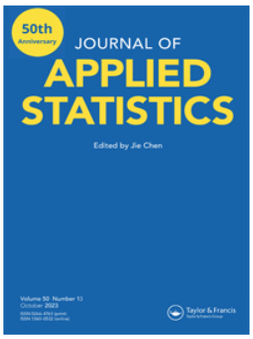


Journal of Applied Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cjas20

Bayesian heterogeneity in a meta–analysis with two studies and binary data

M. Martel, M. A. Negrín & F. J. Vázquez–Polo

To cite this article: M. Martel, M. A. Negrín & F. J. Vázquez–Polo (2023) Bayesian heterogeneity in a meta–analysis with two studies and binary data, Journal of Applied Statistics, 50:13, 2760-2776, DOI: 10.1080/02664763.2022.2084719

To link to this article: https://doi.org/10.1080/02664763.2022.2084719

JOURNAL OF APPLIED STATISTICS
2023, VOL. 50, NO. 13, 2760–2776
https://doi.org/10.1080/02664763.2022.2084719

Taylor & Francis
Taylor & Francis Group

Bayesian heterogeneity in a meta–analysis with two studies and binary data

M. Martel, M. A. Negrín and F. J. Vázquez–Polo

Dpt. of Quantitative Methods and TiDES Institute, U. of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Canary Islands, Spain

ABSTRACT

The meta–analysis of two trials is valuable in many practical situations, such as studies of rare and/or orphan diseases focussed on a single intervention. In this context, additional concerns, like small sample size and/or heterogeneity in the results obtained, might make standard frequentist and Bayesian techniques inappropriate. In a meta–analysis, moreover, the presence of between–sample heterogeneity adds model uncertainty, which must be taken into consideration when drawing inferences. We suggest that the most appropriate way to measure this heterogeneity is by clustering the samples and then determining the posterior probability of the cluster models. The meta–inference is obtained as a mixture of all the meta–inferences for the cluster models, where the mixing distribution is the posterior model probability. We present a simple two–component form of Bayesian model averaging that is unaffected by characteristics such as small study size or zero–cell counts, and which is capable of incorporating uncertainties into the estimation process. Illustrative examples are given and analysed, using real sparse binomial data.

ARTICLE HISTORY
Received 21 October 2021
Accepted 24 May 2022

KEYWORDS
Bayesian model averaging (BMA); binomial data; heterogeneity; meta–analysis; sparse data; two studies

MATHEMATICS SUBJECT CLASSIFICATIONS
62C10; 62F15; 90B50; 91C20

1. Introduction

The main role of meta–analysis is to summarise data from various clinical studies using ad–hoc statistical procedures. One area where a meta–analysis may usefully be conducted is that of rare or orphan diseases, about which few studies have been conducted and where research findings may be heterogeneous. In this context, heterogeneity is understood as the statistical variation present in the collected effect size data, to be identified and analysed after pooling the information drawn from the studies included [33]. In considering Bayesian random–effects for meta–analyses, it is assumed that for each study i with sample size n_i, the observed effect x_i follows a normal distribution with mean parameter θ_i (the treatment effectiveness conditional on study i) and variance τ_i. In addition, the true treatment effect θ_i in trial i is normally distributed $\mathcal{N}(\theta_i \mid \theta, \tau)$, where θ, the meta–parameter, is the true effect of the treatment in question. The *link* distribution $\mathcal{N}(\theta_i \mid \theta, \tau)$ accounts for the uncertainty on θ_i around θ for each study i, and τ for the biased assessment

CONTACT F. J. Vázquez–Polo francisco.vazquezpolo@ulpgc.es

Supplemental data for this article can be accessed here. https://doi.org/10.1080/02664763.2022.2084719

of the treatment effect. Finally, priors on θ and τ are obtained in order to perform a Bayesian estimation for the parameters of interest. The question of the link distribution has received little previous attention in the literature, but it is important to note that it must be coherent with the specified marginals and be able to accommodate considerable between–study heterogeneity. When effectiveness is measured by a discrete 0–1 random variable, this hierarchical normal model is typically applied to the logit transformation of the data $\log(x_i/(n_i - x_i))$ with the reparametrisation $\log(\theta_i/(1 - \theta_i))$ [3,4,11,25,36]. As observed by Thomas *et al.* [37] the performance of a meta–analysis regarding binary data has received much attention. This question is of interest for several reasons. On the one hand, effectiveness is commonly measured in terms of a binary variable according to whether or not a specific objective has been achieved (in the healthcare context, for example, objectives such as survival, non–relapse or achieving a low viral load). Charles *et al.* [6] found that fully half of all trials calculated the necessary sample size according to stipulated binary outcomes. It is important to note that different statistical considerations must be made for binary outcomes than for continuous ones. If the samples contain zeros, a fixed data continuity correction is normally applied, after which a logit approximation is made. However, this practice has been criticised by Sweeting *et al.* [36], who proposed an alternative empirical data correction. Friedrich *et al.* [16] showed that significant aspects of the meta–analysis could be invalidated if trials with no outcome events in the treatment or control arms are either not considered or are modified by the common continuity correction of 0.5. The same problem may arise when the meta–analysis contains a double zero, i.e. when both the treatment and the control arms contain zero events. Weber *et al.* [42] recommended that zero–cell corrections should be avoided due to the unsatisfactory statistical properties thus produced. However, continuity corrections are not the only way option, and other forms of meta–analysis have been proposed, such as those based on binomial models, as described by Günhan *et al.* [18]. IntHout *et al.* [24] noted that small studies present more heterogeneity than large ones. Moreover, between–sample heterogeneity introduces model uncertainty into the process. Rücker *et al.* [33] observed that statistical heterogeneity and small–study effects are major issues that can impair the validity of a meta–analysis. Furthermore, it can be difficult to estimate heterogeneity when the analysis is based on a small number of studies. This could give rise to model uncertainty and hence produce misleading results. In view of these considerations, we believe the degree of heterogeneity present, i.e. the statistical variation in the effect size data, must be identified and assessed to ensure the validity of the meta–analysis performed [33]. The statistical heterogeneity in a meta–analysis is usually determined by one of the following approaches

(a) by estimating the between–study variation, which is characterised as the variance, or heterogeneity parameter τ. Spiegelhalter *et al.* [35] proposed the following ranges of typical τ threshold values for heterogeneity: 0.25 (moderate), 0.5 (substantial), 1 (large) and 2 (very large);
(b) alternatively, by testing the null hypothesis that the true treatment effects across studies are identical [3,5], using the test–statistics Q, or the sum of squared deviations of all effects about the mean, on a standardised scale, given by $Q = \sum_{i=1}^{k}(x_i - \theta_{\text{fe}})^2/v_i$, where θ_{fe} is the mean effect size using fixed effect weights, and v_i is the square standard error of the ith study (assumed to be known). Under the null hypothesis of homogeneity, Q follows a χ^2–distribution with $k-1$ degrees of freedom i.e. the number of studies minus 1). As the Q test only informs practitioners about the presence or absence of

heterogeneity, alternative measures such as the $I^2 = 100 \times (Q - (k-1))/Q$ index have been proposed to quantify the degree of heterogeneity in a meta–analysis, i.e. the total inter–study variation attributable to heterogeneity (Higgins *et al.* [22]). I^2 ranges from 0 to 100%, and a practical guide offered by Higgins *et al.* [23] assigned values of 25%, 50% and 75% to represent low, moderate and high levels of heterogeneity, respectively, and considered a significant degree of heterogeneity to be present when $I^2 > 50\%$.

However, both of these approaches present certain problems when the meta–analysis is based on a small number of studies. This is especially serious when there are only two studies. Thus, the intensive simulations described by Chuang *et al.* [7] suggest that estimates of between–study variance are inaccurate for meta–analyses based on small numbers of studies.

Another important consideration is that tests of heterogeneity have relatively little detection power when data are sparse and/or the meta–analysis is based on a small number of studies [15,27,33]. Such a situation commonly arises, for example, in analyses concerning rare or orphan diseases, in which zero values are often observed in both the trial and the treatment arms considered and in many cases the meta–analysis is based on just two studies.

The problem of statistical heterogeneity in a meta–analysis based on just two studies has been highlighted by Friede *et al.* [15], IntHout *et al.* [24] and Pateras *et al.* [30] among many others. The inherent difficulty in this situation is heightened in small–scale trials by the presence of zero–cell counts, i.e. the non–occurrence of the event being investigated. In such cases, significant between–study heterogeneity is very likely. If this is not properly accounted for, the research conclusions drawn will be unreliable [36,41,42]. Gonnermann *et al.* [17] observed that, in the presence of heterogeneity, the meta–analysis of two studies remains an 'unsolved' problem. The question of meta–analyses based on only two studies has been addressed by the European Medicines Agency, [13,14] in its guides to statistical principles for clinical trials. Among many randomised controlled trials located on the Cochrane Review Database, Crins *et al.* [8] and Miller *et al.* [26] identified real–life datasets on which statistical methods for two–study meta–analyses have been utilised (Friede *et al.* [15]).

1.1. A motivating example

Extracted from Friede *et al.* [15], the data shown in Table 1 correspond to the comparison between treatment (85 patients were treated with Krystexxa 8 mg every 2 weeks) and control (placebo), followed by an analysis of the safety endpoint (infusion reaction), based on US Food and Drug Administration approval of Krystexxa, a treatment for chronic gout in adult patients refractory to conventional treatment, in the context of orphan diseases.

Table 1. Krystexxa dataset.

	Treatment		Control	
Study	Events	Total	Events	Total
Study C405 (x_1)	11	43	1	20
Study C406 (x_2)	11	42	1	23

Although these data correspond to a case–study in which there are no zero cells, an initial consideration is useful to understand our approach to heterogeneity. Section 3 provides a comprehensive analysis of this case–study and of two others, containing single and double–zeros. In a study based on frequentist techniques, Friede *et al.* [15] found no evidence excluding the presence of homogeneity, with estimated $\hat{\tau} = 0.00$, and $I^2 = 0.0\%$. However, this simplistic 0–1 decision (accepting or rejecting the presence of homogeneity) ignores the real import of the heterogeneity within the dataset. As a result, misleading inferences may be drawn. Our paper proposes a statistical model for meta–analyses that may contain zeros and be based on only two studies. In this model: (1) no logit transformation is made of the data and parameters; (2) any uncertainty arising from between–sample heterogeneity is quantified. We suggest these goals can be achieved by means of Bayesian clustering.

The data considered in the meta–analysis of two studies consist of $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$ with $\boldsymbol{x}_i = (x_i, n_i),\ i = 1, 2$. When only two studies are considered, there may be only two data–based heterogeneity structures: (i) the case in which the data observed are from the same sampling model (i.e. homogeneity, in notation $\{\boldsymbol{x}_1\boldsymbol{x}_2\}$) or (ii) the data are from two different sampling models (i.e. heterogeneity, in notation $\{\boldsymbol{x}_1\}\{\boldsymbol{x}_2\}$).

This induced between–sample heterogeneity could be viewed as a simple clustering problem where the model uncertainty is included in the inference, using a two–component mixture model (or equivalently, a two–component Bayesian model averaging).

Obtaining the likelihood function using $\boldsymbol{x}$ is a major problem, because the sample information $\boldsymbol{x}_i$ from study i is related to θ_i but not to the parameter of interest θ. Therefore, the likelihood function required for estimating θ must be strongly dependent on the cluster structure of the samples.

In this context, we compute the posterior probabilities of each cluster structure (representing heterogeneity or homogeneity), and then obtain the meta–inference as a mixture of those derived from the cluster models, with the posterior model probabilities being used as a mixing distribution. Unlike previous work in this field, in which studies have analysed between–study heterogeneity for a given quantity such as mean difference, relative risk or odds ratio, we analyse each treatment separately, thus enabling the heterogeneity structure to vary between the intervention and the control.

The rest of this paper is organised as follows. In order to overcome possible problems with samples containing zero events, as well as the need to perform continuity corrections [4,18,24,36], we first introduce the Bayesian binomial model. In this methodological section, we consider clustering as a possible approach to managing sampling heterogeneity in a meta–analysis of two studies. We then develop the likelihood function needed to draw inferences about the parameter of interest, when cluster structures are considered. Next, Section 3 presents some examples with real data, applying the proposed methodology. Finally, some brief conclusions are drawn.

2. Statistical modelling

In this section, a Bayesian binomial approach is taken to perform a meta–analysis of two studies. For each study i, we have the Binomial–Beta model given by

$$\begin{aligned} X_i &\sim \text{Bin}(x_i \mid n_i, \theta_i), \quad x_i = 0, \ldots, n_i, \\ \theta_i &\sim \text{Beta}(\theta_i \mid 1, 1), \quad i = 1, 2, \end{aligned} \tag{1}$$

where θ_i represents the treatment effectiveness, n_i the number of patients, and x_i the number of successful treatments, conditional on the study $i = 1, 2$, such that

$$\Pr(x_i \mid n_i, \theta_i) = \binom{n_i}{x_i} \theta_i^{x_i} (1 - \theta_i)^{n_i - x_i}, \quad x_i = 0, \ldots, n_i. \tag{2}$$

In most situations, the data contain a preponderance of zeros, and therefore we assume that the prior information on the conditional effectiveness θ_i is weak and assign the uniform prior $\text{Beta}(\theta_i \mid 1, 1)$.

Although Jeffreys and Haldane distributions are possible alternatives that are widely accepted as objective priors for the Bernoulli parameter, in the presence of sparse data, some arguments favour the use of uniform ones, for instance, when there is a preponderance of zeros. Thus, Tuyl *et al.* [38] stated, 'the Jeffreys prior $\text{Beta}(\theta \mid 1/2, 1/2)$ can be too informative concentrating too much probability mass close to 0, thus suppressing the importance of the observed data'. On the other hand, the Haldane prior is $\text{Beta}(\theta \mid 0, 0)$ and improper. Such a link function is not desirable in the meta–analysis context. Moreover, when the data are equal to 0, the Haldane posterior is in fact improper which, as Bernardo [2] noted, 'is less than adequate'. Posterior predictive arguments in favour of the $\text{Beta}(\theta \mid 1, 1)$ are also given in Tuyl *et al.* [39].

Let us now consider a binary 'latent' variable X, representing a treatment result obtained in a virtual study. Obviously, the distribution assigned to X is of the same type as that of X_i, i.e. we have the Bernoulli meta–model $\text{Ber}(x \mid \theta)$, where θ is the true (unconditional) treatment outcome, which is not affected by between–study variability. A Bernoulli meta–model is induced by

$$\Pr(x \mid \theta) = \theta^x (1 - \theta)^{1-x}, \quad \text{and} \quad \pi(\theta) = \text{Beta}(\theta \mid 1, 1). \tag{3}$$

As usual, two arms are compared in each study (labelled treatment (T) and control/placebo (C), respectively). The proposed model is analogous in each case, and so for simplicity we omit the subindex describing the arm in each study. In practical applications with a real data set, a subindex notation will be used in each meta–analysis. The Bayesian meta–analysis is then based on the posterior distribution of the parameter θ, which is derived via Bayes' theorem and given by

$$\pi(\theta \mid \boldsymbol{x}) \propto f(\boldsymbol{x} \mid \theta)\pi(\theta), \quad 0 < \theta < 1, \tag{4}$$

where $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$ with $\boldsymbol{x}_i = (x_i, n_i)$, $i = 1, 2$, and $f(\boldsymbol{x} \mid \theta)$ denotes the likelihood function for estimating θ given the observed data. However, the likelihood function in (4) cannot be obtained with the information on sample $\boldsymbol{x}_i$ on study i, which is related to θ_i but not to θ. Therefore, further steps are required to derive an appropriate likelihood function.

2.1. The linking distribution

A distribution $\pi(\theta_i \mid \theta)$ is needed to link the experimental parameters θ_i with the meta–parameter θ. This hierarchical structure implies a crucial consideration in choosing a linking distribution, one that is sometimes overlooked by practitioners, namely that this linking distribution must ensure there is coherence between the conditional and marginal distributions of the experimental parameter and the meta–parameter. Mathematically, this

requires that the corresponding bivariate distribution belong to the class of bivariate distributions with given marginals, i.e. the bivariate distribution $\pi(\theta_i, \theta) = \pi(\theta_i \mid \theta)\pi(\theta)$ must satisfy the integral equations

$$\int_0^1 \pi(\theta_i, \theta)\, \mathrm{d}\theta_i = \pi(\theta) \quad \text{and} \quad \int_0^1 \pi(\theta_i, \theta)\, \mathrm{d}\theta = \pi(\theta_i), \quad i = 1, 2. \tag{5}$$

Following Moreno *et al.* [28] we consider the conditional intrinsic linking distributions given by

$$\pi(\theta_i \mid \theta) = \sum_{z=0}^{t} \mathrm{Bin}(z \mid t, \theta) \times \mathrm{Beta}(\theta_i \mid z + 1, t - z + 1), \tag{6}$$

that is, a finite mixture of Beta distributions with binomial weights, where t is an integer representing the (imaginary) training sample size used in creating the link distribution. The mean and variance of this distribution are $\mathbb{E}(\theta_i \mid \theta) = \frac{1+t\theta}{2+t}$, and $\mathbb{V}(\theta_i \mid \theta) = \frac{(1+t)(1-2t\theta(\theta-1))}{(2+t)^2(3+t)}$, respectively. Given that the mean converges to θ as $t \to \infty$ and the variance vanishes at infinity, the hyperparameter t indicates how strongly the conditional distribution $\pi(\theta_i \mid \theta)$ concentrates mass around θ. Note that the correlation coefficient between θ_i and θ is $t/t + 1$. In practice, this hyperparameter t is fixed, assuming a large enough correlation. The bidimensional prior $\pi(\theta_i, \theta) = \pi(\theta_i \mid \theta) \times \mathrm{Beta}(\theta \mid 1, 1)$ has two interesting properties. Firstly, it satisfies equations (5) for any t (i.e. it belongs to the Fréchet class with uniform marginals), and therefore, the linking class of intrinsic distributions and the Bayesian models (2) and (3) are coherent. Furthermore, the concentration of $\pi(\theta_i \mid \theta)$ around θ is determined by the size of the training sample t. Thus, the larger the t the greater the concentration.

Assuming that θ_1 and θ_2 are conditional independent given θ, the joint linking distribution of (θ_1, θ_2) conditional on θ is given by

$$\pi(\theta_1, \theta_2 \mid \theta) = \pi(\theta_1 \mid \theta) \times \pi(\theta_2 \mid \theta). \tag{7}$$

Finally, observe that the proposed intrinsic linking distribution of θ_i, conditional on θ, does not require the use of the continuity correction and logit transformation for the sparse data, as is the case with standard random–effect models, whereby several concerns might arise.

2.2. The likelihood

It is clear that the likelihood function strongly depends on the cluster structure of the samples. As in Moreno *et al.* [29] and Vázquez–Polo *et al.* [40], we adopt the following definition of a cluster: two samples $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, from $f(x \mid \theta_1)$ and $f(x \mid \theta_2)$, respectively, are in the same cluster if $\theta_1 = \theta_2$.

When there are only two studies in the meta–analysis, only two cluster structures are possible: either i) $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are in the same cluster, with the notation $\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$, which corresponds to a sample structure of homogeneity; or ii) $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are in different clusters, with the notation $\{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\}$, which corresponds to sample heterogeneity. Therefore, for the data $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$, we need to obtain the likelihood for θ conditional on a given

cluster model. Each cluster model indicates a different heterogeneity structure of the sampling model for $\boldsymbol{x}$, and the posterior probability informs us about the uncertainty for this structure. Finally, the likelihood for estimating θ conditional on data $\boldsymbol{x}$ is obtained as a two–component mixture of the above conditional likelihood functions. Following the partition model approach introduced by Barry and Hartigan [1], the cluster model given by the homogeneity structure $\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$ has the sampling distribution

$$f(\boldsymbol{x} \mid \{\boldsymbol{x}_1, \boldsymbol{x}_2\}, \theta_1) = \binom{n_1 + n_2}{x_1 + x_2} \theta_1^{x_1+x_2} (1-\theta_1)^{n_1+n_2-(x_1+x_2)}, \tag{8}$$

and the heterogeneity configuration $\{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\}$ has the corresponding sampling model

$$f(\boldsymbol{x} \mid \{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\}, (\theta_1, \theta_2)) = \binom{n_1}{x_1} \theta_1^{x_1} (1-\theta_1)^{n_1-x_1} \binom{n_2}{x_2} \theta_2^{x_2} (1-\theta_2)^{n_2-x_2}. \tag{9}$$

From this, we can obtain the likelihood, conditional on each cluster model:

(i) For the homogeneity configuration, we have

$$f(\boldsymbol{x} \mid \{\boldsymbol{x}_1, \boldsymbol{x}_2\}, \theta) = \int_0^1 \binom{n_1 + n_2}{x_1 + x_2} \theta_1^{x_1+x_2} (1-\theta_1)^{n_1+n_2-(x_1+x_2)} \pi(\theta_1 \mid \theta)\, \mathrm{d}\theta_1, \tag{10}$$

where $\pi(\theta_1 \mid \theta)$ is given in (6). After some algebra, expression (10) can be rewritten as

$$\begin{aligned} f(\boldsymbol{x} \mid \{\boldsymbol{x}_1, \boldsymbol{x}_2\}, \theta) = {} & \frac{\Gamma(x_1 + x_2 + 1)\Gamma(n_1 + n_2 + t - x_1 - x_2 + 1)}{\Gamma(n_1 + n_2 + t + 2)} \\ & \times (1+t)(1-\theta)^t {}_3F_2\left(\mathbf{a}, \mathbf{b}, \frac{\theta}{\theta - 1}\right), \end{aligned} \tag{11}$$

where ${}_3F_2(\mathbf{a}, \mathbf{b}, z)$ denotes the generalised hypergeometric function with argument z and vector parameters $\mathbf{a}$ and $\mathbf{b}$ of dimensions 3 and 2, respectively, with $\mathbf{a} = (-t, -t, x_1 + x_2 + 1)$ and $\mathbf{b} = (1, -n_1 - n_2 - t + x_1 + x_2)$, and

(ii) for the heterogeneity configuration:

$$\begin{aligned} & f(\boldsymbol{x} \mid \{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\}, \theta) \\ & = \int_0^1 \binom{n_1}{x_1} \theta_1^{x_1} (1-\theta_1)^{n_1-x_1} \pi(\theta_1 \mid \theta)\, \mathrm{d}\theta_1 \\ & \quad \times \int_0^1 \binom{n_2}{x_2} \theta_2^{x_2} (1-\theta_2)^{n_2-x_2} \pi(\theta_2 \mid \theta)\, \mathrm{d}\theta_2 \\ & = (1+t)^2 (1-\theta)^{2t} \frac{\Gamma(x_1 + 1)\Gamma(n_1 + t - x_1 + 1)}{\Gamma(n_1 + t + 2)} {}_3F_2\left(\mathbf{a}_1, \mathbf{b}_1, \frac{\theta}{\theta - 1}\right) \\ & \quad \times \frac{\Gamma(x_2 + 1)\Gamma(n_2 + t - x_2 + 1)}{\Gamma(n_2 + t + 2)} {}_3F_2\left(\mathbf{a}_2, \mathbf{b}_2, \frac{\theta}{\theta - 1}\right), \end{aligned} \tag{12}$$

where $\mathbf{a}_i = (-t, -t, x_i + 1)$ and $\mathbf{b}_i = (1, -n_i - t + x_i)$, for $i = 1, 2$.

Finally, to derive the likelihood function we sum (11) and (12) with respect to a discrete prior on the cluster structures. An objective prior on the space of heterogeneity structures

is the discrete uniform prior. Thus, the (unconditional) likelihood for estimating θ for the data $\boldsymbol{x}$ is given by

$$f(\boldsymbol{x} \mid \theta) = \frac{1}{2} f(\boldsymbol{x} \mid \{\boldsymbol{x}_1, \boldsymbol{x}_2\}, \theta) + \frac{1}{2} f(\boldsymbol{x} \mid \{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\}, \theta), \tag{13}$$

and the posterior distribution in (4) is obtained via Bayes' theorem by $\pi(\theta \mid \boldsymbol{x}) \propto \pi(\theta) f(\boldsymbol{x} \mid \theta)$.

2.3. The posterior distributions

Given that our approach is based on a simple two–component mixture and averaging the two cluster configurations, when making inferences about the parameter of interest, the required posterior distribution can also be obtained as a mixture of posterior distributions over each cluster configuration. For instance, if θ is the treatment effect under the control treatment, its posterior distribution, given data $\boldsymbol{x}$, is given by

$$\pi(\theta \mid \boldsymbol{x}) = \pi(\theta \mid \{\boldsymbol{x}_1, \boldsymbol{x}_2\}) \pi(\{\boldsymbol{x}_1, \boldsymbol{x}_2\} \mid \boldsymbol{x}) + \pi(\theta \mid \{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\}) \pi(\{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\} \mid \boldsymbol{x}) \tag{14}$$

where

$$\pi(\{\boldsymbol{x}_1, \boldsymbol{x}_2\} \mid \boldsymbol{x}) = \frac{m(\boldsymbol{x} \mid \{\boldsymbol{x}_1, \boldsymbol{x}_2\})}{m(\boldsymbol{x} \mid \{\boldsymbol{x}_1, \boldsymbol{x}_2\}) + m(\boldsymbol{x} \mid \{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\})}, \tag{15}$$

and $\pi(\{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\} \mid \boldsymbol{x}) = 1 - \pi(\{\boldsymbol{x}_1, \boldsymbol{x}_2\} \mid \boldsymbol{x})$, represent the (posterior) weights in the mixture of each cluster configuration, with

$$m(\boldsymbol{x} \mid \{\boldsymbol{x}_1, \boldsymbol{x}_2\}) = \int_0^1 f(\boldsymbol{x} \mid \{\boldsymbol{x}_1, \boldsymbol{x}_2\}, \theta_1) \pi(\theta_1 \mid \{\boldsymbol{x}_1, \boldsymbol{x}_2\}) \, \mathrm{d}\theta_1 \tag{16}$$

$$m(\boldsymbol{x} \mid \{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\}) = \int_0^1 \int_0^1 f(\boldsymbol{x} \mid \{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\}, (\theta_1, \theta_2)) \pi(\theta_1, \theta_2 \mid \{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\}) \, \mathrm{d}\theta_1 \, \mathrm{d}\theta_2 \tag{17}$$

where $f(\boldsymbol{x} \mid \{\boldsymbol{x}_1, \boldsymbol{x}_2\}, \theta_1)$ and $f(\boldsymbol{x} \mid \{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\}, (\theta_1, \theta_2))$ as in (8) and (9), respectively, and $\pi(\theta_1 \mid \{\boldsymbol{x}_1, \boldsymbol{x}_2\})$ and $\pi(\theta_1, \theta_2 \mid \{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\})$ are the intrinsic priors for the corresponding cluster configuration, i.e.

$$\pi(\theta_1 \mid \{\boldsymbol{x}_1, \boldsymbol{x}_2\}) = \int_0^1 \pi(\theta_1 \mid \theta) \pi(\theta) \, \mathrm{d}\theta,$$

and

$$\pi(\theta_1, \theta_2 \mid \{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\}) = \int_0^1 \pi(\theta_1 \mid \theta) \pi(\theta) \, \mathrm{d}\theta \times \int_0^1 \pi(\theta_2 \mid \theta) \pi(\theta) \, \mathrm{d}\theta,$$

The posterior distributions of each cluster configuration in (14), since $\pi(\theta) = \mathrm{Beta}(\theta \mid 1, 1)$, are given by

$$\pi(\theta \mid \{\boldsymbol{x}_1, \boldsymbol{x}_2\}) = \frac{f(\boldsymbol{x} \mid \{\boldsymbol{x}_1, \boldsymbol{x}_2\}, \theta)}{\int_0^1 f(\boldsymbol{x} \mid \{\boldsymbol{x}_1, \boldsymbol{x}_2\}, \theta) \, \mathrm{d}\theta}, \tag{18}$$

and

$$\pi(\theta \mid \{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\}) = \frac{f(\boldsymbol{x} \mid \{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\}, \theta)}{\int_0^1 f(\boldsymbol{x} \mid \{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\}, \theta)\, \mathrm{d}\theta}. \tag{19}$$

The same procedure is applied when θ represents the treatment effect under a new treatment to be compared. These posterior distributions are computed numerically over the parametric space $\Theta = (0, 1)$, using `Wolfram Mathematica`, which has a huge library of ready–to–use functions and the advantage of simple code. Furthermore, once posterior distributions are obtained, the command `ProbabilityDistribution` can be used to generate any sample of the posterior distribution of the parameter of interest, which is transformed as in the following illustrations.

3. Illustrations

To illustrate the arguments developed in the preceding section, we now revisit the motivating example presented in the first section and introduce two new real datasets. For these case studies, we assume in our examples that $t = 49$, which implies a correlation of 0.98, indicating that the intrinsic link distribution concentrates a considerable mass of probability around θ. Other values of t have also been used and the results obtained are very robust. The `Mathematica` code for the case–study data sets can be found in the Supplementary Material online section.

3.1. Motivating example revisited

Returning to the motivating example considered in Section 1, it can be seen from Table 2 that we have different posterior probabilities for the cluster structures and thus for the Bayesian model averaging of the true treatment effect under Treatment (T) or Control (C), θ_T and θ_C.

All quantities of interest can be obtained from the posterior mixture distribution in (14). Moreover, using (14), it is straightforward to simulate the posterior distribution of parameters θ_T and θ_C and their usual transformations, such as the risk and odds ratio, log odds ratio, etc. Table 3 shows the estimated values obtained for the treatment effects θ_T and θ_C and their 95% Bayesian intervals, using the proposed BMA approach. The odds ratio and its 95% Bayesian credible interval (95% CI) are also obtained.

Friede *et al.* [15] proposed point and interval estimators for the OR, based on the DerSimonian and Laird [12] approach (denoted by DL–Normal), the Hartung and Knapp [20,21] and Sidik and Jonkman [34] approach (DL–HKSJ) and the modified KH (DL–mKH) proposal [32]. The estimated odds ratios and 95% confidence intervals obtained are shown at

Table 2. Krystexxa cluster configuration.

	Cluster configuration			
	Treatment		Control	
	Cluster model	Posterior probability	Cluster model	Posterior probability
Homogeneity	$\{\boldsymbol{x}_1\boldsymbol{x}_2\}$	0.62	$\{\boldsymbol{x}_1\boldsymbol{x}_2\}$	0.57
Heterogeneity	$\{\boldsymbol{x}_1\}\{\boldsymbol{x}_2\}$	0.38	$\{\boldsymbol{x}_1\}\{\boldsymbol{x}_2\}$	0.43

Table 3. Estimated treatment effects and 95% Bayesian credible intervals in the Krystexxa dataset.

Treatment		Control	
$\mathbb{E}(\theta_T \mid \boldsymbol{x})$	95% CI	$\mathbb{E}(\theta_C \mid \boldsymbol{x})$	95% CI
0.27	(0.11, 0.46)	0.08	(0.00, 0.22)
OR			
$\mathbb{E}(\text{OR} \mid \boldsymbol{x})$		95% CI	
5.88		(0.87, 89.59)	
Estimated odds ratios and 95% confidence interval obtained in Friede *et al.* [15]			
DL–Normal ($\hat{\tau} = 0.00$):	7.14	(1.59, 32.01)	
KL–HKSJ ($\hat{\tau} = 0.00$):	7.14	(2.30, 22.18)	
DL–mKH ($\hat{\tau} = 0.00$):	7.14	(0.00, 119543.65)	

the end of Table 3. All three methods obtain an estimation of $\hat{\tau} = 0.00$, indicating there is no evidence suggesting we should reject the presence of homogeneity. However, the posterior probabilities shown in Table 2 indicate that other heterogeneity configurations should also be considered, specifically that of $\{\boldsymbol{x}_1\}\{\boldsymbol{x}_2\}$. This example highlights the differences arising with the above estimation (including the unrealistic mKH case).

In an alternative approach, Friede *et al.* [15] presented a Bayesian estimation model using half–normal priors, obtaining the following results: HNormal(1.00) ($\hat{\tau} = 0.55$) 7.14 (95% Highest Density Interval 1.04–49.15) and HNormal(0.50) 7.14 ($\hat{\tau} = 0.31$) (95% HDI 1.39–36.70), where HNormal(ν) refers to a half–normal distribution with scale parameter ν. These values of $\hat{\tau}$ are indicative of heterogeneity, and then the respective Bayesian intervals are much narrower. However, this ignores the case of homogeneity, which accounts for a large proportion of the posterior probability in both treatments. In consequence, misleading inferences may be drawn. In contrast to the frequentist and Bayesian estimations described above, the BMA interval for the OR, shown in Table 3, takes into account all of the uncertainty about between–study variability.

Finally, Figure 1 shows the posterior distribution of the log odds ratio, revealing a positively skewed behaviour that clearly indicates an increase in the infusion reaction ($\Pr(\log(\text{OR}) > 0 \mid \boldsymbol{x}) = \Pr(\text{OR} > 1 \mid \boldsymbol{x}) = 0.963$), as previously noted by Davi *et al.* [9].

3.2. A single zero example

Thrombo–embolic complications are a major cause of morbidity and mortality in hip fracture patients, a substantial proportion of whom will develop deep vein thrombosis. For this condition, the standard treatment is pharmacological thromboprophylaxis, while heparin and other anticoagulants are less commonly used.

The data set in this example is extracted from Handoll *et al.* [19] and corresponds to a meta–analysis conducted to investigate the use of prophylactic subcutaneous unfractionated or low–molecular weight heparin after hip fracture repair to prevent deep vein thrombosis in elderly patients. Among other results, Handoll *et al.* [19] concluded that '*data were insufficient to establish any effect on the incidence of fatal pulmonary embolism and overall mortality*'.

Table 4 contains data from a subanalysis conducted to compare LMW heparin versus control/placebo with the outcome of fatal pulmonary embolism. The table also presents

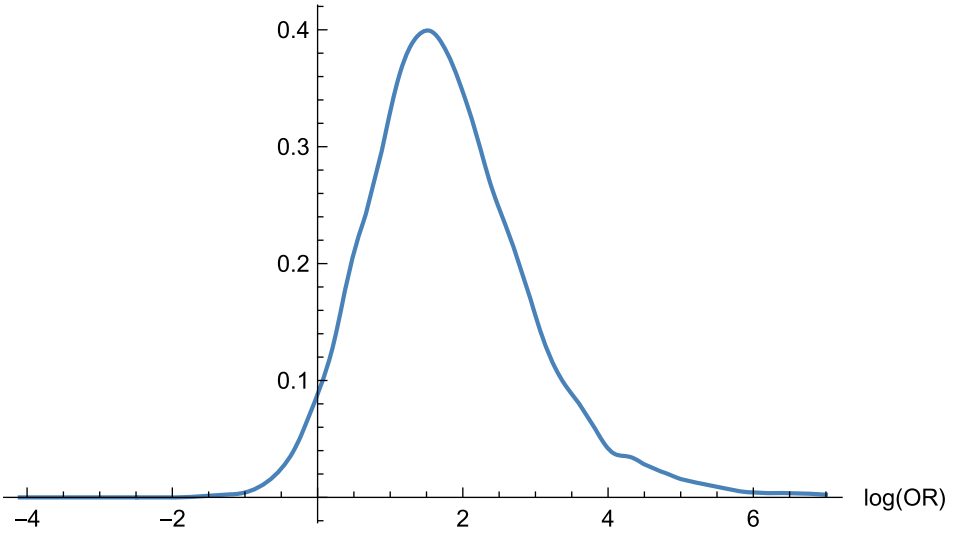


Figure 1. Posterior distribution of the log(OR) in the Krystexxa dataset.

Table 4. Dataset in Handoll *et al.* [19]: LMW heparin vs control/placebo, Outcome: fatal pulmonary embolism.

	Treatment		Control	
Study	Events	Total	Events	Total
Figuerido 1994 (x_1)	1	16	3	25
Jorgensen 1992 (x_2)	0	30	1	38
	Cluster configuration			
	Cluster model	Posterior probability	Cluster model	Posterior probability
	$\{x_1x_2\}$	0.53	$\{x_1x_2\}$	0.56
	$\{x_1\}\{x_2\}$	0.47	$\{x_1\}\{x_2\}$	0.44

the posterior probabilities of the cluster configurations obtained by the method we propose.

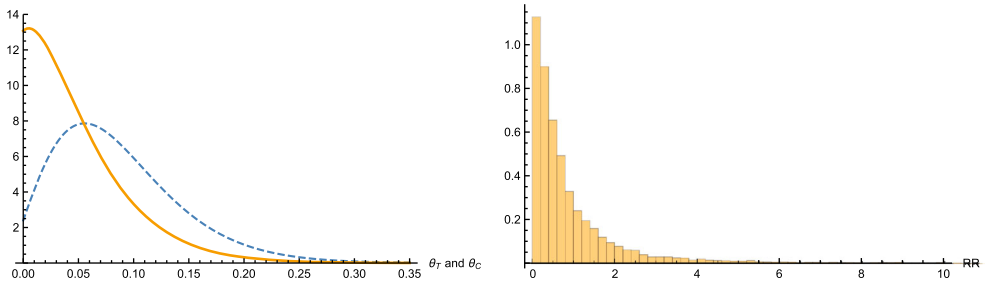
Table 4 shows there are different posterior probabilities for the cluster structures. In both arms, the homogeneity structure $\{x_1, x_1\}$ concentrates the highest posterior probability of being true, but failure to incorporate the heterogeneity structure in the analysis could produce misleading results since this configuration contains a large part of the model uncertainty (more than 0.4 in each case).

Table 5 shows the posterior summaries of the true treatment effect under Treatment (T) or Control (C), θ_T and θ_C, and their 99% Bayesian credible intervals (99% CI). The risk ratio (RR) and its 99% credible intervals are also obtained.

Figure 2 (left panel) shows the BMA posterior density of θ_T and θ_C. It can be seen that the posterior density of θ under the control treatment is slightly shifted towards higher values with respect to the LMW heparin treatment. Moreover, the sparsity of the data induces a reverse J–shaped posterior distribution for θ_T. On the other hand, Figure 2 (right panel) shows the posterior distribution of the risk ratio (RR), where the (posterior) probability $\Pr(RR < 1 \mid \boldsymbol{x}) = 0.71$, revealing an apparent probability in favour of heparin.

Table 5. Estimated treatment effects and 99% Bayesian intervals in the LMW heparin dataset.

Treatment		Control	
$\mathbb{E}(\theta_T \mid \boldsymbol{x})$	99% CI	$\mathbb{E}(\theta_C \mid \boldsymbol{x})$	99% CI
0.05	(0.00, 0.23)	0.09	(0.00, 0.28)
RR			
$\mathbb{E}(RR \mid \boldsymbol{x})$		99% CI	
0.97		(0.00, 8.21)	



(a) (b)

Figure 2. Posterior distributions in the Handoll *et al.* [19] dataset. (a) BMA posterior density of the parameters θ_T (continuous line) and θ_C (dashed line) and (b) Posterior distribution of the *RR*.

For the sake of comparison, we note the following. Handoll *et al.* [19] obtained a Mantel–Haenszel estimate for the risk ratio of 0.47 (99% CI, 0.08−2.90). These estimated values differ from those presented in Table 5, obtained by the methods we propose. Handoll *et al.* [19] found no evidence of heterogeneity with estimated $\hat{\tau} = 0$ while I^2 was estimated at 0.0%. However, these results are slightly different from those given in Table 4. Certainly, the homogeneity model has the largest posterior probability, for both arms, but a large proportion of the uncertainty accumulated in the remaining heterogeneity structures cannot be discounted as is the case with the BMA approach. In consequence, a wider interval is obtained by the mixture procedure containing all the associated uncertainty of the models.

3.3. A case with two and double–zero studies

The prescription of antibiotics for sore throat is an open question in medical practice. This condition is very common and usually remits spontaneously. Nevertheless, primary care physicians commonly prescribe antibiotics. The data set in Table 6 corresponds to a meta–analysis of the use of antibiotics to prevent rheumatic fever, seeking to assess their benefit in the management of sore throat, as reported by Del Mar *et al.* [10]. The data represent antibiotics versus control for the treatment of sore throat (with fever symptoms). The binary outcome is the detection of fever symptoms on day 3 (in children, compared to adults).

According to Del Mar *et al.* [10], the two studies have the following Peto ORs: OR and 95% confidence interval cannot be estimated for the Krober study due to the presence of double zeros; for the Nelson study, these values are 1.87, 95%CI 0.48–7.23; and the same

Table 6. Antibiotics vs control for the treatment of sore throat in the Del Mar *et al.* [10] dataset.

	Treatment		Control	
Study	Events	Total	Events	Total
Krober 1985 ($\boldsymbol{x}_1$)	0	15	0	11
Nelson 1984 ($\boldsymbol{x}_2$)	12	17	10	18
	Cluster configuration			
	Cluster model	Posterior probability	Cluster model	Posterior probability
	$\{\boldsymbol{x}_1\}\{\boldsymbol{x}_2\}$	0.96	$\{\boldsymbol{x}_1\}\{\boldsymbol{x}_2\}$	0.83
	$\{\boldsymbol{x}_1\boldsymbol{x}_2\}$	0.04	$\{\boldsymbol{x}_1\boldsymbol{x}_2\}$	0.17
	Treatment effects			
	$\mathbb{E}(\theta_T \mid \boldsymbol{x})$	95% CI	$\mathbb{E}(\theta_C \mid \boldsymbol{x})$	95% CI
	0.35	(0.15, 0.57)	0.32	(0.12, 0.56)
	OR			
		$\mathbb{E}(\text{OR} \mid \boldsymbol{x})$	95% CI	
	Krober 1985	6.68	(0.02, 34.16)	
	Nelson 1984	2.91	(0.31, 12.07)	
	Overall	1.53	(0.25, 5.10)	

Note: Cluster configuration and estimated treatment effect and 95% Bayesian credible intervals.

values are obtained 1.87, 95%CI 0.48–7.23 for the overall effect estimates. Moreover, heterogeneity measures cannot be applied to a double–zero cell in the data set. Table 6 presents the posterior probabilities of the cluster configurations (heterogeneity and homogeneity partitions).

In contrast to the results reported by Del Mar *et al.* [10], shown in Table 6, the heterogeneity and homogeneity configurations can be computed numerically by considering their posterior probabilities. For both arms, although the heterogeneity structure is the best model, this omits part of the uncertainty associated with the homogeneity configuration and may produce a misleading estimation. Table 6 presents the estimated treatment effects and the 95% Bayesian credible intervals for each treatment, together with the corresponding odds ratio. These findings seem to indicate that the results obtained by Del Mar *et al.* [10] reflect slightly better outcomes for the control vs. the treatment group (1.87 vs 1.53). Nevertheless, there is general agreement between the reports for each procedure (Peto and BMA). A striking difference between the BMA approach and [10] is that the former enables us to obtain the posterior quantities of interest in each study considered, whether or not it contains a double–zero cell.

Obviously, the inclusion of a double–zero cell in the study data increases the uncertainty of the analysis, which is reflected in the ranges obtained. However, when these cells are incorporated into the complete analysis (with both studies) they contribute additional information that must be taken into account. The interval (0.25, 5.10) spans all the uncertainty in the data, whether or not they are derived from a study with a double zero.

Figure 3 shows the BMA posterior density of θ_T and θ_C (left panel) and the posterior distribution of the odds ratio (OR), where the (posterior) probability $\Pr(\text{OR} > 1 \mid \boldsymbol{x}) = 0.74$, which shows there is a certain probability in favour of the control.

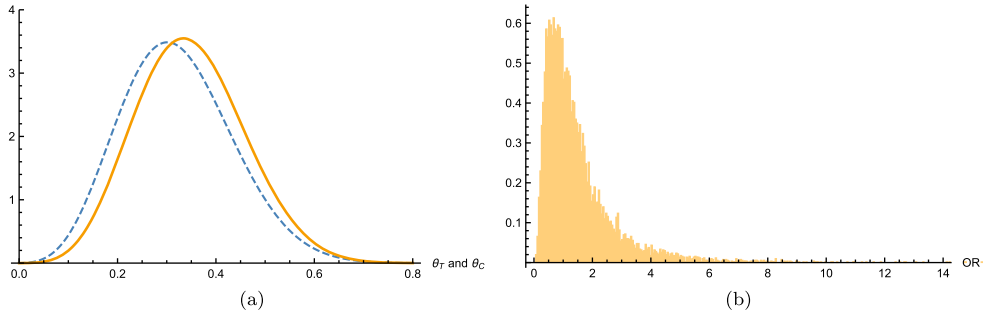


Figure 3. Posterior distributions in the Del Mar *et al.* [10] dataset. (a) BMA posterior density of the parameters θ_T (continuous line) and θ_C (dashed line) and (b) Posterior distribution of the overall OR.

4. Conclusions

When a meta–analysis is based on just a few studies, the possibility of between-study statistical heterogeneity is a factor of major significance [15,24,30]. Moreover, the problem of ensuring statistical validity is exacerbated if this sort of trial includes zero–occurrence events. Additionally, the improper use of continuity correction procedures can result in invalid conclusions being drawn [36]. To overcome these problems, Veroniki *et al.* [41] described various methods that can be used to estimate between–study heterogeneity. However, the case studies we present show that the usual frequentist measures employed to determine the presence or otherwise of heterogeneity, i.e. I^2 and τ, may not address all the scenarios considered in the meta–analysis and therefore might not be appropriate.

Even when a standard Bayesian approach with suitable priors is adopted (see Section 3.1), the estimates obtained may detect some heterogeneity but fail to quantify the model uncertainty. In consequence, inferences regarding the parameters of interest, such as OR or log(OR), may not properly reflect the variations presented in the data.

Our proposal, based on Bayesian model averaging, effectively addresses the problem of accounting for between–study heterogeneity and statistical uncertainty, even in meta–analyses with zero–occurrence events. In the method we present, the heterogeneity structures (models) obtained are averaged by creating a cluster of the data observed in order to draw valid inferences about the treatment effect. This clustering procedure necessarily incorporates any uncertainty present in the models (if it were ignored, invalid conclusions might be drawn). This consideration is of major importance, as the likelihood function is different for each model considered.

Accordingly, the BMA approach we describe is not only conceptually persuasive, it also provides a novel method for probabilistic clustering by which challenging meta–analysis scenarios (such as those involving two studies and single/double–zero occurrence cells) can be addressed and resolved.

The approach we describe can easily be extended to meta–analyses of more than two studies, containing single or double zeros. In such cases, the number of heterogeneity configurations can become very large, according to the Bell number (Rota [31]). However, 75% of meta–analyses contain five or fewer studies, and so the necessary coding is not computationally expensive. The code included in the supplementary section allows these analyses to be performed automatically.

Acknowledgments

The authors are grateful to an Associate Editor and to three anonymous referees for their valuable, constructive comments and suggestions, which have greatly helped us improve this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

M. Martel http://orcid.org/0000-0002-7013-4747
M. A. Negrín http://orcid.org/0000-0002-7074-6268
F. J. Vázquez–Polo http://orcid.org/0000-0002-0632-6138

References

[1] D. Barry and J.A. Hartigan, *Product partition models for change point problems*, Ann. Statist. 20 (1992), pp. 260–279. doi:10.1214/aos/1176348521.
[2] J.M. Bernardo, *Reference posterior distributions for Bayesian inference (with discussion)*, J. R. Stat. Soc. Ser. B 41 (1979), pp. 113–147.
[3] D.K. Bhaumik, A. Amatya, S.L.T. Normand, J. Greenhouse, E. Kaizar, B. Neelon, and R.D. Gibbons, *Meta–analysis of rare binary adverse event*, J. Amer. Statist. Assoc. 107 (2012), pp. 555–567. doi:10.1080/01621459.2012.664484.
[4] M.J. Bradburn, J.J. Deeks JJ, J.A. Berlin, and A.R. Localio, *Much ado about nothing: A comparison of the performance of meta–analytical methods with rare errors*, Stat. Med. 26 (2007), pp. 53–77. doi:10.1002/sim.2528.
[5] P.L. Canner, *An overview of six clinical trials of aspirin in coronary heart desease*, Stat. Med. 6 (1987), pp. 255–263. doi:10.1002/sim.4780060310.
[6] P. Charles, B. Giraudeau, A. Dechartres, G. Baron, and P. Ravaud, *Reporting of sample size calculation in randomised controlled trials: Review*, BMJ 338 (2009), Article ID b1732. doi:10.1136/bmj.b1732.
[7] Y. Chuang, S. Rabe–Hesketh, and I-H. Choi, *Avoiding zero between–study variance estimates in random–effects meta–analysis*, Stat. Med. 32 (2014), pp. 4071–4089. doi:10.1002/sim.5821.
[8] N.D. Crins, C. Röver, A.D. Goralczyk, and T. Friede, *Interleukin-2 receptor antagonists for pediatric liver transplant recipients: A systematic review and meta–analysis of controlled studies*, Pediatr. Transplant. 18 (2014), pp. 839–850. doi:10.1111/petr.12362.
[9] R.C. Davi, J. Buenconsejo, K. Hull, S. Okada, and R. Sista, *Krystexxa*^{TM} *(Pegloticase), PEG–uricase and Puricase*. Statistical Review and Evaluation, U.S. Department of Health and Human Services, Food and Drug Administration (FDA), 2010. STN 125293–0037.
[10] C. Del Mar, P.P. Glasziou, and A. Spink, *Antibiotics for sore throat*, Cochrane Database Syst. Rev. 2 (2004), Article ID CD000023. doi:10.1002/14651858.CD000023.pub2.
[11] R. DerSimonian and R. Kacker, *Random–effects model for meta–analysis in clinical trials: An update*, Contemp. Clin. Trials 28 (12007), pp. 105–114. doi:10.1016/j.cct.2006.04.004.
[12] R. DerSimonian and N. Laird, *Meta–analysis in clinical trials*, Control Clin. Trials 7 (1986), pp. 177–188. doi:10.1016/0197-2456(86)90046-2.
[13] European Medicines Agency, *Note for guidance on statistical principles for clinical trials*, CPMP/ICH/363/96, 1998.
[14] European Medicines Agency, *Points to consider on application with 1. Meta-analyses; 2. One pivotal study*, CPMP/EWP/2330/99, 2001.
[15] T. Friede, C. Röver, S. Wandel, and B. Neuenschwander, *Meta–analysis of two studies in the presence of heterogeneity with applications in rare diseases*, Biom. J. 59 (2017), pp. 658–671. doi:10.1002/bimj.201500236.

[16] J.O. Friedrich, N.K.J. Adhikari, and J. Beyene, *Inclusion of zero total event trials in meta–analyses maintains analytic consistency and incorporates all available data*, BMC Med. Res. Methodol. 7 (2007), p. 5. doi:10.1186/1471-2288-7-5.
[17] A. Gonnermann, T. Framke, A. Grosshening, and A. Koch, *No solution yet for combining two independent studies in the presence of heterogeneity*, Stat. Med. 34 (2015), pp. 2476–2480. doi:10.1002/sim.6473.
[18] B.K. Günhan, C. Röver C, and T. Friede, *Random–effects meta–analysis of few studies involving rare events*, Res. Synth. Methods 11 (2020), pp. 74–90. doi:10.1002/jrsm.1370..
[19] H.H.G. Handoll, M.J. Farrar, J. McBirnie, G. Tytherleigh–Strong, A.A. Milne, and W.J. Gillespie, *Heparin, low molecular weight heparin and physical methods for preventing deep vein thrombosis and pulmonary embolism following surgery for hip fractures*, Cochrane Database Syst. Rev. 4 (2002), Article ID CD000305. doi:10.1002/14651858.CD000305.
[20] J. Hartung and G. Knapp, *A refined method for the meta–analysis of controlled clinical trials with binary outcome*, Stat. Med. 20 (2001), pp. 3875–3889. doi:10.1002/sim.1009.
[21] J. Hartung and G. Knapp, *On tests of the overall treatment effect in meta–analysis with normally distributed responses*, Stat. Med. 20 (2001), pp. 1771–1782. doi:10.1002/sim.791.
[22] J.P. Higgins and S.G. Thompson, *Quantifying heterogeneity in a meta–analysis*, Stat. Med. 21 (2002), pp. 1539–1558. doi:10.1002/sim.1186.
[23] J.P. Higgins, S.G. Thompson, J.J. Deeks, and D.G. Altman, *Measuring inconsistency in meta–analyses*, BMJ 327 (2003), pp. 557–560. doi:10.1136/bjm.327.7414.557.
[24] J. IntHout, J.P.A. Ioannidis, G.F. Borm, and J.J. Goeman, *Small studies are more heterogeneous than large ones: A meta–meta–analysis*, J. Clin. Epidemiol. 68 (2015), pp. 860–869. doi:10.1016/j.jclinepi.2015.03.017.
[25] D.R. Larose and D.K. Dey, *Grouped random effects models for Bayesian meta–analysis*, Stat. Med. 16 (1997), pp. 1817–1829.
[26] R. Miller, J. Mitchell, and D. Moore, *Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND)*, Cochrane Database Syst. Rev. 3 (2012), Article ID CD001447.
[27] M. Mittlböck and H. Heinzl, *A simulation study comparing properties of heterogeneity measures in meta–analyses*, Stat. Med. 25 (2006), pp. 4321–4333. doi:10.1002/sim.2692.
[28] E. Moreno, F.J. Vázquez–Polo, and M.A. Negrín, *Objective Bayesian meta–analysis for sparse discrete data*, Stat. Med. 33 (2014), pp. 3676–3692. doi:10.1002/sim.6163.
[29] E. Moreno, F.J. Vázquez–Polo, and M.A. Negrín, *Bayesian meta–analysis: The role of the between–sample heterogeneity*, Stat. Methods Med. Res. 27 (2018), pp. 3643–3657. doi:10.1177/0962280217709837.
[30] K. Pateras, S. Nikolakopoulos, D. Mavridis, and K.C.B. Roes, *Interval estimation of the overall treatment effect in a meta–analysis of a few small studies with zero events*, Contemp. Clin. Trials. Comm. 9 (2018), pp. 98–107. doi:10.1016/j.conctc.2017.11.012.
[31] G.-C. Rota, *The number of partitions of a set*, Amer. Math. Monthly 71 (1964), pp. 498–504.
[32] C. Röver, G. Knapp, and T. Friede, *Hartung–Knapp–Sidik–Jonkman approach and its modification for random–effects meta–analysis with few studies*, BMC Med. Res. Methodol. 15 (2015), p. 99. doi:10.1186/s12874-015-0091-1.
[33] G. Rücker, G. Schwarzer, J.R. Carpenter, H. Binder, and M. Schumacher, *Treatment–effect estimates adjusted for small–study effects via a limit meta–analysis*, Biostatistics 12 (2011), pp. 122–142. doi:10.1093/biostatistics/kxq046.
[34] K. Sidik and J.N. Jonkman, *On constructing confidence intervals for a standardized mean difference in meta–analysis*, Comm. Statist. Simulation Comput. 32 (2003), pp. 1191–1203. doi:10.1081/SAC-120023885.
[35] D.J. Spiegelhalter, K.R. Abrams, and J.P. Myles, *Bayesian Approaches to Clinical Trials and Health–Care Evaluation*, Wiley & Sons, Chichester, UK, 2004.
[36] M.J. Sweeting, A.J. Sutton, and P.C. Lambert, *What to add to nothing? Use and avoidance of continuity corrections in meta–analysis of sparse data*, Stat. Med. 23 (2004), pp. 1351–1375. doi:10.1002/sim.1761.

[37] D. Thomas, S. Radji, and A. Benedetti, *Systematic review of methods for individual patient data meta–analysis with binary outcomes*, BMC Med. Res. Methodol. 14 (2014), p. 79. doi:10.1186/1471-2288-14-79.
[38] F. Tuyl, R. Gerlach, and K. Mengersen, *A comparison of Bayes–Laplace, Jeffreys, and other priors: The case of zero events*, Amer. Statist. 62 (2008), pp. 40–44. doi:10.1198/000313008X267839.
[39] F. Tuyl, R. Gerlach, and K. Mengersen, *Posterior predictive arguments in favor of the Bayes–Laplace prior as the consensus prior for binomial and multinomial parameters*, Bayesian Anal. 4 (2009), pp. 151–158. doi:10.1214/09-BA405.
[40] F.J. Vázquez–Polo, M.A. Negrín–Hernández, and M. Martel–Escobar, *Meta–anlysis with few studies and binary data: A Bayesian model averaging approach*, Mathematics 8 (2020), p. 2159. doi:10.3390/math8122159.
[41] A.A. Veroniki, D. Jackson, W. Viechtbauer, R. Bender, J. Bowden, G. Knapp, O. Kuss, J.P.T. Higgins, D. Langan, and G. Salanti, *Methods to estimate the between–study variance and its uncertainty in meta–analysis*, Res. Synth. Methods 7 (2016), pp. 55–79. doi:10.1002/jrsm.1164.
[42] F. Weber, G. Knapp, A. Glass A, G. Kundt, and K. Ickstadt, *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 12 (2020), pp. 291–315. doi:10.1002/jrsm.1471.