Quantification of empirical determinacy: the impact of likelihood weighting on posterior location and spread in Bayesian meta-analysis estimated with JAGS and INLA

Sona Hunanyan[1], Håvard Rue[2], Martyn Plummer[3], and Małgorzata Roos[1]

[1]Department of Biostatistics at the Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Hirschengraben 84, CH-8001 Zurich, Switzerland
[2]CEMSE Division, King Abdullah University of Science and Technology, Saudi Arabia
[3]Department of Statistics, University of Warwick, United Kingdom

September 27, 2021

Abstract

The popular Bayesian meta-analysis expressed by Bayesian normal-normal hierarchical model (NNHM) synthesizes knowledge from several studies and is highly relevant in practice. Moreover, NNHM is the simplest Bayesian hierarchical model (BHM), which illustrates problems typical in more complex BHMs. Until now, it has been unclear to what extent the data determines the marginal posterior distributions of the parameters in NNHM. To address this issue we computed the second derivative of the Bhattacharyya coefficient with respect to the weighted likelihood, defined the total empirical determinacy (TED), the proportion of the empirical determinacy of location to TED (pEDL), and the proportion of the empirical determinacy of spread to TED (pEDS). We implemented this method in the R package `ed4bhm` and considered two case studies and one simulation study. We quantified TED, pEDL and pEDS under different modeling conditions such as model parametrization, the primary outcome, and the prior. This clarified to what extent the location and spread of the marginal posterior distributions of the parameters are determined by the data. Although these investigations focused on Bayesian NNHM, the method proposed is applicable more generally to complex BHMs.

Keywords: Empirical determinacy, likelihood weighting, Bayesian meta-analysis, Bayesian hierarchical models, identification

1 Introduction

Bayesian meta-analysis, which synthesizes the evidence from several already published studies, is an indispensable tool for evidence-based medicine. Usually, Bayesian meta-analysis is based on a Bayesian normal-normal hierarchical model (NNHM) [Bender et al., 2018, Friede et al., 2017a,b, Röver, 2020]. This is the simplest Bayesian hierarchical model (BHM), but presents problems typical in more complex BHMs [Vehtari et al., 2021].

If different values of the parameters in a likelihood map to a single data model, then these values are indeterminate, or in other words non-identified, and are not estimable from the likelihood alone [Gustafson, 2015, Lewbel, 2019, Poirier, 1998]. The Bayesian approach is known to mitigate this problem, because a likelihood combined with the proper priors for the parameters that are not likelihood identifiable still yields posterior estimates [Eberly and Carlin, 2000, Gelfand and Sahu, 1999, Gustafson, 2015, Kadane, 1975]. This raises a legitimate question: to what extent are Bayesian posterior parameter estimates determined by the data?

In classical statistical models, non-identified parameters in the likelihood are known to cause problems [Lele, 2010, Lele et al., 2007, 2010, Lewbel, 2019, Skrondal and Rabe-Hesketh, 2004, Sólymos, 2010]. One method that detects non-identified parameters in classical statistical models is data cloning [Lele, 2010, Lele et al., 2007, 2010, Sólymos, 2010]. Data cloning increases the total sample size by artificially replicating the data and focuses on the standard errors of the parameters. If the standard error of a parameter does not decrease at a rate of $1/\sqrt{n}$ with increasing sample size n, such a parameter is deemed to be non-identified [Gustafson, 2015, Lele, 2010, Lele et al., 2007, 2010, Sólymos, 2010].

In contrast, non-identification of parameters is not a problem for complex BHMs, because proper priors lead to proper posteriors [Eberly and Carlin, 2000, Gelfand and Sahu, 1999, Gustafson, 2015]. However, it is unknown to what extent the posterior estimates of such parameters are determined by the data. In complex BHMs, it can happen that the posterior is not determined by the data at all but is completely determined by the prior. Because the detection of non-identified parameters in BHMs is very challenging, Carlin and Louis [1996] and Eberly and Carlin [2000] suggest that one should avoid BHMs for which the identifiability issues have not been clarified. For this reason, a formal method applicable to Bayesian NNHM and BHMs is needed that can quantify the empirical determinacy of the parameters, thus answering the question of to what extent the posterior estimates are determined by the data.

Recently, Roos et al. [2021] considered Bayesian NNHM and proposed a two-dimensional method for sensitivity assessment based on the second derivative of the Bhattacharyya coefficient (BC). In one dimension, they used a factor f to scale the within-study standard deviation provided by the data. This scaling perturbed the total number of individual patient data provided for Bayesian meta-analysis, which is related to data cloning [Lele et al., 2010, Sólymos, 2010]. Alternatively, perturbation of the total number of observations could be expressed by a weighted likelihood with $w = 1/f^2$ [Bissiri et al., 2016, Holmes and Walker, 2017]. An open question remains whether the applicability of the scaling proposed by Roos et al. [2021] could be generalized to weighting other likelihoods, or extended to general-purpose Bayesian estimation techniques such as the Bayesian numerical approximation INLA [Rue et al., 2009], and Bayesian MCMC sampling with JAGS [Plummer, 2016].

We expect that likelihood weighting will affect both the location and the spread of the marginal posterior distributions. Because the method proposed by Roos et al. [2021] quantifies the total impact of the scaling on the posteriors without focusing on location and spread, a refined method is needed. Ideally, such a refined method should quantify what proportion of the total impact of the data on the posterior affects the location and what proportion affects the spread.

To provide a refinement, we consider the second derivative of the BC induced by the weighted likelihood, which quantifies the total empirical determinacy (TED) for each parameter. Based on the properties of both the BC and the rules for the computation of the second derivative, we split TED into two parts: one for location (EDL) and one for the spread (EDS). This enables the quantification of pEDL, the proportion of EDL to TED, and pEDS, the proportion of EDS to TED. This method is implemented in the R package `ed4bhm` and works with JAGS and INLA. The method proposed quantifies how much weighted data impact posterior parameter estimates and whether location or spread is more affected by the weighting.

To demonstrate how the proposed method works in applications, we consider two case studies and one simulation study. We use INLA and JAGS for the estimation. We consider both centered and non-centered parametrizations for Bayesian NNHM and demonstrate how TED, pEDL, and pEDS depend on the amount of pooling induced by the heterogeneity prior. Moreover, we consider both the Bayesian NNHM applied to log(OR) and a logit model. To demonstrate that the proposed method is useful in applications, for clarifying the empirical determinacy of parameters, we consider three variants of the NNHM combined with different prior assumptions in a simulation study.

In Section 2, we review the theoretical background behind our method and its implementation through R package `ed4bhm`. In Section 3, we review the case studies and the models assumed. Then, in Section 4 we present the results which demonstrate the performance of the measures TED, pEDL and pEDS in applications. We conclude with a discussion in Section 5.

2   Methods

In this section, we review the Bayesian NNHM, present the idea of weighting the likelihood, and define the TED measure, which quantifies the total empirical determinacy of posterior parameters. Moreover, we define the proportion of empirical determinacy for posterior location (pEDL) and spread (pEDS). In addition, we specify the implementation of TED, pEDL and pEDS in INLA and JAGS. Finally, a short description of the R package `ed4bhm` is provided.

2.1   Bayesian meta-analysis

The Bayesian normal-normal hierarchical model (NNHM) consists of three levels: the sampling model, the latent random-effects field, and the priors. We consider the Bayesian NNHM with data $y_j, j = 1, \ldots, k$, within-study standard deviations σ_j and between-study standard

deviation τ

$$
\begin{aligned}
y_j \mid \theta_j, \sigma_j &\sim \mathcal{N}(\theta_j, \sigma_j^2), \\
\theta_j \mid \mu, \tau &\overset{\text{iid}}{\sim} \mathrm{N}(\mu, \tau^2), \quad j = 1, \ldots, k, \\
\mu &\sim \pi(\mu), \quad \tau \sim \pi(\tau).
\end{aligned}
\tag{1}
$$

In this model, σ_j are fixed, μ is the overall effect, τ is the heterogeneity, and θ_j are the random effects. For case studies, we assume a weakly informative normal prior $\pi(\mu)$ and either a half-normal (HN) or a half-Cauchy (HC) prior $\pi(\tau)$. We use HN and HC with domain on $[0, \infty)$ that emerge after taking absolute value of normal and Cauchy distributions located at 0 with domain on $(-\infty, \infty)$. We estimate marginal posterior distributions for all the parameters $\psi \in \{\mu, \tau, \theta_1, \ldots, \theta_k\}$ with both INLA [Rue et al., 2009] and JAGS [Plummer, 2016].

2.2 Likelihood weighting

Likelihood weighting is applicable to BHMs with various primary outcomes. A base model is the model without weighting ($w = 1$). Posterior distribution in the base model can be obtained as

$$
\pi_1(\eta \mid y) \propto \pi(y \mid \eta, \theta)\pi(\theta \mid \eta)\pi(\eta),
\tag{2}
$$

where $\pi(y \mid \eta, \theta)$ is the likelihood, $\pi(\theta \mid \eta)$ is the latent field, $\pi(\eta)$ is the assumed prior distribution, and η is the set of all the parameters in the model except the random effects θ.

The posterior distribution in a model with weighted likelihood can be written as

$$
\pi_w(\eta \mid y) \propto (\pi(y \mid \eta, \theta))^w \pi(\theta \mid \eta)\pi(\eta).
\tag{3}
$$

Note that, $w > 1$ gives more weight to the likelihood (corresponds to the case of having more data) and $w < 1$ gives less weight to the likelihood (corresponds to the case of having less data). For example, for the Bayesian NNHM given by (1), the posterior distribution from a weighted model reads as

$$
\pi_w(\mu, \tau, \theta_1, \ldots, \theta_k \mid y) \propto (\pi(y \mid \mu, \theta_1, \ldots, \theta_k))^w \pi(\theta_1, \ldots, \theta_k \mid \tau)\pi(\mu)\pi(\tau),
$$

where $\theta_1, \ldots, \theta_k$ are random effects.

2.3 Affinity measure Bhattacharyya coefficient

Following Roos et al. [2021], we consider the Bhattacharyya coefficient (BC) [Bhattacharyya, 1943] between two probability density functions, which is given by

$$
\mathrm{BC}(\pi_1(\psi \mid y), \pi_w(\psi \mid y)) = \int_{-\infty}^{\infty} \sqrt{\pi_1(\psi \mid y)\pi_w(\psi \mid y)}d\psi,
\tag{4}
$$

where $\psi = (\eta, \theta)$. BC attains values in $[0, 1]$. Moreover, BC is 1 if and only if the two densities completely overlap and is 0 when the domains of the two densities have no overlap. BC has convenient numerical properties and is invariant under any one-to-one transformation

3

(for example, logarithmic) [Jeffreys, 1961, Roos and Held, 2011, Roos et al., 2015, 2021]. Furthermore, BC is directly connected to the Hellinger distance (H), by $H^2 = 1 - BC$.

For two normal densities, BC defined in Equation (4) reads

$$BC(\pi_1^N, \pi_w^N) = \sqrt{\frac{2\sigma_1\sigma_w}{\sigma_1^2 + \sigma_w^2}} \exp\left[-\frac{(\mu_w - \mu_1)^2}{4(\sigma_w^2 + \sigma_1^2)}\right], \tag{5}$$

where π_j^N is the density of a $\mathcal{N}(\mu_j, \sigma_j^2)$ distribution with $j = \{1, w\}$. In general, any density can be approximated to the first order by a normal distribution [Hjort and Glad, 1995, Johnson et al., 1994]. Following Roos et al. [2021], we approximate the marginal posterior distribution $\pi(\psi \mid y)$ for $\psi \in \{\mu, \tau, \theta_1, \ldots, \theta_k\}$ by $\pi^N(\psi \mid y)$ and apply moment matching together with the Equation (5) to obtain

$$BC(\pi_1(\psi \mid y), \pi_w(\psi \mid y)) \approx BC(\pi_1^N, \pi_w^N).$$

For posteriors of precision parameters, we apply this approximation to log-transformed posteriors. This approach is justified by the invariance of BC.

2.4 Quantification of the total empirical determinacy

Following Roos et al. [2021], we define the $TED(\psi)$ measure as the negative second derivative of the Bhattacharyya coefficient (BC) between the base and weighted posterior distributions for each parameter of the model. Here,

$$TED(\psi) = -\left.\frac{d^2 BC(\pi_1(\psi \mid y), \pi_w(\psi \mid y))}{dw^2}\right|_{w=1}, \tag{6}$$

where $\pi_w(\psi \mid y)$ is the marginal posterior distribution from the model with weighted likelihood with $\psi \in \{\mu, \tau, \theta_1, \ldots, \theta_k\}$. By $(H^2)'' = (1 - BC)'' = -BC''$, TED evaluates the second derivative of the squared Hellinger distance H^2. Note that, $w = 1$ denotes the base model with the original likelihood, which induces the base marginal posterior density $\pi_1(\psi \mid y)$. We quantify the total empirical determinacy (TED) of the marginal posterior of the parameter. TED can be used to compare the empirical determinacy between different parameters ψ and ϕ by computing a ratio $TED(\psi)/TED(\phi)$ [Roos et al., 2021].

Note that TED in Equation (6) measures the impact of likelihood weighting on the whole marginal posterior distribution. We refine this approach and consider the impact of likelihood weighting on location (L) and spread (S) of the marginal posterior distribution in the next section.

2.5 Empirical determinacy of location and spread

In this section, we show how to quantify the impact of likelihood weighting on posterior location (L) and spread (S). This gives rise to empirical determinacy of location (EDL) and spread (EDS).

Given the BC between two normal densities in Equation (5), we consider $\mathrm{BC}(w) = \mathrm{BC_S}(w)\mathrm{BC_L}(w)$, where

$$\mathrm{BC_S}(w) = \sqrt{\frac{2\sigma_1\sigma_w}{\sigma_1^2 + \sigma_w^2}}$$

and

$$\mathrm{BC_L}(w) = \exp\left\{-\frac{(\mu_w - \mu_1)^2}{4(\sigma_w^2 + \sigma_1^2)}\right\}.$$

Equations (2)–(5) of the Supplementary Material show that

$$(\mathrm{BC})''\Big|_{w=1} = (\mathrm{BC_L})''\Big|_{w=1} + (\mathrm{BC_S})''\Big|_{w=1},$$

which is equivalent to

$$\mathrm{TED}(\psi) = \mathrm{EDL}(\psi) + \mathrm{EDS}(\psi), \tag{7}$$

where

$$\mathrm{EDL}(\psi) = -\frac{d^2\mathrm{BC_L}(\pi_1(\psi \mid y), \pi_w(\psi \mid y))}{dw^2}\Bigg|_{w=1},$$

and

$$\mathrm{EDS}(\psi) = -\frac{d^2\mathrm{BC_S}(\pi_1(\psi \mid y), \pi_w(\psi \mid y))}{dw^2}\Bigg|_{w=1},$$

for $\psi \in \{\mu, \tau, \theta_1, \ldots, \theta_k\}$. Equation (7) facilitates computation of the proportion of the empirical determinacy of the location (pEDL), which is the proportion of EDL to TED, and the empirical determinacy of the spread (pEDS), which is the proportion of the EDS to TED (Equation (8))

$$\mathrm{pEDL}(\psi) = \frac{\mathrm{EDL}(\psi)}{\mathrm{TED}(\psi)} \quad \text{and} \quad \mathrm{pEDS}(\psi) = \frac{\mathrm{EDS}(\psi)}{\mathrm{TED}(\psi)}, \tag{8}$$

where $\psi \in \{\mu, \tau, \theta_1, \ldots, \theta_k\}$.

2.6 Computational issues

In practice, we approximate the derivatives numerically by the second-order central difference quotient formula. Note that $\mathrm{BC}(\pi_1^{\mathrm{N}}(\psi \mid y), \pi_1^{\mathrm{N}}(\psi \mid y)) = 1$. Thus, for Equation (6), we obtain

$$\begin{aligned}\mathrm{TED}(\psi) &\approx -\frac{d^2\mathrm{BC}(\pi_1^{\mathrm{N}}(\psi \mid y), \pi_w^{\mathrm{N}}(\psi \mid y))}{dw^2} \\ &\approx \frac{\mathrm{BC}(\pi_1^{\mathrm{N}}(\psi \mid y), \pi_{1+\delta}^{\mathrm{N}}(\psi \mid y)) - 2 + \mathrm{BC}(\pi_1^{\mathrm{N}}(\psi \mid y), \pi_{1-\delta}^{\mathrm{N}}(\psi \mid y))}{\delta^2}.\end{aligned}$$

For computations, we consider weights $w = 1 \pm \delta$, with $\delta = 0.01$. For precisions, we conduct all computations on the logarithmic scale.

2.6.1 Implementation of TED in INLA

In order to implement $\text{TED}(\psi)$ in INLA, we use the INLA object from the base model $(w = 1)$. The function `inla.rerun` fits an INLA model with weighted likelihood. We use this function for $w = 1 - \delta$ and $w = 1 + \delta$. Moreover, we extract the summary statistics (mean and standard deviation) of marginal posterior distributions directly from INLA objects. For log-precision, we extract the estimates on internal scale provided in the INLA object.

2.6.2 Implementation of TED in JAGS

For JAGS, we derive a general formula for weighting the likelihood in BHMs fit by MCMC sampling. The posterior distributions for the base and weighted models are given by Equations (2) and (3). From Equation (2) we obtain

$$\pi(\eta) \propto \frac{\pi_1(\eta \mid y)}{\pi(y \mid \eta, \theta)\pi(\theta \mid \eta)}. \tag{9}$$

We then plug in the formula (9) into (3) to get

$$\pi_w(\eta \mid y) \propto \left(\pi(y \mid \eta, \theta)\right)^w \pi(\theta \mid \eta) \frac{\pi_1(\eta \mid y)}{\pi(y \mid \eta, \theta)\pi(\theta \mid \eta)}.$$

Then,

$$\pi_w(\eta \mid y) \propto \left(\pi(y \mid \eta, \theta)\right)^{w-1} \pi_1(\eta \mid y). \tag{10}$$

To estimate the posterior from a weighted model without re-running MCMC sampling we use Equation (10). To evaluate the likelihood values $\pi(y \mid \eta, \theta)$ in JAGS, we activate the `dic` module, which monitors

$$\text{dev}(\psi) = -2\log(\pi(y|\psi)),$$

where $\psi = (\eta, \theta)$. Hence,

$$\pi(y|\psi) = \exp(-\text{dev}(\psi)/2).$$

The MCMC samples $\psi^{(m)}$ from $\pi_1(\psi \mid y)$ and $\text{dev}(\psi)$ values can be extracted from the JAGS object provided by JAGS. To estimate the mean and standard deviation of $\pi_w(\psi \mid y)$, we compute

$$\widehat{\mathsf{E}}(\psi \mid y) = \frac{1}{\sum_{m=1}^{M} c_m} \sum_{m=1}^{M} c_m \psi^{(m)} \tag{11}$$

and

$$\widehat{\mathsf{SD}}(\psi \mid y) = \sqrt{\widehat{\mathrm{Var}}(\psi \mid y)} = \sqrt{\frac{1}{\sum_{m=1}^{M} c_m} \sum_{m=1}^{M} c_m \left(\psi^{(m)} - \widehat{\mathsf{E}}(\psi \mid y)\right)^2}, \tag{12}$$

where $c_m = (\pi(y \mid \psi))^{w-1}$ [Held and Sabanés Bové, 2020]. For log-precisions, we take the logarithmic transformation of the MCMC sample for the precision parameter and apply Equations (11) and (12) on that transformed sample.

2.7 Relative latent model complexity

In the Bayesian NNHM, the interplay between the within-study standard deviation σ_i values provided by the data from k studies and the between-study standard deviation τ can be characterized by the effective number of parameters in the model [Spiegelhalter et al., 2002]. Like Roos et al. [2021], we consider the standardized ratio, the relative latent model complexity (RLMC)

$$\text{RLMC} = \frac{1}{k} \sum_{i=1}^{k} \frac{\tau^2}{\tau^2 + \sigma_i^2}. \tag{13}$$

RLMC defined in Equation (13) attains values between 0 and 1 and expresses the amount of pooling induced by a heterogeneity prior [Gelman and Hill, 2007].

In Section 4.2, a grid of scale parameter values (4.1, 10.4, 18, 31.2, 78.4) for the heterogeneity prior τ induced by the grid of RLMC values fixed at 0.05, 0.25, 0.5, 0.75, 0.95. Whereas HN(4.1), which corresponds to RLMC=0.05, assigns much probability mass to τ values close to 0 and induces much pooling, HN(78.4), which corresponds to RLMC=0.95, assigns less probability mass to values of τ close to 0 and induces little pooling.

2.8 Bayesian computation and convergence diagnostics

The MCMC simulations performed in this paper are based on four parallel chains, with a length of 4×10^5 iterations. In each chain, we removed the first half of iterations as a burn-in period and from the remaining 2×10^5 iterations we kept every 100th iteration in each of the four chains. Our choice of these parameters was guided by `raftery.diag` for Model C1 (Section 4.4) and supported by Vehtari et al. [2021].

To assess the convergence to a stationary distribution, we applied Convergence Diagnostics and Output Analysis implemented in the package `coda` [Plummer et al., 2006]. Moreover, we implemented the rank plots proposed by Vehtari et al. [2021], which are histograms of posterior draws ranked over all chains and plotted separately for each chain. Nearly uniform rank plots for each chain indicate good mixing of chains. In addition, we used the function `n.eff` from the package `stableGR` [Vats and Knudson, 2021], which calculates the effective sample size using the lugsail variance estimators and determines whether Markov chains have converged.

2.9 The R package `ed4bhm`

The open-access R package `ed4bhm` *Empirical determinacy for Bayesian hierarchical models* (`https://github.com/hunansona/ed4bhm`) uses the proposed method to quantify the empirical determinacy of BHMs implemented in INLA and in JAGS. The two main functions in this package are called `ed.inla` and `ed.jags`. These functions were used to generate the results reported in Sections 4.1 – 4.4, and in Sections 3 – 4 and 7 of the Supplementary Material.

3 Data and models

In this section, we review data and models for two case studies. Moreover, we review the design of a simulation study.

3.1 Eight schools

To quantify the effect of a coaching program on SAT-V (Scholastic Aptitude Test-Verbal) scores in eight high schools (Table 3 on page 18 of the Supplementary Material), data from a randomized study was pre-analyzed and used for a Bayesian meta-analysis. The data from these eight schools has been used to study the performance of the Bayesian NNHM and to demonstrate typical issues which arise for BHMs [Gelman and Hill, 2007, Gelman et al., 2014, Vehtari et al., 2021].

We consider two parametrizations of the Bayesian NNHM: the centered and non-centered parametrizations [Vehtari et al., 2021]. The model with centered parametrization is defined as

$$
\begin{aligned}
y_j &\sim \mathcal{N}(\theta_j, \sigma_j^2), \\
\theta_j &\sim \mathcal{N}(\mu, \tau^2), \\
\mu &\sim \mathcal{N}(0, 16), \\
\tau &\sim \mathrm{HN}(5),
\end{aligned}
\tag{14}
$$

where $j = 1, \ldots, 8$. This parametrization is used for both INLA and JAGS. On the other hand, the model with the non-centered parametrization [Gelman et al., 2014, Vehtari et al., 2021] reads

$$
\begin{aligned}
y_j &\sim \mathcal{N}(\theta_j, \sigma_j^2), \\
\theta_j &= \mu + \tau\tilde{\theta}, \\
\tilde{\theta}_j &\sim \mathcal{N}(0, 1), \\
\mu &\sim \mathcal{N}(0, 16), \\
\tau &\sim \mathrm{HN}(5),
\end{aligned}
\tag{15}
$$

for $j = 1, \ldots, 8$, which we implemented in JAGS.

3.2 Respiratory tract infections

In this section, we review the meta-analysis data set including 22 randomized, controlled clinical trials on the prevention of respiratory tract infections (RTI) by selective decontamination of the digestive tract in intensive care unit patients [Bodnar et al., 2017] that are presented in Table 4 on page 19 of the Supplementary Material.

For the RTI data set, we consider two different models. First, a Bayesian NNHM model

$$
\begin{aligned}
y_j &\sim \mathcal{N}(\theta_j, \sigma_j^2), \\
\theta_j &= \mu + \eta_j, \quad j = 1, \ldots, 22, \\
\mu &\sim \mathcal{N}(0, 16), \\
\eta_j &\sim \mathcal{N}(0, \tau^2), \\
\tau &\sim \mathrm{HC}(1),
\end{aligned}
\tag{16}
$$

and second, a binomial model with logistic transformation

$$
\begin{aligned}
z_j &\sim \mathrm{Bin}(p_j, n_j), \\
\mathrm{logit}(p_j) &= \alpha + \beta x_j + \eta_j, \quad j = 1, \ldots, 44, \\
\alpha &\sim \mathcal{N}(0, 16), \\
\beta &\sim \mathcal{N}(0, 16), \\
\eta_j &\sim \mathcal{N}(0, \tau^2), \\
\tau &\sim \mathrm{HC}(1),
\end{aligned}
\tag{17}
$$

where x is a stacked vector attaining value 1 for the treatment and 0 for the control group, $z = (\mathrm{f.t}, \mathrm{f.c})$, $n = (\mathrm{n.t}, \mathrm{n.c})$, where f.t, f.c, n.t and f.c are vectors of length 22 (see in Table 4 of the Supplementary Material). Originally, the weakly informative HC(1) prior assumption for τ was suggested by Bodnar et al. [2017]. Ott et al. [2021] suggested a prior N(0, 16) for μ and used the HC(1) prior for τ. We assume priors for α and β similar to the prior for μ.

3.3 Simulation study

Our simulation study extends the original simulation suggested by Sólymos [2010]. We simulate a sample of random observations of size $n = 50$ under NNHM with parameters $\mu = 2.5, \sigma = 0.2, \tau = 0.5$. See Section 7 of the Supplementary Material for further details.

For the analysis of the simulated data, we use three types of models A, B, and C. Model A is a normal model that does not assume random effects. Models B and C assume a Bayesian NNHM with random effects. Model B assumes that the within-study standard deviation is known and assigns a prior to the between-study standard deviation τ. This defines the usual Bayesian NNHM. In contrast, model C assigns priors to both the within-study (σ) and between-study (τ) standard deviations. Note that the parameters σ and τ in model C are non-identified, because only the sum of the within-study and between-study variances is identified by the likelihood [Bayarri and Berger, 2004, Gelman et al., 2014, Lele, 2010, Sólymos, 2010].

For the parameter μ, all models A, B, and C assume a zero mean normal prior with variance fixed at 10^3. For model B, three different priors are assigned to the between-study standard deviation: $\exp(\mathrm{N}(0, 10^3))$ for B1 [Sólymos, 2010], SqrtIG(0.001, 0.001) for B2, and SqrtIG(4, 1) for B3. For model C3, a SqrtIG(150, 6) prior is assigned to the within-study standard deviation σ. Models C1 and C3 assume identical priors for τ and σ. Table 5 on page 22 of the Supplementary Material reports the properties of all priors assumed for the standard deviations. Whereas $\exp(\mathrm{N}(0, 10^3))$ and SqrtIG(0.001, 0.001) show very large medians, SqrtIG(4, 1) and SqrtIG(150, 6) show medians close to both true parameters $\tau = 0.5$ and $\sigma = 0.2$ chosen for the simulation.

4 Results

4.1 Motivating examples

Sections 3 and 4 of the Supplementary Material consider two motivating examples. Both examples use a normal likelihood with identified parameters. One example considers the posterior of the mean. The other example focuses on the posterior of the precision. Because the parameters in the likelihoods of both motivating examples are identified, the rate of the decrease of the sample size is close to $1/\sqrt{w}$, which matches the result provided by data cloning [Lele, 2010, Lele et al., 2007, 2010, Sólymos, 2010]. In both motivating examples, the estimates of the pEDS are close to 1. This indicates that likelihood weighting affects mostly the spread of the parameters. In fact, both examples motivate that the $1/\sqrt{w}$ rate used by cloning translates into the properties of the pEDS measure.

4.2 Eight schools example

We analyze the data from the eight schools (Table 3 on page 18 of the Supplementary Material) with INLA and JAGS and show the posterior descriptive statistics and estimates of the empirical determinacy for the parameters μ and $\log(\tau^{-2})$ in Table 1 on page 11. For JAGS, both the centered (JAGSc, Equation (14)) and non-centered (JAGSnc, Equation (15)) parametrizations showed high values of ESS. Although the values of the posterior descriptive statistics provided by INLA, JAGSc, and JAGSnc are similar, it is unknown to what extent the posteriors of μ and $\log(\tau^{-2})$ are impacted by the data.

This issue is addressed by the estimates of both the total empirical determinacy (TED) and the within-parameter empirical determinacy (EDL and EDS). For example, the values of TED(μ) are larger than those of TED($\log(\tau^{-2})$). This indicates that the data impacts the posterior of μ more than the posterior of $\log(\tau^{-2})$. pEDL and pEDS provide further details and demonstrate that for both μ and $\log(\tau^{-2})$ the location of the marginal posterior distribution is more impacted by the data than its spread. Whereas the estimates of pEDS(μ) provided by INLA, JAGSc and JAGSnc are close to 0.18, those of pEDS($\log(\tau^{-2})$) differ depending on the centered (JAGSc and INLA) and non-centered (JAGSnc) parametrization. This result indicates that the parametrization used for MCMC sampling may affect the impact of the data on the parameter estimates.

Figure 1 puts the results of Table 1 on page 11 in a broader context and shows a transition phase plot for the data of the eight schools (Table 3 on page 18 of the Supplementary Material) fit by INLA, JAGSc, and JAGSnc across a grid of HN heterogeneity priors with scale parameters equal to 4.1, 10.4, 18, 31.2, 78.4. As explained in Section 2.7, this grid specifies an RLMC grid (0.05, 0.25, 0.5, 0.75, 0.95) of pooling induced by the heterogeneity prior. The top panel of Figure 1 demonstrates that TED is always larger for μ (red) than for $\log(\tau^{-2})$ (blue) across both the grid of RLMC values and the estimation techniques (INLA, JAGSc, and JAGSnc). Independently of the amount of pooling induced by the heterogeneity prior, the posterior of μ is more determined by the data than is the posterior of $\log(\tau^{-2})$.

The proportion of the estimates of the within-parameter empirical determinacy (pEDS) for the scale of μ (middle panel) and $\log(\tau^{-2})$ (bottom panel) of Figure 1 provide more insight. For HN heterogeneity priors with values of the scale parameter equal to 4.1, 10.4, 18, 31.2,

param	method	ESS	mean	sd	q0.025	q0.5	q0.975	TED	EDL	EDS	pEDL	pEDS
μ	INLA		3.58	2.95	-2.24	3.58	9.34	0.11	0.09	0.02	0.82	0.18
	JAGSc	1.5e+04	3.58	2.94	-2.30	3.56	9.36	0.10	0.08	0.02	0.82	0.18
	JAGSnc	1.8e+04	3.59	2.96	-2.18	3.56	9.36	0.11	0.09	0.02	0.83	0.17
$\log(\tau^{-2})$	INLA		-1.60	2.18	-4.49	-2.07	4.00	7e-04	7e-04	3e-05	0.96	0.04
	JAGSc	1.6e+04	-1.60	2.22	-4.48	-2.07	4.06	7e-04	5e-04	2e-04	0.69	0.31
	JAGSnc	1.6e+04	-1.59	2.18	-4.44	-2.07	4.06	6e-04	5e-04	7e-05	0.88	0.12

Table 1: Model eight schools: mean, sd, 95%CrI, median and TED, EDL, EDS, pEDL and pEDS estimates for marginal posterior distributions for the parameters μ and $\log(\tau^{-2})$ calculated in INLA and JAGS (centered and non-centered parametrizations) with ESS of MCMC samples for the data provided in Table 3 on page 18 of the Supplementary Material.
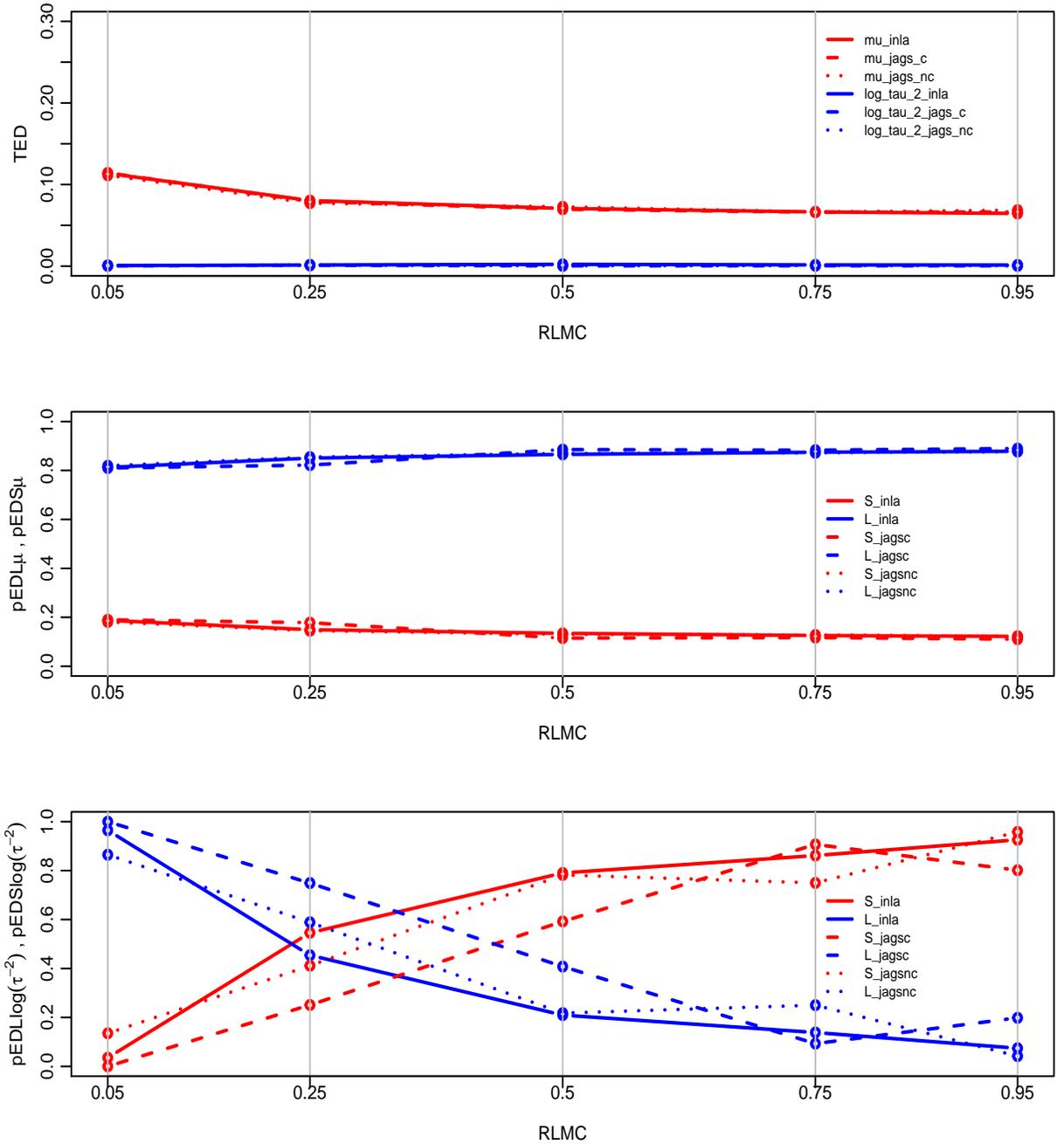
Figure 1: Model eight schools: Transition phase plots for the total TED measure (top) and for the pEDL and pEDS of μ (middle) and of $\log(\tau^{-2})$ (bottom) for the data in Table 3 on page 18 of the Supplementary Material modeled according to Equations (14) and (15) (centered and non-centered parametrizations) and analyzed with INLA and JAGS. The effective median RLMC for eight schools data with HN(5) heterogeneity prior is 0.08. The scale parameters for the HN prior across the grid of 0.05, 0.25, 0.5, 0.75, 0.95 RLMC values are equal to 4.1, 10.4, 18, 31.2, 78.4 (Section 2.7).

78.4 and for three estimation techniques (INLA, JAGSc, JAGSnc), the values of $\text{pEDS}(\mu)$ remain at a low level of at most 20%. This indicates that the data determine the spread of the posterior of μ less than its location. In contrast, the estimates of $\text{pEDS}(\log(\tau^{-2}))$ highly depend on the amount of pooling induced by the heterogeneity prior. The estimates of $\text{pEDS}(\log(\tau^{-2}))$ increase from 10% to 90% across the values of RLMC. This means that heterogeneity priors that induce little pooling lead to posteriors of $\log(\tau^{-2})$ with spreads more determined by the data than for heterogeneity priors that induce much pooling. Although the estimates of $\text{pEDS}(\mu)$ are similar for the three estimation techniques (INLA, JAGSc, JAGSnc), the values of $\text{pEDS}(\log(\tau^{-2}))$ depend on the estimation technique.

4.3   RTI example

Table 2 on page 14 provides posterior descriptive statistics and estimates of the empirical determinacy for the RTI data (Table 4 of the Supplementary Material) fit by INLA and JAGS. We considered two different models: a Bayesian NNHM defined in Equation (16) with a normal primary outcome and a Bayesian logit model defined in Equation (17) with a binomial primary outcome. The MCMC chains provided by JAGS show high values of ESS (Table 2 on page 14). The marginal posteriors provided by INLA and JAGS match well and provide similar descriptive statistics (see Figures 6 and 7 of the Supplementary Material). The descriptive statistics for the parameter $\log(\tau^{-2})$ differ between the NNHM and logit models. For both the NNHM and logit models, the estimates of TED indicate that the posterior of $\log(\tau^{-2})$ is more impacted by the data than are the posteriors of μ, α, and β. NNHM provides lower values of $\text{pEDS}(\mu)$ than the values of $\text{pEDS}(\beta)$ provided by the logit model. This indicates that the structure of the primary outcome and the model can have a great impact on the empirical determinacy of the parameters.

4.4   Simulation: empirical determinacy

Table 3 on page 15 provides estimates of the empirical determinacy for the simulation study described in Section 3.3, which considers three types of NNHM models (A, B and C), which are fit by INLA and JAGS. The results for JAGS are based on MCMC samples with ESS reported in Table 4 on page 16.

For model A, the estimates of $\text{TED}(\mu)$ and $\text{TED}(\log(\gamma^{-2}))$ in Table 3 on page 15 are similar. Moreover, $\text{pEDS}(\mu)$ and $\text{pEDS}(\log(\gamma^{-2}))$ are high. For example, $\text{pEDS}(\log(\gamma^{-2})) \geq 0.86$ demonstrates that the spread of the posterior of $\log(\gamma^{-2})$ is highly impacted by the data.

For models B1, B2 and B3, the values of $\text{TED}(\log(\tau^{-2}))$ are larger than the values of $\text{TED}(\mu)$ in Table 3 on page 15 and indicate that the posterior of $\log(\tau^{-2})$ is more impacted by the data than is the posterior of μ. Similarly to the phase transition in the example of the eight schools discussed in Section 4.2, the values of $\text{TED}(\log(\tau^{-2}))$ depend on the heterogeneity prior. Moreover, the large estimates of $\text{pEDS}(\mu)$ demonstrate that a large proportion of the change in the posterior distribution of μ is due to the change in spread rather than in location. In contrast, the low estimates of $\text{pEDS}(\log(\tau^{-2}))$ indicate that the posterior spread is not much determined by the data.

param	method	model	ESS	mean	sd	q0.025	q0.5	q0.975	TED	EDL	EDS	pEDL	pEDS
μ	INLA			-1.29	0.21	-1.74	-1.28	-0.91	0.39	0.39	7e-04	1.00	2e-03
	JAGS		1.6e+04	-1.29	0.21	-1.74	-1.28	-0.91	0.40	0.40	6e-04	1.00	2e-03
	NNHM												
$\log(\tau^{-2})$	INLA			0.79	0.57	-0.27	0.76	2.00	1.04	0.78	0.27	0.74	0.26
	JAGS		1.6e+04	0.79	0.59	-0.27	0.76	2.04	1.23	0.78	0.45	0.63	0.37
α	INLA			-0.59	0.26	-1.11	-0.60	-0.07	0.02	0.02	0.01	0.75	0.25
	JAGS		1e+04	-0.60	0.26	-1.11	-0.60	-0.08	0.02	0.01	6e-03	0.61	0.39
	logit												
β	INLA			-1.50	0.38	-2.25	-1.49	-0.76	0.03	0.03	3e-03	0.90	0.10
	JAGS		1e+04	-1.49	0.38	-2.26	-1.48	-0.74	0.02	0.02	3e-03	0.89	0.11
	logit												
$\log(\tau^{-2})$	INLA			-0.30	0.27	-0.85	-0.30	0.21	0.33	0.32	6e-03	0.98	0.02
	JAGS		1.6e+04	-0.31	0.27	-0.86	-0.30	0.21	0.31	0.30	0.01	0.96	0.04

Table 2: Models for RTI: NNHM (above) and logit (below). Mean, sd, 95% CrI, median and TED, EDL, EDS, pEDL and pEDS for marginal posterior distributions for the parameters μ and $\log(\tau^{-2})$ for NNHM, and α, β and $\log(\tau^{-2})$ for logit models defined in Equations (16) and (17) calculated in INLA and JAGS for data provided in Table 4 on page 19 of the Supplementary Material with ESS of MCMC samples.

	μ_{TED}	σ^*_{TED}	τ^*_{TED}	γ^*_{TED}	μ_{pEDL}	μ_{pEDS}	σ^*_{pEDL}	σ^*_{pEDS}	τ^*_{pEDL}	τ^*_{pEDS}	γ^*_{pEDL}	γ^*_{pEDS}
A : INLA	0.15			0.16	6e-08	1.00					0.14	0.86
A : JAGS	0.15			0.15	5e-04	1.00					0.09	0.91
B1 : INLA	1e-06		0.21		0.00	1.00			0.89	0.11		
B1 : JAGS	7e-04		0.21		0.35	0.65			0.92	0.08		
B2 : INLA	8e-06		0.24		0.00	1.00			0.90	0.10		
B2 : JAGS	3e-05		0.27		0.05	0.95			0.89	0.11		
B3 : INLA	2e-04		0.13		6e-08	1.00			0.93	0.07		
B3 : JAGS	3e-03		0.12		0.05	0.95			0.95	0.05		
C1 : INLA	0.45	3e+03	2e+03		6e-08	1.00	0.90	0.12	0.88	0.13		
C1 : JAGS	182.48	1e+04	9e+03		0.70	0.25	0.57	0.65	0.2	0.92		
C2 : INLA	0.01	1e+04	1e+04		6e-08	1.00	0.83	0.37	0.94	0.12		
C2 : JAGS	0.06	6e+02	2e+02		0.8	0.2	0.67	0.34	0.87	0.13		
C3 : INLA	3e-04	1.02	0.18		6e-08	1.00	1.00	0.00	0.92	0.08		
C3 : JAGS	4e-03	1.13	0.2		1e-03	1.00	0.99	0.01	0.97	0.03		

Table 3: TED, pEDL, and pEDS values from models A, B.1, B.2, B.3, C.1, C.2 and C.3 for the simulated data described in Section 3.3. σ^*, τ^* and γ^* stand for $\log(\sigma^{-2})$, $\log(\tau^{-2})$ and $\log(\gamma^{-2})$, respectively.

	μ	σ	τ	γ
A	16000.00			16000.00
B1	16000.00		16000.00	
B2	16000.00		16000.00	
B3	16000.00		16000.00	
C1	78.65	2203.84	2826.11	
C2	15429.11	9314.01	9079.20	
C3	16000.00	16000.00	15678.21	

Table 4: ESS for the JAGS samples for the models A, B.1, B.2, B.3, C.1, C.2 and C.3 for the simulation study described in Section 3.3. For the MCMC sampling the total number of iterations used is 4×10^5, the burn-in is 2×10^5, thinning $= 100$.

Models of type C assume priors on the between-study standard deviation and on the within-study standard deviation. The estimation of models C1, C2 and C3 is very challenging. For example, the ESS of MCMC samples for model C1 provided by JAGS (Table 4 on page 16) are very small. Although the parameters σ and τ are non-identified in models C2 and C3, ESS is reasonably high. Table 3 on page 15 shows that TED of the posteriors of $\log(\sigma^{-2})$ and $\log(\tau^{-2})$ is larger than that of the parameter μ. Again, the values of TED($\log(\sigma^{-2})$) and TED($\log(\tau^{-2})$) depend on heterogeneity priors. For C1 and C2, the estimates of pEDS(μ) differ between INLA and JAGS attaining large values for INLA and low values for JAGS. This indicates that numerical Bayesian approximation (INLA) and MCMC sampling (JAGS) can react differently to models with a non-identified likelihood. For model C1, pEDS($\log(\sigma^{-2})$) and pEDS($\log(\tau^{-2})$) values shown by INLA and JAGS disagree. This is due to the small values of the ESS value of the MCMC samples provided by JAGS. In contrast, these estimates of pEDS agree well for models C2 and C3. The difference between the values of pEDS($\log(\sigma^{-2})$) and pEDS($\log(\tau^{-2})$) for models C2 and C3 indicates that the prior assumptions impact the values of pEDS.

5   Discussion

We considered two case studies and one simulation study. For the well-known eight-school example we applied Bayesian NNHM, considering INLA analysis and both centered and non-centered parametrizations for JAGS. Moreover, we provided a transition phase plot for the estimates of TED, pEDL and pEDS across a grid of heterogeneity prior scale parameters, which govern the amount of pooling induced by the heterogeneity prior. This provided insights into how TED, pEDL and pEDS change depending on the properties of the heterogeneity prior. For the RTI data set, we used both the Bayesian NNHM applied to log(OR) and a logit model providing insights into how TED, pEDL and pEDS change depending on the primary outcome. To challenge the method proposed, we relaxed the assumption that the within-study standard deviation is known and assumed priors for both the within-study and between-study standard deviation. This provided an insight into how TED, pEDL and pEDS perform when the underlying model has two non-identified parameters for both INLA and JAGS. The proposed method provided novel insight and proved useful in clarifying the

empirical determinacy of the parameters in the Bayesian NNHM.

The analysis of two simple motivating examples, normal mean and normal precision, translated the results provided by data cloning, which are based on the $1/\sqrt{n}$ criterion, to the unified pEDS measure. They showed that for identified parameters of non-hierarchical likelihoods the spread of the posterior is mainly affected by the likelihood weighting and leads to pEDS estimates close to 1. We prefer the use of pEDS, because the application of the $1/\sqrt{n}$ criterion to BHMs is not well justified [Jiang, 2017, Lewbel, 2019].

This method considerably refines and generalizes the original method proposed by Roos et al. [2021]. We proposed a unified method that quantifies what proportion of the total impact of likelihood weighting on the posterior is due to the change in the location (pEDL) and what proportion is due to the change in the spread (pEDS). This was achieved by matching posteriors moments with those of a normal distribution for the computation of the BC. Note that the normal distribution encapsulates the information about its location and spread in two unrelated parameters. This property proved useful for the definition of pEDL and pEDS. Because likelihood weighting affects both the location and the spread of the marginal posterior distributions, the proposed method is better suited for the quantification of empirical determinacy than other methods that are focused only on the total impact [Kallioinen et al., 2021, Roos et al., 2021].

We successfully applied the proposed method to a non-normal likelihood. This shows that the proposed method can also be applied to other BHMs with different primary outcomes. Moreover, we implemented this method in general-purpose Bayesian estimation programs, such as INLA and JAGS. For JAGS, we developed and applied a technique for the fast and efficient computation of likelihood weighting, which dispenses with re-runs of MCMC chains. All these refinements and generalizations enable future extensions of the method proposed to complex BHMs and to other general-purpose Bayesian estimation programs, such as Stan [Stan Development Team, 2016]. The estimates of the empirical determinacy will help researchers understand to what extent posterior estimates are determined by the data in complex BHMs.

Currently, the weighting approach focuses on weights that are very close to 1. This is very useful to study the impact of the total number of patients, which is induced by within-study sample size changes, on the posteriors of parameters in Bayesian NNHM. However, the data cloning approach, which changes the total number of studies included for the Bayesian meta-analysis, indicates that integer weights that are larger than 1 could also be very useful in assessing the impact of the data on BHMs. More work is necessary to extend the proposed method to cope with integer weights.

The proposed method is based on estimates of the mean and standard deviation for location and spread. These posterior descriptive statistics are provided by default by general-purpose software for Bayesian computation. Moreover, they are used by other modern approaches to quantify the impact of priors on posteriors [Reimherr et al., 2021]. However, for stability reasons, Vehtari et al. [2021] recommend the use of location and spread estimates based on quantiles. In addition, Vehtari et al. [2017] recommend Pareto smoothed importance sampling (PSIS) to regularize importance weights, which flow into the computation of the location and spread of the weighted posteriors. Therefore, anchoring the method proposed on quantiles and the use of PSIS are two additional goals that need to be addressed by our future research.

This paper focuses mainly on Bayesian NNHM. However, the application of the method proposed to other complex BHMs would provide a further insight into the empirical determinacy of posterior parameter estimates in other applications. The openly accessible R package `ed4bhm` (`https://github.com/hunansona/ed4bhm`) conveniently facilitates the application of the proposed method in other settings.

Acknowledgments

References

M.J. Bayarri and J.O. Berger. An interplay of Bayesian and frequentist analysis. *Statistical Science*, 19(1):58–80, 2004.

R. Bender, T. Friede, A. Koch, O. Kuss, P. Schlattmann, G. Schwarzer, and G. Skipka. Methods for evidence synthesis in the case of very few studies. *Research Synthesis Methods*, 9(3):382–392, 2018. URL `https://doi.org/10.1002/jrsm.1297`.

A. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society*, 35:99–109, 1943.

P.G. Bissiri, C.C. Holmes, and S.G. Walker. A general framework for updating belief distributions. *Journal of the Royal Statistical Society. Series B*, 78(5):1103–1130, 2016.

O. Bodnar, A. Link, B. Arendacká, A. Possolo, and C. Elster. Bayesian estimation in random effects meta-analysis using a non-informative prior. *Statistics in Medicine*, 36: 378–399, 2017. doi: 10.1002/sim.7156. URL `https://doi.org/10.1002/sim.7156`.

B.P. Carlin and T.A. Louis. Identifying prior distributions that produce specific decisions, with application to monitoring clinical trials. In D. Berry, K. Chaloner, and J. Geweke, editors, *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner*, pages 493–503. New York: Wiley, 1996.

L.E. Eberly and B.P. Carlin. Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. *Statistics in Medicine*, 19(17-18):2279–2294, 2000. URL `https://doi.org/10.1002/1097-0258(20000915/30)19:17/18<2279:: AID-SIM569>3.0.CO;2-R`.

T. Friede, C. Röver, S. Wandel, and B. Neuenschwander. Meta-analysis of few small studies in orphan diseases. *Research Synthesis Methods*, 8(1):79–91, 2017a. URL `https://doi. org/10.1002/jrsm.1217`.

T. Friede, C. Röver, S. Wandel, and B. Neuenschwander. Meta-analysis of two studies in the presence of heterogeneity with applications in rare diseases. *Biometrical Journal*, 59 (4):658–671, 2017b. URL `https://doi.org/10.1002/bimj.201500236`.

A.E. Gelfand and S.K. Sahu. Identifiability, improper priors, and Gibbs sampling for Generalized Linear Models. *Journal of the American Statistical Association*, 94(445):247–253, 1999. URL `https://www.jstor.org/stable/2669699`.

A. Gelman and J. Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press, Cambridge, 2007.

A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian Data Analysis. 3rd Edition.* Chapman & Hall/CRC Press, Boca Raton, 2014.

P. Gustafson. *Bayesian Inference for Partially Identified Models. Exploring Limits of Limited Data.* Chapman & Hall/CRC Press, Boca Raton, 2015.

L. Held and D. Sabanés Bové. *Likelihood and Bayesian Inference: With Applications in Biology and Medicine.* Springer, 2020. URL `https://www.springer.com/gp/book/9783662607916`.

N.L. Hjort and I.K. Glad. Nonparametric density estimation with a parametric start. *The Annals of Statistics*, 23(3):882–904, 1995. URL `https://www.jstor.org/stable/2242427`.

C.C. Holmes and S.G. Walker. Assigning a value to a power likelihood in a general Bayesian model. *Biometrika*, 104(2):497–503, 2017.

H. Jeffreys. *Theory of Probability. Third Edition.* Oxford University Press, Oxford, 1961.

J. Jiang. *Asymptotic Analysis of Mixed Effects Models: Theory, Applications, and Open Problems.* CRC Press, 2017.

N.L. Johnson, S. Kotz, and N. Balkrishnan. *Continuous Univariate Distributions. Volume 1. Second Edition.* John Wiley & Sons, New York, 1994.

J.B. Kadane. The role of identification in Bayesian theory. In S.E. Feinberg and A. Zellner, editors, *Studies in Bayesian Econometrics and Statistics*, pages 175–191. Noth-Holland Publishing Company, 1975.

N. Kallioinen, T. Paananen, P.-C. Bürkner, and A. Vehtari. Detecting and diagnosing prior and likelihood sensitivity with power-scaling. *arXiv:2107.14054*, 2021. URL `https://arxiv.org/pdf/2107.14054.pdf`.

S.R. Lele. Model complexity and information in the data: Could it be house built on sand? *Ecology*, 91(12):3493–3496, 2010. URL `https://doi.org/10.1890/10-0099.1`.

S.R. Lele, B. Dennis, and F. Lutscher. Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. *Ecology Letters*, 10:551–563, 2007.

S.R. Lele, K. Nadeem, and B. Schmuland. Estimability and likelihood inference for Generalized Linear Mixed Models using data cloning. *Journal of the American Statistical Association*, 105(492):1617–1625, 2010. URL `https://www.jstor.org/stable/27920191`.

A. Lewbel. The identification zoo: meanings of identification in econometrics. *Journal of Economic Literature*, 57(4):835–903, 2019. URL `https://doi.org/10.1257/jel.20181361`.

M. Ott, M. Plummer, and M. Roos. How vague is vague? How informative is informative? Reference analysis for Bayesian meta-analysis. *Statistics in Medicine*, 40(20):4505–4521, 2021. doi: https://doi.org/10.1002/sim.9076. URL `https://onlinelibrary.wiley.com/doi/10.1002/sim.9076`.

M. Plummer. *JAGS: Just Another Gibbs Sampler*, 2016. URL `http://mcmc-jags.sourceforge.net`. version 4.2.0.

M. Plummer, N. Best, K. Cowles, and K. Vines. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, 6(1):7–11, 2006.

D.J. Poirier. Revising beliefs in nonidentified models. *Econometric Theory*, 14:483–509, 1998.

M. Reimherr, X. Meng, and D.L. Nicolae. Prior sample size extensions for assessing prior impact and prior-likelihood discordance. *Journal of the Royal Statistical Society Series B Statistical Methodology*, Early View, 2021. URL `https://doi.org/10.1111/rssb.12414`.

M. Roos and L. Held. Sensitivity analysis in Bayesian generalized linear mixed models for binary data. *Bayesian Analysis*, 6(2):259–278, 2011. URL `https://projecteuclid.org/euclid.ba/1339612046`.

M. Roos, T.G. Martins, L. Held, and H. Rue. Sensitivity analysis for Bayesian hierarchical models. *Bayesian Analysis*, 10(2):321–349, 2015. doi: 10.1214/14-BA909. URL `https://projecteuclid.org/euclid.ba/1422884977`.

M. Roos, S. Hunanyan, H. Bakka, and H. Rue. Sensitivity and identification quantification by a relative latent model complexity perturbation in Bayesian meta-analysis. *Biometrical Journal*, (to appear), 2021. doi: https://doi.org/10.1002/bimj.202000193. URL `https://onlinelibrary.wiley.com/doi/epdf/10.1002/bimj.202000193`.

C. Röver. Bayesian random-effects meta-analysis using the bayesmeta R package. *Journal of Statistical Software*, 93(6):1–51, 2020. doi: 10.18637/jss.v093.i06. URL `https://www.jstatsoft.org/article/view/v093i06`.

H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B.*, 71(2):319–392, 2009. URL `https://doi.org/10.1111/j.1467-9868.2008.00700.x`.

A. Skrondal and S. Rabe-Hesketh. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equations Models.* Chapman & Hall/CRC, 2004.

P. Sólymos. dclone: Data cloning in R. *The R Journal*, 2(2):29–37, 2010.

D.J. Spiegelhalter, N.G. Best, B.P. Carlin, and A. van der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B.*, 64(4):583–616, 2002. ISSN 1369-7412. URL `https://doi.org/10.1111/1467-9868.00353`.

Stan Development Team. *Stan Modeling Language Users Guide and Reference Manual. Version 2.14.0.* 2016. URL `http://mc-stan.org`.

D. Vats and C. Knudson. Revisiting the Gelman-Rubin diagnostic. *Statistical Science*, (to appear), 2021. URL `https://arxiv.org/abs/1812.09384`.

A. Vehtari, A. Gelman, and J. Gabry. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27:1413–1432, 2017.

A. Vehtari, A. Gelman, D. Simpson, B. Carpenter, and P. Bürkner. Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC (with Discussion). *Bayesian Analysis*, 16(2):667–718, 2021. doi: 10.1214/20-BA1221. URL `https://doi.org/10.1214/20-BA1221`.