*, 485–514. [[CrossRef](#)]
5. DerSimonian, R.; Laird, N. Meta-analysis in clinical trials. *Control. Clin. Trials* **1986**, *7*, 177–188. [[CrossRef](#)]
6. Cornell, J.E.; Mulrow, C.D.; Localio, R.; Stack, C.B.; Meibohm, A.R.; Guallar, E.; Goodman, S.N. Random-effects meta-analysis of inconsistent effects: A time for change. *Ann. Intern. Med.* **2014**, *160*, 267–270. [[CrossRef](#)] [[PubMed](#)]
7. Langan, D.; Higgins, J.P.T.; Jackson, D.; Bowden, J.; Veroniki, A.A.; Kontopantelis, E.; Viechtbauer, W.; Simmonds, M. A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Res. Synth. Methods* **2019**, *10*, 83–98. [[CrossRef](#)] [[PubMed](#)]
8. Doncaster, C.P.; Spake, R. Correction for bias in meta-analysis of little-replicated studies. *Methods Ecol. Evol.* **2018**, *9*, 634–644. [[CrossRef](#)]
9. Lin, L. Bias caused by sampling error in meta-analysis with small sample sizes. *PLoS ONE* **2018**, *13*, e0204056. [[CrossRef](#)]
10. Jackson, D.; Law, M.; Stijnen, T.; Viechtbauer, W.; White, I.R. A comparison of seven random-effects models for meta-analyses that estimate the summary odds ratio. *Stat. Med.* **2018**, *37*, 1059–1085. [[CrossRef](#)] [[PubMed](#)]
11. Hartung, J.; Knapp, G. A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Stat. Med.* **2001**, *20*, 3875–3889. [[CrossRef](#)]
12. Bhaumik, D.K.; Amatya, A.; Normand, S.-L.T.; Greenhouse, J.; Kaizar, E.; Neelon, B.; Gibbons, R.D. Meta-analysis of rare binary adverse event data. *J. Am. Stat. Assoc.* **2012**, *107*, 555–567. [[CrossRef](#)]
13. Riley, R.D.; Steyerberg, E.W. Meta-analysis of a binary outcome using individual participant data and aggregate data. *Res. Synth. Methods* **2010**, *1*, 2–19. [[CrossRef](#)] [[PubMed](#)]

14. Mathes, T.; Kuss, O. A comparison of methods for meta-analysis of a small number of studies with binary outcomes. *Res. Synth. Methods* **2018**, *9*, 366–381. [[CrossRef](#)]
15. Bakbergenuly, I.; Kulinskaya, E. Meta-analysis of binary outcomes via generalized linear mixed models: A simulation study. *BMC Med. Res. Methodol.* **2018**, *18*, 70. [[CrossRef](#)] [[PubMed](#)]
16. Beisemann, M.; Doebler, P.; Holling, H. Comparison of random-effects meta-analysis models for the relative risk in the case of rare events: A simulation study. *Biom. J.* **2020**, *62*, 1597–1630. [[CrossRef](#)]
17. 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*, 1097–1116. [[CrossRef](#)] [[PubMed](#)]
18. Schmid, C.H. Using Bayesian inference to perform meta-analysis. *Eval. Health Prof.* **2001**, *24*, 165–189. [[CrossRef](#)]
19. McGlothlin, A.E.; Viele, K. Bayesian hierarchical models. *JAMA* **2018**, *320*, 2365–2366. [[CrossRef](#)] [[PubMed](#)]
20. Ashby, D. Bayesian statistics in medicine: A 25 year review. *Stat. Med.* **2006**, *25*, 3589–3631. [[CrossRef](#)]
21. Pullenayegum, E.M.; Thabane, L. Teaching Bayesian statistics in a health research methodology program. *J. Stat. Educ.* **2009**, *17*, 3. [[CrossRef](#)]
22. Bittl, J.A.; He, Y. Bayesian analysis: A practical approach to interpret clinical trials and create clinical practice guidelines. *Circ. Cardiovasc. Qual. Outcomes* **2017**, *10*, e003563. [[CrossRef](#)]
23. Negrín-Hernández, M.-A.; Martel-Escobar, M.; Vázquez-Polo, F.-J. Bayesian meta-analysis for binary data and prior distribution on models. *Int. J. Environ. Res. Public Health* **2021**, *18*, 809. [[CrossRef](#)]
24. Pullenayegum, E.M. An informed reference prior for between-study heterogeneity in meta-analyses of binary outcomes. *Stat. Med.* **2011**, *30*, 3082–3094. [[CrossRef](#)]
25. Quintana, M.; Viele, K.; Lewis, R.J. Bayesian analysis: Using prior information to interpret the results of clinical trials. *JAMA* **2017**, *318*, 1605–1606. [[CrossRef](#)]
26. Wei, Y.; Higgins, J.P.T. Bayesian multivariate meta-analysis with multiple outcomes. *Stat. Med.* **2013**, *32*, 2911–2934. [[CrossRef](#)] [[PubMed](#)]
27. Lin, L.; Chu, H. Bayesian multivariate meta-analysis of multiple factors. *Res. Synth. Methods* **2018**, *9*, 261–272. [[CrossRef](#)]
28. Grant, R.L. The uptake of Bayesian methods in biomedical meta-analyses: A scoping review (2005–2016). *J. Evid. Based Med.* **2019**, *12*, 69–75. [[CrossRef](#)]
29. Borenstein, M.; Hedges, L.V.; Higgins, J.P.T.; Rothstein, H.R. *Introduction to Meta-Analysis*; John Wiley & Sons: Chichester, UK, 2009.
30. Röver, C. Bayesian random-effects meta-analysis using the bayesmeta R package. *J. Stat. Softw.* **2020**, *93*, 1–51. [[CrossRef](#)]
31. Bürkner, P.-C. brms: An R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **2017**, *80*, 1–28. [[CrossRef](#)]
32. Van Valkenhoef, G.; Lu, G.; de Brock, B.; Hillege, H.; Ades, A.E.; Welton, N.J. Automating network meta-analysis. *Res. Synth. Methods* **2012**, *3*, 285–299. [[CrossRef](#)]
33. Lin, L.; Zhang, J.; Hodges, J.S.; Chu, H. Performing arm-based network meta-analysis in R with the pcnetmeta package. *J. Stat. Softw.* **2017**, *80*, 1–25. [[CrossRef](#)] [[PubMed](#)]
34. Muthén, B.; Asparouhov, T. Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychol. Methods* **2012**, *17*, 313–335. [[CrossRef](#)]
35. Higgins, J.P.T.; Thompson, S.G.; Spiegelhalter, D.J. A re-evaluation of random-effects meta-analysis. *J. R. Stat. Soc. Ser. A* **2009**, *172*, 137–159. [[CrossRef](#)]
36. Depaoli, S.; Clifton, J.P. A Bayesian approach to multilevel structural equation modeling with continuous and dichotomous outcomes. *Struct. Equ. Modeling Multidiscip. J.* **2015**, *22*, 327–351. [[CrossRef](#)]
37. Depaoli, S. Mixture class recovery in GMM under varying degrees of class separation: Frequentist versus Bayesian estimation. *Psychol. Methods* **2013**, *18*, 186–219. [[CrossRef](#)] [[PubMed](#)]
38. Friede, T.; Röver, C.; Wandel, S.; Neuenschwander, B. Meta-analysis of few small studies in orphan diseases. *Res. Synth. Methods* **2017**, *8*, 79–91. [[CrossRef](#)] [[PubMed](#)]
39. Kruschke, J.K.; Liddell, T.M. The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychon. Bull. Rev.* **2018**, *25*, 178–206. [[CrossRef](#)] [[PubMed](#)]
40. Carlin, B.P.; Louis, T.A. *Bayesian Methods for Data Analysis*, 3rd ed.; CRC Press: Boca Raton, FL, USA, 2009.
41. Smith, T.C.; Spiegelhalter, D.J.; Thomas, A. Bayesian approaches to random-effects meta-analysis: A comparative study. *Stat. Med.* **1995**, *14*, 2685–2699. [[CrossRef](#)] [[PubMed](#)]
42. Warn, D.E.; Thompson, S.G.; Spiegelhalter, D.J. Bayesian random effects meta-analysis of trials with binary outcomes: Methods for the absolute risk difference and relative risk scales. *Stat. Med.* **2002**, *21*, 1601–1623. [[CrossRef](#)]
43. Thompson, S.G.; Smith, T.C.; Sharp, S.J. Investigating underlying risk as a source of heterogeneity in meta-analysis. *Stat. Med.* **1997**, *16*, 2741–2758. [[CrossRef](#)]
44. Sutton, A.J.; Abrams, K.R. Bayesian methods in meta-analysis and evidence synthesis. *Stat. Methods Med. Res.* **2001**, *10*, 277–303. [[CrossRef](#)]
45. Dias, S.; Ades, A.E. Absolute or relative effects? Arm-based synthesis of trial data. *Res. Synth. Methods* **2016**, *7*, 23–28. [[CrossRef](#)]
46. Hong, H.; Chu, H.; Zhang, J.; Carlin, B.P. Rejoinder to the discussion of “a Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons” by S. Dias and A. E. Ades. *Res. Synth. Methods* **2016**, *7*, 29–33. [[CrossRef](#)]

47. White, I.R.; Turner, R.M.; Karahalios, A.; Salanti, G. A comparison of arm-based and contrast-based models for network meta-analysis. *Stat. Med.* **2019**, *38*, 5197–5213. [[CrossRef](#)]
48. Turner, R.M.; Jackson, D.; Wei, Y.; Thompson, S.G.; Higgins, J.P.T. Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. *Stat. Med.* **2015**, *34*, 984–998. [[CrossRef](#)]
49. Elmariah, S.; Mauri, L.; Doros, G.; Galper, B.Z.; O'Neill, K.E.; Steg, P.G.; Kereiakes, D.J.; Yeh, R.W. Extended duration dual antiplatelet therapy and mortality: A systematic review and meta-analysis. *Lancet* **2015**, *385*, 792–798. [[CrossRef](#)]
50. Greco, T.; Landoni, G.; Biondi-Zoccai, G.; D'Ascenzo, F.; Zangrillo, A. A Bayesian network meta-analysis for binary outcome: How to do it. *Stat. Methods Med. Res.* **2016**, *25*, 1757–1773. [[CrossRef](#)]
51. Gelman, A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.* **2006**, *1*, 515–534. [[CrossRef](#)]
52. Hartling, L.; Fernandes, R.M.; Bialy, L.; Milne, A.; Johnson, D.; Plint, A.; Klassen, T.P.; Vandermeer, B. Steroids and bronchodilators for acute bronchiolitis in the first two years of life: Systematic review and meta-analysis. *BMJ* **2011**, *342*, d1714. [[CrossRef](#)] [[PubMed](#)]
53. Spiegelhalter, D.J.; Abrams, K.R.; Myles, J.P. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*; John Wiley & Sons: Chichester, UK, 2004.
54. Turner, R.M.; Davey, J.; Clarke, M.J.; Thompson, S.G.; Higgins, J.P.T. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *Int. J. Epidemiol.* **2012**, *41*, 818–827. [[CrossRef](#)]
55. Lamont, K.; Scott, N.W.; Jones, G.T.; Bhattacharya, S. Risk of recurrent stillbirth: Systematic review and meta-analysis. *BMJ* **2015**, *350*, h3080. [[CrossRef](#)]
56. Crocker, J.C.; Ricci-Cabello, I.; Parker, A.; Hirst, J.A.; Chant, A.; Petit-Zeman, S.; Evans, D.; Rees, S. Impact of patient and public involvement on enrolment and retention in clinical trials: Systematic review and meta-analysis. *BMJ* **2018**, *363*, k4738. [[CrossRef](#)] [[PubMed](#)]
57. Baxi, S.; Yang, A.; Gennarelli, R.L.; Khan, N.; Wang, Z.; Boyce, L.; Korenstein, D. Immune-related adverse events for anti-PD-1 and anti-PD-L1 drugs: Systematic review and meta-analysis. *BMJ* **2018**, *360*, k793. [[CrossRef](#)]
58. Martineau, A.R.; Jolliffe, D.A.; Hooper, R.L.; Greenberg, L.; Aloia, J.F.; Bergman, P.; Dubnov-Raz, G.; Esposito, S.; Ganmaa, D.; Ginde, A.A.; et al. Vitamin D supplementation to prevent acute respiratory tract infections: Systematic review and meta-analysis of individual participant data. *BMJ* **2017**, *356*, i6583. [[CrossRef](#)]
59. Normand, S.-L.T. Meta-analysis: Formulating, evaluating, combining, and reporting. *Stat. Med.* **1999**, *18*, 321–359. [[CrossRef](#)]
60. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* **2010**, *36*, 3. [[CrossRef](#)]
61. Viechtbauer, W. Confidence intervals for the amount of heterogeneity in meta-analysis. *Stat. Med.* **2007**, *26*, 37–52. [[CrossRef](#)]
62. Veroniki, A.A.; Jackson, D.; Viechtbauer, W.; Bender, R.; Bowden, J.; Knapp, G.; Kuss, O.; Higgins, J.P.T.; Langan, D.; Salanti, G. Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res. Synth. Methods* **2016**, *7*, 55–79. [[CrossRef](#)] [[PubMed](#)]
63. Ju, K.; Lin, L.; Chu, H.; Cheng, L.-L.; Xu, C. Laplace approximation, penalized quasi-likelihood, and adaptive Gauss–Hermite quadrature for generalized linear mixed models: Towards meta-analysis of binary outcome with sparse data. *BMC Med. Res. Methodol.* **2020**, *20*, 152. [[CrossRef](#)] [[PubMed](#)]
64. Bender, R.; Friede, T.; Koch, A.; Kuss, O.; Schlattmann, P.; Schwarzer, G.; Skipka, G. Methods for evidence synthesis in the case of very few studies. *Res. Synth. Methods* **2018**, *9*, 382–392. [[CrossRef](#)] [[PubMed](#)]
65. Michael, H.; Thornton, S.; Xie, M.; Tian, L. Exact inference on the random-effects model for meta-analyses with few studies. *Biometrics* **2019**, *75*, 485–493. [[CrossRef](#)] [[PubMed](#)]
66. Seide, S.E.; Röver, C.; Friede, T. Likelihood-based random-effects meta-analysis with few studies: Empirical and simulation studies. *BMC Med. Res. Methodol.* **2019**, *19*, 16. [[CrossRef](#)] [[PubMed](#)]
67. Röver, C.; Knapp, G.; Friede, T. Hartung-Knapp-Sidik-Jonkman approach and its modification for random-effects meta-analysis with few studies. *BMC Med. Res. Methodol.* **2015**, *15*, 99. [[CrossRef](#)] [[PubMed](#)]
68. Günhan, B.K.; Röver, C.; Friede, T. Random-effects meta-analysis of few studies involving rare events. *Res. Synth. Methods* **2020**, *11*, 74–90. [[CrossRef](#)]
69. Cai, T.; Parast, L.; Ryan, L. Meta-analysis for rare events. *Stat. Med.* **2010**, *29*, 2078–2089. [[CrossRef](#)]
70. Friede, T.; Röver, C.; Wandel, S.; Neuenschwander, B. Meta-analysis of two studies in the presence of heterogeneity with applications in rare diseases. *Biom. J.* **2017**, *59*, 658–671. [[CrossRef](#)] [[PubMed](#)]
71. Gronsbell, J.; Hong, C.; Nie, L.; Lu, Y.; Tian, L. Exact inference for the random-effect model for meta-analyses with rare events. *Stat. Med.* **2020**, *39*, 252–264. [[CrossRef](#)]
72. Ren, Y.; Lin, L.; Lian, Q.; Zou, H.; Chu, H. Real-world performance of meta-analysis methods for rare events using the Cochrane Database of Systematic Reviews. *J. Gen. Intern. Med.* **2019**, *34*, 960–968. [[CrossRef](#)]
73. Xu, C.; Furuya-Kanamori, L.; Zorzela, L.; Lin, L.; Vohra, S. A proposed framework to guide evidence synthesis practice for meta-analysis with zero-events studies. *J. Clin. Epidemiol.* **2021**, *135*, 70–78. [[CrossRef](#)] [[PubMed](#)]
74. Hong, H.; Wang, C.; Rosner, G.L. Meta-analysis of rare adverse events in randomized clinical trials: Bayesian and frequentist methods. *Clin. Trials* **2021**, *18*, 3–16. [[CrossRef](#)]
75. Efthimiou, O. Practical guide to the meta-analysis of rare events. *Evid. Based Ment. Health* **2018**, *21*, 72–76. [[CrossRef](#)]

76. Riley, R.D.; Simmonds, M.C.; Look, M.P. Evidence synthesis combining individual patient data and aggregate data: A systematic review identified current practice and possible methods. *J. Clin. Epidemiol.* **2007**, *60*, 431–439. [[CrossRef](#)]
77. Riley, R.D.; Lambert, P.C.; Abo-Zaid, G. Meta-analysis of individual participant data: Rationale, conduct, and reporting. *BMJ* **2010**, *340*, c221. [[CrossRef](#)]
78. Tierney, J.F.; Fisher, D.J.; Burdett, S.; Stewart, L.A.; Parmar, M.K.B. Comparison of aggregate and individual participant data approaches to meta-analysis of randomised trials: An observational study. *PLOS Med.* **2020**, *17*, e1003019. [[CrossRef](#)]
79. Hong, H.; Carlin, B.P.; Shamliyan, T.A.; Wyman, J.F.; Ramakrishnan, R.; Sainfort, F.; Kane, R.L. Comparing Bayesian and frequentist approaches for multiple outcome mixed treatment comparisons. *Med. Decis. Mak.* **2013**, *33*, 702–714. [[CrossRef](#)] [[PubMed](#)]
80. Seide, S.E.; Jensen, K.; Kieser, M. A comparison of Bayesian and frequentist methods in random-effects network meta-analysis of binary data. *Res. Synth. Methods* **2020**, *11*, 363–378. [[CrossRef](#)] [[PubMed](#)]
81. Thompson, C.G.; Semma, B. An alternative approach to frequentist meta-analysis: A demonstration of Bayesian meta-analysis in adolescent development research. *J. Adolesc.* **2020**, *82*, 86–102. [[CrossRef](#)] [[PubMed](#)]
82. Bennett, M.M.; Crowe, B.J.; Price, K.L.; Stamey, J.D.; Seaman, J.W. Comparison of Bayesian and frequentist meta-analytical approaches for analyzing time to event data. *J. Biopharm. Stat.* **2013**, *23*, 129–145. [[CrossRef](#)]
83. Pappalardo, P.; Ogle, K.; Hamman, E.A.; Bence, J.R.; Hungate, B.A.; Osenberg, C.W. Comparing traditional and Bayesian approaches to ecological meta-analysis. *Methods Ecol. Evol.* **2020**, *11*, 1286–1295. [[CrossRef](#)]
84. Weber, F.; Knapp, G.; Glass, Å.; Kundt, G.; Ickstadt, K. Interval estimation of the overall treatment effect in random-effects meta-analyses: Recommendations from a simulation study comparing frequentist, Bayesian, and bootstrap methods. *Res. Synth. Methods* **2021**, in press.
85. Rhodes, K.M.; Turner, R.M.; Higgins, J.P.T. Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data. *J. Clin. Epidemiol.* **2015**, *68*, 52–60. [[CrossRef](#)] [[PubMed](#)]