Bayesian Analysis (2023) **18**, Number 3, pp. 723–751

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†, Håvard Rue‡, Martyn Plummer§, and Małgorzata Roos¶

Abstract. The popular Bayesian meta-analysis expressed by the normal-normal hierarchical model synthesizes knowledge from several studies and is highly relevant in practice. The normal-normal hierarchical model is the simplest Bayesian hierarchical model, but illustrates problems typical in more complex Bayesian hierarchical models. Until now, it has been unclear to what extent the data determines the marginal posterior distributions of the parameters in the normal-normal hierarchical model. To address this issue we computed the second derivative of the Bhattacharyya coefficient with respect to the weighted likelihood. This quantity, which we define as the total empirical determinacy (TED), can be written as the sum of two terms: the empirical determinacy of location (EDL), and the empirical determinacy of spread (EDS). We implemented this method in the R package ed4bhm and considered two case studies and one simulation study. We quantified TED, EDL and EDS under different modeling conditions such as model parametrization, the primary outcome, and the prior. This clarifies 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 normal-normal hierarchical model, the method proposed is applicable more generally to complex Bayesian hierarchical models.

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 normal-normal hierarchical model (NNHM) (Friede et al., 2017a,b;

*Support by the Swiss National Science Foundation (no. 175933) granted to Małgorzata Roos is gratefully acknowledged.

†Department of Biostatistics at the Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Hirschengraben 84, CH-8001 Zurich, Switzerland, hunan.sona@gmail.com

‡CEMSE Division, King Abdullah University of Science and Technology, Saudi Arabia, haavard.rue@kaust.edu.sa

§Department of Statistics, University of Warwick, United Kingdom, martyn.plummer@warwick.ac.uk

¶Department of Biostatistics at the Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Hirschengraben 84, CH-8001 Zurich, Switzerland, malgorzata.roos@uzh.ch

 https://doi.org/10.1214/22-BA1325

Röver, 2020; Bender et al., 2018). This is the simplest Bayesian hierarchical model (BHM), but presents problems typical in more complex BHMs (Gelman and Hill, 2007; Gelman et al., 2014; Talts et al., 2020; 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 (Poirier, 1998; Gustafson, 2015; Lewbel, 2019). 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 (Kadane, 1975; Gelfand and Sahu, 1999; Eberly and Carlin, 2000; Gustafson, 2015). 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 (Skrondal and Rabe-Hesketh, 2004; Lele et al., 2007; Lele, 2010; Lele et al., 2010; Sólymos, 2010; Lewbel, 2019). One method that detects non-identified parameters in classical statistical models is data cloning (Lele et al., 2007; Lele, 2010; Lele et al., 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 (Lele et al., 2007; Lele, 2010; Lele et al., 2010; Sólymos, 2010; Gustafson, 2015).

In contrast, non-identification of parameters is not a problem for complex BHMs, because proper priors lead to proper posteriors (Gelfand and Sahu, 1999; Eberly and Carlin, 2000; Gustafson, 2015). As accurately stated by Bernardo and Smith, "Identifiability is a property of the parametric model, but a Bayesian analysis of a non-identified model is always possible if a proper prior on all the parameters is specified" (Bernardo and Smith, 2000, p. 239). See also Gustafson (2015, Chapter 8) for a very nuanced and in-depth discussion of this issue. However, it is unknown to what extent the posterior estimates of parameters in BHMs 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 the 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 provided by the Integrated Nested Laplace Approximation (INLA) (Rue et al., 2009), and Bayesian Markov chain Monte Carlo (MCMC) sampling with the Just Another Gibbs Sampler (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 present the case studies, the models assumed, and the results which demonstrate the performance of the measures TED, pEDL and pEDS in applications. We conclude with a discussion in Section 4.

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

$$\begin{aligned} y_j \mid \theta_j, \sigma_j &\sim \mathcal{N}(\theta_j, \sigma_j^2), \\ \theta_j \mid \mu, \tau &\overset{\text{iid}}{\sim} \mathrm{N}(\mu, \tau^2), \quad j = 1, \ldots, k, \\ \mu \sim \pi(\mu), &\quad \tau \sim \pi(\tau). \end{aligned} \tag{2.1}$$

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

$$\pi_1(\eta \mid y) \propto \pi(y \mid \eta, \theta)\pi(\theta \mid \eta)\pi(\eta), \tag{2.2}$$

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

$$\pi_w(\eta \mid y) \propto (\pi(y \mid \eta, \theta))^w \, \pi(\theta \mid \eta)\pi(\eta). \tag{2.3}$$

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 (2.1), 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. The details provided below will serve to embed the likelihood weighting in a broader context.

We assume that the observation (or response) variable y_i belongs to an exponential family, which provides a solid basis for a variety of Bayesian hierarchical models based on Latent Gaussian models fit by INLA and JAGS (Rue et al., 2009; Clayton, 1996). Although raising the likelihood to a known nonnegative scalar power w changes the normalizing constant of individual densities, this is not a concern for Bayesian approaches that determine the posterior up to a proportionality constant (Lunn et al., 2013, p. 35–36). Because the interpretation of model parameters remains unchanged, priors remain the same for varying values of w. The weighted likelihood terminology used in this paper is motivated by the role that the weight parameter, the nonnegative scalar w, plays in the exponential family (McCullagh and Nelder, 1989; Clayton, 1996). Alternatively, a likelihood raised to a power can be referred to as "power-scaling of the likelihood" (Kallioinen et al., 2021).

The technique of weighting the likelihood to generalize the usual Bayesian update has been well-established. As shown by Bissiri et al. (2016), it is related to the notion of pseudolikelihood and pseudoposterior (Walker and Hjort, 2001), to the concept of power priors (Ibrahim et al., 1998; Ibrahim and Chen, 2000; Neuenschwander et al., 2009), and to classical robust statistics provided by the general class of M estimators (Huber, 1964, 1981). In addition, the formal framework for general Bayesian inference proposed by Bissiri et al. (2016) also uses weighted likelihood, which expresses model misspecification in the context of Bayesian robustness (Holmes and Walker, 2017). In this context, a misspecified likelihood is equivalent to a likelihood perturbed by weighting.

With a main focus on the base likelihood ($w = 1$), likelihood weighting with $w < 1$ and $w > 1$ creates perturbed proxy models of this base likelihood. Whereas likelihood weighting with $w < 1$ gives less prominence to data, flattens the likelihood and effectively downweights the influence of the data in the posterior, $w > 1$ gives more prominence to data, makes the likelihood more peaked and increases the impact of the data on the posterior. If the spread of the marginal posterior of a parameter is not affected by likelihood weighting, this parameter cannot be well determined by the data (Gustafson, 2015). Therefore, to assess the empirical determinacy of parameters in Bayesian hierarchical models, we focus on the original base likelihood ($w = 1$), deliberately perturb the likelihood by weighting it slightly with w close to 1, and carefully examine the impact of these perturbations on the location and spread of marginal posteriors by computing derivatives evaluated at $w = 1$.

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

$$\mathrm{BC}(\pi_1(\psi \mid y), \pi_w(\psi \mid y)) = \int_{-\infty}^{\infty} \sqrt{\pi_1(\psi \mid y)\pi_w(\psi \mid y)}d\psi, \tag{2.4}$$

where $\psi = (\eta, \theta)$. BC attains values in [0, 1]. Moreover, BC is 1 if and only if the two densities are equal 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 (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 $\mathrm{H}^2 = 1 - \mathrm{BC}$.

For two normal densities, BC defined in (2.4) reads

$$\mathrm{BC}(\pi_1^{\mathrm{N}}(\psi \mid y), \pi_w^{\mathrm{N}}(\psi \mid y)) = \sqrt{\frac{2s_1 s_w}{s_1^2 + s_w^2}} \exp\left[-\frac{(m_w - m_1)^2}{4(s_w^2 + s_1^2)}\right], \tag{2.5}$$

where π_j^{N} is the density of a $\mathcal{N}(m_j, s_j^2)$ distribution with $j = \{1, w\}$. 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 $\pi^{\mathrm{N}}(\psi \mid y)$ and apply moment matching together with the (2.5) to obtain

$$\mathrm{BC}(\pi_1(\psi \mid y), \pi_w(\psi \mid y)) \approx \mathrm{BC}(\pi_1^{\mathrm{N}}(\psi \mid y), \pi_w^{\mathrm{N}}(\psi \mid y)). \tag{2.6}$$

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

The BC in (2.5) is a simple measure of agreement that is based on two familiar summary statistics, mean and standard deviation, that are routinely used to summarize the location and the spread of marginal posterior distributions generated by both MCMC samples and Bayesian numerical approximations. To compute BC in (2.6), we focus on these summary statistics, assume that they are suitable, and match them with moments of a normal distribution. Thus, the approximation in (2.6) translates the general BC measure of affinity between the base ($w = 1$) and the slightly weighted posterior (w close to 1) in (2.4) into a measure of agreement between the means and standard deviations of base and weighted posteriors in (2.5). Because the meaning of the measure in (2.5) is always the same regardless of the representation of marginal posteriors, this measure efficiently unifies MCMC sampling and Bayesian numerical approximations. This unification is particularly important, because we aim to provide empirical determinacy measures which are independent of the technique used to fit Bayesian hierarchical models. However, this unification requires marginal posterior distributions (possibly after an appropriate transformation) that are approximately normal.

The normal distribution is the easiest distribution known to encapsulate the information about location and spread in two unrelated parameters. Therefore, the normal approximation in (2.6) is a convenient technical tool that efficiently disentangles the impact of the weighted likelihood on the location and on the spread of marginal posteriors. To assess the empirical determinacy of parameters, we focus on the standard deviation of posteriors and check whether they are impacted by likelihood weighting (Gustafson, 2015). Although the BC in (2.4) cannot disentangle location and spread in general, the specific BC in (2.5) based on a normal distribution is more effective. As we will show in Section 2.5, a local sensitivity measure based on the normal approximation in (2.6) can be decomposed into one part that measures the impact of data on the location and one part measuring the effect on the spread, thus clarifying what proportion of the total sensitivity of parameters in Bayesian hierarchical models is accounted for by the spread. These proportions can be misleadingly low if the standard deviation of a marginal posterior is a poor estimate of the spread. Nonetheless, reliable proportions can be computed if approximate normality holds and both summary statistics suitably represent the location and spread of marginal posteriors.

There are several justifications for using the normal approximation of a raw and, if necessary, a log-transformed marginal posterior in (2.6). First, if the marginal posterior density has finite mean and standard deviation estimates, it can be approximated to the first order by a normal distribution (Hjort and Glad, 1995). Second, posteriors possess asymptotic expansions in powers of $n^{-1/2}$ that have a normal distribution as a leading term (Johnson, 1967, 1970). Third, normality improves for regular models and in typical settings where the asymptotic normality of the posterior distribution applies (Held and Sabanés Bové, 2020). Finally, Rue et al. (2009) investigated the appropriateness of the normal approximation based on mean and standard deviation in Latent Gaussian models and found that normal approximation provides reasonable results for most real problems and data sets.

2.4 Quantification of the total empirical determinacy

Following Roos et al. (2021), we define the TED(ψ) 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,

$$\mathrm{TED}(\psi) = -\left.\frac{d^2\mathrm{BC}(\pi_1(\psi \mid y), \pi_w(\psi \mid y))}{dw^2}\right|_{w=1} \approx -\left.\frac{d^2\mathrm{BC}(\pi_1^{\mathrm{N}}(\psi \mid y), \pi_w^{\mathrm{N}}(\psi \mid y))}{dw^2}\right|_{w=1}, \tag{2.7}$$

where $\pi_w(\psi \mid y)$ is the marginal posterior distribution from the model with weighted likelihood and $\pi_w^{\mathrm{N}}(\psi \mid y)$ is its normal approximation with $\psi \in \{\mu, \tau, \theta_1, \ldots, \theta_k\}$. By $(\mathrm{H}^2)'' = (1 - \mathrm{BC})'' = -\mathrm{BC}''$, 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 $\mathrm{TED}(\psi)/\mathrm{TED}(\phi)$ (Roos et al., 2021).

For each parameter ψ in a Bayesian hierarchical model, BC in (2.5) measures the agreement between the means and standard deviations of base ($w = 1$) and weighted posteriors. Typically, a Bayesian hierarchical model has many parameters, which operate on different scales. For example, the Bayesian NNHM involves $\psi \in \{\mu, \tau, \theta_1, \ldots, \theta_k\}$, so that the posterior of the overall mean μ operates on a different scale than the posterior of heterogeneity τ. Note that all parameters are simultaneously impacted by the same amount of weighting and we quantify how all parameters are impacted by this perturbation. In this context, BC handles all parameters in a Bayesian hierarchical model on an equal footing regardless of their scale and unifies the quantification of the impact of likelihood weighting on posteriors of all parameters.

The TED sensitivity measure in (2.7) is based on this unifying BC. It fits in well with the general framework of formal sensitivity quantification, which is usually based on differential calculus or its numerical approximations (McCulloch, 1989; Dey and Birmiwal, 1994; Xie and Carlin, 2006; Roos et al., 2015). The negative second derivatives of BC in (2.7) assesses the curvature of the squared Hellinger distance (H^2) and quantifies the acceleration with which the marginal posterior changes locally around the base posterior ($w = 1$) when likelihood perturbations are induced by weighting. Although it is difficult to establish the meaning of the particular sensitivity values provided by differential calculus, these values do provide information about the order of sensitivity estimates (Roos et al., 2021, Section 2.3 of the Supplementary Material) and facilitate a unified comparison of different specifications of Bayesian models (McCulloch, 1989; Roos et al., 2021), so that TED values can be directly compared across parameters.

Note that TED in (2.7) 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 (2.5), we consider $\mathrm{BC}(w) = \mathrm{BC}_{\mathrm{S}}(w)\mathrm{BC}_{\mathrm{L}}(w)$, where

$$\mathrm{BC}_{\mathrm{S}}(w) = \sqrt{\frac{2s_1 s_w}{s_1^2 + s_w^2}} \tag{2.8}$$

and

$$\mathrm{BC}_{\mathrm{L}}(w) = \exp\left\{-\frac{(m_w - m_1)^2}{4(s_w^2 + s_1^2)}\right\}. \tag{2.9}$$

Whereas $\mathrm{BC}_{\mathrm{S}}(w)$ in (2.8) measures the distance between spreads (S), $\mathrm{BC}_{\mathrm{L}}(w)$ in (2.9) is the one-dimensional Mahalanobis distance between locations (L).

Equations (2)–(5) of the Supplementary Material (Hunanyan et al., 2022) show that

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

which is equivalent to

$$\mathrm{TED}(\psi) = \mathrm{EDL}(\psi) + \mathrm{EDS}(\psi), \tag{2.10}$$

where

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

and

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

for $\psi \in \{\mu, \tau, \theta_1, \ldots, \theta_k\}$. For each parameter ψ of a Bayesian hierarchical model, $\mathrm{EDS}(\psi)$ quantifies the sensitivity of the spread of the posterior locally around the base posterior ($w = 1$) induced by likelihood weighting perturbations. Likely, $\mathrm{EDL}(\psi)$ provides the corresponding sensitivity estimate for location.

Equation (2.10) 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

$$\mathrm{pEDL}(\psi) = \frac{\mathrm{EDL}(\psi)}{\mathrm{TED}(\psi)} \quad \text{and} \quad \mathrm{pEDS}(\psi) = \frac{\mathrm{EDS}(\psi)}{\mathrm{TED}(\psi)}, \tag{2.11}$$

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

Inspired by Gustafson (2015), the above derivations focus on likelihood weighting and establish a unified empirical determinacy measure for all parameters of a Bayesian hierarchical model fit either by MCMC sampling or by Bayesian numerical approximations. We consider only small likelihood perturbations from the value $w = 1$ used for standard Bayesian inference. Moreover, we focus on the rates of change (i.e. how does BC change for infinitesimal perturbations of w away from 1), for BC based on the location and spread of marginal posteriors in (2.5). Equation (2.10) shows that the total local sensitivity TED based on BC in (2.5) splits into two sensitivity parts of location (EDL) and spread (EDS). If the spread of the marginal posterior of a parameter is not affected by likelihood weighting, this parameter is non-identified and cannot be well determined by the data (Gustafson, 2015). The pEDS estimate in (2.11) rephrases this condition in terms of the proportion of two local sensitivity estimates (EDS/TED). Whereas large value of pEDS close to 1 indicate that likelihood weighting mostly affects the spread of the marginal posterior and the parameter is well determined by the data, small values of pEDS close to 0 indicate that likelihood weighting barely affects the spread of the marginal posterior and that the parameter is not well determined by the data. Notably, although TED values can only show the order of sensitivity estimates, both pEDL and pEDS estimates are dimensionless and can be easily interpreted.

To the best of our knowledge, the factorization of the BC into two parts for location ($\mathrm{BC_L}$) and spread ($\mathrm{BC_S}$) in (2.5) is unique to the normal distribution. This factorization is the basis of the decomposition of the TED into two additive components EDL and EDS in (2.10). Future research could focus on whether the theory can be generalized to a larger family of distributions. We speculate that it will always be possible to define quantities analogous to EDL and EDS, but they will not be additive, i.e. we will in general not have the decomposition which is shown in (2.10). However, the decomposition will remain valid to the same extent as the normal approximation applies to the posterior distribution, an argument that has worked well in practice. For example, the log-Gamma posterior obtained analytically and for JAGS and INLA (Section 3.1) showed similar empirical determinacy estimates. Given the conceptual utility of this decomposition and the wide applicability of the normal approximation to the posterior, we believe that the current theory is useful despite its apparently narrow theoretical foundation.

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 (2.7), 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.

Implementation of TED in INLA

In order to implement $\mathrm{TED}(\psi)$ in INLA, we use INLA output 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 output. For log-precision, we extract the estimates on internal scale provided by INLA.

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 (2.2) and (2.3). From (2.2) we obtain

$$\pi(\eta) \propto \frac{\pi_1(\eta \mid y)}{\pi(y \mid \eta, \theta)\pi(\theta \mid \eta)}. \tag{2.12}$$

We then plug in the formula (2.12) into (2.3) to get

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

Then,

$$\pi_w(\eta \mid y) \propto (\pi(y \mid \eta, \theta))^{w-1} \, \pi_1(\eta \mid y). \tag{2.13}$$

To estimate the posterior from a weighted model without re-running MCMC sampling we use (2.13). To evaluate the likelihood values $\pi(y \mid \eta, \theta)$ in JAGS, we use the ability of JAGS to monitor the deviance

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

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

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

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

$$\widehat{\mathsf{E}}(\psi \mid y) = \frac{1}{\sum_{m=1}^{M} c_m} \sum_{m=1}^{M} c_m \psi^{(m)} \tag{2.14}$$

and

$$\widehat{\mathrm{SD}}(\psi \mid y) = \sqrt{\widehat{\mathrm{Var}}(\psi \mid y)} = \sqrt{\frac{1}{\sum_{m=1}^{M} c_m} \sum_{m=1}^{M} c_m \left(\psi^{(m)} - \widehat{\mathsf{E}}(\psi \mid y)\right)^2}, \tag{2.15}$$

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 (2.14) and (2.15) on that transformed sample.

Equation (2.13) demonstrates that switching from $\pi_1(\eta|y)$, the posterior based on the unweighted likelihood, to $\pi_w(\eta|y)$, the posterior based on the weighted likelihood, is computationally very efficient, because it only requires reweighting of existing MCMC simulation draws. Importance sampling is a well-known, efficient method to convert one probability measure into another probability measure (Chopin and Papaspiliopoulos, 2020, Section 8.4). Not surprisingly, the mean and the standard deviation estimates of the posterior based on the weighted-likelihood in (2.14) and (2.15) are equivalent to auto-normalized importance sampling estimates (Hastings, 1970; Chopin and Papaspiliopoulos, 2020, Section 8.3), where c_m denotes the importance weights. Moment estimates can be affected by extreme values of importance sampling weights (Tierney, 1994), thus necessitating stabilization of these importance weights (Vehtari et al., 2017). Nonetheless, importance sampling has promising asymptotic properties (Geweke, 1989) and has proved very useful for efficient assessment of Bayesian sensitivity (O'Neill, 2009; Tsai et al., 2011; Kallioinen et al., 2021).

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)

$$\mathrm{RLMC} = \frac{1}{k} \sum_{i=1}^{k} \frac{\tau^2}{\tau^2 + \sigma_i^2}. \tag{2.16}$$

RLMC defined in (2.16) attains values between 0 and 1 and expresses the amount of pooling induced by a heterogeneity prior (Gelman and Hill, 2007).

In Section 3.2, a grid of scale parameter values (4.1, 10.4, 18, 31.2, 78.4) for the half-normal heterogeneity prior for τ is 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 3.4) 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 (ESS) using the lugsail variance estimators and determines whether Markov chains have converged. Alternatively, the effective sample size of Markov chains could be computed based on the adaptive truncation rule of monotonically decreasing autocorrelations proposed by Geyer (1992), as recommended by Vehtari et al. (2021) and substantiated on https://avehtari.github.io/rhat_ess/ess_comparison.html.

2.9 The R package `ed4bhm`

The open source 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 3.1–3.4, and in Sections 3–4 and 7 of the Supplementary Material (Hunanyan et al., 2022).

3 Results

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

3.1 Motivating examples

Sections 3 and 4 of the Supplementary Material (Hunanyan et al., 2022) 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. These examples consider a normal and a log-Gamma posterior and demonstrate similar empirical determinacy estimates obtained analytically and for JAGS and INLA. 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 et al., 2007; Lele, 2010; Lele et al., 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.

3.2 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 (Hunanyan et al., 2022)), 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

$$\begin{aligned} y_j &\sim \mathcal{N}(\theta_j, \sigma_j^2), \\ \theta_j &\sim \mathcal{N}(\mu, \tau^2), \\ \mu &\sim \mathcal{N}(0, 16), \\ \tau &\sim \text{HN}(5), \end{aligned} \tag{3.1}$$

where $j = 1, \dots, 8$. 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

$$\begin{aligned} y_j &\sim \mathcal{N}(\theta_j, \sigma_j^2), \\ \theta_j &= \mu + \tau\tilde{\theta}, \\ \tilde{\theta_j} &\sim \mathcal{N}(0, 1), \\ \mu &\sim \mathcal{N}(0, 16), \\ \tau &\sim \text{HN}(5), \end{aligned} \tag{3.2}$$

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

We analyze the data from the eight schools 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 736. For JAGS, both the centered (JAGSc, (3.1)) and non-centered (JAGSnc, (3.2)) 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

param	method	ESS	mean	sd	q0.025	q0.5	q0.975	TED	EDL	EDS	pEDL	pEDS
	INLA		3.6	3.0	−2.2	3.6	9.3	0.11	0.09	0.02	0.82	0.18
μ	JAGSc	1.5e+04	3.6	2.9	−2.3	3.6	9.4	0.10	0.08	0.02	0.82	0.18
	JAGSnc	1.8e+04	3.6	3.0	−2.2	3.6	9.4	0.11	0.09	0.02	0.83	0.17
	INLA		−1.6	2.2	−4.5	−2.1	4.0	7e-04	7e-04	3e-05	0.96	0.04
$\log(\tau^{-2})$	JAGSc	1.6e+04	−1.6	2.2	−4.5	−2.1	4.1	7e-04	5e-04	2e-04	0.69	0.31
	JAGSnc	1.6e+04	−1.6	2.2	−4.4	−2.1	4.1	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 (Hunanyan et al., 2022).

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 736 in a broader context and shows a transition phase plot for the data of the eight schools 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, 78.4 and for three estimation techniques (INLA, JAGSc, JAGSnc), the values of pEDS(μ) 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 pEDS($\log(\tau^{-2})$) highly depend on the amount of pooling induced by the heterogeneity prior. The estimates of 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 pEDS(μ) are similar for the three estimation techniques (INLA, JAGSc, JAGSnc), the values of pEDS($\log(\tau^{-2})$) depend on the estimation technique.

3.3 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 (Hunanyan et al., 2022).

For the RTI data set, we consider two different models. First, a Bayesian NNHM model for 22 trial-specific log odds ratios

$$\begin{aligned} y_j &\sim \mathcal{N}(\theta_j, \sigma_j^2), \\ \theta_j &= \mu + \eta_j, \quad j = 1, \dots, 22, \\ \mu &\sim \mathcal{N}(0, 16), \\ \eta_j &\sim \mathcal{N}(0, \tau^2), \\ \tau &\sim \mathrm{HC}(1), \end{aligned} \tag{3.3}$$

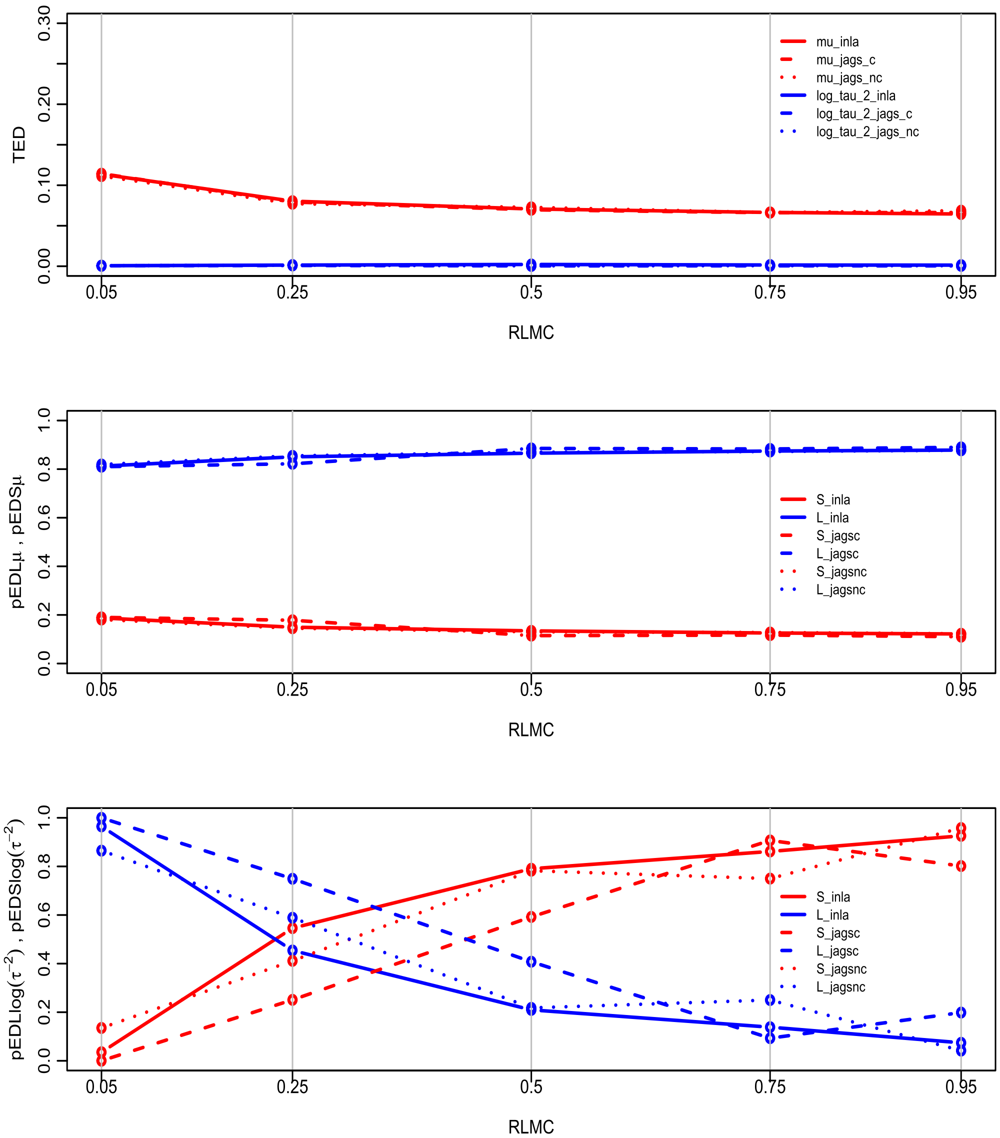

Figure 1: Model eight schools: Transition phase plots for the total empirical determinacy (TED) measure (top) and for the proportion of the empirical determinacy of location and spread to TED (pEDL and pEDS) of μ (middle) and of $\log(\tau^{-2})$ (bottom) for the data in Table 3 on page 18 of the Supplementary Material (Hunanyan et al., 2022) modeled according to (3.1) and (3.2) (centered and non-centered parametrizations) and analyzed with INLA and JAGS. The effective median relative latent model complexity (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).

and second, a binomial model with logistic transformation

$$\begin{aligned} z_j &\sim \mathrm{Bin}(p_j, n_j), \\ \mathrm{logit}(p_j) &= \alpha + \beta x_j + \eta_j, \quad j = 1, \ldots, 44, \\ \alpha &\sim \mathcal{N}(0, 16), \\ \beta &\sim \mathcal{N}(0, 16), \\ \eta_j &\sim \mathcal{N}(0, \tau^2), \\ \tau &\sim \mathrm{HC}(1), \end{aligned} \tag{3.4}$$

where x is a stacked vector attaining value 1 for the treatment and 0 for the control group, z and n contain the corresponding number of cases and the total number of patients in trials in both groups, respectively. 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.

Table 2 on page 740 provides posterior descriptive statistics and estimates of the empirical determinacy for the RTI data fit by INLA and JAGS. We considered two different models: a Bayesian NNHM defined in (3.3) with a normal primary outcome and a Bayesian logit model defined in (3.4) with a binomial primary outcome. The MCMC chains provided by JAGS show high values of ESS (Table 2 on page 740). The marginal posteriors provided by INLA and JAGS match well and provide similar descriptive statistics (see Figures 6 and 7 of the Supplementary Material (Hunanyan et al., 2022)). 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 pEDS(μ) than the values of pEDS(β) 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.

3.4 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 (Hunanyan et al., 2022) for further details. For the analysis of the simulated data, we use three types of models A, B, and C, which are summarized in Table 3 on page 741. 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; Lele, 2010; Sólymos, 2010; Gelman et al., 2014).

param	method	model	ESS	mean	sd	q0.025	q0.5	q0.975	TED	EDL	EDS	pEDL	pEDS
μ	NNHM	INLA		-1.3	0.2	-1.7	-1.3	-0.9	0.39	0.39	7e-04	1.00	2e-03
		JAGS	1.6e+04	-1.3	0.2	-1.7	-1.3	-0.9	0.40	0.40	6e-04	1.00	2e-03
$\log(\tau^{-2})$	NNHM	INLA		0.8	0.6	-0.3	0.8	2.0	1.04	0.78	0.27	0.74	0.26
		JAGS	1.6e+04	0.8	0.6	-0.3	0.8	2.0	1.23	0.78	0.45	0.63	0.37
α	logit	INLA		-0.6	0.3	-1.1	-0.6	-0.1	0.02	0.02	0.01	0.75	0.25
		JAGS	1e+04	-0.6	0.3	-1.1	-0.6	-0.1	0.02	0.01	6e-03	0.61	0.39
β	logit	INLA		-1.5	0.4	-2.3	-1.5	-0.8	0.03	0.03	3e-03	0.90	0.10
		JAGS	1e+04	-1.5	0.4	-2.3	-1.5	-0.7	0.02	0.02	3e-03	0.89	0.11
$\log(\tau^{-2})$	logit	INLA		-0.3	0.3	-0.8	-0.3	0.2	0.33	0.32	6e-03	0.98	0.02
		JAGS	1.6e+04	-0.3	0.3	-0.9	-0.3	0.2	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 (3.3) and (3.4) calculated in INLA and JAGS for data provided in Table 4 on page 19 of the Supplementary Material (Hunanyan et al., 2022) with ESS of MCMC samples.

model	likelihood	random effects	prior overall mean	prior within-study	prior between-study
A	$y_i \sim \mathcal{N}(\mu, \gamma^2)$				
A			$\mu \sim \mathcal{N}(0, 10^3)$	$\log \gamma \sim \mathcal{N}(0, 10^3)$	
B	$y_i \mid \theta_i \sim \mathcal{N}(\theta_i, \sigma^2)$	$\theta_i \sim \mathcal{N}(\mu, \tau^2)$			
B1			$\mu \sim \mathcal{N}(0, 10^3)$		$\log \tau \sim \mathcal{N}(0, 10^3)$
B2			$\mu \sim \mathcal{N}(0, 10^3)$		$1/\tau^2 \sim \mathrm{G}(0.001, 0.001)$
B3			$\mu \sim \mathcal{N}(0, 10^3)$		$1/\tau^2 \sim \mathrm{G}(4, 1)$
C	$y_i \mid \theta_i \sim \mathcal{N}(\theta_i, \sigma^2)$	$\theta_i \sim \mathcal{N}(\mu, \tau^2)$			
C1			$\mu \sim \mathcal{N}(0, 10^3)$	$\log \sigma \sim \mathcal{N}(0, 10^3)$	$\log \tau \sim \mathcal{N}(0, 10^3)$
C2			$\mu \sim \mathcal{N}(0, 10^3)$	$1/\sigma^2 \sim \mathrm{G}(0.001, 0.001)$	$1/\tau^2 \sim \mathrm{G}(0.001, 0.001)$
C3			$\mu \sim \mathcal{N}(0, 10^3)$	$1/\sigma^2 \sim \mathrm{G}(150, 6)$	$1/\tau^2 \sim \mathrm{G}(4, 1)$

Table 3: Summary of models used for the simulation study described in Section 3.4. Section 7 of the Supplementary Material (Hunanyan et al., 2022) provides more details.

For the parameter μ, all models A, B, and C assume a zero mean normal prior with variance fixed at 10^3 (Table 3 on page 741). 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 (Hunanyan et al., 2022) 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.

Table 4 on page 743 provides estimates of the empirical determinacy for the simulation study described above, 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 5 on page 744.

For model A, the estimates of TED(μ) and TED($\log(\gamma^{-2})$) in Table 4 on page 743 are similar. Moreover, pEDS(μ) and pEDS($\log(\gamma^{-2})$) are high. For example, pEDS($\log(\gamma^{-2})$) ≥ 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 TED($\log(\tau^{-2})$) are larger than the values of TED(μ) in Table 4 on page 743 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 3.2, the values of TED($\log(\tau^{-2})$) depend on the heterogeneity prior. Moreover, the large estimates of pEDS(μ) 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 pEDS($\log(\tau^{-2})$) indicate that the posterior spread is not much determined by the data.

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 5 on page 744) is very small. Although the parameters σ and τ are non-identified in models C2 and C3, ESS is reasonably high. Table 4 on page 743 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 ESS values 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.

	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	0.00	1.00					0.14	0.86
A : JAGS	0.15			0.15	0.00	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		0.00	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		0.00	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	0.20	0.92		
C2 : INLA	0.01	1e+04	1e+04		0.00	1.00	0.83	0.37	0.94	0.12		
C2 : JAGS	0.06	6e+02	2e+02		0.80	0.20	0.67	0.34	0.87	0.13		
C3 : INLA	3e-04	1.02	0.18		0.00	1.00	1.00	0.00	0.92	0.08		
C3 : JAGS	4e-03	1.13	0.20		0.00	1.00	0.99	0.01	0.97	0.03		

Table 4: 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 3.4. σ^*, τ^* and γ^* stand for $\log(\sigma^{-2})$, $\log(\tau^{-2})$ and $\log(\gamma^{-2})$, respectively.

	μ	σ	τ	γ
A	16000			16000
B1	16000		16000	
B2	16000		16000	
B3	16000		16000	
C1	79	2204	2826	
C2	15429	9314	9079	
C3	16000	16000	15678	

Table 5: 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 3.4. For the MCMC sampling the total number of iterations used is 4×10^5, the burn-in is 2×10^5, thinning = 100.

4 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 posterior 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 (Roos et al., 2021; Kallioinen 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 general Bayesian inference framework imposes no restrictions on the likelihood function (Berger and Srinivasan, 1978; Walker and Hjort, 2001; Chernozhukov and Hong, 2003; Bissiri et al., 2016), thus indicating that the approach to likelihood weighting could be extended beyond the exponential family assumption. More work is needed to systematically investigate such extensions.

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 open source R package `ed4bhm` (https://github.com/hunansona/ed4bhm) conveniently facilitates the application of the proposed method in other settings.

Supplementary Material

Supplementary Material (DOI: 10.1214/22-BA1325SUPP; .pdf).

References

Bayarri, M. and Berger, J. (2004). "An interplay of Bayesian and frequentist analysis." *Statistical Science*, 19(1): 58–80. MR2082147. doi: https://doi.org/10.1214/088342304000000116. 739

Bender, R., Friede, T., Koch, A., Kuss, O., Schlattmann, P., Schwarzer, G., and Skipka, G. (2018). "Methods for evidence synthesis in the case of very few studies." *Research Synthesis Methods*, 9(3): 382–392. doi: https://doi.org/10.1002/jrsm.1297. 723

Berger, J. and Srinivasan, C. (1978). "Generalized Bayes estimators in multivariate problems." *The Annals of Statistics*, 6(4): 783–801. URL https://www.jstor.org/stable/2958855. MR0478426. 745

Bernardo, J. and Smith, A. (2000). *Bayesian Theory*. John Wiley & Sons, Ltd. MR1274699. doi: https://doi.org/10.1002/9780470316870. 724

Bhattacharyya, A. (1943). "On a measure of divergence between two statistical populations defined by their probability distributions." *Bulletin of the Calcutta Mathematical Society*, 35: 99–109. MR0010358. 727

Bissiri, P., Holmes, C., and Walker, S. (2016). "A general framework for updating belief distributions." *Journal of the Royal Statistical Society. Series B*, 78(5): 1103–1130. MR3557191. doi: https://doi.org/10.1111/rssb.12158. 724, 727, 745

Bodnar, O., Link, A., Arendacká, B., Possolo, A., and Elster, C. (2017). "Bayesian estimation in random effects meta-analysis using a non-informative prior." *Statistics in Medicine*, 36: 378–399. MR3582981. doi: https://doi.org/10.1002/sim.7156. 737, 739

Carlin, B. and Louis, T. (1996). "Identifying prior distributions that produce specific decisions, with application to monitoring clinical trials." In Berry, D., Chaloner, K., and Geweke, J. (eds.), *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner*, 493–503. New York: Wiley. MR1367306. 724

Chernozhukov, V. and Hong, H. (2003). "An MCMC approach to classical estimation." *Journal of Econometrics*, 115(2): 293–346. MR1984779. doi: https://doi.org/10.1016/S0304-4076(03)00100-3. 745

Chopin, N. and Papaspiliopoulos, O. (2020). *An Introduction to Sequential Monte Carlo*. Springer. MR4215639. doi: https://doi.org/10.1007/978-3-030-47845-2. 733

Clayton, D. G. (1996). "Generalized linear mixed models." In Gilks, W., Richardson, S., and Spiegelhalter, D. (eds.), *Markov Chain Monte Carlo in Practice*, 275–301. Chapman & Hall. MR1397972. 726, 727

Dey, D. and Birmiwal, L. (1994). "Robust Bayesian analysis using divergence measures."

Statistics & Probability Letters, 20(4): 287–294. MR1294607. doi: https://doi.org/10.1016/0167-7152(94)90016-7. 729

Eberly, L. and Carlin, B. (2000). "Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models." *Statistics in Medicine*, 19(17-18): 2279–2294. doi: https://doi.org/10.1002/1097-0258(20000915/30)19:17/18<2279::AID-SIM569>3.0.CO;2-R. 724

Friede, T., Röver, C., Wandel, S., and Neuenschwander, B. (2017a). "Meta-analysis of few small studies in orphan diseases." *Research Synthesis Methods*, 8(1): 79–91. doi: https://doi.org/10.1002/jrsm.1217. 723

Friede, T., Röver, C., Wandel, S., and Neuenschwander, B. (2017b). "Meta-analysis of two studies in the presence of heterogeneity with applications in rare diseases." *Biometrical Journal*, 59(4): 658–671. doi: https://doi.org/10.1002/bimj.201500236. 723

Gelfand, A. and Sahu, S. (1999). "Identifiability, improper priors, and Gibbs sampling for Generalized Linear Models." *Journal of the American Statistical Association*, 94(445): 247–253. MR1689229. doi: https://doi.org/10.2307/2669699. 724

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2014). *Bayesian Data Analysis*. 3rd Edition. Chapman & Hall/CRC Press, Boca Raton. MR3235677. 724, 735, 739

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge. 724, 733, 735

Geweke, J. (1989). "Bayesian inference in econometric models using Monte Carlo integration." *Econometrica*, 57(6): 1317–1339. MR1035115. doi: https://doi.org/10.2307/1913710. 733

Geyer, C. (1992). "Practical Markov Chain Monte Carlo." *Statistical Science*, 7(4): 473–483. URL https://www.jstor.org/stable/2246094. 734

Gustafson, P. (2015). *Bayesian Inference for Partially Identified Models. Exploring Limits of Limited Data*. Chapman & Hall/CRC Press, Boca Raton. MR3642458. 724, 727, 728, 731

Hastings, W. (1970). "Monte Carlo sampling methods using Markov chains and their applications." *Biometrika*, 57(1): 97–109. URL https://www.jstor.org/stable/2334940. MR3363437. doi: https://doi.org/10.1093/biomet/57.1.97. 733

Held, L. and Sabanés Bové, D. (2020). *Likelihood and Bayesian Inference: With Applications in Biology and Medicine*. Springer. MR4350635. doi: https://doi.org/10.1007/978-3-662-60792-3. 729, 733

Hjort, N. and Glad, I. (1995). "Nonparametric density estimation with a parametric start." *The Annals of Statistics*, 23(3): 882–904. URL https://www.jstor.org/stable/2242427. MR1345205. doi: https://doi.org/10.1214/aos/1176324627. 729

Holmes, C. and Walker, S. (2017). "Assigning a value to a power likelihood in a general Bayesian model." *Biometrika*, 104(2): 497–503. MR3698270. doi: https://doi.org/10.1093/biomet/asx010. 724, 727

Huber, P. (1964). "Robust estimation of a location parameter." *Annals of Mathematical Statistics*, 35(1): 73–101. MR0161415. doi: https://doi.org/10.1214/aoms/1177703732. 727

Huber, P. (1981). *Robust Statistics*. John Wiley & Sons, Inc., New York. 727

Hunanyan, S., Rue, H., Plummer, M., and Roos, M. (2022). "Supplementary Material: Quantification of empirical determinacy: the impact of likelihood weighting on posterior location and spread in Bayesian meta-analysis estimated with JAGS and INLA." *Bayesian Analysis*. doi: https://doi.org/10.1214/22-BA1325SUPP. 730, 734, 735, 736, 737, 738, 739, 740, 741, 742

Ibrahim, J. and Chen, M. (2000). "Power prior distributions for regression models." *Statistical Science*, 15(1): 46–60. URL https://www.jstor.org/stable/2676676. MR1842236. doi: https://doi.org/10.1214/ss/1009212673. 727

Ibrahim, J., Ryan, L., and Chen, M. (1998). "Using historical controls to adjust for covariates in trend tests for binary data." *Journal of the American Statistical Association*, 93(444): 1282–1293. URL https://www.jstor.org/stable/2670043. 727

Jeffreys, H. (1961). *Theory of Probability*. Third Edition. Oxford University Press, Oxford. MR0187257. 727

Jiang, J. (2017). *Asymptotic Analysis of Mixed Effects Models: Theory, Applications, and Open Problems*. CRC Press. MR3676375. 744

Johnson, R. (1967). "An asymptotic expansion for posterior distributions." *The Annals of Mathematical Statistics*, 38(6): 1899–1906. MR0219161. doi: https://doi.org/10.1214/aoms/1177698624. 729

Johnson, R. (1970). "Asymptotic expansions associated with posterior distributions." *The Annals of Mathematical Statistics*, 41(3): 851–864. URL https://www.jstor.org/stable/2239239. MR0263198. doi: https://doi.org/10.1214/aoms/1177696963. 729

Kadane, J. (1975). "The role of identification in Bayesian theory." In Feinberg, S. and Zellner, A. (eds.), *Studies in Bayesian Econometrics and Statistics*, 175–191. Noth-Holland Publishing Company. MR0483124. 724

Kallioinen, N., Paananen, T., Bürkner, P., and Vehtari, A. (2021). "Detecting and diagnosing prior and likelihood sensitivity with power-scaling." arXiv:2107.14054. 727, 733, 745

Lele, S. (2010). "Model complexity and information in the data: Could it be house built on sand?" *Ecology*, 91(12): 3493–3496. doi: https://doi.org/10.1890/10-0099.1. 724, 734, 739

Lele, S., Dennis, B., and Lutscher, F. (2007). "Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods." *Ecology Letters*, 10: 551–563. 724, 734

Lele, S., Nadeem, K., and Schmuland, B. (2010). "Estimability and likelihood inference for Generalized Linear Mixed Models using data cloning." *Journal of the American Statistical Association*, 105(492): 1617–1625. URL https://www.jstor.org/stable/27920191. MR2796576. doi: https://doi.org/10.1198/jasa.2010.tm09757. 724, 734

Lewbel, A. (2019). "The identification zoo: meanings of identification in econometrics." *Journal of Economic Literature*, 57(4): 835–903. doi: https://doi.org/10.1257/jel.20181361. 724, 744

Lunn, D., Jackson, C., Best, N., Thomas, A., and Spiegelhalter, D. (2013). *The BUGS Book: A Practical Introduction to Bayesian Analysis*. Chapman & Hall/CRC. 727

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Second Edition. Chapman & Hall/CRC. MR3223057. doi: https://doi.org/10.1007/978-1-4899-3242-6. 727

McCulloch, R. (1989). "Local model influence." *Journal of the American Statistical Association*, 84(406): 473–478. URL https://www.jstor.org/stable/2289932. 729, 730

Neuenschwander, B., Branson, M., and Spiegelhalter, D. (2009). "A note on the power prior." *Statistics in Medicine*, 28(28): 3562–3566. MR2744383. doi: https://doi.org/10.1002/sim.3722. 727

O'Neill, B. (2009). "Importance sampling for Bayesian sensitivity analysis." *International Journal of Approximate Reasoning*, 50(2): 270–278. MR2514500. doi: https://doi.org/10.1016/j.ijar.2008.03.015. 733

Ott, M., Plummer, M., and Roos, M. (2021). "How vague is vague? How informative is informative? Reference analysis for Bayesian meta-analysis." *Statistics in Medicine*, 40(20): 4505–4521. MR4300127. doi: https://doi.org/10.1002/sim.9076. 739

Plummer, M. (2016). *JAGS: Just Another Gibbs Sampler*. Version 4.2.0. URL http://mcmc-jags.sourceforge.net. 725, 726

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). "CODA: Convergence Diagnosis and Output Analysis for MCMC." *R News*, 6(1): 7–11. 734

Poirier, D. (1998). "Revising beliefs in nonidentified models." *Econometric Theory*, 14: 483–509. MR1650041. doi: https://doi.org/10.1017/S0266466698144043. 724

Reimherr, M., Meng, X., and Nicolae, D. (2021). "Prior sample size extensions for assessing prior impact and prior-likelihood discordance." *Journal of the Royal Statistical Society Series B Statistical Methodology*, Early View. MR4294538. doi: https://doi.org/10.1111/rssb.12414. 745

Roos, M. and Held, L. (2011). "Sensitivity analysis in Bayesian generalized linear mixed models for binary data." *Bayesian Analysis*, 6(2): 259–278. MR2806244. doi: https://doi.org/10.1214/11-BA609. 727

Roos, M., Hunanyan, S., Bakka, H., and Rue, H. (2021). "Sensitivity and identification quantification by a relative latent model complexity perturbation in Bayesian meta-analysis." *Biometrical Journal*, 63(8): 1555–1574. MR4362586. doi: https://doi.org/10.1002/bimj.202000193. 724, 725, 727, 728, 729, 730, 733, 744, 745

Roos, M., Martins, T., Held, L., and Rue, H. (2015). "Sensitivity analysis for Bayesian hierarchical models." *Bayesian Analysis*, 10(2): 321–349. MR3420885. doi: https://doi.org/10.1214/14-BA909. 727, 729

Röver, C. (2020). "Bayesian random-effects meta-analysis using the bayesmeta R package." *Journal of Statistical Software*, 93(6): 1–51. URL https://www.jstatsoft.org/article/view/v093i06. 723

Rue, H., Martino, S., and Chopin, N. (2009). "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. MR2649602. doi: https://doi.org/10.1111/j.1467-9868.2008.00700.x. 725, 726, 729

Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equations Models*. Chapman & Hall/CRC. MR2059021. doi: https://doi.org/10.1201/9780203489437. 724

Sólymos, P. (2010). "dclone: Data cloning in R." *The R Journal*, 2(2): 29–37. 724, 734, 739, 742

Spiegelhalter, D., Best, N., Carlin, B., and van der Linde, A. (2002). "Bayesian measures of model complexity and fit." *Journal of the Royal Statistical Society, Series B.*, 64(4): 583–616. MR1979380. doi: https://doi.org/10.1111/1467-9868.00353. 733

Stan Development Team (2016). *Stan Modeling Language Users Guide and Reference Manual. Version 2.14.0*. URL http://mc-stan.org. 745

Talts, S., Betancourt, M., Simpson, D., Vehtari, A., and Gelman, A. (2020). "Validating Bayesian inference algorithms with simulation-based calibration." arXiv:1804.06788v2. 724

Tierney, L. (1994). "Markov chains for exploring posterior distributions." *The Annals of Statistics*, 22(4): 1701–1728. URL https://www.jstor.org/stable/2242477. MR1329166. doi: https://doi.org/10.1214/aos/1176325750. 733

Tsai, Y., Murdoch, D., and Dupuis, D. (2011). "Influence measures and robust estimators of dependence in multivariate extremes." *Extremes*, 14(4): 343–363. MR2853106. doi: https://doi.org/10.1007/s10687-010-0114-6. 733

Vats, D. and Knudson, C. (2021). "Revisiting the Gelman-Rubin diagnostic." *Statistical Science*, 36(4): 518–529. arxiv:1812.09384. MR4323050. doi: https://doi.org/10.1214/20-sts812. 734

Vehtari, A., Gelman, A., and Gabry, J. (2017). "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC." *Statistics and Computing*, 27: 1413–1432. MR3647105. doi: https://doi.org/10.1007/s11222-016-9696-4. 733, 745

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P. (2021). "Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC (with Discussion)." *Bayesian Analysis*, 16(2): 667–718. MR4298989. doi: https://doi.org/10.1214/20-ba1221. 724, 734, 735, 745

Walker, S. and Hjort, N. (2001). "On Bayesian consistency." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 63(4): 811–821. MR1872068. doi: https://doi.org/10.1111/1467-9868.00314. 727, 745

Xie, Y. and Carlin, B. (2006). "Measures of Bayesian learning and identifiability in hierarchical models." *Journal of Statistical Planning and Inference*, 136(10): 3458–3477. MR2256284. doi: https://doi.org/10.1016/j.jspi.2005.04.003. 729

Acknowledgments

We thank both reviewers and the associate editor for their thoughtful comments and constructive suggestions.