

## RESEARCH ARTICLE

# Random-effects meta-analysis of few studies involving rare events

Burak Kürsad Günhan<sup>1</sup> | Christian Röver<sup>1</sup> | Tim Friede<sup>1</sup>

Department of Medical Statistics,  
University Medical Center Göttingen,  
Göttingen, Germany

**Correspondence**

Burak Kürsad Günhan, Department of  
Medical Statistics, University Medical  
Center Göttingen, Göttingen, Germany.  
Email:  
burak.gunhan@med.uni-goettingen.de

Meta-analyses of clinical trials targeting rare events face particular challenges when the data lack adequate numbers of events for all treatment arms. Especially when the number of studies is low, standard random-effects meta-analysis methods can lead to serious distortions because of such data sparsity. To overcome this, we suggest the use of *weakly informative priors* (WIPs) for the treatment effect parameter of a Bayesian meta-analysis model, which may also be seen as a form of penalization. As a data model, we use a binomial-normal hierarchical model (BNHM) that does not require continuity corrections in case of zero counts in one or both arms. We suggest a normal prior for the log-odds ratio with mean 0 and standard deviation 2.82, which is motivated (a) as a symmetric prior centered around unity and constraining the odds ratio within a range from 1/250 to 250 with 95% probability and (b) as consistent with empirically observed effect estimates from a set of 37 773 meta-analyses from the Cochrane Database of Systematic Reviews. In a simulation study with rare events and few studies, our BNHM with a WIP outperformed a Bayesian method without a WIP and a maximum likelihood estimator in terms of smaller bias and shorter interval estimates with similar coverage. Furthermore, the methods are illustrated by a systematic review in immunosuppression of rare safety events following pediatric transplantation. A publicly available R package, *MetaStan*, is developed to automate a Bayesian implementation of meta-analysis models using WIPs.

**KEYWORDS**

Bayes, few studies, random-effects meta-analysis, rare events, weakly informative priors

## 1 | INTRODUCTION

Individual clinical studies are often underpowered to detect difference of probabilities or rates of rare events, for example, safety events, and thus, meta-analysis may be the only way to obtain reliable evidence of treatment differences with regard to the rare events.<sup>1</sup> On the other hand, meta-analysis of clinical studies for rare events faces particular challenges, since the numbers of events might be very small in some treatment arms. The problem

is even more pronounced when some studies have no events either in one or in both treatment arms (so-called single-zero or double-zero studies).

The exclusion of the double-zero studies from the analysis can bias the treatment effect parameter estimate away from the null (especially for the unbalanced design)<sup>2</sup> and also causes loss of information, since double-zero studies contain information through their sample sizes.<sup>3</sup> Hence, we consider methods that do not remove double-zero studies from the analysis. Two established fixed-effect

meta-analysis methods exist for rare events, namely, Peto's method<sup>4</sup> and the Mantel-Haenszel (MH) method.<sup>5</sup> On the other hand, an assumption of homogeneity, that is, a single common parameter for all studies, is typically unrealistic for studies in the biomedical sciences.<sup>6-8</sup> Therefore, we focus on random-effects methods in this paper.

Standard (approximate) random-effects meta-analysis methods, for example, the normal-normal hierarchical model,<sup>9</sup> require a *continuity correction* in case of single-zero or double-zero studies, that is, the addition of a fixed value (typically 0.5) to all cells of the contingency table for studies with no events or with 100% events (no nonevents). Such simple approaches have been found problematic for meta-analyses involving rare events.<sup>10</sup> Therefore, statistical models based on exact distributional assumptions have been suggested. These include different parametrizations of the binomial-normal hierarchical model (BNHM),<sup>11</sup> a mixed effects conditional logistic model,<sup>12</sup> a Poisson-normal hierarchical model,<sup>13</sup> a Poisson-Gamma hierarchical model,<sup>14</sup> or a beta-binomial model (BBM).<sup>3</sup> In this paper, we focus on a parametrization of the BNHM that was suggested by Smith et al.<sup>15</sup>

Consider an extreme case of meta-analysis of rare events, where all studies include no events for the same treatment arm. These data sparsity problem in a meta-analysis can be seen as a *separation* problem in the logistic regression context<sup>16</sup> in which case a maximum likelihood estimate (MLE) for the treatment effect parameter does not exist. A very useful way to deal with separation problems, or, more generally, data sparsity in logistic regression is *penalization*, that is, adding a penalty (adjustment) term to the original likelihood function to regularize (or stabilize) the estimates.<sup>17</sup> In a frequentist framework, penalty terms may be specified so that these *nudge* the MLE into a desired direction if the maximum is not or poorly defined; one such example is Firth penalization.<sup>17-19</sup> From a Bayesian viewpoint, penalization may often be motivated as *weakly informative priors* (WIPs) that are multiplied to the likelihood function.<sup>20</sup>

Numbers of studies included in meta-analyses are typically small, posing additional challenges.<sup>21</sup> For Bayesian meta-analysis of few studies, different WIPs have been suggested for the heterogeneity parameter; see Chung et al<sup>22</sup> for penalized MLE approach and also see Gelman<sup>23</sup> and Friede and Röver<sup>24</sup> for Bayesian inference. Here, we con-

sider the meta-analysis of few studies targeting rare events. To deal with data sparsity present in the meta-analysis of few studies with rare events, we suggest the use of WIPs for the treatment effect parameter in a fully Bayesian context inspired by penalization ideas.<sup>17,20</sup> We use a BNHM that is parameterized in terms of baseline risks and a treatment effect for the data. Note that this is a contrast-based model meaning that relative treatment effects are assumed to be exchangeable across trials.<sup>25</sup> Our suggested default WIP for the treatment effect parameter is motivated via the consideration of the prior expected range of treatment effect values. Furthermore, it is consistent with effect estimates empirically observed in a large set of meta-analyses from the Cochrane Database of Systematic Reviews (CDSR) with binary endpoints.

The main contribution of this paper is the introduction of default WIPs as penalization for treatment effect parameters to deal with data sparsity in the meta-analysis of few studies involving rare events. Another contribution is the introduction of an **R** package, **MetaStan** (<https://CRAN.R-project.org/package=MetaStan>), which is developed to automate a Bayesian implementation of meta-analysis models using WIPs as described in the paper and which is publicly available from CRAN. In Section 2, we describe a systematic review concerning rare safety events associated with immunosuppressive therapy following pediatric transplantation. In Section 3, we describe the application of WIPs for the treatment effect parameter. We review a BNHM for meta-analysis, discuss the derivation of WIP, and an empirical investigation of treatment effect parameter estimates from the CDSR. Long-run properties of different methods including the proposed one are investigated in the simulation studies in Section 5. In Section 6, the example is revisited to illustrate the proposed method and its implementation. We close with some conclusions and provide a discussion.

## 2 | AN APPLICATION IN PEDIATRIC TRANSPLANTATION

Several rare pediatric liver diseases can nowadays be successfully treated by liver transplantation with good long-term outcomes. Crins et al<sup>26</sup> conducted a systematic review of controlled but not necessarily random-

**TABLE 1** Data on patient deaths and posttransplant lymphoproliferative disease (PTLD) from the meta-analysis in pediatric transplantation conducted by Crins et al<sup>26</sup>

	Outcome: Death				Outcome: PTLD			
	Control		Experimental		Control		Experimental	
	Events	Total	Events	Total	Events	Total	Events	Total
Heffron et al <sup>29</sup>	3	20	4	61	-	-	-	-
Schuller et al <sup>30</sup>	-	-	-	-	0	12	0	18
Ganschow et al <sup>31</sup>	3	54	1	54	0	54	1	54
Spada et al <sup>32</sup>	3	36	4	36	1	36	1	36
Gras et al <sup>33</sup>	3	34	2	50	-	-	-	-

ized studies of the Interleukin-2 receptor antibodies (IL-2RA) basiliximab and daclizumab in pediatric liver transplantation. Primary outcomes were acute rejections (ARs), steroid-resistant rejections (SRRs), graft loss, and death. Their analyses were based on a random-effects meta-analysis using a restricted maximum likelihood approach (REML).<sup>27</sup> Crins et al<sup>26</sup> used risk ratios as effect measures, while we use the odds ratios here. With rare events, however, these should be very similar. Heterogeneity was assessed using Cochrane's Q test.<sup>28</sup> Secondary outcomes included renal dysfunction and lymphoproliferative disease (PTLD). For illustrative purposes, here, we focus on death and PTLDs, and these outcomes are displayed in Table 1.

The specific problems with meta-analyses concerning rare events outlined in the introduction are prominent here. Firstly, the numbers of events are very small. For the PTLD dataset, there is one single-zero study and one double-zero study. Secondly, there are few studies available, only four for deaths and three for PTLD. Empirical event rates are lower in three of the four experimental groups in the data on patient deaths. For PTLD, the data appear inconclusive for Schuller et al<sup>30</sup> and Spada et al,<sup>32</sup> and only a single event observed in the experimental group suggests an increased risk in the study by Ganschow et al.<sup>31</sup>

### 3 | WIPS FOR THE TREATMENT EFFECT

In this section, we present the usage of WIPs for the treatment effect parameter to conduct random-effects meta-analysis of rare events with few studies. As a data model, we review a BNHM and then show how to derive a WIP for a treatment effect parameter. Then, empirical evidence obtained from the CDSR supporting the choice of WIPs is illustrated.

#### 3.1 | Data model

The BNHM has been introduced by Smith et al.<sup>15</sup> In the BNHM, for each trial  $i \in \{1, \dots, k\}$  and treatment arm  $j \in \{0, 1\}$ , the event counts  $r_{ij}$  are modeled using a binomial distribution, that is,  $r_{ij} \sim \text{Bin}(\pi_{ij}, n_{ij})$ . The logit link is used to transform  $\pi_{ij}$  onto the log odds scale where effects can be assumed to be additive

$$\begin{aligned} \text{logit}(\pi_{ij}) &= \mu_i + \theta_i x_{ij} \\ \theta_i &\sim \mathcal{N}(\theta, \tau^2), \end{aligned} \quad (1)$$

where  $x_{ij}$  is a treatment indicator, namely,  $+0.5 = \text{experimental}$  ( $j = 1$ ) and  $-0.5 = \text{control}$  ( $j = 0$ ). The  $\mu_i$  are fixed effects denoting the baseline risks of the event in each study  $i$ ,  $\theta$  is the mean treatment effect, and  $\tau$  is the het-

erogeneity in treatment effects between trials. The BNHM belongs to the family of generalized linear mixed models (GLMMs); this family also includes models for other types of data including continuous or count outcomes. It is important to note here that by treating the baseline risks  $\mu_i$  as fixed effects, the analysis effectively stratifies the risk by study, as pooling of risks might compromise the studies' randomization. In this sense, it constitutes a contrast-based model.<sup>25</sup> Unlike the normal-normal hierarchical model, the BNHM does not rely on a normal approximation, since it builds on the binomial nature of the data directly.

The BNHM can be fitted using frequentist approaches, for example, via maximum likelihood estimation (MLE).<sup>11</sup> Alternatively, Bayesian methods are commonly used. In a fully Bayesian approach, prior distributions for parameters  $\theta$ ,  $\mu_i$ , and  $\tau$  need to be specified. Note that the parameter  $\theta$  is on the log-odds ratio scale whereas  $\mu_i$  are on the log-odds scale. Baseline risks ( $\mu_i$ ) may be seen as intercept parameters in a standard logistic regression model. For  $\mu_i$ , we use a vague normal prior with mean 0 and standard deviation 10, following the recommendation by Gelman et al.<sup>20</sup> The prior choice for  $\theta$  is our main focus and will be discussed in Section 3.2. The prior choice for the heterogeneity parameter  $\tau$ , which is a standard deviation parameter, has gained much attention in the literature as discussed in the introduction. Friede et al<sup>24</sup> have shown that for meta-analysis of few studies, the use of WIPs for  $\tau$  displays desirable long-run properties in comparison with frequentist alternatives. Following their suggestions, we use a half-normal prior with scale of 0.5 ( $\mathcal{HN}(0.5)$ ) for  $\tau$  which has the median of 0.337 with an upper 95% quantile of 0.98. Values for  $\tau$  of 0.25, 0.5, 1, and 2 represent moderate, substantial, large, and very large heterogeneity.<sup>34</sup> Thus, a  $\mathcal{HN}(0.5)$  prior captures heterogeneity values for log-OR typically seen in meta-analyses of log-ORs and will therefore be a sensible choice in many applications.

#### 3.2 | Derivation of a WIP for the treatment effect

A common prior choice for the treatment effect parameter  $\theta$  is a noninformative (vague) prior such as normal distribution with a large variance, for example,  $\mathcal{N}(0, 100^2)$ . One way of constructing a WIP works via consideration of the prior expected range of treatment effect values.<sup>35</sup> Before the derivation of the WIP for treatment effect parameter  $\theta$ , recall that  $\theta$  is on the log-odds ratio scale. Thus, a value of  $\theta = 0$  means an odds ratio of 1, ie, *no effect*, and a value of  $\theta = 1$  means that odds differ by a factor (ratio) of  $\exp(1) = 2.7$ .

We assume a symmetric prior centered around zero, implying equal probabilities for positive or negative

treatment effects. Symmetry then implies (on the log-odds ratio scale) that

$$P(\theta > q) = P(\theta < -q), \quad (2)$$

where (on the odds ratio scale)

$$\exp(-q) = \frac{1}{\exp(q)}. \quad (3)$$

The prior's scale parameter  $\sigma_{\text{prior}}$  then may be set such that a priori the odds ratio is with 95% probability confined to a certain range:

$$P(1/\delta < \exp(\theta) < \delta) = 95\%. \quad (4)$$

In case of a normal prior with standard deviation  $\sigma_{\text{prior}}$ , we can then simply specify

$$\sigma_{\text{prior}} = \frac{\log(\delta)}{1.96}. \quad (5)$$

We conservatively specify  $\delta$  as 250, meaning that we consider it unlikely that the odds ratio will be larger than 250 or smaller than  $1/250$ . By plugging in this number into (5), we obtain  $\sigma_{\text{prior}} = 2.82$ .

Another way to motivate the prior standard deviation is by using the idea of *unit information priors*.<sup>36,37</sup> When the treatment effect parameter is on the log-odds ratio scale (as in the BNHM), then the standard error is given by  $\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$ . Assuming equal allocation, a neutral effect, and equal counts of events and nonevents, we can simply set the table allocation to  $a = b = c = d = \frac{N}{4}$ . Therefore, if we (heuristically) reverse the argument, a prior

for the log-odds ratio with zero mean and 2.82 standard deviation gives<sup>37</sup>

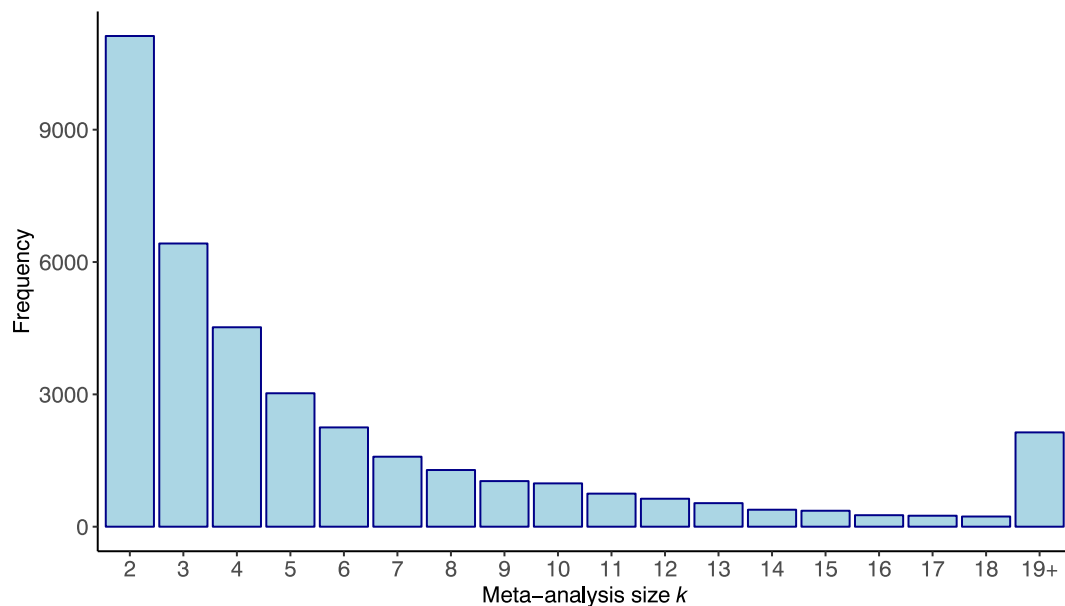
$$2.82 \approx \sqrt{8} = \sqrt{\frac{1}{\frac{N}{4}} + \frac{1}{\frac{N}{4}} + \frac{1}{\frac{N}{4}} + \frac{1}{\frac{N}{4}}}. \quad (6)$$

Hence,  $N = 2$ . In other words, in terms of prior's effective sample size, the choice of  $\sigma_{\text{prior}} = 2.82$  is equivalent to adding two patients to the dataset. From this, it follows that  $\mathcal{N}(0, 2.82^2)$  is a reasonable choice as a WIP for  $\theta$ .

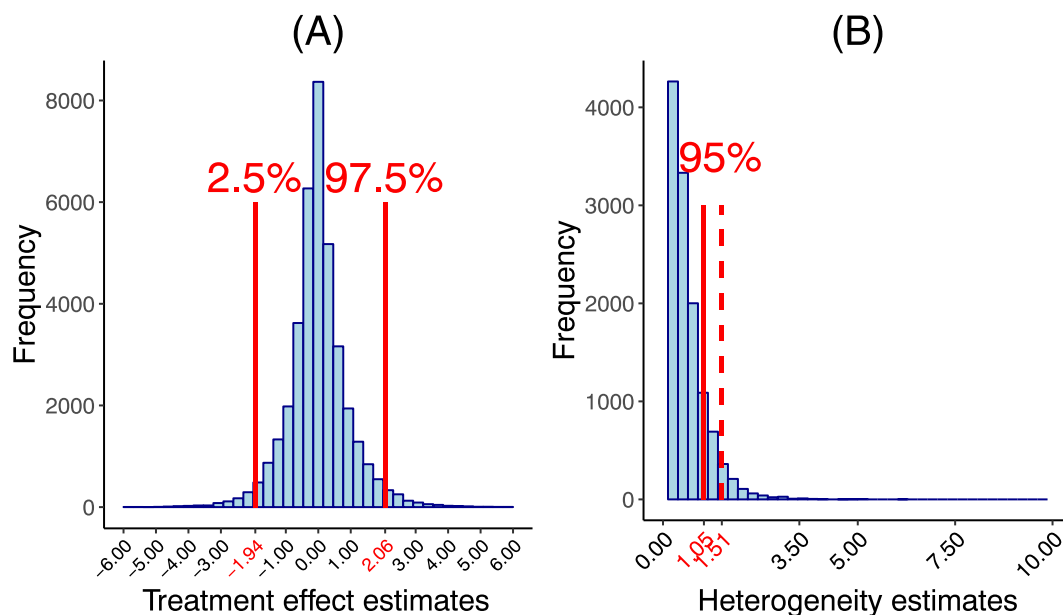
Note also the analogy between this WIP and commonly used continuity corrections: Zero entries in a contingency tables are commonly *fixed* by adding a correction term of 0.5 to each table cell of the single-zero or double-zero study, which also amounts to a total of two patients *added* to the data. This way of conducting continuity correction adds two patients to each single-zero or double-zero study, while the use of WIP is equivalent to adding two patients to the whole dataset.

### 3.3 | Empirical evidence supporting the WIP for the treatment effect

For an empirical investigation of the WIP for treatment effect parameter, we consider the meta-analysis datasets archived in the CDSR. All systematic reviews in the CDSR are available on the Cochrane Library website,<sup>38</sup> and personal or institutional access is required. For downloading the data from the CDSR and converting to CSV files, we use the program `Cochrane_scraper` (version 1.1.0).<sup>39</sup> We were able to access all Cochrane systematic reviews available in March 2018 (CD000004 to CD012788).



**FIGURE 1** The distribution of numbers of studies included in each meta-analysis obtained from the Cochrane Database of Systematic Reviews (CDSR). The category labelled *19+* corresponds to meta analyses of size 19 or larger [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 2** The distribution of the estimates of the mean treatment effect parameter  $\theta$  A, and the distribution of the estimates of the (nonzero) heterogeneity standard deviation parameter  $\tau$  B, obtained from the reanalysis of meta-analysis datasets in Cochrane Database of Systematic Reviews (CDSR) when the binomial-normal hierarchical model (BNHM) via maximum likelihood estimate (MLE) is used for estimation. In A, two red lines ( $-1.94$  and  $2.06$ ) show the 2.5% and 97.5% quantiles of the  $\theta$  estimates, respectively. In B, the solid red line ( $1.05$ ) and the dashed red line ( $1.51$ ) indicate the 95% quantiles of the  $\tau$  estimates including zero-estimates and excluding zero-estimates, respectively. The fraction of zero-estimates of  $\tau$  is 63% [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

Meta-analyses were excluded if they included only one study, if the analysis was labelled as a subgroup or sensitivity analysis or there was insufficient information for classification, or if all data within the meta-analysis appeared to be erroneous. Finally, we only consider meta-analyses with dichotomous outcomes. In total, 37 773 meta-analysis datasets from 4712 reviews are included. Note that we did not distinguish regarding efficacy or safety analyses.

The frequency of the number of studies  $k$  considered for each meta-analysis is illustrated in Figure 1. The percentage of the meta-analyses including five or less studies is 66%. This figure is consistent with other re-analyses of the CDSR (see, eg, previous works<sup>8,21,40</sup>). We re-analyzed the meta-analysis datasets from the CDSR using the BNHM via an MLE approach. This procedure is implemented using the **R** package `lme4`.<sup>41</sup> A histogram of the estimates of  $\theta$  is illustrated in Figure 2A; 2.5% and 97.5% quantiles of the estimates of  $\theta$  are  $-1.94$  and  $2.06$ , respectively. By following Turner et al,<sup>40</sup> we exclude the zero heterogeneity estimates; nonzero estimates of  $\tau$  are shown in Figure 2B. The fraction of nonzero heterogeneity estimates is 63%, which is also consistent with previous findings.<sup>40</sup> The 95% quantile of nonzero estimates of  $\tau$  is  $1.51$ , while the 95% quantile of  $\tau$  estimates including zeroes is  $1.05$ . The underlying distribution of the estimates of  $\theta$  and  $\tau$  and their variability are useful to see how large these estimates are in some general population, in this case the CDSR. Thus, these give us a rough sense of what would be a reasonable

default prior distribution. Therefore, we suggest the use of WIPs,  $\mathcal{N}(0, 2.82^2)$  for  $\theta$  and  $\mathcal{HN}(0.5)$  for  $\tau$ , which are consistent with estimates of  $\theta$  and  $\tau$  empirically observed in the CDSR, meaning that both indicate odds ratios within reasonable ranges, and heterogeneity mostly below  $1.0$ .

#### 4 | IMPLEMENTATION OF THE PROPOSED PROCEDURE IN R USING STAN

The Bayesian implementation of the BNHM can be fitted with the probabilistic programming language **Stan**,<sup>42</sup> which employs a modern Markov chain Monte Carlo (MCMC) algorithm, namely, Hamiltonian Monte Carlo with the No-U-Turn Sampler.<sup>43</sup> It is known that the parametrization of the model can affect the performance of an MCMC algorithm. In the presence of sparse data such as in the meta-analysis of few studies involving rare events, Betancourt et al<sup>44</sup> showed that centered parametrization of a hierarchical model (such as the BNHM) brings computational issues compared with a noncentered parametrization. Thus, we use the noncentered reparametrized version of the BNHM for our implementations. Specifically, applying both location and scale reparametrization, (3.1) becomes  $\mu_i + \theta x_{ij} + u_i \tau x_{ij}$  where  $u_i \sim \mathcal{N}(0, 1)$  and  $x_{ij} = +0.5$  (experimental) or  $x_{ij} = -0.5$  (control). (Correction added on 06 January 2020, after first online publication: The preceding equation has been updated from “ $\mu_i + \theta_i x_{ij} + u_i \tau^2$ ” to “ $\mu_i + \theta x_{ij} + u_i \tau x_{ij}$ ”.)

```

1  data {
2    int<lower=1> N;                // num studies
3    int<lower=0> rctrl[N];        // num events, control
4    int<lower=1> nctrl[N];       // num patients, control
5    int<lower=0> rtrt[N];        // num events, treatment
6    int<lower=1> ntrt[N];       // num patients, treatment
7    vector[2] mu_prior;         // Prior parameters for mu (mean and stdev)
8    vector[2] theta_prior;      // Prior parameters for theta (mean and stdev)
9    real tau_prior;             // Prior scale for tau
10   int tau_prior_dist;         // Indicator for prior distribution of tau
11 }
12
13 parameters {
14   vector[N] mu;                // baseline risks (log odds)
15   real theta;                 // relative trt effect (log OR)
16   vector[N] zeta;             // individual trt effects
17   real<lower=0> tau;          // heterogeneity stdev.
18 }
19
20 transformed parameters {
21   real pctrl[N];
22   real ptrt[N];
23
24   for(i in 1:N) {
25     pctrl[i] = inv_logit(mu[i] - theta * 0.5 - zeta[i] * tau * 0.5);
26     ptrt[i] = inv_logit(mu[i] + theta * 0.5 + zeta[i] * tau * 0.5);
27   }
28 }
29
30 model {
31   // latent variable (random-effects)
32   zeta ~ normal(0, 1);
33   // prior distributions
34   mu ~ normal(mu_prior[1], mu_prior[2]);
35   theta ~ normal(theta_prior[1], theta_prior[2]);
36   if(tau_prior_dist == 1) tau ~ normal(0, tau_prior)T[0,];
37   if(tau_prior_dist == 2) tau ~ uniform(0, tau_prior);
38   if(tau_prior_dist == 3) tau ~ cauchy(0, tau_prior)T[0,];
39   // likelihood
40   rctrl ~ binomial(nctrl, pctrl); // control event count
41   rtrt ~ binomial(ntrt, ptrt);   // treatment event count
42 }

```

Listing 1: **Stan** code defining the binomial-normal hierarchical model. (Correction added on 06 January 2020, after first online publication: Listing 1 has been updated)

For practical applications, learning **Stan's** syntax and the required knowledge of available features in **Stan** might present a hurdle preventing application of **Stan**. To this end, we developed a new **R** package **MetaStan** which is a purpose-built package defined on top of **Rstan**, the **R** interface for **Stan**. Our package **MetaStan** (<https://CRAN.R-project.org/package=MetaStan>) includes the precompiled **Stan** model of the BNHM, which eliminates the compilation time and the need of

learning **Stan's** syntax. The **Stan** code for the BNHM is shown in Listing 1. **MetaStan** includes different options for WIPs of the model parameters of the BNHM. **MetaStan** syntax is similar to the syntax of the popular meta-analysis package **metafor**<sup>27</sup> so that it should be easy for a **metafor** user to utilize our package. The syntax of **MetaStan** is displayed for the pediatric transplantation example in Section 6, and in Appendix A, we show how to install and use **MetaStan**.

## 5 | SIMULATION STUDY

In order to assess the long-run properties of the proposed approach and compare it with some alternatives, we conducted a simulation study.

### 5.1 | Simulation setup

The simulation scenarios are similar to those considered by Friede et al,<sup>24</sup> but with the important difference that we focus on rare events. The datasets are generated under the BNHM, more specifically (3.1). Numbers of studies ( $k \in \{2, 3, 5\}$ ) and true treatment effects ( $\theta = \{-5, -4, -3, -2, -1, -0.5, 0, 0.5, 1, 2, 3, 4, 5\}$ ) are varied, resulting in a total of 39 simulation scenarios. To reflect the rare-event cases, true baseline risks on the probability scale are taken uniformly between 0.005 and 0.05. Following Kuss,<sup>3</sup> a log-normal distribution is fitted to the sample sizes obtained from the CDSR data, resulting in a log-normal distribution with parameters  $\mu = 5$  and  $\sigma = 1$ . Hence, sample sizes are generated from  $\mathcal{LN}(5, 1)$ , the minimum sample size is restricted to two patients (values below 2 are rounded up to 2), and at least one patient in each treatment arm is assumed. The degree of heterogeneity ( $\tau$ ) is taken as  $\tau = 0.28$  (moderate heterogeneity), which is the median value of the predictive distribution for between-study heterogeneity in a meta-analysis in a general setting as estimated by Turner et al.<sup>40</sup> According to a binomial probability of 0.5, patients were allocated to the treatment groups, thus mimicking randomization. The simulations were carried out with 10 000 replications per scenario. The data sparsity is reflected in the average fractions of single-zero or double-zero studies in a simulated meta-analysis dataset, which are shown in Figure 3A. Notice that the fractions of the single-zero and double-zero studies are the highest when true treatment effect is  $-5$ ,

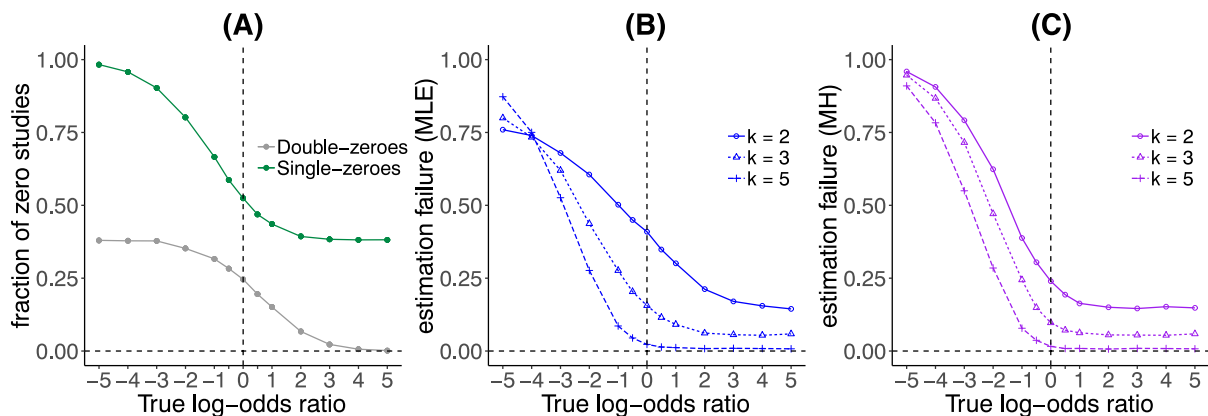
and they are decreasing with the increase of the treatment effect.

The proposed approach (BNHM using a WIP, that is  $\mathcal{N}(0, 2.82^2)$ , for  $\theta$ : WIP) and four comparators are included in the analysis, namely, BNHM using a vague prior ( $\mathcal{N}(0, 100^2)$ ) for  $\theta$  (Vague), BNHM using MLE (MLE), the Mantel-Haenszel (MH) method<sup>5</sup> and a Bayesian implementation of the beta-binomial model (BBM).<sup>3</sup> It is important to note the differences of the MH and BBM from the BNHM methods. MH is a fixed-effect meta-analysis method, and BBM has a different underlying data generating process than the BNHM. For both Vague and WIP approaches, the prior for  $\tau$  and  $\mu$  are taken as  $\mathcal{HN}(0.5)$ , and  $\mathcal{N}(0, 10^2)$ , respectively. The MH estimator of the treatment effect parameter is given by

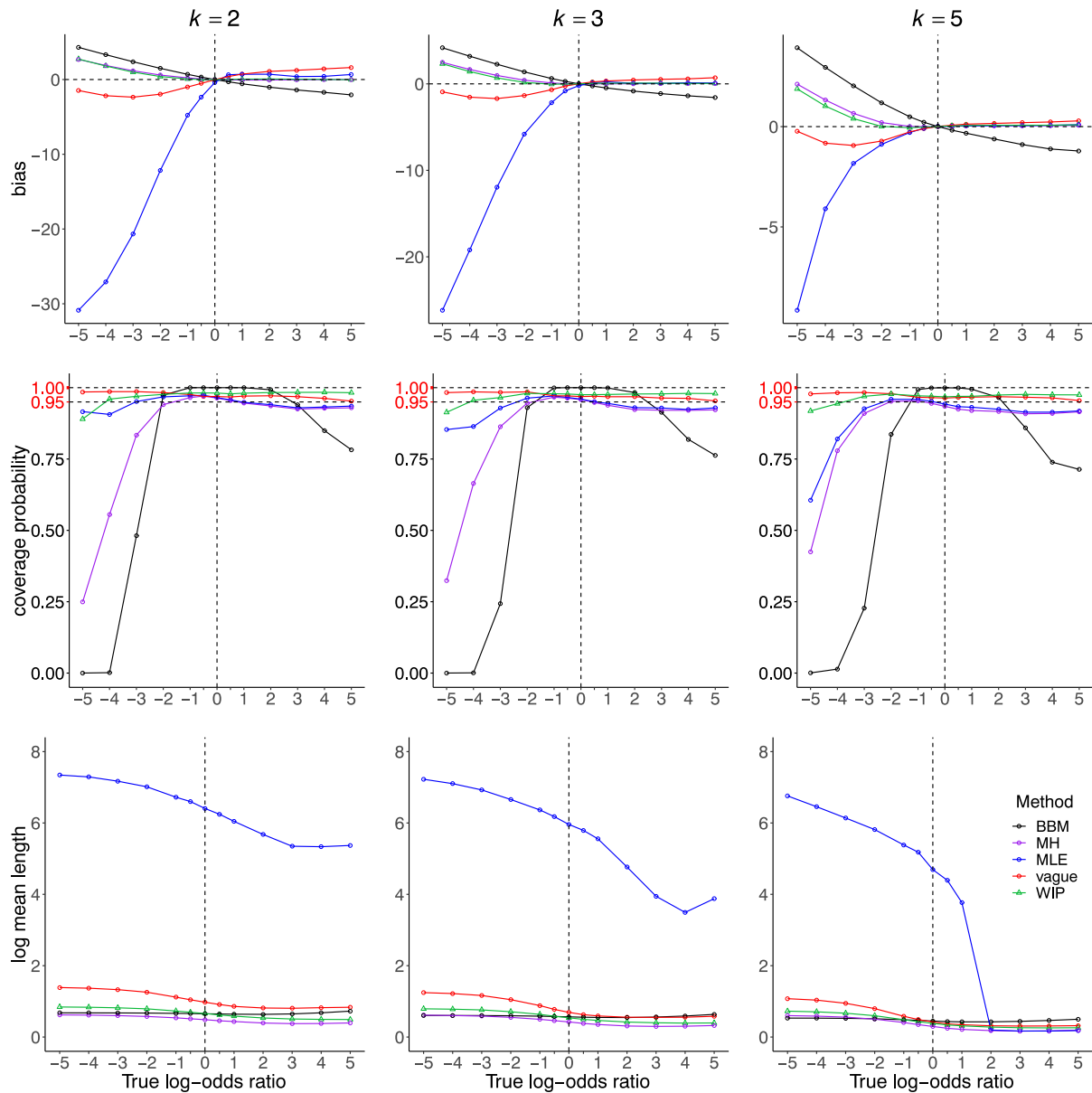
$$\hat{\theta}_{\text{MH}} = \frac{\sum_{i=1}^k \frac{r_{i1}(n_{i0}-r_{i0})}{n_i}}{\sum_{i=1}^k \frac{r_{i0}(n_{i1}-r_{i1})}{n_i}},$$

where  $n_i = n_{i0} + n_{i1}$ . In the BBM, the event counts  $r_{ij}$  are modeled using a binomial distribution,  $r_{ij} \sim \text{Bin}(\pi_{ij}, n_{ij})$ , as in the BNHM. The probabilities of event are assumed to be beta distributed:  $\pi_{ij} \sim \text{Beta}(\alpha_j, \beta_j)$  where both arms share the same correlation parameter  $\rho = \frac{1}{\alpha_0 + \beta_0 + 1} = \frac{1}{\alpha_1 + \beta_1 + 1}$ , implying  $\alpha_0 + \beta_0 = \alpha_1 + \beta_1$ . It is common to reparametrize the model using mean parameters  $\Phi_j$  such that  $\Phi_j = \frac{\alpha_j}{\alpha_j + \beta_j}$ . Finally, the linear predictor can be written as  $\text{logit}(\Phi_j) = \mu + \theta x_j$  where  $\theta$  is the parameter for the treatment effect, and  $x_j$  is a treatment indicator, 1 = experimental ( $j = 1$ ) and 0 = control ( $j = 0$ ). Vague priors are chosen for all parameters, namely, uniform priors across the interval  $[0, 1]$  for all three parameters:  $\Phi_0$ ,  $\Phi_1$ , and  $\rho$ .

Three MCMC chains were run in parallel for a total of 2000 iterations including 1000 iterations of burn-in. These values are tested in some replications; convergence



**FIGURE 3** The average fraction of single-zero or double-zero studies in a simulated meta-analysis dataset A, and the fraction of the estimation failure for maximum likelihood estimate (MLE) and Mantel-Haenszel (MH) with different numbers of studies  $k$  used in the simulations (B and C) are shown [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



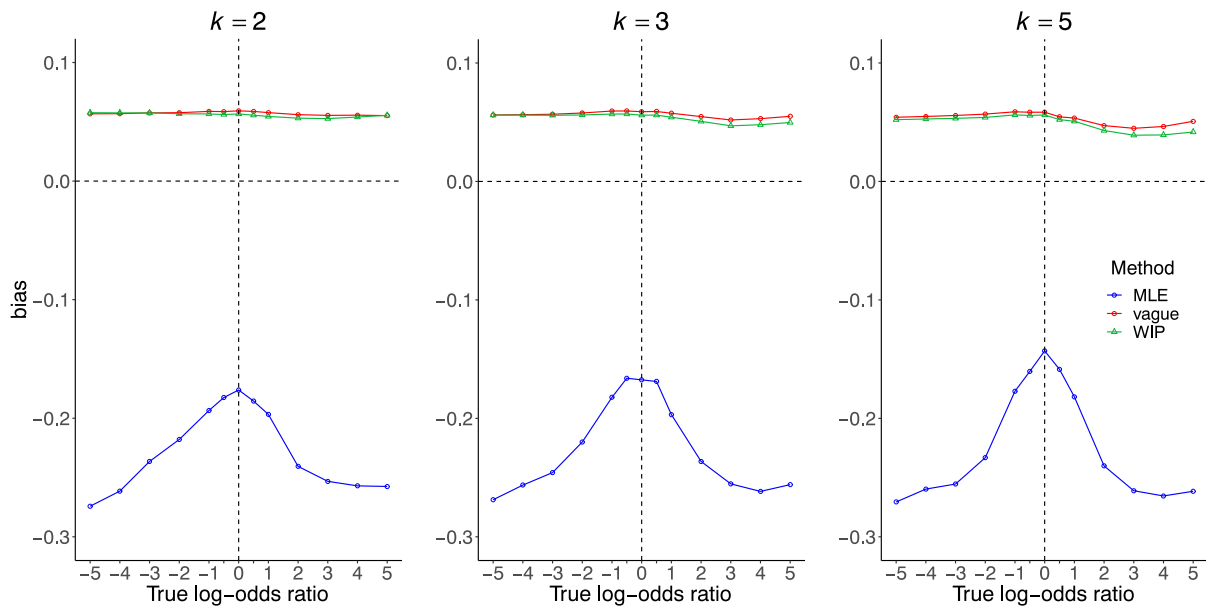
**FIGURE 4** The bias for the mean treatment effect  $\theta$ , coverage probabilities, and log mean length of the interval estimates for  $\theta$  obtained by five methods (beta-binomial model [BBM], Mantel-Haenszel [MH], maximum likelihood estimate [MLE], Vague, and weakly informative prior [WIP]) are shown [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

diagnostics are assessed and chosen accordingly. All chains were assumed to have reached convergence (no estimation failure). We used the package `lme4` for the MLE (using the adaptive Gauss-Hermite approximation to the maximum log-likelihood) and `metafor` for the MH (without using any continuity corrections) whereas the Vague, WIP, and BBM methods were fitted with our **MetaStan** package. Note that we use highest density intervals (HDI), which are the shortest credible intervals, as opposed to the commonly used equal-tailed credible intervals. The HDI were obtained using the `HDInterval`<sup>45</sup> package. All computations were performed using **R**.<sup>46</sup> The code for the computations for all methods used in the simulations is provided in Appendices A to D.

## 5.2 | Simulation results

For the MLE and the MH, the fractions of estimation failures are shown in Figure 3B and 3C. Estimation failure occurred for the MLE when the Gauss-Hermite approximation does not converge to the maximum log-likelihood, or when the MH estimator is not defined. The MLE and MH methods show very similar behavior of the estimation failure. Estimation failure is closely related to the fraction of meta-analysis datasets including single-zero or double-zero studies in the dataset, which can be seen by comparing Figure 3A and 3B,C. This is because when the data are highly sparse, estimation becomes more challenging for both MLE and MH. As a performance measure, we





**FIGURE 5** The bias for the heterogeneity parameter  $\tau$  obtained by three methods (maximum likelihood estimate [MLE], Vague, and weakly informative prior [WIP]) is shown. True heterogeneity standard deviation is assumed to be  $\tau=0.28$  [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

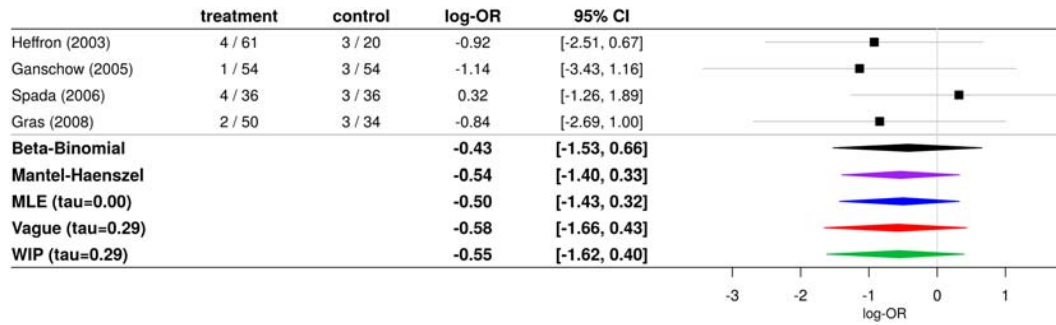
use the bias  $(\frac{1}{N} \sum_{i=1}^N (\hat{\theta} - \theta))$  based on the MLE, the MH estimator, and posterior medians. The direction of the bias is also important, since depending on the nature of the outcome (safety or efficacy), a positive or a negative bias may be considered *conservative*. Moreover, the coverage probability and the mean length of interval estimates for  $\theta$  are reported. The coverage probability of 95% for interval estimates and shorter interval estimates are desirable.

The bias of posterior medians from the Vague, the WIP and the BBM, and for MLEs from the MLE and for the MH estimator from the MH across scenarios is displayed in the first row of Figure 4. Note that failed runs were excluded from the calculation of performance measures that is relevant only for MLE and MH. The MLE shows unacceptably high bias for the scenarios with  $\theta \leq 0$ , corresponding to the scenarios in which the fraction of zero studies is also very high. On the other hand, the MH estimator clearly outperforms the MLE and exhibit bias very close to the WIP. The WIP displays somewhat positive bias whereas the Vague shows negative bias for the scenarios with  $\theta \leq 0$ . This behavior of WIP is expected, since the WIP shrinks the posterior towards zero. For safety analyses, a positive bias commonly means a more conservative behavior and may hence be considered less harmful than a negative bias. It is important to note that the results of the bias behave similar to the fraction of zero studies and the fraction of estimation failure of the MLE, meaning that the bias is higher in scenarios with more sparse data. Since the Vague approach uses a vague prior on  $\theta$ , one might expect a somewhat similar behavior of bias from the Vague and the MLE approaches. However, the fact that the Vague approach includes a WIP for  $\tau$  and that estimation is based on inte-

gration rather than maximization may be explanations of the better performance of the Vague method in comparison with the MLE. The WIP and the MH outperforms the BBM in terms of bias across all scenarios. Performance in terms of bias is improving for all methods when the number of studies  $k$  is increasing. For Figures 3 and 4, the curves are not symmetric around zero. This asymmetry is due to the fact that while the true treatment effect (log-OR) is varied between  $-5$  and  $+5$ , the true baseline risk (probability) is drawn uniformly between 0.005 and 0.05 in the simulations.

Figure 4 also shows coverage probabilities and log mean lengths for 95% HDI obtained by the Vague, the WIP, the BBM, and for 95% Wald confidence intervals (CIs) obtained by the MLE and the MH. The CI and HDI obtained by the MH and the BBM show unacceptably low coverage especially for  $\theta < -2$ . However, the undercoverage of the BBM and somewhat relative poor performance in terms of bias may stem from the fact that data are generated under the BNHM. Also, the CI obtained by the MLE displays low coverage especially for  $k = 5$ . We will revisit the coverage of the MLE in the discussion. The WIP method shows higher coverage than nominal level across all different true treatment effects except for  $\theta = -5$ . On the other hand, the HDI obtained by WIP are shorter in comparison with HDI obtained by the Vague and CI obtained by the MLE approaches.

Lastly, the bias for the heterogeneity parameter  $\tau$  obtained by three methods (the MLE, the Vague, and the WIP) are demonstrated in Figure 5. For Bayesian methods, posterior medians are used as the point estimates. Recall that the prior used for  $\tau$  both in the Vague and the



**FIGURE 6** The motivating pediatric transplantation application when the outcome is death: Top panel displays the observed log-odds ratios (computed using a continuity correction in case of zero counts). The bottom panel shows mean treatment effect estimates of  $\theta$  obtained by beta-binomial model (BBM), Mantel-Haenszel (MH), maximum likelihood estimate (MLE), Vague, and weakly informative prior (WIP). Heterogeneity parameter estimates  $\tau$  are also given on the left [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

WIP is weakly informative ( $\mathcal{HN}(0.5)$ ). The MLE underestimates the true heterogeneity, whereas the Vague and the WIP methods slightly overestimate it. The Vague and the WIP produce very similar bias. These observations are in alignment with the conclusions made by Friede et al.<sup>24</sup>

## 6 | EXAMPLE REVISITED

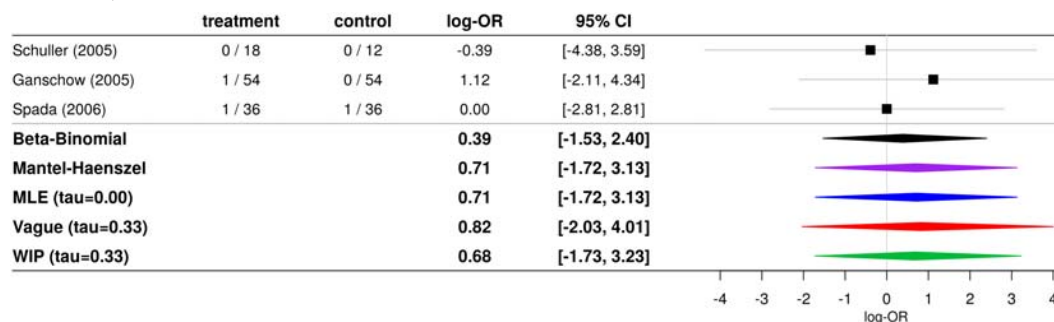
Returning to the dataset described in Section 2, we consider the data on death and PTLD outcomes shown in Table 1. The observed log-odds ratios are displayed in Figures 6 and 7. To be able to visualize the observed log-odds ratios when there is a single-zero or double-zero study, a continuity correction of 0.5 is added to all cells of the single-zero or double-zero study's contingency table. The wide CI for observed log-odds ratios reflect the rather small sample sizes in the datasets. Furthermore, the variability in the point estimates may be reflected upon to assess the degree of heterogeneity between trial estimates.

We analyze the datasets using the five methods investigated in the simulation studies, namely, the Vague, WIP, MLE, MH, and BBM approaches. The code to implement the MLE and the MH are given in Appendix B. Recall that the only difference between Vague and WIP is the prior used for the treatment effect parameter  $\theta$  in the model, namely,  $\mathcal{N}(0, 100^2)$  for the former and  $\mathcal{N}(0, 2.82^2)$  for the latter. WIP can be implemented in a routine data analysis using our `MetaStan` package as follows:

```
bnhm.wip.CrinsPTLD.stan = meta_stan(ntrt = exp.total,
                                   nctrl = cont.total,
                                   rtrt = exp.PTLD.events,
                                   rctrl = cont.PTLD.event,
                                   data = dat.Crins2014.PTLD,
                                   tau_prior_dist = "half-normal",
                                   tau_prior = 0.5,
                                   delta = 250)
```

The argument `delta` corresponds to  $\delta$  from (5) and thus is used to calculate the WIP for  $\theta$ . Alternatively, one can directly specify the prior parameters for  $\theta$ , in our case, equivalently, we can have `theta_prior = c(0, 2.82)`. The Vague method is simply implemented by omitting the argument `delta` and specifying `theta_prior = c(0, 100)`. The BBM is also implemented in `MetaStan`, and the required syntax is shown in Appendix C. To check MCMC convergence, we use the Gelman-Rubin statistics and traceplots. For the WIP approach, the corresponding traceplots are shown in Figures A1 and A2 for death and PTLD outcomes, respectively. There was no divergence reported for both datasets. The MLE fit and the MH estimation for the dataset where death is the outcome does not cause any warning from `lme4` and `metafor`, respectively. For the PTLD outcome, `lme4` gives a warning suggesting that the estimates may not be reliable. Nevertheless, it produces the MLE estimate and CI for treatment effect parameter, and we report them. For PTLD outcomes, when computing the MH estimator, `metafor` gives a warning due to double-zero studies (double-zero studies are removed from the analysis by default) but still returns an estimate. Note that both MLE and MH ignore the double-zero study (Schuller et al<sup>30</sup>); hence, the analyses are based on two studies only.

The results for the death and PTLD outcomes from the five methods are shown in Figures 6 and 7, respectively. For MLE and MH, the estimates and 95% CI are given. For



**FIGURE 7** The motivating pediatric transplantation application when the outcome is posttransplant lymphoproliferative disease (PTLD): Top panel displays the observed log-odds ratios (computed using a continuity correction in case of zero counts). The bottom panel shows mean treatment effect estimates of  $\theta$  obtained by beta-binomial model (BBM), Mantel-Haenszel (MH), maximum likelihood estimate (MLE), Vague, and weakly informative prior (WIP). Heterogeneity parameter estimates  $\tau$  are also given on the left [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

Vague, WIP, and BBM, posterior medians and 95% HDI are shown. Both for PTLD and death outcomes, apart from the BBM, the point estimates of  $\theta$  from the four methods look quite similar. The differing behavior of the BBM was also observed in the simulations. The PTLD data are similar to the scenarios when the number of studies is three, and the true treatment effect is in the range from 0 to 1. Negative bias obtained by the BBM can be seen in Figure 5 (in the corresponding scenario). The death data are similar to the scenarios when number of studies is five (since it is not highly sparse), and true treatment effect is in the range from  $-1$  to 0. Here, positive bias obtained by the BBM can be seen in Figure 5 (in the corresponding scenario). Furthermore, the point estimates obtained by the WIP and the MH are very close as in the simulations. MLE gives shorter interval estimates compared with Bayesian alternatives, this is (partly) because  $\tau$  was estimated as 0. In the original paper, Crins et al<sup>26</sup> fitted a normal-normal hierarchical model using REML,<sup>27</sup> and the risk ratio was used as the measure of the treatment effect. They concluded that treatment IL-2RA failed to show statistically significant result for reducing death. We obtained similar point estimates with somewhat wider interval estimates to Crins et al,<sup>26</sup> specifically their risk ratio estimate was 0.61 (CI, 0.27-1.37), and we obtained the odds ratio estimate 0.58 (HDI 0.20-1.49) using the WIP method. Concerning the PTLD, the risk ratio was estimated as 1.60 (CI, 0.20-12.67) by Crins et al,<sup>26</sup> the odds ratio is estimated 1.98 (HDI 0.18-25.18) using the WIP method. The wider interval estimates obtained by WIP may stem from the fact that the uncertainty in  $\tau$  is taken into account.

The estimates of the between-trial heterogeneity  $\hat{\tau}$  are also included in the figure, which are only available for the Vague, WIP, and MLE. Considering death as outcomes, the heterogeneity parameter  $\tau$  is estimated 0.29, 0.29, and 0.00 using WIP, Vague, and MLE, respectively. Similarly for the PTLD outcomes, for  $\hat{\tau}$ , we obtained 0.33, 0.33, and

0.00 using WIP, Vague, and MLE, respectively. The heterogeneity parameter of the BBM  $\rho$  is estimated as 0.34 and 0.03 for PTLD and death outcomes, respectively. Moreover, Crins et al<sup>26</sup> concluded that there is no evidence for heterogeneity between trials using using Cochran's Q test for both death and PTLD outcomes. Since the prior used for  $\tau$  is the same for WIP and Vague, similar heterogeneity estimates are expected. The similar  $\tau$  estimates by WIP and Vague were also observed in the simulations (Figure 5). On the other hand, the MLE estimate ( $\hat{\tau} = 0.00$ ) is most probably underestimating the actual amount of heterogeneity. The underestimation of  $\tau$  by MLE and slightly lower bias of the WIP compared with the Vague was observed in the simulations (Figure 5).

## 7 | CONCLUSIONS AND DISCUSSION

An assumption of the homogeneity is often considered unrealistic for meta-analyses in biomedical sciences; hence, random-effects meta-analysis models are suggested.<sup>6</sup> Furthermore, as can be seen in the CDSR, a substantial fraction of published meta-analyses is based on few studies only. On the other hand, fitting a random-effects models based on only few studies often poses problems for inference, as certain asymptotics cannot be relied upon.<sup>47</sup> Additional issues arise for binary outcomes when only few or no events are observed in some of the studies or study arms. To deal with such data sparsity in the meta-analysis, we have proposed the use of WIPs for the treatment effect parameter  $\theta$  in a BNHM. We demonstrated how a normal WIP for  $\theta$  can be derived by considering an a priori interval for the treatment effect on a log-odds ratio scale. Also, the empirical evidence obtained from 37 773 meta-analyses with binomial outcomes from the CDSR supports the proposed WIP. In simulation studies, the suggested method displays lower bias for  $\theta$  and

substantially shorter interval estimates for  $\theta$  with somewhat higher coverage than nominal level in comparison to alternative methods.

The use of a Bayesian approach exhibits analogy of some degree to the use of continuity corrections. While continuity corrections might to some extent be perceived as ad hoc makeshift fixes, they have also quite doubtlessly proven very useful in practice. A Bayesian approach tackles the problem from a very different angle, but it is not so surprising that the resulting procedure again exhibits some similarity to continuity corrections. The relation to current common practice may in fact also be seen as somewhat comforting. Use of an (informative) prior within a Bayesian analysis on the other hand is not a desperate measure; it is rather an integral part of a coherent model specification that may also be subjected to checks of plausibility and operating characteristics; this is what we have tried to demonstrate in the present paper.

The simulation results displayed in Figure 4 are somewhat in contrast to the results given by Friede et al.,<sup>24</sup> who observed lower coverage than nominal level of MLE methods in a similar setting, but not based on rare events. We also investigated a scenario closer to their setup by considering higher baseline risks between 0.05 and 0.20. The results are shown in Figure D1, and indeed, here, the MLE method exhibits lower coverage than nominal level, as reported by Chung et al.<sup>22</sup> and Friede et al.<sup>24</sup> The high bias and too wide interval estimates obtained by the Vague and the MLE are still present, but not as high as in the results of the simulations in which true baseline risks are lower.

Jackson et al.<sup>11</sup> investigated seven random-effects meta-analysis models including the BNHM which we consider in this paper (*model 4* in Jackson et al.<sup>11</sup>) and another parametrization of the BNHM (*model 2* in Jackson et al.<sup>11</sup>). The only difference in the specification between the two models is that in their Model 2, the treatment indicator  $x_{ik}$  of (3.1) is +1 for the experimental arm, and 0 for the control arm. Note that commonly used network meta-analysis models, for example,<sup>48</sup> are generalizations of Model 2 in Jackson et al.<sup>11</sup> As reported by Jackson et al.,<sup>11</sup> we also observe the underestimation of the heterogeneity parameter  $\tau$  and hence decided to only consider their model 4. On the other hand, it is important to note that the usage of a WIP for  $\theta$  also improves the performance in model 2, as we have seen for the model 4.

This investigation has some limitations. One crucial limitation is that we only considered the BNHM as a data-generating process in our simulation study. Hence, we did not investigate the robustness of the BNHM under model misspecification. Also, the design of the simulation study constitutes a model misspecification problem for the MH method, which is a fixed-effect model, and for the BBM, which assumes a different underlying

data-generating process. Moreover, we did not consider other parametrizations of the BNHM as described, eg, in Jackson et al.<sup>11</sup> Lastly, one may find it too restrictive to have a normal prior for  $\theta$  as we have in our proposed model, it may be worth exploring alternatives like Cauchy or log- $F$  distributions<sup>17,20</sup> for penalization.

The proposed approach is not restricted to the BNHM; similar approaches may analogously be defined in other models, eg, a Poisson-normal hierarchical model. However, a crucial point is that the treatment effect parameter is explicitly parameterized in the model, so that it can directly be *penalized* via the prior specification. Hence, so-called contrast-based models<sup>25</sup> (in which relative treatment effects are assumed to be exchangeable across trials) are suitable for this purpose unlike arm-based models. Note that this is also related to the inclusion of baseline risks as fixed effects with vague priors. This was on purpose as we consider this closest to the idea of stratifying the analyses by study, a common feature of meta-analyses regardless of fixed or random-effects. Furthermore, the contrast-based models such as the BNHM preserve the randomization, in contrast to the arm-based models as explained in Dias and Ades.<sup>25</sup>

The BNHM can be extended to a network meta-analysis model,<sup>49</sup> which is desirable if there are multiple treatments, and/or multiarm trials in the dataset. Even if the dataset in a network meta-analysis consists of many studies overall, some of the treatment effects may still be informed by few studies only. Thus, the use of WIPs for treatment effect parameters in the context of network meta-analysis with rare events can be very helpful. Different distributions as WIP for  $\theta$ , different parametrizations of BNHM, or different data models can be implemented in **Stan** or MCMC methods in general. Although, currently, our package **MetaStan** is restricted to use a BNHM and BBM for pairwise meta-analysis, we will consider to extend it to conduct meta-analysis and network meta-analysis with flexible data model and prior options in the future.

## ACKNOWLEDGMENT

We thank Leonhard Held who contributed valuable comments and pointed us to several important references.

## HIGHLIGHTS

**What is already known:** Standard random-effects meta-analysis methods are not suitable for meta-analysis of few studies with rare events.

**What is new:** To deal with data sparsity present in the random-effects meta-analysis of few studies with rare events, we suggest the use of weakly informative priors as penalization for the treatment effect parameter.

**Potential impact for RSM readers outside the authors' field:** To make it more accessible to meta-analysts, a publicly available R package, *MetaStan*, is developed for fitting Bayesian meta-analysis models using weakly informative priors.

## CONFLICT OF INTEREST

The author reported no conflict of interest.

## ORCID

Burak Kürsüd Günhan  <https://orcid.org/0000-0002-7454-8680>

Christian Röver  <https://orcid.org/0000-0002-6911-698X>

Tim Friede  <https://orcid.org/0000-0001-5347-7441>

## REFERENCES

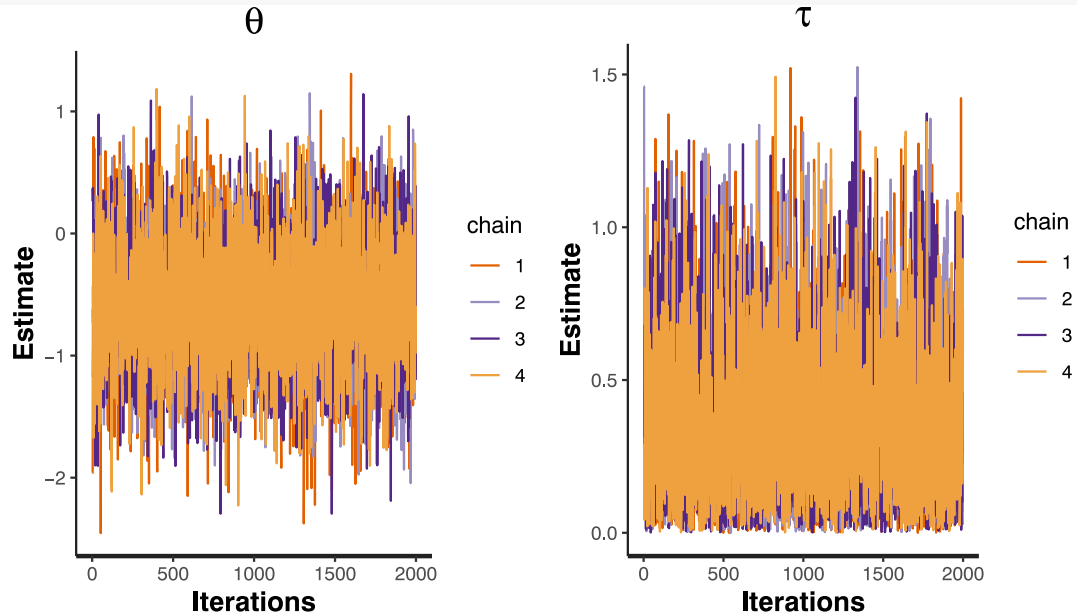
- Higgins JPT, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester: Wiley; 2008.
- Friedrich JO, Adhikari NKJ, Beyene J. Inclusion of zero total event trials in meta-analyses maintains analytic consistency and incorporates all available data. *BMC Med Res Methodol*. 2007;7(1):5. <https://doi.org/10.1186/1471-2288-7-5>
- Kuss O. Statistical methods for meta-analyses including information from studies without any events—add nothing to nothing and succeed nevertheless. *Stat Med*. 2015;34(7):1097-1116.
- Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Prog Cardiovasc Dis*. 1985;27(5):335-371.
- Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst Monogr*. 1959;22(4):719-748. <https://doi.org/10.1093/jnci/22.4.719>
- Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc*. 2009;172(1):137-159. <https://doi.org/10.1111/j.1467-985X.2008.00552.x>
- Turner RM, Jackson D, Wei Y, Thompson SG, Higgins JPT. Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. *Stat Med*. 2015;34(6):984-998.
- Kontopantelis E, Springate DA, Reeves D. A re-analysis of the cochrane library data: the dangers of unobserved heterogeneity in meta-analyses. *PLoS One*. 2013;8(7):1-14.
- Hedges LV, Olkin I. *Random effects models for effect sizes. Statistical methods for meta-analysis*; 1985.
- Bradburn MJ, Deeks JJ, Berlin JA, Russell Localio A. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Stat Med*. 2007;26(1):53-77. <https://doi.org/10.1002/sim.2528>
- Jackson D, Law M, Stijnen T, Viechtbauer W, White IR. A comparison of seven random-effects models for meta-analyses that estimate the summary odds ratio. *Stat Med*. 2018;37(7):1059-1085.
- Stijnen T, Hamza TH, Özdemir P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Stat Med*. 2010;29(29):3046-3067.
- Böhning D, Mylona K, Kimber A. Meta-analysis of clinical trials with rare events. *Biom J*. 2015;57(4):633-648. <https://doi.org/10.1002/bimj.201400184>
- Cai T, Parast L, Ryan L. Meta-analysis for rare events. *Stat Med*. 2010;29(20):2078-2089. <https://doi.org/10.1002/sim.3964>
- Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. *Stat Med*. 1995;14(24):2685-2699. <https://doi.org/10.1002/sim.4780142408>
- Albert A, Anderson JA. On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*. 1984;71(1):1-10.
- Greenland S, Mansournia MA. Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions. *Stat Med*. 2015;34(23):3133-3143.
- Firth D. Bias reduction of maximum likelihood estimates. *Biometrika*. 1993;80(1):27-38.
- Heinze G, Schemper M. A solution to the problem of separation in logistic regression. *Stat Med*. 2002;21(16):2409-2419.
- Gelman A, Jakulin A, Pittau MG, Su Y. A weakly informative default prior distribution for logistic and other regression models. *Ann Appl Stat*. 2008;2(4):1360-1383.
- Davey J, Turner RM, Clarke MJ, Higgins JPT. 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(1):160.
- Chung Y, Rabe-Hesketh S, Choi IH. Avoiding zero between-study variance estimates in random-effects meta-analysis. *Stat Med*. 2013;32(23):4071-4089.
- Gelman A. Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian Anal*. 2006;1(3):515-534. <https://doi.org/10.1214/06-BA117A>
- Friede T, Röver C, Wandel S, Neuenschwander B. Meta-analysis of few small studies in orphan diseases. *Res Synth Methods*. 2017;8(1):79-91. <https://doi.org/10.1002/jrsm.1217>
- Dias S, Ades AE. Absolute or relative effects? Arm-based synthesis of trial data. *Res Synth Methods*. 2016;7(1):23-28. <https://doi.org/10.1002/jrsm.1184>
- Crins ND, Röver C, Goralczyk AD, Friede T. Interleukin-2 receptor antagonists for pediatric liver transplant recipients: a systematic review and meta-analysis of controlled studies. *Pediatr Transplant*. 2014;18(8):839-850. <https://doi.org/10.1111/ptr.12362>
- Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw*. 2010;36(3):1-48. <http://www.jstatsoft.org/v36/i03/>
- Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327:557-560.
- Heffron TG, Pillen T, Smallwood GA, Welch D, Oakley B, Romero R. Pediatric liver transplantation with daclizumab induction therapy. *Transplant*. 2003;75:2040-2043.
- Schuller S, Wiederkehr JC, Coelho-Lemos IM, Avilla SG, Schultz C. Daclizumab induction therapy associated with tacrolimus-mmf has better outcome compared with tacrolimus-MMF alone in pediatric living donor

- liver transplantation. *Transplant Proc.* 2005;37(2): 1151-1152.
31. Ganschow R, Grabhorn E, Schulz A, Hugo AV, Rogiers X, Burdelski M. Long-term results of basiliximab induction immunosuppression in pediatric liver transplant recipients. *Pediatr Transplant.* 2005;9(6):741-745.
  32. Spada M, Petz W, Bertani A, et al. Randomized trial of basiliximab induction versus steroid therapy in pediatric liver allograft recipients under tacrolimus immunosuppression. *Am J Transplant.* 2006;6(8):1913-1921.
  33. Gras JM, Gerkens S, Beguin C, et al. Steroid-free, tacrolimus-basiliximab immunosuppression in pediatric liver transplantation: clinical and pharmacoeconomic study in 50 children. *Liver Transpl.* 2008;14(4):469-477.
  34. Spiegelhalter DJ, Abrams KR, Myles JP. *Prior distributions. Bayesian Approaches to Clinical Trials and Health-Care Evaluation.* West Sussex: CRC Press; 2004.
  35. Greenland S. Bayesian perspectives for epidemiological research: I. Foundations and basic methods. *Int J Epidemiol.* 2006;35(3):765-775.
  36. Kass RE, Wasserman L. A reference Bayesian test for nested hypotheses and its relationship to the schwarz criterion. *J Am Stat Assoc.* 1995;90(431):928-934.
  37. Röver C. Bayesian random-effects meta-analysis using the bayesmeta R package. arxiv.org e-print archive. <https://arxiv.org/abs/1711.08683> updated november 23, 2017. accessed August 30, 2018.
  38. The cochrane. cochrane database of systematic reviews. <https://www.cochranelibrary.com> accessed april 15, 2018.
  39. Springate D. Cochrane\_scraper v1.1.0. <https://doi.org/10.5281/zenodo.10782> updated july, 2014. accessed april, 2018.
  40. Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JPT. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the cochrane Database of Systematic Reviews. *Int J Epidemiol.* 2012;41(3):818-827. <https://doi.org/10.1093/ije/dys041>
  41. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw.* 2015;67(1):1-48.
  42. Stan Development Team. Stan modeling language user's guide and reference manual, version 2.17.0.; 2017.
- ```
install.packages("MetaStan")
```
43. Betancourt M. A conceptual introduction to Hamiltonian Monte Carlo. arxiv.org e-print archive. <https://arxiv.org/abs/1701.02434>. updated july 16, 2018. accessed august 24, 2018.
- ```
library("MetaStan")
data("dat.Crins2014")
## Subset of dataset where PTLD outcomes available
dat.Crins2014.PTLD = subset(dat.Crins2014, is.finite(exp.PTLD.events))
```
44. Betancourt M, Girolami M. Hamiltonian Monte Carlo for hierarchical models. *Current Trends in Bayesian Methodology With Applications*; 2015.
  45. Meredith M, kruschke J. HDInterval: Highest (posterior) density intervals, R package version 0.2.0.<https://CRAN.R-project.org/package=HDInterval>; 2018.
  46. R core team. *R: A language and environment for statistical computing*; R foundation for statistical computing. <https://www.R-project.org/> Vienna, Austria. 2018.
  47. Jackson D, White IR. When should meta-analysis avoid making hidden normality assumptions?. *Biom J.* 2018;60(6):1040-1058. <https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.201800071>
  48. Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU Technical support document 2: A generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. <http://www.nicedsu.org.uk> last updated September 2016; 2011.
  49. Günhan BK, Friede T, Held L. A design-by-treatment interaction model for network meta-analysis and meta-regression with integrated nested laplace approximations. *Res Synth Methods.* 2018;9(2):179-194.
- How to cite this article:** Günhan BK, Röver C, Friede T. Random-effects meta-analysis of few studies involving rare events. *Res Syn Meth.* 2020;11: 74–90. <https://doi.org/10.1002/jrsm.1370>
- ### APPENDIX A: HOW TO USE THE METASTAN R PACKAGE
- The stable version of **MetaStan** is available on CRAN (<https://CRAN.R-project.org/package=MetaStan>) and can be installed as follows:
- The example described in the text (Crins dataset) is available in the package, and it can be loaded as follows:
- Additional information can be obtained by typing `?dat.Crins2014` (for any dataset and function in the package).
- `meta_stan` is the main fitting function of this package. The main computations are executed via the **rstan** package's `sampling` function. We can fit the binomial-normal hierarchical using a WIP for treatment effect as follows:

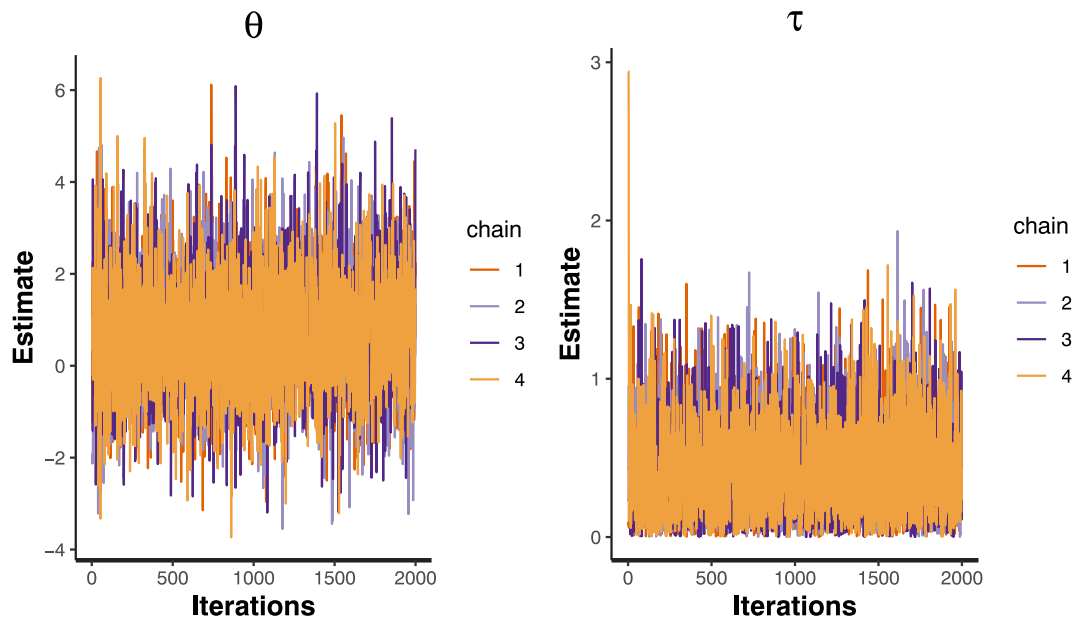
```

bnhm.wip.CrinsPTLD.stan = meta_stan(ntrt = exp.total,
                                   nctrl = cont.total,
                                   rtrt = exp.PTLD.events,
                                   rctrl = cont.PTLD.event,
                                   data = dat.Crins2014.PTLD,
                                   tau_prior_dist = "half-normal",
                                   tau_prior = 0.5,
                                   delta = 250)
chains = 4,
iter = 2000,
warmup = 1000)

```



**FIGURE A1** Traceplots for the estimated parameters  $\theta$  and  $\tau$  including burn-in for death outcomes [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE A2** Traceplots for the estimated parameters  $\theta$  and  $\tau$  including burn-in for posttransplant lymphoproliferative disease (PTLD) outcomes [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

Convergence diagnostics and the results can be very conveniently obtained using the **shinystan** package as follows:

```
library("shinystan")
## Firstly convert "stan" object to a "shinystan" object
bnhm.wip.CrinsPTLD.shinystan = as.shinystan(bnhm.wip.CrinsPTLD.stan$fit)
launch_shinystan(bnhm.wip.CrinsPTLD.shinystan)
```

Traceplots for the estimated parameters  $\theta$  and  $\tau$  including burn-in are shown in Figures A1 and A2 for death and PTLD outcomes, respectively.

Lastly, the posterior summary statistics can be obtained using the following command:

```
bnhm.wip.CrinsPTLD.stan$fit_sum
```

## APPENDIX B: R CODE TO IMPLEMENT BNHM USING THE MLE AND THE MH METHODS

Firstly, the BNHM using the MLE:

```
library("lme4")
## Firstly convert dataset to a long format
## using MetaStan::convert_data_arm function
dat.Crins2014.PTLD.long = convert_data_arm(dat.Crins2014.PTLD$exp.total,
                                           dat.Crins2014.PTLD$cont.total,
                                           dat.Crins2014.PTLD$exp.PTLD.events,
                                           dat.Crins2014.PTLD$cont.PTLD.events)

glmer(cbind(r, sampleSize - r) ~ factor(mu) + factor(theta) + (theta12 - 1|mu),
      data = dat.Crins2014.PTLD.long, family = binomial(link = "logit"), nAGQ = 7)
```

Secondly, the MH method:

```
library("metafor")
rma.mh(n1i = exp.total, n2i = cont.total,
      ai = exp.PTLD.events, ci = cont.PTLD.event,
      data = dat.Crins2014.PTLD, measure = "OR")
```

## APPENDIX C: R CODE TO IMPLEMENT THE BBM METHOD

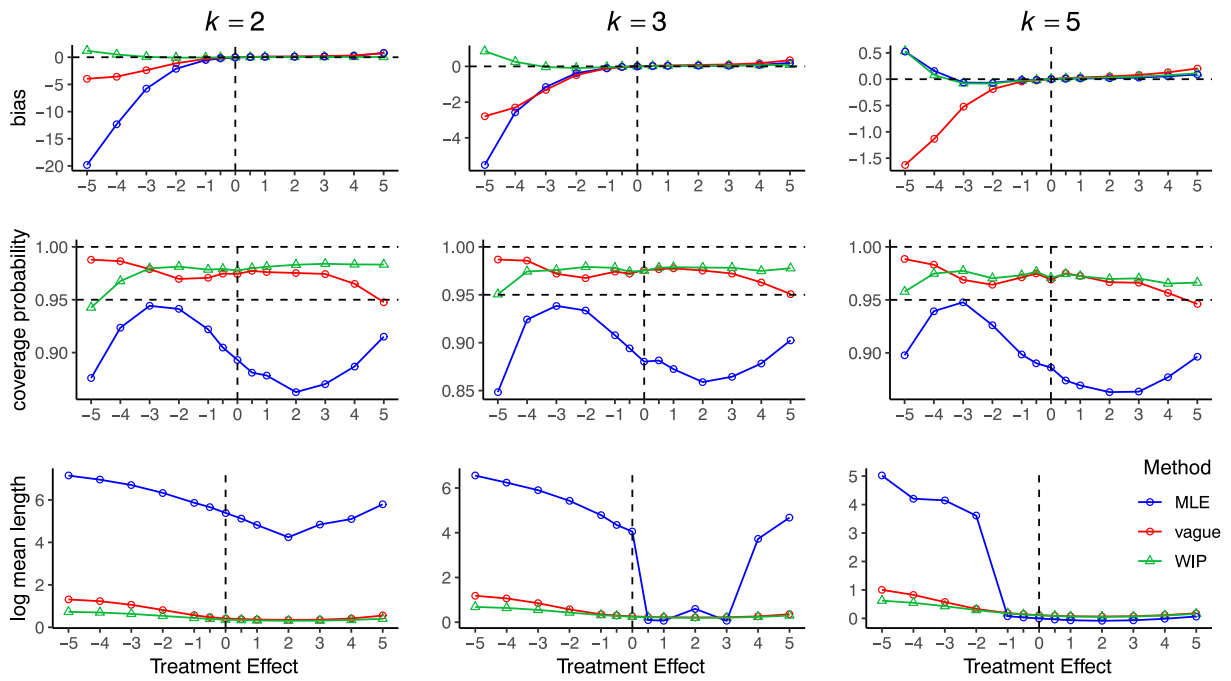
```
bbm.CrinsPTLD.stan = meta_stan(ntrt = exp.total,
                              nctrl = cont.total,
                              rtrt = exp.PTLD.events,
                              rctrl = cont.PTLD.event,
                              data = dat.Crins2014.PTLD,
                              model = "Beta-binomial")
```

## APPENDIX D: ADDITIONAL SIMULATION RESULTS

We also conducted simulations using the same settings as described in Section 5 under BNHM, but using higher

baseline risk probabilities, specifically, baseline risks ( $\mu_i$ ) are uniformly taken between 0.05 and 0.2. Results are illustrated in Figure D1 (analogous to Figure 5).





**FIGURE D1** Simulations with high baseline risks: The bias for the mean treatment effect  $\theta$ , coverage probabilities, and log mean length of the interval estimates for  $\theta$  obtained by three methods (maximum likelihood estimate [MLE], Vague, and weakly informative prior [WIP]) are shown [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]