Meta-Analysis with Few Studies and Binary Data: A Bayesian Model Averaging Approach

Francisco-José Vázquez-Polo * [ID]**, Miguel-Ángel Negrín-Hernández** [ID] **and María Martel-Escobar** [ID]

Department of Quantitative Methods & TiDES Institute, University of Las Palmas de Gran Canaria,
35017 Las Palmas, Spain; miguel.negrin@ulpgc.es (M.-Á.N.-H.); maria.martel@ulpgc.es (M.M.-E.)
* Correspondence: francisco.vazquezpolo@ulpgc.es; Tel.: +34-928-451-800

Abstract: In meta-analysis, the existence of between-sample heterogeneity introduces model uncertainty, which must be incorporated into the inference. We argue that an alternative 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 probabilities. When there are few studies, the number of cluster configurations is manageable, and the meta-inferences can be drawn with BMA techniques. Although this topic has been relatively neglected in the meta-analysis literature, the inference thus obtained accurately reflects the cluster structure of the samples used. In this paper, illustrative examples are given and analysed, using real binary data.

---

1. Introduction

Statistical methods for meta-analysis have been widely applied in many research areas and are of particular importance in healthcare studies. When randomised controlled clinical trials (or studies) are made of a medical treatment for patients with a given disease and heterogeneous samples are compiled, a meta-analysis may be conducted to determine what can fairly be concluded about the fundamental question addressed in each such study, i.e., the effectiveness of the treatment. This parameter is usually estimated on the basis of a prior estimate of the between-study variation, for which purposes random-effects meta-analysis techniques are employed. Thus, meta-analysis addresses the problem of drawing inferences concerning the treatment effect based on k samples. Between-sample heterogeneity introduces additional uncertainty into the process, namely that of model uncertainty. Rücker et al. [1] commented that statistical heterogeneity and small study effects are major issues affecting the validity of meta-analysis. Heterogeneity is difficult to estimate if few studies are considered, and the resulting model uncertainty might lead to misleading results being obtained. In this paper, we define heterogeneity as the statistical variation found in the collected effect size data, which has to be assessed after the studies have been pooled [1].

According to Sutton and Abrams [2], the Bayesian random-effects model for meta-analyses is given by:

$$x_i \sim \mathcal{N}(\theta_i, \tau_i), \ i = 1, \ldots, k$$
$$\theta_i \sim \mathcal{N}(\theta, \tau),$$
$$\theta \sim [-, -] \ \tau \sim [-, -], \tag{1}$$

where for each study $i \in \{1, \ldots, k\}$, the observed effect x_i is assumed to be normally distributed with mean parameter θ_i (the treatment effectiveness conditional on study i) and variance τ_i. Similarly, θ represents the unconditional treatment effectiveness, i.e., the pooled effect size; τ is its variance; and $[-, -]$ indicates a prior density to be assigned. When the effectiveness is measured by a discrete 0–1 random variable, the above hierarchical normal model is typically applied to the logit transformation of the data $\log(x_i/(n_i - x_i))$ with the reparametrisation $\log(\theta_i/(1 - \theta_i))$ [3–8]. If the samples contain zeros, a fixed data continuity correction is applied, and a logit approximation is then used. However, this practice has been criticised by Sweeting et al. [9], who proposed an alternative empirical data correction. Furthermore, small studies present more heterogeneity than large ones [10]. Recently, Weber et al. (2020) [11] recommended that zero-cell corrections should be avoided due to the poor performance of their statistical properties. Obviously, continuity corrections are not the only way to proceed. A number of meta-analysis methods based, e.g., on binomial models were proposed in [12] and the references therein.

In the meta-analysis literature, determining statistical heterogeneity mainly consists of either estimating the between-study variation, characterised as the variance, or heterogeneity parameter, or otherwise testing the null hypothesis that the true treatment effects across studies are identical $H_0 : \theta_1 = \theta_2 = \ldots = \theta_k = \theta$ [8,13]. A measure of evidence for heterogeneity is obtained by means of the Q test, a chi-squared test with $k - 1$ degrees of freedom, first defined by Cochran [14]. Alternatively, and in order to overcome the dependence on the number of studies considered for the meta-analysis, which is the case with the Q test, Higgings and Thompson [15] proposed the dimensionless I^2 index, which describes the percentage of variation across studies that is due to heterogeneity rather than chance. Mittlböck and Heinzl [16] compared these two relative measures of heterogeneity jointly with the index $H_M^2 = \dfrac{Q}{k-1} - 1$, through an intensive simulation study. Basically, I^2 and H_M^2 have similar properties, although H_M^2 is preferred when few studies are considered. As shown in [16], the number of studies affects the power of the heterogeneity test and the heterogeneity measures I^2, but not H_M^2.

It is well known that all tests of heterogeneity have relatively little power to detect heterogeneity when data are sparse and/or when meta-analyses are based on a small number of studies [1,9,15,17]. These situations are very common in daily practice, for instance with respect to rare or orphan diseases, where zero values are often observed in both the trial and the treatment arms considered.

Statistical heterogeneity is a characteristic of the studies that compose a meta-analysis and can be present at different levels of intensity, ranging from homogeneity to heterogeneity. There are also many intermediate situations that should be considered in order to correctly estimate the quantities of interest. Higgins et al. [18] proposed a tentative naive categorisation of values for I^2, relating values of 25%, 50% and 75% with low, moderate and high levels of heterogeneity, respectively, and considering a significant degree of heterogeneity to be present when $I^2 > 50\%$.

In the present analysis, we address the question from a different standpoint, arguing that between-sample heterogeneity is a clustering problem and that model uncertainty can be incorporated into the inference using a Bayesian procedure. Under this procedure, the posterior probabilities of the cluster models are computed, and the meta-inference is then obtained as a mixture of the meta-inferences obtained for the cluster models, using the posterior model probabilities as a mixing distribution. Another point of difference is that most previous studies analysed between-study heterogeneity for a quantity of interest in the comparison of two treatments, such as the mean difference, relative risk or odds ratio. In our proposal, the configuration of between-study heterogeneity is analysed for each treatment separately, which allows the heterogeneity structure to differ between the experimental treatment and the control.

The proposed methodology is known as Bayesian model averaging and has recently begun to be used in meta-analyses. For instance, Bayesian meta-analyses modelling averages over fixed and random effect models were considered in Gronau et al. [19] and Scheibehenne et al. [20].

1.1. A Motivating Example

Günhan et al. [12] suggested the use of weakly informative priors for the treatment effect parameter of a Bayesian meta-analysis model, to be applied in a paediatric transplant dataset. In the present study, and for illustrative purposes, we focus on a dataset corresponding to renal post-transplant lymphoproliferative diseases (PTLD). The results obtained are shown in Table 1.

Table 1. Data on post-transplant lymphoproliferative diseases (PTLD) in Günhan et al.

Trial	Treatment			
	Control		Experimental	
	Events	Total	Events	Total
Schuller et al. (x_1)	0	12	0	18
Ganschow et al. (x_2)	0	54	1	54
Spada et al. (x_3)	1	36	1	36

In this respect, Crins et al. [21] conducted Cochrane's Q test and found no evidence of heterogeneity between the trials. Moreover, Günhan et al. obtained low values for the estimated heterogeneity parameter under weakly informative or vague priors and maximum likelihood estimation. However, the structure of the partition clustering induced by the PTLD dataset in Table 1 shown in Figure 1 indicates that different forms of inter-trial variability could be present.
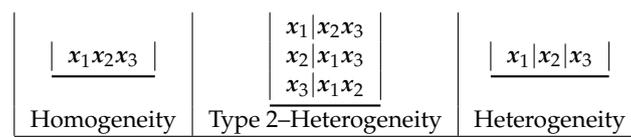

Figure 1. Clustering structure and different heterogeneity structures in the Günhan et al. dataset.

There are $|\mathcal{M}| = 5$ possible partitions (models or heterogeneity configurations), and for each of the cluster classes, there are $1, 3, 1$ partitions, respectively. Thus, the selection of any one model (homogeneity, for example) among the five that are possible may ignore the uncertainty underlying model selection and misrepresent the uncertainty concerning the quantities of interest. In view of these considerations, we believe that Bayesian model averaging (BMA) is an appropriate means of accounting for model uncertainty and, thus, heterogeneity. Expectations and quantities of interest are obtained by weighted averages over the set of models, rather than by selecting a single model and drawing inferences as if it were the true model.

For each of the treatments considered (control and experimental), Table 2 shows the (posterior) probabilities of each of the five models induced by the clustering applied to the data in this example. These posterior probabilities constitute the weights to be used for averaging the set of models. Their expressions are given in Section 3. As an example, let us observe the control treatment. The case of homogeneity has the highest posterior probability of being true, 0.39. However, this option does not consider the remaining heterogeneity structures and therefore would result in misleading inferences being drawn, by rejecting all other possible models, which constitute a combined probability of 0.61. The same reasoning is applicable to the experimental treatment.

Table 2. Cluster models and their posterior probabilities.

Cluster Model	Posterior Probabilities	
	Control	Experimental
M_1: Homogeneity $\{x_1 x_2 x_3\}$	0.39	0.45
M_2: Heterogeneity $\{\{x_1\}, \{x_2\}, \{x_3\}\}$	0.29	0.25
$M_3 : \{\{x_1\}, \{x_2 x_3\}\}$	0.09	0.11
$M_4 : \{\{x_2\}, \{x_1 x_3\}\}$	0.10	0.09
$M_5 : \{\{x_3\}, \{x_1 x_2\}\}$	0.13	0.10

The BMA approach to meta-analysis involves averaging over all the possible models when making inferences about the quantities of interest. For example, for the dataset considered in this case study, if θ is the treatment effect under the control treatment, its posterior distribution, for the data given, is:

$$\pi(\theta|\text{data}) = \sum_{m=1}^{|\mathcal{M}|} \pi(\theta|M_m, \text{data}) \Pr(M_m|\text{data}), \tag{2}$$

i.e., the average of the posterior distributions under each of the models ("type of heterogeneity") considered, weighted by its posterior model probability. For instance, the posterior mean for θ is a weighted average of the posterior means in each model:

$$\mathbb{E}(\theta|\text{data}) = \sum_{m=1}^{|\mathcal{M}|} \mathbb{E}(\theta|M_m, \text{data}) \Pr(M_m|\text{data}), \tag{3}$$

where the posterior probability for model M_m can be obtained by:

$$\Pr(M_m|\text{data}) = \frac{\Pr(M_m)\Pr(\text{data}|M_m)}{\sum_{m=1}^{|\mathcal{M}|} \Pr(M_m)\Pr(\text{data}|M_m)}, \quad m = 1, \ldots, |\mathcal{M}|, \tag{4}$$

where:

$$\Pr(\text{data}|M_m) = \int \Pr(\text{data}|\theta_m, M_m)\Pr(\theta_m|M_m)\, d\theta_m, \tag{5}$$

is the integrated likelihood of the model M_m, θ_m is the vector of the parameters of model M_m, $\Pr(\theta_m|M_m)$ is the prior density of θ_m under model M_m, $\Pr(\text{data}|\theta_m, M_m)$ is the likelihood and $\Pr(M_m)$ is the prior probability of model M_m. All probabilities are implicitly conditional on $\mathcal{M}$, the set of all models being considered, where $|\mathcal{M}|$ represents its cardinal.

1.2. Summary

The rest of this paper is organised as follows. The binomial Bayesian models and the conditional distributions linking the experimental parameters θ_i and the meta-parameter θ are presented in Section 2, where the Bayesian procedure for clustering the samples and the likelihood of the meta-parameter are also given. The Bayesian model averaging procedure for estimating the meta-parameters is described in Section 3. Three illustrative examples are provided in the following section, with real datasets. Finally, Section 5 summarises the main conclusions drawn and presents some concluding remarks.

2. The Bayesian Binomial Model

Assume a meta-analysis involves k studies that provide k independent discrete samples with the binomial distribution $\{\text{Bin}(x_i|n_i, \theta_i), i = 1, \ldots, k\}$, 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. Assume, moreover, that the prior information on the conditional treatment effectiveness θ_i is weak.

Accordingly, the uniform prior $\mathrm{Unif}(\theta_i|0,1)$ is used in a context where the data contain zeros [22,23]. Therefore, for $i = 1, \ldots, k$, we consider the Bayesian sampling models:

$$M_i : \left\{ \mathrm{Bin}(x_i|n_i, \theta_i), \ \pi(\theta_i) = \mathbf{1}_{(0,1)}(\theta_i) \right\}, \tag{6}$$

where:

$$\mathrm{Bin}(x_i|n_i, \theta_i) = \mathrm{Pr}(x_i|n_i, \theta_i) = \binom{n_i}{x_i} \theta_i^{x_i} (1 - \theta_i)^{n_i - x_i}, \quad x_i = 0, 1, \ldots, n_i, \tag{7}$$

and $\mathbf{1}_A$ is the indicator function. The value of one is assigned to all elements of A, and zero elsewhere. In this analysis, it is convenient to consider a $0 - 1$ latent variable x, the meta-variable, which is defined as the result obtained when a treatment with meta-effectiveness θ is applied to a patient in a virtual study, which is not affected by between-study variability. The distribution of this meta-variable x is of the same type as that of x_i, and hence, we have the Bernoulli meta-model $\mathrm{Ber}(x|\theta)$, where the meta-parameter θ represents the true (unconditional) treatment effect. This meta-model gives a precise meaning to the meta-parameter θ. The objective Bayesian meta-model M is then given by:

$$M : \left\{ \mathrm{Pr}(x|\theta) = \theta^x (1 - \theta)^{1-x}, \ \pi(\theta) = \mathbf{1}_{(0,1)}(\theta) \right\}. \tag{8}$$

The Bayesian meta-analysis is then based on the posterior distribution of the parameter θ, which is given by:

$$\pi(\theta|\boldsymbol{x}) \propto f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k|\theta)\pi(\theta), \quad 0 < \theta < 1, \tag{9}$$

where $\boldsymbol{x}_i = (x_i, n_i), \ i = 1, \ldots, k$. The likelihood function in (9) 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 for θ:

1. A distribution $\pi(\theta_i|\theta)$ is needed to link the experimental parameters θ_i and the meta-parameter θ. 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.
2. It is clear that the likelihood of θ strongly depends on the cluster structure of the samples. Therefore, before formulating the likelihood of θ for the samples $\boldsymbol{x}_i, \ i = 1, \ldots, k$, we must address the important question of whether k, the dimension of the model, can be reduced by clustering homogeneous samples.
3. In other words, 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. For the data $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$, we need to obtain the likelihood of θ conditional on a given cluster model. Finally, the likelihood of θ for the data $\boldsymbol{x}$ is obtained as a mixture of the above conditional likelihood functions.

We now proceed step-by-step.

2.1. Linking the Experimental Parameters with the Meta-Parameter

A conditional distribution $\pi(\theta_i|\theta)$, i.e., the linking distribution, is chosen to represent the likelihood of θ for the available samples. To ensure that this conditional distribution is compatible with the marginal priors $\pi(\theta_i)$ and $\pi(\theta)$ given in (6) and (8), the bivariate distribution $\pi(\theta_i, \theta) = \pi(\theta_i|\theta)\pi(\theta)$ must satisfy the integral equations:

$$\int_0^1 \pi(\theta_i, \theta)d\theta_i = \pi(\theta) \ \text{ and } \ \int_0^1 \pi(\theta_i, \theta)d\theta = \pi(\theta_i). \tag{10}$$

Following Moreno et al. [24], we consider the conditional intrinsic linking distributions $\{\pi^I(\theta_i|\theta, t), \ t = 1, 2, \ldots\}$ obtained by comparing model M with model M_i, which presents interesting properties as a linking distribution. For any positive integer t, the intrinsic method gives the conditional intrinsic prior as the following mixture of Beta distributions,

$$\pi^I(\theta_i|\theta) = \sum_{z=0}^{t} \text{Bin}(z|t, \theta) \times \text{Beta}(\theta_i|z + 1, t - z + 1). \tag{11}$$

From (11), we obtain:

$$\mathbb{E}(\theta_i|\theta) = \frac{1 + t\,\theta}{2 + t}, \text{ and } \mathbb{V}(\theta_i|\theta) = \frac{(1 + t)(1 - 2t\theta(\theta - 1))}{(2 + t)^2(3 + t)},$$

and hence, the hyperparameter t indicates how strongly the conditional distribution $\pi^I(\theta_i|\theta)$ concentrates mass around θ. In practice, hyperparameter t is fixed. Hence, for the sake of simplicity in notation, we refer to the linking distribution $\pi(\theta_i|\theta)$ rather than $\pi(\theta_i, |\theta, t)$.

The bidimensional prior $\pi^I(\theta_i, \theta) = \pi^I(\theta_i|\theta)\mathbf{1}_{(0,1)}(\theta)$ satisfies Equation (10) for any t, and therefore, the linking class of intrinsic distributions and the Bayesian models (6) and (8) are coherent. In the following, we assume that θ_i, $i = 1, \ldots, k$, are conditional independent given θ. Therefore, the linking distribution of $\theta_1, \ldots, \theta_k$ conditional on θ is given by:

$$\pi^I(\theta_1, \ldots, \theta_k|\theta) = \prod_{i=1}^{k} \pi^I(\theta_i|\theta). \tag{12}$$

Finally, observe that the proposed linking intrinsic distribution of θ_i, conditional on θ, does not require the use of the logit transformation for the sparse data as is the case with standard random-effect models.

2.2. Clustering the Experimental Samples

Following Moreno et al. [25], we first define what is meant by cluster. The samples x_i and x_j, $i \neq j$, from $f(x|\theta_i)$ and $f(x|\theta_j)$, respectively, are said to be in the same cluster if $\theta_i = \theta_j$. The between-study heterogeneity is then determined by the number of clusters and by the location of the samples $x_1, \ldots, x_k$ within these clusters. The posterior distribution of the clusters is needed in order to draw inferences from these quantities.

To cluster the samples, we adopt the product partition model approach proposed by Barry and Hartigan [26], together with a Bayesian model selection procedure based on Bayes factors for the intrinsic priors for the model parameters and on hierarchical uniform priors for the cluster models.

Following Casella et al. [27], we employ the following notations and expressions in the meta-analysis conducted. For a given p, we define a partition of the samples into p clusters by the vector $\mathbf{r}_p = (r_1, \ldots, r_k)$, where r_i, $i = 1, \ldots, k$, is an integer between one and p denoting the cluster to which x_i is assigned. For example, for $k = 3$ as in Figure 1, the possible $\mathbf{r}$ vectors are shown in Table 3.

Table 3. Different cluster configurations in a meta-analysis with $k = 3$ trials.

r	Cluster Model (Heterogeneity Structure)
$(1, 1, 1)$	$\theta_1 = \theta_2 = \theta_3$ (homogeneity)
$(1, 2, 2)$	$\theta_2 = \theta_3$ (type 2–heterogeneity)
$(1, 2, 1)$	$\theta_1 = \theta_3$ (type 2–heterogeneity)
$(1, 1, 2)$	$\theta_1 = \theta_2$ (type 2–heterogeneity)
$(1, 2, 3)$	all θs are different (heterogeneity)

2.3. The Likelihood of θ

From (7), given a partition $\mathbf{r}_p = (r_1, \ldots, r_k)$, the sampling distribution of $\mathbf{x} = (x_1, \ldots, x_k)$ is:

$$f(\mathbf{x}|p, \mathbf{r}_p, \boldsymbol{\theta}_p) = \prod_{j=1}^{p} \binom{m_j}{s_j} \theta_j^{s_j} (1 - \theta_j)^{m_j - s_j}, \tag{13}$$

where $\boldsymbol{\theta}_p = (\theta_{r_1}, \ldots, \theta_{r_k})$ is an unknown parameter of dimension p and the component θ_j in (13) corresponds to $r_i = j$, $m_j = \sum_{i: r_i = j} n_i$ and $s_j = \sum_{i: r_i = j} x_i$.

For example, the cluster model given by the heterogeneity structure $(1, 2, 2)$ in Table 3 has the sampling distribution:

$$\binom{n_1}{x_1} \theta_1^{x_1} (1 - \theta_1)^{n_1 - x_1} \binom{n_2 + n_3}{x_2 + x_3} \theta_2^{x_2 + x_3} (1 - \theta_2)^{(n_2 + n_3) - (x_2 + x_3)}.$$

The heterogeneity partition $\mathbf{r}_k = (1, 2, \ldots, k)$ has the corresponding likelihood function given by:

$$f(\mathbf{x}|k, \mathbf{r}_k, \boldsymbol{\theta}_k) = \prod_{i=1}^{k} \binom{n_i}{x_i} \theta_i^{x_i} (1 - \theta_i)^{n_i - x_i}.$$

Now, following Moreno et al. [25], integrating out $\boldsymbol{\theta}_p$ with the intrinsic prior $\pi(\boldsymbol{\theta}_p|p, \mathbf{r}_p) = \int \pi^I(\theta_1, \ldots, \theta_p|\theta) \mathbf{1}_{(0,1)}(\theta) \, d\theta$, we obtain the likelihood of θ, conditional on the cluster model $(p, \mathbf{r}_p)$. After some algebra and using the assistance `Wolfram Mathematica`, the likelihood of θ, conditional on the cluster model $(p, \mathbf{r}_p)$, is given by:

$$\begin{aligned} f(\mathbf{x}|p, \mathbf{r}_p, \theta) &= \prod_{j=1}^{p} \int_0^1 f(\mathbf{x}|p, \mathbf{r}_p, \boldsymbol{\theta}_p) \pi(\theta_j|\theta) \, d\theta_j \\ &= (1 + t)^p (1 - \theta)^{tp} \prod_{j=1}^{p} \frac{\Gamma(s_j + 1)\Gamma(m_j + t - s_j + 1)}{\Gamma(m_j + t + 2)} \, {}_3F_2(\mathbf{a}_j, \mathbf{b}_j, \frac{\theta}{\theta - 1}), \end{aligned} \tag{14}$$

where $_3F_2(\mathbf{v}, \mathbf{w}, z)$ denotes the generalised hypergeometric function with argument z and vector parameters $\mathbf{v}$ and $\mathbf{w}$ of dimensions two and three, respectively, $\mathbf{a}_j = (-t, -t, s_j + 1)$ and $\mathbf{b}_j = (1, -m_j - t + s_j)$.

To derive the necessary likelihood function of θ, we then integrate out (14) with respect to a discrete prior on $(p, \mathbf{r}_p)$. As recommended by Moreno et al., an appropriate uniform hierarchical prior is used, in which each factorised prior is uniform.

$$\pi(p, \mathbf{r}_p) = \pi(p|\mathbf{r}_p)\pi(p) = \frac{k_1! \cdot \ldots \cdot k_p! \prod_{i=1}^{k} \left(\sum_{j=1}^{p} \mathbf{1}(k_j = i)\right)!}{k!} \frac{1}{b(k, p)} \frac{1}{k}. \tag{15}$$

For example, for the cluster configuration $(1, 2, 2)$ in Table 3, the uniform hierarchical prior is:

$$\pi(2, (1, 2, 2)) = \frac{1!2!}{3!} \cdot \frac{1!}{1} \cdot \frac{1}{3} = \frac{1}{9}.$$

The complexity in this task lies in knowing how many elements are in each heterogeneity subclass. According to Casella et al. [27], who presented a comprehensive description of these calculations, the first product factor in (15) corresponds to the inverse of the multinomial factor, where $k_1 \leq k_2 \leq \ldots \leq k_p$ are integers assigned to cluster $1, \ldots, p$ such that $k_1 + \ldots + k_p = k$. The second product factor takes into account the redundancies in the number of samples $k_1, \ldots, k_p$, and $b(k, p)$ is the number of these possible sizes of k_j that satisfy the recursive equation $b(k, p) = b(k - 1, p - 1) + b(k - p, p)$, $1 \leq p \leq k$ with $b(k, 1) = b(k, k) = 1$. For each value given for k and p, this number is readily obtained with

the `Mathematica` package `PartitionFunctionP`. The third factor corresponds to the discrete uniform distribution of the number of studies in the meta-analysis.

Finally, from (14) and (15), the (unconditional) likelihood of θ for the data x is given by:

$$f(x|\theta) = \sum_{p=1}^{k} \left(\sum_{\mathbf{r}_p} f(x|p, \mathbf{r}_p, \theta) \pi(p, \mathbf{r}_p) \right). \tag{16}$$

3. Bayesian Model Averaging in the Meta-Analysis

As mentioned in Section 1, the BMA approach to meta-analysis requires us to average all possible models (heterogeneity structures) when making inferences on θ. In this case, the posterior probabilities in (4) correspond to those of any heterogeneity structure given by a pair $(p, \mathbf{r}_p)$, which is represented by:

$$\Pr(p, \mathbf{r}_p|x) = \frac{m_{\mathbf{r}_p}(x|p, \mathbf{r}_p)\pi(p, \mathbf{r}_p)}{\sum_{p=1}^{k} \left(\sum_{\mathbf{r}_p} m_{\mathbf{r}_p}(x|p, \mathbf{r}_p)\pi(p, \mathbf{r}_p) \right)}, \tag{17}$$

where $m_{\mathbf{r}_p}(x|p, \mathbf{r}_p) = \int f(x|p, \mathbf{r}_p, \boldsymbol{\theta}_p)\pi(\boldsymbol{\theta}_p|p, \mathbf{r}_p)\, d\boldsymbol{\theta}_p$ is the marginal of the data x conditional on model $(p, \mathbf{r}_p)$, with $f(x|p, \mathbf{r}_p, \boldsymbol{\theta}_p)$ and $\pi(\boldsymbol{\theta}_p|p, \mathbf{r}_p)$ as in (13) and (15), respectively. These posterior model probabilities $\Pr(p, \mathbf{r}_p|x)$ are the weights for the meta inference.

The posterior distribution in (2) becomes:

$$\pi(\theta|x) = \sum_{\mathbf{r}_p} \pi(\theta|x, p, \mathbf{r}_p) \Pr(p, \mathbf{r}_p), \tag{18}$$

where, since as in (8) $\pi(\theta) = \mathbf{1}_{(0,1)}(\theta)$,

$$\pi(\theta|x, p, \mathbf{r}_p) = \frac{f(x|p, \mathbf{r}_p, \theta)}{\int_0^1 f(x|p, \mathbf{r}_p, \theta)\, d\theta}, \quad 0 < \theta < 1. \tag{19}$$

The posterior distribution in (18) is computed numerically over the parametric space $\Theta = (0, 1)$ using `Wolfram Mathematica`, which has a huge library of ready-to-use functions and, moreover, provides a simple code. Furthermore, once the posterior distribution has been obtained, the command `ProbabilityDistribution` can be used to generate any sample of the posterior distribution of the parameter of interest and its transforms, as shown in the examples given in Section 4.

4. Illustrations

To illustrate the arguments developed in the preceding section, we now analyse two real datasets. For these case studies, we assume that $t = 48$, 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 datasets is available from the authors upon request.

4.1. Motivating Example Revisited

Returning to the dataset considered in Section 1, from Table 2, observe that we have different posterior probabilities for the cluster structures and thus for the Bayesian model average of the true treatment effect under treatment (T) or control (C), θ_T and θ_C. All the measures of interest can be obtained from the BMA posterior distribution in (18). Table 4 shows the posterior summaries of the PTLD rate for both cases (mean and highest density interval (HDI)).

Table 4. Posterior PTLD rates $\mathbb{E}(\theta|\boldsymbol{x})$ under control and experimental treatments. HDI, highest density interval; RR, risks ratio.

Control		Treatment	
$\mathbb{E}(\theta_C|\boldsymbol{x})$	95% HDI	$\mathbb{E}(\theta_T|\boldsymbol{x})$	95% HDI
0.030	(0, 0.109)	0.038	(0, 0.127)
log(RR)			
$\mathbb{E}(\log(RR)|\boldsymbol{x})$		95% HDI	
0.26		(−3.27, 3.91)	

Figure 2 (left panel) shows the BMA posterior density of θ_C and θ_T given by (18). It is apparent that the posterior density of θ under the experimental treatment is slightly shifted towards higher PTLD rates with respect to the control treatment. Observe, moreover, that the sparsity of the data yields reverse J-shaped posterior distributions with a mode at $\theta = 0$.

On the other hand, since we have an explicit expression for the posterior distribution of θ_C and θ_T from (18), it is straightforward to simulate these independent distributions. The usual parameters of interest, such as the risks ratio or the odds ratio and their log, are obtained immediately using the corresponding transform of the above-simulated values. Figure 2 (right panel) shows the posterior distribution of the log risks ratio (RR), where the (posterior) probability $\Pr(\log(RR) > 0|\boldsymbol{x}) = \Pr(RR > 1|\boldsymbol{x}) = 0.57$.
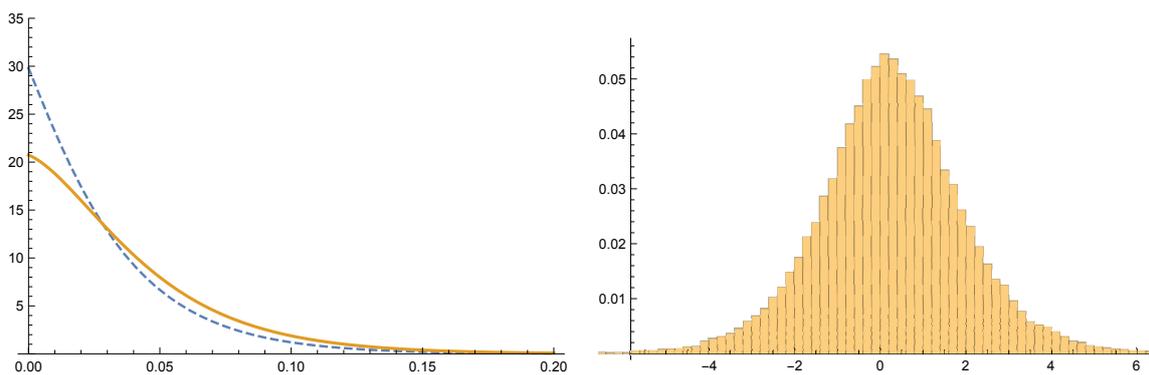

Figure 2. (**Left**) BMA posterior density of the meta-parameters θ_C (dashed line) and θ_T (continuous line). (**Right**) Posterior distribution of the logRR.

For comparison, Table 5 shows the estimations of the risk and odds ratio from Crins et al. [21] and from Günhan et al. [12], together with those from our BMA approach.

Table 5. Estimated risks and odds ratios in the motivating example (revisited).

Quantity	Crins et al. [21] & Günhan et al. [12]	BMA
Risks ratio	1.60 (95% HDI 0.22–29.96)	1.30 (95% HDI 0.04–44.66)
Odds ratio	1.98 (95% HDI 0.20–25.18)	1.32 (95% HDI 0.038–49.899)

These estimated values show some differences from those obtained by Crins et al. and Günhan et al. Although Crins et al. asserted that there is no evidence for between-trial heterogeneity for PTLD outcomes, this conclusion does not follow from Table 2. Certainly, the homogeneity model has the largest posterior probability, for both treatment options, but a large proportion of the uncertainty accumulated in the remaining heterogeneity structures must be considered, in line with the BMA

approach. This, together with the very small number of informative studies conducted in this respect, no more than three, accounts for the wider intervals obtained by the BMA procedure.

4.2. Another Case with Few and Double-Zero Studies

The dataset in Table 6 is extracted from Cosmi et al. [28] and corresponds to the comparison of total mortality in four studies ($k = 4$) of patients treated with ticlopidine (T) versus oral anticoagulation for coronary stenting (C). Cosmi et al. found no evidence of heterogeneity, and I^2 was estimated at 0.0%. This fact occurs frequently for meta-analyses with a small number of studies [29]. Furthermore, the Q test for this situation possesses a very low power [30]. As a consequence, we have little information on homogeneity (heterogeneity) either way.

Table 6. Data in Cosmi et al.

Study	Treatment		Control	
	Events	Total	Events	Total
FANTASTIC (x_1)	2	243	4	230
ISAR (x_2)	1	257	2	260
MATTIS (x_3)	3	177	2	173
STARS (x_4)	0	546	0	550

In this case, there are 15 models. These are all averaged in the BMA posterior distributions, but for simplicity, the top cluster models (posterior probability greater than 0.05) for both treatments are shown in Table 7. The homogeneity cluster model is the best model for the control treatment, but only has a 0.221 mass of probability, while the remaining probability is distributed among the other 14 models. The outcome is similar for the experimental treatment, for which the best option is now the heterogeneity cluster model. Obviously, these between-study variabilities should be considered in any meta-analysis.

Table 7. Top cluster models in Cosmi et al.

Treatment		Control	
Cluster Model	Post. Prob.	Cluster Model	Post. Prob.
$\{x_1\}\{x_2\}\{x_3\}\{x_4\}$	0.303	$\{x_1 x_2 x_3 x_4\}$	0.221
$\{x_1 x_2 x_3 x_4\}$	0.124	$\{x_1\}\{x_2\}\{x_3\}\{x_4\}$	0.216
$\{x_1 x_2 x_3\}\{x_4\}$	0.115	$\{x_1 x_3\}\{x_2 x_4\}$	0.122
$\{x_1\}\{x_2 x_3\}\{x_4\}$	0.094	$\{x_1\}\{x_2 x_4\}\{x_3\}$	0.079
$\{x_1 x_3\}\{x_2\}\{x_4\}$	0.085	$\{x_1 x_2\}\{x_3\}\{x_4\}$	0.073
$\{x_1 x_3\}\{x_2\}x_4\}$	0.072	$\{x_1 x_2 x_4\}\{x_3\}$	0.068
$\{x_1 x_2\}\{x_3\}\{x_4\}$	0.069	$\{x_1 x_2 x_3\}\{x_4\}$	0.060
the rest	<0.05	$\{x_1 x_3\}\{x_2\}\{x_4\}$	0.051
		the rest	<0.05

Cosmi et al. used the Mantel-Haenszel method under random-effects to estimate the risks ratio, for which a value of 0.73 (95% CI 0.25–2.18) was obtained. The main conclusion derived from these data was that ticlopidine plus aspirin versus oral anticoagulation did not affect all-cause mortality. Using the proposed BMA approach, a risks ratio of 0.959 is obtained, which is in line with the value reported by Cosmi et al. However, once again, a realistic and larger interval becomes apparent when all of the uncertainty is managed via the BMA perspective (95% HDI 0.02–41.86).

5. Conclusions

The question of statistical heterogeneity in a meta-analysis based on a relatively small number of studies is important and has been addressed by Friede et al. [17], IntHout et al. [10] and

Pateras et al. [31], among many others. Moreover, if zero events occur in these small trials, the challenge is multiplied. There has been much debate about the presence of significant between-study heterogeneity in these cases, and it has been reported that the misuse of continuity correction procedures can lead to incorrect inferences being drawn [9]. In this respect, Veroniki et al. [32] catalogued various methods for estimating between-study heterogeneity. The examples given in the present paper also show that the usual measures employed to determine the presence or otherwise of heterogeneity, I^2 and τ, may not consider all the scenarios contained in the meta-analysis, and therefore, it is currently recommended to accompany the I^2 statistic and the estimate of τ with a confidence interval. Such a confidence interval will be very wide, especially in the case of a small number of effect sizes in a meta-analysis.

The Bayesian model averaging that we propose is an intuitively attractive approach to the problem of accounting for heterogeneity uncertainty, even in meta-analyses with zero events. This method involves averaging all the heterogeneity structures (models) obtained by clustering the observed sample when making inferences about the true treatment effect. The number of models (types of heterogeneity) considered in BMA rises in line with the number k of studies, according to Bell's number [27]. However, this number is easily manageable for the case of few studies (2 for $k = 2$, 5 for $k = 3$, 15 for $k = 4$, and 52 for the limit case $k = 5$), and therefore, the computation of the posterior probabilities of each model and the derivation of the posterior BMA distribution of the parameter of interest are immediate and computationally feasible. In this respect, Rhodes et al. [33] conducted an extensive review of the Cochrane Database of Systematic Reviews (CDSR), developed in Davey et al. [34], and pointed out that: "Of 22453 meta-analyses from CDSR, containing at least two studies, just under 75% contained five or fewer studies". We recommend our method when the meta-analysis has 10 or fewer studies.

The clustering procedure considered takes into account all uncertainty in the heterogeneity structures. This fact has important consequences for treatment comparison, due to the likelihood of θ being different for each heterogeneity configuration.

On the other hand, observed heterogeneity could be a starting point to investigate sources of heterogeneity. In this paper, every potential cluster model is estimated. However, this might also be a starting point to investigate sources of heterogeneity before the simple model averaging takes place.

From these considerations, we conclude that the BMA procedure proposed is not only more compelling in a conceptual sense, but also provides novel probabilistic clustering methods that are capable of resolving challenging meta-analysis scenarios, in which there are few studies and sparse data.

References

1. Rücker, G.; Schwarzer, G.; Carpenter, J.R.; Binder, H.; Schumacher, M. Treatment-effect estimates adjusted for small-study effects via a limit meta-analysis. *Biostatistics* **2011**, *12*, 122–142. [CrossRef] [PubMed]
2. Sutton, A.J.; Abrams, K.R. Bayesian methods in meta-analysis and evidence synthesis. *Stat. Methods Med. Res.* **2001**, *10*, 277–303. [CrossRef] [PubMed]

3. Larose, D.R.; Dey, D.K. Grouped random effects models for Bayesian meta-analysis. *Stat. Med.* **1997**, *16*, 1817–1829. [CrossRef]

4. Larose, D.R.; Dey, D.K. Modeling publication bias using weighted distributions in a Bayesian framework. *Comput. Stat. Data Anal.* **1998**, *26*, 279–302. [CrossRef]

5. DerSimonian, R.; Kacker, R. Random–effects model for meta-analysis of clinical trials: An update. *Contemp. Clin. Trials* **2007**, *28*, 105–114. [CrossRef]

6. Bradburn, M.J.; Deeks, J.J.; Berlin, J.A.; Localio, A.R. Much ado about nothing: A comparison of the performance of meta-analytical methods with rare errors. *Stat. Med.* **2007**, *26*, 53–77. [CrossRef]

7. Hemming, K.; Hutton, J.L.; Maguire, M.G.; Marson, A.G. Meta-regression with partial information on summary trial or patient characteristics. *Stat. Med.* **2010**, *29*, 1312–1324. [CrossRef]

8. Bhaumik, D.K.; Amatya, A.; Norm, S.L.T.; Greenhouse, J.; Kaizar, E.; Neelon, B.; Gibbons, R.D. Meta-analysis of rare binary adverse event. *J. Am. Stat. Assoc.* **2012**, *107*, 555–567. [CrossRef]

9. Sweeting, M.J.; Sutton, A.J.; Lambert, P.C. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Stat. Med.* **2004**, *23*, 1351–1375. [CrossRef]

10. IntHout, J.; Ioannidis, J.P.A.; Borm, G.F.; Goeman, J.J. Small studies are more heterogeneous than large ones: A meta-meta-analysis. *J. Clin. Epidemiol.* **2015**, *68*, 860–869. [CrossRef]

11. Weber, F.; Knapp, G.; Ickstadt, K.; Kundt, G.; Glass, A. Zero-cell corrections in random-effects meta-analyses. *Res. Synth. Methods* **2020**, *11*, 913–919. [CrossRef] [PubMed]

12. 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] [PubMed]

13. Canner, P.L. An overview of six clinical trials of aspirin in coronary heart desease. *Stat. Med.* **1987**, *6*, 255–263. [CrossRef]

14. Cochran, W.G. The combination of estimates from different experiments. *Biometrics* **1954**, *10*, 101–129. [CrossRef]

15. Higgins, J.P.T.; Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **2002**, *21*, 1539–1558. [CrossRef]

16. Mittlböck, M.; Heinzl, H. A simulation study comparing properties of heterogeneity measures in meta–analyses. *Stat. Med.* **2006**, *25*, 4321–4333. [CrossRef] [PubMed]

17. Friede, T.; Röver, C.; Wandel, S.; Neuenschawander, B. Meta-analysis of two studies in the presence of heterogeneity with applications in rare diseases. *Biom. J.* **2017**, *59*, 658–671. [CrossRef]

18. Higgins, J.P.; Thompson, S.G.; Deeks, J.J.; Altman, D.G. Measuring inconsistency in meta-analyses. *BMJ* **2003**, *327*, 557–560. [CrossRef]

19. Gronau, Q.F.; Van Erp, S.; Heck, D.W.; Cesario, J.; Jonas, K.J.; Wagenmakers, E.J. A Bayesian model-averaged meta-analysis of the power pose effect with informed and default priors: The case of felt power. *Compr. Results Soc. Psychol.* **2017**, *2*, 123–138. [CrossRef]

20. Scheibehenne, B.; Gronau, Q.F.; Jamil, T.; Wagenmakers, E.J. Fixed or random? A resolution through model averaging. *Psychol. Sci.* **2017**, *28*, 1698–1701. [CrossRef]

21. Crins, N.D.; Röver, C.; Goralczyk, A.D.; Friede, T. Interleukin-2 receptor antagonists for pediatric liver transplant recipients: A systematic review and meta-analysis of controlled studies. *Pediatr. Transplant.* **2014**, *18*, 839–850. [CrossRef] [PubMed]

22. Tuyl, F.; Gerlach, R.; Mengersen, K. A comparison of Bayes-Laplace, Jeffreys, and other priors: The case of zero events. *Am. Stat.* **2008**, *62*, 40–44. [CrossRef]

23. Tuyl, F.; Gerlach, R.; Mengersen, K. Posterior predictive arguments in favor of the Bayes-Laplace prior as the consensus prior for binomial and multinomial parameters. *Bayesian Anal.* **2009**, *4*, 151–158. [CrossRef]

24. Moreno, E.; Vázquez-Polo, F.J.; Negrín, M.A. Objective Bayesian meta-analysis for sparse discrete data. *Stat. Med.* **2014**, *33*, 3676–3692. [CrossRef]

25. Moreno, E.; Vázquez-Polo, F.J.; Negrín, M.A. Bayesian meta-analysis: The role of the between-sample heterogeneity. *Stat. Methods Med. Res.* **2018**, *27*, 3643–3657. [CrossRef]

26. Barry, D.; Hartigan, J.A. Product partition models for change point problems. *Ann. Stat.* **1992**, *20*, 260–279. [CrossRef]

27. Casella, G.; Moreno, E.; Girón, F.J. Cluster analysis, model selection, and prior distributions on models. *Bayesian Anal.* **2014**, *9*, 613–658. [CrossRef]

28. Cosmi, B.; Rubboli, A.; Castelvetri, C.C.; Mil, ri, M. Ticlopidine versus oral anticoagulation for coronary stenting. *Cochrane Database Syst. Rev.* **2001**, *4*, CD002133. [CrossRef]

29. Friede, T.; Röver, C.; Wandel, S.; Neuenschwander, B. Meta-analysis of few small studies in orphan diseases. *Res. Syn. Methods* **2017**, *8*, 79–91. [CrossRef]

30. Hardy, R.J.; Thompson, S.G. Detecting and describing heterogeneity in meta-analysis. *Stat. Med.* **1998**, *17*, 841–856. [CrossRef]

31. Pateras, K.; Nikolakopoulos, S.; Mavridis, D.; Roes, K.C.B. Interval estimation of the overall treatment effect in a meta-analysis of a few small studies with zero events. *Contemp. Clin. Trials Comm.* **2018**, *9*, 98–107. [CrossRef] [PubMed]

32. 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. Syn. Methods* **2016**, *7*, 55–79. [CrossRef] [PubMed]

33. Rhodes, K.M.; Turner, R.M.; Higgins, J.P. 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]

34. Davey, J.; Turner, R.M.; Clarke, M.J.; Higgins, J.P. Characteristics of meta-analyses and their component studies in the Cochrane Database of Systematic Reviews: A cross-sectional, descriptive analysis. *BMC Med. Res. Methodol.* **2011**, *11*, 160. [CrossRef]

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