Received: 22 June 2020 | Revised: 13 May 2021 | Accepted: 22 May 2021

DOI: 10.1002/bimj.202000193

Biometrical Journal

RESEARCH ARTICLES

Sensitivity and identification quantification by a relative latent model complexity perturbation in Bayesian meta-analysis

Małgorzata Roos[1] | **Sona Hunanyan**[1] | **Haakon Bakka**[2] | **Håvard Rue**[3]

[1] Department of Biostatistics, EBPI, University of Zurich, Zurich, Switzerland

[2] Department of Mathematics, University of Oslo, Oslo, Norway

[3] CEMSE Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

Correspondence
Małgorzata Roos, Department of Biostatistics, EBPI, University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland.
Email: malgorzata.roos@uzh.ch

Funding information
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: 175933

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

Abstract

In recent years, Bayesian meta-analysis expressed by a normal–normal hierarchical model (NNHM) has been widely used for combining evidence from multiple studies. Data provided for the NNHM are frequently based on a small number of studies and on uncertain within-study standard deviation values. Despite the widespread use of Bayesian NNHM, it has always been unclear to what extent the posterior inference is impacted by the heterogeneity prior (sensitivity *S*) and by the uncertainty in the within-study standard deviation values (identification *I*). Thus, to answer this question, we developed a unified method to simultaneously quantify both sensitivity and identification (*S-I*) for all model parameters in a Bayesian NNHM, based on derivatives of the Bhattacharyya coefficient with respect to relative latent model complexity (RLMC) perturbations. Three case studies exemplify the applicability of the method proposed: historical data for a conventional therapy, data from which one large study is first included and then excluded, and two subgroup meta-analyses specified by their randomization status. We analyzed six scenarios, crossing three RLMC targets with two heterogeneity priors (half-normal, half-Cauchy). The results show that *S-I* explicitly reveals which parameters are affected by the heterogeneity prior and by the uncertainty in the within-study standard deviation values. In addition, we compare the impact of both heterogeneity priors and quantify how *S-I* values are affected by omitting one large study and by the randomization status. Finally, the range of applicability of *S-I* is extended to Bayesian NtHM. A dedicated R package facilitates automatic *S-I* quantification in applied Bayesian meta-analyses.

KEYWORDS

Bayesian meta-analysis, formal sensitivity and identification diagnostics, normal–normal hierarchical model, normal-t hierarchical model, relative latent model complexity

1 | INTRODUCTION

To establish reliable evidence regarding the benefits of various healthcare interventions, we need statistical methods that can cope with varying results from different studies and allow for a combined evaluation of evidence. Under these circumstances, meta-analysis is the statistical method of choice in evidence-based medicine, because it shifts the focus away from the evidence in a single study toward a combined view of evidence provided by several different studies. Such analysis takes into account the between study heterogeneity that, when ignored, can result in overoptimistically small standard errors of estimates, meaning that the significance of an estimated effect may actually be invalid. Frequently, however, only a very small number of studies are available for meta-analysis, and thus, in these cases, the results can also be misleading due to unobserved heterogeneity. The Bayesian methodology mitigates this problem by incorporating priors on heterogeneity standard deviation into a normal–normal hierarchical model (NNHM).

However, the heterogeneity in a meta-analysis based on k studies can only be well understood when both the between-study heterogeneity τ and the within-study standard deviation σ_i values are simultaneously considered (Hardy & Thompson, 1998). The mutual interplay between the between-study heterogeneity τ and the within-study standard deviation σ_i values is also active in the Bayesian setting and impacts the posterior estimates (Gustafson et al., 2006). However, this interplay is prone to two sources of uncertainty. While it is well known that the choice of the prior on the heterogeneity parameter τ is uncertain, it seems to be less well known that the within-study standard deviation σ_i values are actually uncertain estimates. Both sources of uncertainty act simultaneously and can affect posteriors. Therefore, we need a method that can simultaneously quantify the impact of both sources of uncertainty.

Although only two studies are necessary for a Bayesian NNHM to be applicable, in cases when only a small number of studies are supplied for a meta-analysis, the impact of the heterogeneity prior on the posterior can become a concern. In many cases, various selection criteria and exclusion of old and outlying studies can generate subgroups of the original data set, thus further reducing the number of studies included (Bjordal et al., 2004; Crins et al., 2014; Smith et al., 1995). Indeed, over 30% of meta-analyses extracted from the Cochrane Database of Systematic Reviews (CDSR) are based on only two studies and nearly 75% are based on five or fewer studies (Davey et al., 2011; IntHout et al., 2015; Turner et al., 2012). When only such a small number of studies are included in a Bayesian NNHM, the sensitivity of the posterior to the heterogeneity prior can become an issue.

Bayesian meta-analysis expressed by an NNHM is also used to analyze studies when the data are too limited to provide a precise estimate of the within-study standard deviation σ_i. There are studies that contain only two participants and at least 25% of the studies of the CDSR contain less than 50 participants (Davey et al., 2011; Turner et al., 2012). When studies are not sufficiently large, then the estimation of the within-study standard deviation is unstable (Hoaglin, 2015). Therefore, when such studies with only a small number of participants are included in a Bayesian NNHM, the sensitivity of the posterior to the uncertainty in the within-study standard deviation values is a concern.

Practical applications of the Bayesian NNHM thus need to be aware of both the sensitivity of the posterior to the heterogeneity prior (called here sensitivity S) and the sensitivity of the posterior to the within-study standard deviation uncertainty (called here identification I). Moreover, these concerns are intertwined and cannot be separately assessed, thus necessitating a simultaneous two-dimensional sensitivity-identification quantification. However, we currently still lack a systematic account of how sensitivity and identification affect the posterior inference in a Bayesian NNHM.

As it is impossible to elicit a unique prior (Gelman et al., 2017), a method is needed that perturbs the specific prior and assesses the impact of prior perturbation on posterior inference. The research area for formal sensitivity quantification is very broad and comprises, for example, the robust Bayesian analysis (Berger & Berliner, 1986; Berger et al., 2000; Ríos Insua et al., 2000), local robustness (Gustafson, 2000) and the epsilon-local sensitivity (Roos et al., 2015). The local sensitivity approach deals with the rate of change of posteriors with respect to prior perturbation, usually using differential calculus to quantify local sensitivity.

Identification is a concern in a Bayesian setting, because a model expressed by a nonidentified likelihood can be still estimable within the Bayesian framework, given prior assumptions. This means that in the Bayesian setting, the posterior inference for nonidentified parameters is not impacted by the data but is completely determined by the priors. This can be problematic for scientific inference and policy decisions, which endeavor to be determined by data, because in a Bayesian hierarchical model, it is difficult to know which parameters are informed by the data and which are nonidentified.

The broad and complex research area of formal identification investigation includes Bayesian identifiability (Gelfand & Sahu, 1999; Xie & Carlin, 2006) and Bayesian inference for partially identified models (Gustafson, 2015). Identification addresses the impact of the number of observations in a data set (likelihood) on the posterior inference. Therefore, some

approaches to identification explicitly change (increase or reduce) the number of observations in the analysis. For example, in the context of classical statistical analysis, both cloning and subsetting of observations have been suggested (Jiang, 2013; Lele et al., 2010; Normand, 1999). Neither data cloning nor the subset argument is, however, applicable if there are only two studies included for a medical meta-analysis. Thus, in order to quantify identification in a medical Bayesian meta-analysis, the likelihood must be perturbed such that its impact on the posterior system is changed, but without explicitly adding or removing observations.

To address these concerns, we develop a principled approach to combined sensitivity and identification quantification in a Bayesian NNHM. The novel two-dimensional sensitivity-identification (*S-I*) measure addresses sensitivity and identification in a Bayesian NNHM for all model parameters on an equal footing. We keep the prior for the main effect μ fixed and focus on perturbations of the heterogeneity prior for τ and the within-study standard deviation σ_i. Using the approach to aligning heterogeneity priors for a Bayesian NNHM proposed by Ott et al. (2021) as a starting point, we suggest systematic likelihood and heterogeneity prior perturbations that are based on perturbations of the relative latent model complexity (RLMC). While sensitivity quantifies the impact of the heterogeneity prior controlled by its scale parameter, identification is concerned with the impact of the data (σ_i-uncertainty) on the marginal posterior inference. We provide *S-I* estimates for one light-tailed half-normal (HN) and one heavy-tailed half-Cauchy (HC) heterogeneity prior and three RLMC targets. Finally, we provide the `R` code for automatic *S-I* quantification in an `R` package `si4bayesmeta` on R-Forge.

We apply the proposed framework for simultaneous *S-I* quantification to three medical case studies: historical data for a conventional therapy (Table 1), data with one large study included and excluded (Table 2), and two subgroup meta-analyses specified by their randomization status (Tables S4 and S5 of the Supporting Information). Providing concrete *S-I* estimates helps to pinpoint the severity of the impact of both the heterogeneity prior and the σ_i uncertainty on posterior estimates, which allows the impact of HN and HC heterogeneity priors to be assessed.

In the second section of this paper, we review two challenging and motivating case studies. In Section 3, we introduce the general theoretical framework supporting our method, specify the numerical computation of both raw *S-I* and relative *qS-qI* estimates, and introduce the `R` package `si4bayesmeta`. Results presented in the tables in Section 4 demonstrate the relevance of both *S-I* and *qS-qI* estimates for medical case studies. For the sake of brevity, some relevant material is then included only in the Supporting Information. Primarily, the unified *S-I* approach is extended to cover Bayesian normal-t hierarchical model (NtHM) fitted by MCMC in `JAGS`. The Bayesian NtHM addresses the important case when the assumption of the normality of random effects is not satisfied and thus provides a robustification of the popular but simple Bayesian NNHM against outliers in the data. This extension proves useful in one medical case study and in a simulation study. Moreover, we demonstrate how the randomization status of two subgroup meta-analyses affects the *S-I* estimates. We conclude with a discussion in Section 5.

2 | DATA

This section reviews two challenging and motivating case studies. The historical data set for a conventional periodontal therapy is focused on the mean pocket reduction (MPR), which is a continuous outcome. The second application, which has a binary outcome, deals with the association between diabetes mellitus and the severe course of Coronavirus Disease 2019 (COVID-19). It consists of two data sets, one of which is a subset of another data set from which one large study has been excluded. One additional case study, which consists of two subgroup meta-analyses specified by their randomization status is provided in Section 3 of the Supporting Information.

2.1 | Continuous outcome: Mean pocket reduction

Periodontitis is a complex and increasingly prevalent infective disease that affects the tissue surrounding teeth (GBD, 2016). As a clinical consequence, tissue inflammation leads to pocket formation, bone resorption, tooth loosening, and tooth loss. The conventional nonsurgical periodontal therapy involves removal of dental plaque and calculus (scaling), followed by smoothing (planning) of the affected surfaces of the roots. Usually, the success of this therapy is quantified by the reduction of the pocket depth after treatment. High values of MPR indicate benefit to the patient. Table 1 shows historical, aggregated MPR values after the conventional periodontal therapy in $k = 13$ studies from Zaugg et al. (2014). These data provide a basis for applying common methods of learning from historical data (Neuenschwander et al., 2010; Neuenschwander & Schmidli, 2020; Schmidli et al., 2014).

TABLE 1 Data for the mean pocket reduction (MPR) after a conventional nonsurgical periodontal treatment (Zaugg et al., 2014): within-study sample size (total), MPR (mean), standard deviation (sd), and data supplied for the Bayesian meta-analysis: MPR (y) and SE(MPR) (σ). The reference standard deviation for this data set is $\sigma_{\text{ref}} = 0.190$ (Equation (4))

	Conventional			y	σ
Study	Total	Mean	sd	MPR	SE(MPR)
1	10	0.28	0.32	0.28	0.10
2	11	2.25	2.47	2.25	0.74
3	14	3.00	1.57	3.00	0.42
4	10	0.80	0.86	0.80	0.27
5	42	0.47	0.11	0.47	0.02
6	84	1.20	1.33	1.20	0.15
7	17	0.10	0.30	0.10	0.07
8	24	1.62	0.58	1.62	0.12
9	8	1.18	0.81	1.18	0.29
10	30	0.84	0.55	0.84	0.10
11	30	1.67	2.61	1.67	0.48
12	19	1.95	1.94	1.95	0.45
13	27	0.90	2.00	0.90	0.38

Sample sizes reported in Table 1 are problematic, because they lead to long 95% confidence intervals (95%CI) for the true mean and the true variance. For example, if the true standard deviation is fixed at 2, sample sizes (10, 50, 100) lead to approximate lengths of 95%CIs equal to (2.5, 1.1, 0.8) for the true mean and (7.4, 3.2, 2.2) for the true variance. As expected, these lengths decrease with increasing sample size. However, this decrease is faster for the mean than for the variance, indicating that the standard deviation (sd) estimates are more uncertain than the mean estimates of Table 1, especially for small total sample sizes. Consequently, the standard errors of MPR (σ_i) supplied for the Bayesian meta-analysis are also affected by this substantial uncertainty.

2.2 | Binary outcome: Diabetes mellitus

Each meta-analysis shows specific information allocation across individual studies and specific total information, which is quantified by the sum of within-study precisions (Hardy & Thompson, 1998). Various inclusion and exclusion criteria can generate subgroups of the original data set, with some studies excluded (Bjordal et al., 2004; Crins et al., 2014; Smith et al., 1995). This can change both the total information and the information allocation across individual studies. In the Bayesian setting, the altered information allocation across individual studies can impact both sensitivity and identification estimates. However, quantification of such an impact is challenging.

To illustrate this challenge, we examine the data on the association between diabetes mellitus and the severe course of COVID-19 (Kumar et al., 2020). To assess this association, Kumar et al. (2020) conducted a systematic review and identified $k = 24$ observational studies, which included at least 100 adult COVID-19 patients and reported the severity of COVID-19. Kumar et al. (2020) determined the number of patients with the severe course of COVID-19 in each study. They categorized the course of COVID-19 as severe either by standard predefined criteria or when patients suffered from progressive or refractory disease, showed acute respiratory distress syndrome, required ICU care, or required invasive mechanical ventilation. Moreover, Kumar et al. (2020) determined the number of diabetic patients and computed the log-odds ratio $y_i = \log(\text{OR}_i)$ of severe course of COVID-19 between diabetic and nondiabetic COVID-19 patients and the corresponding $\sigma_i = \text{SE}(\log(\text{OR}_i))$ in each study. The median number of patients included in these $k = 24$ studies was 208 (min = 116, Q1 = 137, Q3 = 479, max = 6637) with a total of 13,954 COVID-19 patients. According to Kumar et al. (2020), 23 studies had good quality, but the largest study based on registry data from 6637 patients (Study 20 in Table 2) was only classified as a fair quality study.

We follow the approach taken by Kumar et al. (2020) and consider two data sets: one based on $k = 24$ studies (DM24, Table 2) and one based on $k = 23$ studies (DM23) with Study 20 removed from the DM24 data set. For DM24, the total information computed as the sum of within-study precisions $1/\sigma_i^2$ of individual studies is equal to 208.1 (Hardy & Thompson,

TABLE 2 Data for the association between diabetes mellitus (DM24) and the severe course of Coronavirus Disease 2019 in $k = 24$ studies (Kumar et al., 2020): total within-study sample size (n), log(OR) (y), and SE(log(OR)) (σ). The reference standard deviation of the DM24 data set is $\sigma_{\text{ref}} = 0.475$ (Equation (4)). All studies with exception of the largest study based on registry data from 6637 patients (Study 20) had good quality. Study 20 was only classified as a fair quality study. Removal of Study 20 leads to a smaller data set (DM23) based on $k = 23$ studies with the reference standard deviation $\sigma_{\text{ref}} = 0.507$ (Equation (4))

Study	n Total	y log (OR)	σ SE (log(OR))
1	393	0.200	0.244
2	140	0.261	0.519
3	116	0.305	0.516
4	221	0.384	0.487
5	214	0.419	0.395
6	476	0.463	0.320
7	161	0.588	0.862
8	273	0.641	0.505
9	548	0.651	0.245
10	597	0.798	0.398
11	124	0.925	0.538
12	487	1.148	0.463
13	1099	1.157	0.250
14	1012	1.211	0.453
15	155	1.301	0.667
16	298	1.303	0.499
17	201	1.471	0.503
18	119	1.477	0.648
19	138	1.520	0.581
20	**6637**	**1.528**	**0.109**
21	135	2.186	0.698
22	167	2.315	0.666
23	123	2.333	0.778
24	120	4.056	1.479

1998). Study 20 comprises a large proportion (40.5%) of this total information. After removing Study 20, the total information of DM23 decreases to 123.9 with the maximal information contribution smaller than 14%. Considering both DM24 and DM23 data sets allows us to address the important question of how sensitivity and identification estimates are impacted by different information allocations across both data sets. However, a method is required that can quantify such an impact.

3 | METHODS

In this section, we introduce the Bayesian NNHM and summarize the idea behind RLMC, which explicitly controls the interplay between the study-specific within-study standard deviation σ_i contained in the data and the impact of the heterogeneity prior in a Bayesian NNHM. In addition, we define a measure of affinity (BC) and systematic perturbations for prior (P) and likelihood (L). Moreover, we define two combined (raw S-I and relative qS-qI) sensitivity-identification measures. Finally, we describe the implementation of the method as an `R` package.

3.1 | Normal–normal hierarchical model

A Bayesian NNHM for k studies is a Bayesian hierarchical model that consists of three parts: the sampling model (likelihood), the latent random-effects model, and priors (Normand, 1999). The likelihood is determined by data

$\boldsymbol{y} = \{(y_i, \sigma_i),\ i = 1, \ldots, k\}$ and by the assumption that observations y_i are realizations of a normal random variable Y_i, where

$$Y_i|\theta_i, \sigma_i \ \sim \ \mathrm{N}(\theta_i, \sigma_i^2) \tag{1}$$

and the within-study standard deviations denoted by σ_i, for $i = 1, \ldots, k$, are fixed, that is, assumed to be known. Exchangeability of the latent random effects θ_i is imposed by assuming that they follow a normal distribution with mean μ and a heterogeneity (between-study) standard deviation τ

$$\theta_i|\mu, \tau \ \sim \ \mathrm{N}(\mu, \tau^2), \quad \text{for} \quad i = 1, \ldots, k. \tag{2}$$

Finally, priors for $\mu \ \sim \ \pi(\mu)$ and $\tau \ \sim \ \pi(\tau)$ are assumed. Given priors $\pi(\mu)$ and $\pi(\tau)$, a Bayesian NNHM updates the information in the data $\boldsymbol{y}$ and arrives at marginal posterior estimates $\pi(\psi|\boldsymbol{y})$ for parameters $\psi \in \{\mu, \tau, \theta_1, \ldots, \theta_k, \theta_{\text{new}}\}$.

The interplay between σ_i and τ in a Bayesian NNHM can be conveniently expressed by the RLMC (Ott et al., 2021), the latent model complexity (Spiegelhalter et al., 2002) per study:

$$\text{RLMC} \ = \ \frac{1}{k} \sum_{i=1}^{k} \frac{\tau^2}{\tau^2 + \sigma_i^2} \ \approx \ \frac{\tau^2}{\tau^2 + \sigma_{\text{ref}}^2}, \tag{3}$$

where the reference within-study standard deviation σ_{ref} is defined by the geometric mean,

$$\sigma_{\text{ref}} = \left(\prod_{i=1}^{k} \sigma_i \right)^{\frac{1}{k}} = \exp\left(\frac{1}{k} \sum_{i=1}^{k} \log(\sigma_i) \right). \tag{4}$$

The RLMC for a Bayesian NNHM defined in Equation (3) attains values in $[0, 1]$ independently of the primary outcome and the number of studies involved in the meta-analysis. RLMC is a relative measure of heterogeneity that depends on both the within-study standard deviations σ_i and a particular value of the between-study standard deviation τ, which is related to I^2 (Higgins & Thompson, 2002) and shrinkage (Gelman et al., 2014; Gelman & Hill, 2007; Efron & Hastie, 2016). Moreover, RLMC is a function of both the pooling factor (Gelman & Hill, 2007) and the degree of heterogeneity (Neuenschwander et al., 2010).

The meaning of all parameters in a Bayesian NNHM in medical applications can be summarized as follows: The parameter μ represents the main effect. Decisions are guided by whether or not the 95% credible interval for μ (95%CI(μ)) contains 0. The parameter τ describes the between-study heterogeneity. This is a key parameter directly impacting the model complexity of a Bayesian NNHM. The posterior distributions of latent random effects $\theta_1, \ldots, \theta_k$ specify study-specific effects. They have the potential to impact design and analyses of future studies when a method is applied to dynamically borrow strength from another study (Röver & Friede, 2020). The predicted true effect in a new study θ_{new} expresses the posterior knowledge about that study (Röver, 2020). The θ_{new} distribution can impact the meta-analytic-predictive approach to incorporating relevant historical information into the design or analysis of clinical trials (Neuenschwander & Schmidli, 2020; Schmidli et al., 2014).

To evaluate the combined evidence, we apply the Bayesian NNHM model to aggregated observations. Therefore, y_i denotes the value of the aggregated outcome of interest in the ith study and σ_i denotes its standard error. We compute Bayesian NNHMs with the `R` package `bayesmeta` (Röver, 2020).

To the prior $\pi(\mu)$, we assign $\mathrm{N}(m, \gamma^2)$ distribution with $m = 0$ and standard deviation $\gamma = 4$, the so-called unit-information prior (Günhan et al., 2020; Kass & Wasserman, 1995; Röver, 2020). Note that both parameters for the $\pi(\mu)$ prior will be kept fixed for all analyses in this paper.

For the heterogeneity prior $\pi(\tau)$, we assume either a light-tailed HN or a heavy-tailed HC distribution. The use of both HN and HC heterogeneity priors appropriately reflects the current best practice recommended for the Bayesian NNHM: HN with the base-scale parameter A_{b}^{HN} fixed either at 0.5 or 1 (HN(0.5), HN(1)) (Friede et al., 2017a, 2017b; Bender et al., 2018) and HC with the base-scale parameter A_{b}^{HC} fixed at 1 (HC(1)) (Gelman, 2006; Polson & Scott, 2012; Gelman et al., 2017). In the main manuscript, the choice of the base-scale parameter A_{b} for HN and HC is guided by this recommendation. In order to enable comparisons, we will align HN and HC heterogeneity priors. The base-scale parameter values of the

TABLE 3 Mean pocket reduction: base-scale parameter values A_b of 50%-RLMC-adjusted HN and HC heterogeneity priors across three RLMC targets 0.76, 0.93, and 0.96. Targets 0.76, 0.93, and 0.96 were obtained for HN(0.5), HN(1), and HC(1) applied to MPR

Prior	RLMC = 0.76	RLMC = 0.93	RLMC = 0.96
HN	0.50	1.00	1.48
HC	0.34	0.67	1.00

aligned heterogeneity priors will be obtained using an appropriate 50%-RLMC-based adjustment of heterogeneity priors (Ott et al., 2021), reviewed in the next section. Note that the choice of base-scale parameter values in Sections 4.1, 4.3, and 4.4 of the Supporting Information is based on the grid of 0.25, 0.5, and 0.75 RLMC values combined with the 50%-RLMC-based adjustment.

3.2 | 50%-RLMC-based adjustment of heterogeneity priors

The heterogeneity standard deviation τ in Equation (2) is a very important parameter for quantifying the similarity of studies. The value $\tau = 0$ means that the sampling mean of all studies is μ (no heterogeneity and no random effects are available). In contrast, $\tau = \infty$ indicates an extremely large between-study heterogeneity. In such a case, the k studies are considered distinct, they must be analyzed in isolation, and k different random effects are necessary to adequately model the data. Parameter τ is also an essential part of σ/τ, the so-called degree of heterogeneity, which is specified by the ratio of the within-study to the between-study standard deviations and is the key parameter determining the impact of the heterogeneity prior on the Bayesian NNHM (Neuenschwander et al., 2010).

In this paper, we will assign either an HN or an HC heterogeneity prior to τ and focus on the interplay between the between-study heterogeneity prior for τ and the within-study standard deviations σ_i. HN(0.5), HN(1), and HC(1) heterogeneity priors recommended in the literature induce median RLMC values 0.76, 0.93, and 0.96 for MPR (Table 3). Therefore, we anchor analyses in Section 4.1 on these three RLMC targets. In contrast, HN(0.5), HN(1), and HC(1) induce median 0.33, 0.67, and 0.82 RLMC values for DM24 (Table S16 of the Supporting Information), which we use as RLMC targets for analyses in Section 4.2. Because the HN(0.5), HN(1), and HC(1) heterogeneity priors recommended in the literature induce data-dependent RLMC targets, we provide an alternative analysis in Sections 4.1, 4.3, and 4.4 of the Supporting Information that keeps the mutual interplay between the between-study heterogeneity τ and the within-study standard deviation σ_i values expressed by RLMC fixed on the grid of three 0.25, 0.5, and 0.75 values.

Given a target RLMC value, Ott et al. (2021) introduced a 50%-RLMC-based adjustment for heterogeneity priors, which aligns the medians of different heterogeneity priors at the reference threshold U_{ref}. Thus,

$$\mathrm{P}[\tau > U_{\text{ref}}] = \alpha = 50\%, \tag{5}$$

where the reference threshold U_{ref} satisfies

$$U_{\text{ref}} \;=\; \sigma_{\text{ref}}\sqrt{\text{RLMC}/(1-\text{RLMC})}. \tag{6}$$

Note that the condition for the tail adjustment in Equation (5) aligns the medians of different heterogeneity priors at a dynamic reference threshold U_{ref} in Equation (6). U_{ref} is dynamic, because it depends on both the data through σ_{ref} and the base target RLMC. Because the densities of HN and HC for τ are two different mathematical functions, it is impossible to get a perfect match. Nevertheless, the 50%-RLMC-based adjustment uses U_{ref} to align the medians of both the HN and HC heterogeneity priors.

Given the above prerequisites, the heterogeneity prior base-scale parameter A_b values in Table 3 and in Table S16 of the Supporting Information are found numerically to fit the condition imposed by Equation (5). For any data set at hand, there is a one-to-one correspondence between the base-scale parameter A_b value of the heterogeneity prior and the target RLMC value. For HN and HC, formulas connecting the tail constraint in Equation (5) and the base-scale parameter value A_b of the heterogeneity prior can be derived analytically to get $A_b^{\text{HN}} = U_{\text{ref}}/\Phi^{-1}(1-\alpha/2)$ for HN and $A_b^{\text{HC}} = U_{\text{ref}}/\tan(\pi(1-\alpha)/2)$ for HC (Ott et al., 2021). For example, if we assume base $\text{RLMC}_b = 0.76$ for the MPR data with $\sigma_{\text{ref}} = 0.19$ in Table 1, the scale parameter for a 50%-RLMC-adjusted HN is 0.5 (Table 3). For the base $\text{RLMC}_b = 0.76$ and the same MPR data set, the scale parameter for a 50%-RLMC-adjusted HC is specified as 0.34. This approach enables a coherent choice of the scale

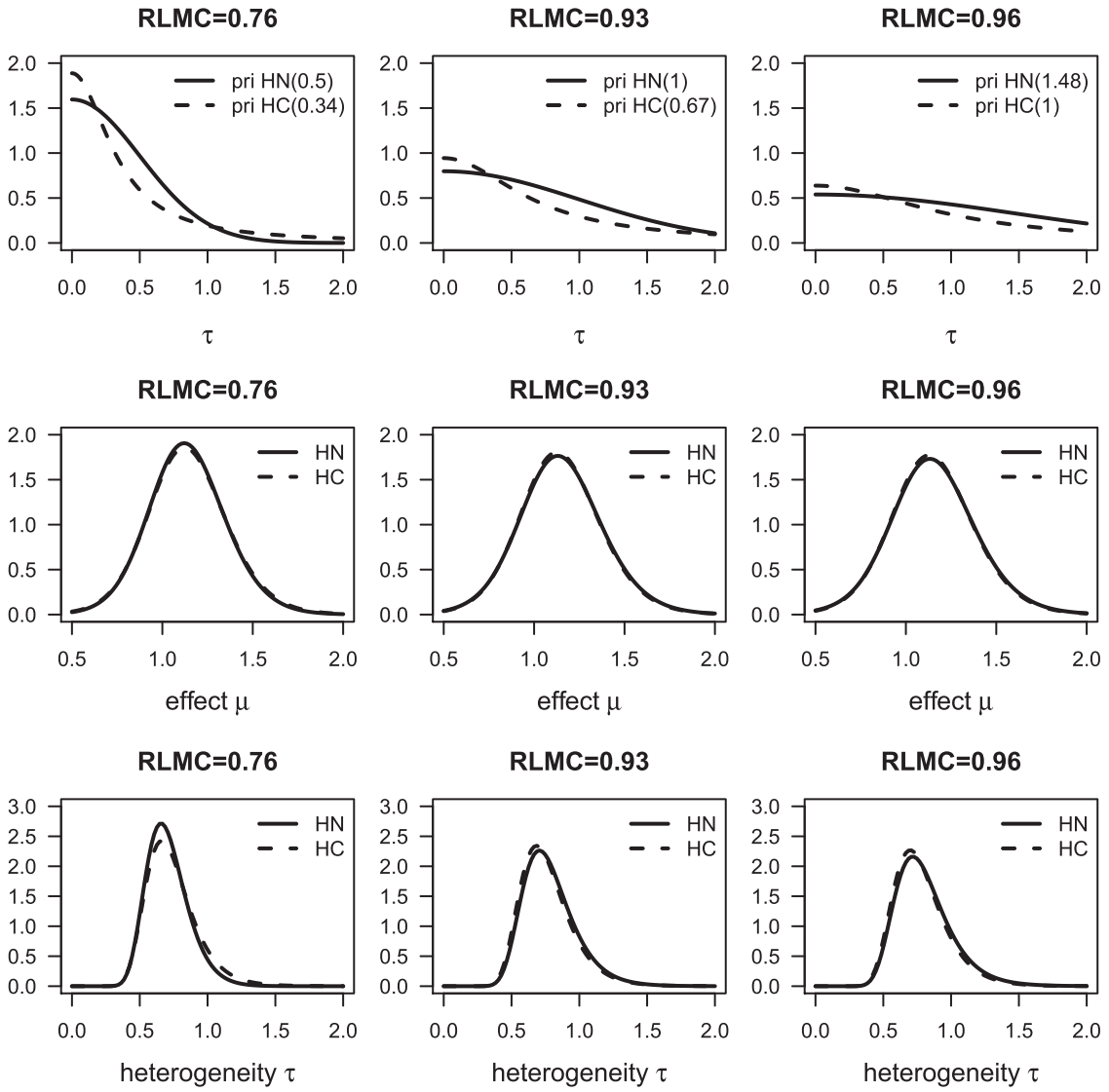

FIGURE 1 Mean pocket reduction: prior densities for τ (top row) and posterior densities for μ (middle row) and τ (bottom row) induced by 50%-RLMC-adjusted HN (solid line) and HC (dashed line) heterogeneity priors with target RLMC values fixed at 0.76 (left column), 0.93 (middle column), and 0.96 (right column). See Table 3 for the corresponding base-scale parameter values of HN and HC heterogeneity priors

parameter values of two different HN and HC heterogeneity priors (Ott et al., 2021). Our experience so far reveals that it is more meaningful to compare posteriors induced by 50%-RLMC-adjusted heterogeneity priors, which represent a similar model complexity. Therefore, Figure 1 is split into panels that show posteriors across different RLMC targets.

3.3 | Affinity measure BC

To quantify the impact of data and heterogeneity prior perturbations on the posterior inference in a Bayesian NNHM, we need a formal measure to assess the affinity of two distributions. We suggest using the Bhattacharyya coefficient (BC), which is a symmetric measure of affinity (Roos et al., 2015; Roos & Held, 2011). The BC between a base (b) and an altered (a) density function is given by

$$\mathrm{BC}(\pi_{\mathrm{b}}(\psi|\boldsymbol{y}), \pi_{\mathrm{a}}(\psi|\boldsymbol{y})) = \int_{-\infty}^{\infty} \sqrt{\pi_{\mathrm{b}}(\psi|\boldsymbol{y})\pi_{\mathrm{a}}(\psi|\boldsymbol{y})}d\psi, \tag{7}$$

with $\psi \in \{\mu, \tau, \theta_1, \ldots, \theta_k, \theta_{\text{new}}\}$. BC is equal to the maximal value 1 if and only if the densities are identical. It takes the minimal value 0 when the densities do not overlap due to disjoint supports.

BC has convenient numerical properties and is invariant to any one-to-one transformation of both densities (Roos & Held, 2011; Jeffreys, 1961). Particularly significant is that the value of BC computed for two densities and two log-transformed densities is equal. BC is also the main component of the Hellinger distance (H), which naturally fits into the framework of Bayesian geometry (de Carvalho et al., 2019). More specifically, $\text{H}^2 = 1 - \text{BC}$.

The BC in Equation (7) between two normal densities can be derived analytically. Denote by π_j^{N} the density of a normal $\text{N}(\mu_j, \sigma_j^2)$ distribution, for $j = \text{b}, \text{a}$. Then, the total BC for two normal densities is

$$\text{BC}(\pi_{\text{b}}^{\text{N}}, \pi_{\text{a}}^{\text{N}}) = \int_{-\infty}^{\infty} \sqrt{\pi_{\text{b}}^{\text{N}}(u)\pi_{\text{a}}^{\text{N}}(u)}du = \sqrt{\frac{\sigma_{\text{b}}\sigma_{\text{a}}}{\left(\frac{\sigma_{\text{b}}^2+\sigma_{\text{a}}^2}{2}\right)}} \exp\left\{-\frac{(\mu_{\text{b}}-\mu_{\text{a}})^2}{8\left(\frac{\sigma_{\text{b}}^2+\sigma_{\text{a}}^2}{2}\right)}\right\}. \tag{8}$$

When computing BC in practice, we utilize the general idea that any density can be approximated to the first order by a parametric normal distribution (Johnson et al., 1994; Hjort & Glad, 1995). We thus focus our attention on the parametric normal approximation of the marginal posterior distribution and approximate $\pi(\psi|\boldsymbol{y})$ for $\psi \in \{\mu, \tau, \theta_1, \ldots, \theta_k, \theta_{\text{new}}\}$ by an appropriate $\pi^{\text{N}}(\psi|\boldsymbol{y})$ normal distribution combined with moment matching and use Equation (8) to get

$$\text{BC}(\pi_{\text{b}}(\psi|\boldsymbol{y}), \pi_{\text{a}}(\psi|\boldsymbol{y})) \approx \text{BC}(\pi_{\text{b}}^{\text{N}}(\psi|\boldsymbol{y}), \pi_{\text{a}}^{\text{N}}(\psi|\boldsymbol{y})). \tag{9}$$

Moment matching equates the mean and standard deviation estimates of a marginal posterior distribution $\pi(\psi|\boldsymbol{y})$ and the approximating normal distribution $\pi^{\text{N}}(\psi|\boldsymbol{y})$. Note that for the parameter τ, we apply the normal approximation to the log-transformed marginal posterior for τ. This approach is justified by the invariance of BC to any one-to-one transformation. For brevity of notation, we refer to this parameter as τ, although all numerical computations are conducted based on $\log(\tau)$.

Note that Equation (8) applied to $\text{BC}(\pi_{\text{b}}^{\text{N}}(\psi|\boldsymbol{y}), \pi_{\text{a}}^{\text{N}}(\psi|\boldsymbol{y}))$ in Equation (9) focuses on changes in posterior moments of base and altered marginal posterior distributions. In fact, the total BC in Equation (8) is based on common descriptive statistics summarizing posterior inference: posterior mean and posterior standard deviation. Equation (8) quantifies the impact of perturbations on both posterior moments in two steps. While the multiplicator measures the discrepancy between two standard deviations, the multiplicand measures the discrepancy between two means that is scaled by the pooled variance. Note that the multiplicand is the well-known Mahalanobis distance in one dimension (Mardia et al., 1979).

3.4 | Systematic heterogeneity prior and likelihood perturbations

Sensitivity (S) assesses to what extent systematic heterogeneity prior perturbations cause posterior variation. In contrast, identification (I) reflects how perturbations of the likelihood (L) impact posterior inference, which can be achieved by suitably scaling the σ_i values in log(Likelihood).

The posterior distribution underlying the Bayesian meta-analysis expressed by an NNHM is a self-contained system in Equation (10). This system consists of the likelihood (L), which encompasses the data, the latent random-effects model and priors (P) (see Section 2.1 of the Supporting Information for details):

$$\log(\text{Posterior}) \approx \log(\text{Likelihood}) + \log(\text{Random-effects model}) + \log(\text{Prior}). \tag{10}$$

Note that the within-study standard deviation values $(\sigma_1, \ldots, \sigma_k)$ in the data $\boldsymbol{y}$ specify the impact of the likelihood L defined in Equation (1) on the posterior in Equation (10). The heterogeneity prior $\pi(\tau)$ specification governs the impact of the heterogeneity prior on the posterior in Equation (10).

The main objective of our method for a Bayesian NNHM is to coherently assess the impact of two different sources of uncertainty: the uncertainty of heterogeneity prior choice and the uncertainty in within-study standard deviation σ_i

values. The impact of both sources is assessed by appropriately chosen systematic perturbations. In the first dimension (S), the uncertainty inherent in a specific choice of the scale parameter value of a heterogeneity prior is assessed using appropriate scale parameter value perturbations. In the second dimension (I), the uncertainty inherent in within-study standard deviation σ_i values provided by the data is assessed using appropriate σ_i value perturbations. Moreover, perturbations in both dimensions are unified and made coherent, because the impact of a heterogeneity prior for τ and the within-study standard deviation σ_i values is connected through the RLMC in a Bayesian NNHM (see Section 2.2 of the Supporting Information for details). For example, lower values of RLMC in Equation (3) can be obtained either by perturbing the heterogeneity prior (P) with fixed data or perturbing data in the likelihood (L) with a fixed heterogeneity prior. This section describes systematic perturbations of heterogeneity priors (P) and the likelihood (L).

3.5 | Systematic heterogeneity prior (P) perturbation

Systematic perturbations of the heterogeneity prior (P) in a Bayesian NNHM are conducted given a fixed likelihood. In particular, both the within-study standard deviation σ_i values in the data and consequently the reference within-study standard deviation σ_{ref} in Equation (4) are kept fixed. If RLMC changes while the likelihood is kept fixed, then the scale parameter A_{b} of a heterogeneity prior for τ must change. Systematic perturbations of the heterogeneity prior (P) are carried out using the 50%-RLMC-based adjustment. Perturbing, that is, slightly changing the value of the RLMC, leads to slight changes of the reference threshold U_{ref} according to Equation (6). Because the tail adjustment with the tail fixed at $\alpha = 50\%$ in Equation (5) is directly connected through an analytical formula to the scale parameter A_{b} of heterogeneity priors, slight perturbations of U_{ref} imply slight perturbations of the scale parameter A_{b} of heterogeneity priors.

As an example, we demonstrate the method described above on the MPR data in Table 1 and show how the scale parameter of an HN heterogeneity prior is perturbed through an RLMC perturbation with fixed σ_i values. If we assume the base $\text{RLMC}_{\text{b}} = 0.76$, then we obtain the base-scale parameter value 0.502 for an HN heterogeneity prior: HN(0.502). In addition, with $h = 0.0044$, we obtain $\text{RLMC}_{\text{b}} - h = 0.7556$, which implies HN(0.496), and $\text{RLMC}_{\text{b}} + h = 0.7644$, which implies HN(0.508).

3.6 | Systematic likelihood (L) perturbation

Here, we show how systematic perturbations of RLMC provoke systematic perturbations of the likelihood (L) (data), given a fixed RLMC-adjusted heterogeneity prior. In order to perturb the likelihood in a Bayesian NNHM subject to the RLMC constraint, we keep the heterogeneity prior fixed according to the base 50%-RLMC-adjustment and multiply the within-study standard deviations σ_i by a factor f. Thus,

$$\sigma_{\text{ref,RLMC}_{\text{a}}} = f\sigma_{\text{ref,RLMC}_{\text{b}}} = \left(\prod_{i=1}^{k}(f\sigma_i)\right)^{1/k}, \tag{11}$$

where

$$f = \sqrt{\frac{1-\text{RLMC}_{\text{a}}}{\text{RLMC}_{\text{a}}}} \Big/ \sqrt{\frac{1-\text{RLMC}_{\text{b}}}{\text{RLMC}_{\text{b}}}}, \tag{12}$$

with altered RLMC_{a} values: either $\text{RLMC}_{\text{a}} = \text{RLMC}_{\text{b}} - h$ or $\text{RLMC}_{\text{a}} = \text{RLMC}_{\text{b}} + h$. Scaling the data by a factor f according to Equations (11) and (12) while keeping the heterogeneity prior fixed at a base target RLMC_{b} makes the likelihood more or less pronounced. This corresponds to changing the effective sample size in each study supplied for a Bayesian meta-analysis.

For example, if we set the base $\text{RLMC}_{\text{b}} = 0.76$ and take $h = 0.0044$, then we obtain the lower $\text{RLMC}_{\text{b}} - h = 0.7556$, which implies a factor $f = 1.0121$ and the upper $\text{RLMC}_{\text{b}} + h = 0.7644$, which implies a factor $f = 0.9879$. Changing RLMC with a fixed heterogeneity prior will induce scaling of the study-specific within-study standard deviation σ_i values by the factor f, and data with $f\sigma_i$ will be supplied for a Bayesian meta-analysis.

Both perturbed heterogeneity priors and perturbed study-specific within-study standard deviation σ_i values used for Bayesian NNHM computations produce perturbed marginal posterior distributions of all parameters in the Bayesian NNHM $\{\mu, \tau, \theta_1, \ldots, \theta_k, \theta_{\text{new}}\}$. Therefore, we are able to compute BC between base and altered marginal posteriors of $\{\mu, \tau, \theta_1, \ldots, \theta_k, \theta_{\text{new}}\}$ with respect to the RLMC perturbations affecting either the heterogeneity prior P or the within-study standard deviation L. This property leads to the unified two-dimensional sensitivity-identification (S-I) measure, which is defined in the next section as a numerical approximation to the second derivative of BC with respect to systematic RLMC perturbations evaluated at a base target RLMC. Note that RLMC targets are fixed at 0.76, 0.93, and 0.96 for MPR in Table 3, at 0.33, 0.67, and 0.82 for DM in Table S16 of the Supporting Information, and at 0.25, 0.5, and 0.75 for analyses reported in Sections 4.1, 4.3, and 4.4 of the Supporting Information.

3.7 | Formal combined sensitivity-identification (S-I) measure

Formal approaches to sensitivity and identification quantification are usually based on differential calculus or its numerical approximations (McCulloch, 1989; Xie & Carlin, 2006). For the BC, the second derivative is a practical quantity (McCulloch, 1989; Dey & Birmiwal, 1994).

We define the unified, formal, two-dimensional sensitivity-identification (S-I) measure as negative second derivatives of the BC with respect to the RLMC perturbations evaluated at a base RLMC_b value when perturbing either the heterogeneity prior (P)

$$S_P(\psi) = -\frac{d^2 \text{BC}_P(\psi; \text{RLMC})}{d\text{RLMC}^2}\bigg|_{\text{RLMC}=\text{RLMC}_\text{b}} \tag{13}$$

or the σ_i, $i = 1, \ldots, k$, values in the likelihood (L)

$$I_L(\psi) = -\frac{d^2 \text{BC}_L(\psi; \text{RLMC})}{d\text{RLMC}^2}\bigg|_{\text{RLMC}=\text{RLMC}_\text{b}}, \tag{14}$$

for $\psi \in \{\mu, \tau, \theta_1, \ldots, \theta_k, \theta_{\text{new}}\}$. The negative second derivatives of BC in Equations (13) and (14) with respect to RLMC perturbations assess the curvature of the squared Hellinger distance (H^2). They quantify the acceleration with which the marginal posterior changes locally around RLMC_b when RLMC perturbations are induced either by P or by L perturbations. We compute numerical approximations to the derivatives and approximate the posterior distribution with a normal distribution when computing BC as shown in Equation (9). Thus, we substitute the true second derivative

$$\frac{d^2 \text{BC}(\psi; \text{RLMC})}{d\text{RLMC}^2}\bigg|_{\text{RLMC}=\text{RLMC}_\text{b}}$$

with its numerical approximation

$$\approx \frac{\text{BC}(\pi^{\text{N}}_{\text{RLMC}_\text{b}}(\psi|\boldsymbol{y}), \pi^{\text{N}}_{\text{RLMC}_\text{b}+h}(\psi|\boldsymbol{y})) - 2 + \text{BC}(\pi^{\text{N}}_{\text{RLMC}_\text{b}}(\psi|\boldsymbol{y}), \pi^{\text{N}}_{\text{RLMC}_\text{b}-h}(\psi|\boldsymbol{y}))}{h^2}.$$

Note that the $S_P(\psi)$ and $I_L(\psi)$ estimates are always positive, because BC attains a maximum 1 at $h = 0$. Large values of both S-I estimates indicate that marginal posteriors are highly affected both by the uncertainty in the choice of the heterogeneity prior scale parameter (S dimension) and by the uncertainty of the within-study standard deviations σ_i (I dimension).

To compute the raw $S_P(\psi)$ and $I_L(\psi)$ estimates in practice, we compute the Bayesian NNHM several times. First, for a fixed base RLMC_b value, the base Bayesian NNHM for the original base data with a 50%-RLMC-adjusted heterogeneity prior is computed. This leads to base $\pi_{\text{RLMC}_\text{b}}(\psi|\boldsymbol{y})$ marginal posterior distributions for all parameters $\psi \in \{\mu, \tau, \theta_1, \ldots, \theta_k, \theta_{\text{new}}\}$. Second, we systematically perturb the RLMC and compute the $\pi_{\text{RLMC}_\text{b}\pm h}(\psi|\boldsymbol{y})$ posteriors that are induced, either when the systematic RLMC perturbation impacts the likelihood L (through perturbed σ_i, $i = 1, \ldots, k$

values) with a fixed heterogeneity prior or when it impacts the heterogeneity prior P with fixed data. Note that the value of the affinity measure $\mathrm{BC}(\pi^{\mathrm{N}}_{\mathrm{RLMC_b}}(\psi|\boldsymbol{y}), \pi^{\mathrm{N}}_{\mathrm{RLMC_b}}(\psi|\boldsymbol{y})) = 1$ for two nonperturbed posteriors.

We then apply the combined *S-I* measure, anchored at a base RLMC, to both HN and HC heterogeneity priors. For numerical computations in this paper, we perturb each RLMC target by a value $h = 0.0044$. For example, if we set the base target $\mathrm{RLMC_b} = 0.76$ and perturb it by the perturbation 0.0044, then we obtain two perturbed RLMC values: $\mathrm{RLMC_b} - h = 0.7556$ and $\mathrm{RLMC_b} + h = 0.7644$. We postpone the justification of this particular h value choice and a discussion of the stability of the resulting *S-I* values until Section 6 of the Supporting Information. Although the analytical second derivative has been shown to be a well-defined and practical quantity (McCulloch, 1989; Dey & Birmiwal, 1994), its numerical approximation can potentially attain a value of 0. However, this issue did not turn out to be a problem in the medical applications considered in this paper.

3.8 | Quotients of sensitivity identification (*qS-qI*)

The unified *S-I* measure defined in Section 3.7 allows for a wide range of direct comparisons relevant for applications. Such comparisons are expressed by quotients of *S-I* values leading to relative *qS-qI* estimates. For example, one set of relevant *qS-qI* values emerges when we wish to compare the impact of two different heterogeneity priors on each parameter in the Bayesian NNHM. In such a case, the unified two-dimensional relative *qS-qI* measure is defined by

$$qS^{\mathrm{HN/HC}}(\psi) = \frac{S_P(\psi, \mathrm{HN})}{S_P(\psi, \mathrm{HC})} \quad \text{and} \quad qI^{\mathrm{HN/HC}}(\psi) = \frac{I_L(\psi, \mathrm{HN})}{I_L(\psi, \mathrm{HC})} \tag{15}$$

with $\psi \in \{\mu, \tau, \theta_1, \ldots, \theta_k, \theta_{\mathrm{new}}\}$ and a fixed RLMC target.

Another interesting *qS-qI* measure is obtained when raw *S-I* values are compared between two different data sets. For example, *S-I* values for $\psi \in \{\mu, \tau, \theta_{\mathrm{new}}\}$ obtained for a data set based on k studies can be compared with those obtained for a subgroup based on $k-1$ studies. Thus, the formal unified two-dimensional relative *qS-qI* measure is defined by

$$qS^{(k-1)/k}(\psi) = \frac{S_P(\psi, k-1)}{S_P(\psi, k)} \quad \text{and} \quad qI^{(k-1)/k}(\psi) = \frac{I_L(\psi, k-1)}{I_L(\psi, k)} \tag{16}$$

with $\psi \in \{\mu, \tau, \theta_{\mathrm{new}}\}$ for a fixed target RLMC and a fixed heterogeneity prior.

Depending on the research question, other relative *qS-qI* estimates can be computed. For example, it can be relevant to compare the impact of two different Bayesian NNHM and NtHM models for meta-analysis or the impact of randomized and nonrandomized subgroups (Section 2.6 of the Supporting Information). Alternatively, it can be of interest to compare the impact of two different RLMC targets. Moreover, raw *S-I* values for each parameter ψ can be compared within one scenario, with both the target RLMC and the heterogeneity prior kept fixed, as defined in Section 2.6 of the Supporting Information.

3.9 | Package: `si4bayesmeta`

The method described above is freely accessible as an `R` package `si4bayesmeta`. The `si4bayesmeta` package (`https://github.com/hunansona/si4bayesmeta`) contains several wrapper functions built on two `R` packages `bayesmeta` (Röver, 2020) and `rjags` (Plummer, 2016). Functions `median_rlmc`, `pri_par_adjust_HN`, and `pri_par_adjust_HC` compute RLMC targets and base-scale parameters of HN and HC heterogeneity priors shown in Table 3 and in Table S16 of the Supporting Information. Two functions, `d2BC_S_I_HN_raw` and `d2BC_S_I_HC_raw`, produce the main result for Bayesian NNHM fitted by `bayesmeta` shown in Table 5 below and in Table S18 of Supporting Information for a specified RLMC target. Posterior estimates for a specified RLMC target in Table 4 below and in Table S17 of the Supporting Information are computed by the functions `raw_estimates_HN` and `raw_estimates_HC`. Additional functions necessary to compute *S-I* estimates for Bayesian NNHM and Bayesian NtHM fitted by MCMC sampling in `JAGS` are specified in Section 2.5 of the Supporting Information. Moreover, the function `h_choice_all` facilitates the choice of the numerical RLMC perturbation h, which is adjusted for the epsilon-local sensitivity (Roos et al., 2015), as outlined in Section 6 of the Supporting Information. All functions are applicable to data frames in the

TABLE 4 Mean pocket reduction: posterior mean and standard deviation (sd) estimates together with the shortest 95% credible interval (CI) and the length it reaches (LCI) for parameters $\{\mu, \tau, \theta_{\text{new}}\}$ obtained with target RLMC values fixed at 0.76, 0.93, and 0.96 and with an HN (upper batch) and an HC (lower batch) 50%-RLMC-adjusted heterogeneity prior. See Table 3 for the corresponding base-scale parameter values of HN and HC heterogeneity priors

	RLMC = 0.76					RLMC = 0.93					RLMC = 0.96				
Par	**Mean**	**sd**	**CI_{low}**	**CI_{up}**	**LCI**	**Mean**	**sd**	**CI_{low}**	**CI_{up}**	**LCI**	**Mean**	**sd**	**CI_{low}**	**CI_{up}**	**LCI**
$\mu^{(\text{HN})}$	1.14	0.22	0.71	1.58	0.87	1.15	0.24	0.68	1.64	0.96	1.15	0.25	0.67	1.65	0.99
$\tau^{(\text{HN})}$	0.71	0.16	0.43	1.02	0.60	0.79	0.20	0.44	1.18	0.74	0.81	0.21	0.44	1.23	0.79
$\theta_{\text{new}}^{(\text{HN})}$	1.14	0.76	−0.37	2.66	3.03	1.15	0.85	−0.53	2.85	3.38	1.15	0.87	−0.58	2.91	3.48
$\mu^{(\text{HC})}$	1.14	0.23	0.69	1.61	0.91	1.15	0.24	0.69	1.62	0.94	1.15	0.24	0.68	1.64	0.96
$\tau^{(\text{HC})}$	0.74	0.19	0.42	1.12	0.70	0.76	0.20	0.43	1.16	0.73	0.78	0.20	0.44	1.19	0.75
$\theta_{\text{new}}^{(\text{HC})}$	1.14	0.80	−0.44	2.75	3.19	1.15	0.82	−0.49	2.81	3.30	1.15	0.84	−0.53	2.85	3.38

TABLE 5 Mean pocket reduction *S-I*: sensitivity and identification estimates for parameters $\{\mu, \tau, \theta_{\text{new}}\}$ in Bayesian NNHMs obtained for six scenarios across target RLMC values fixed at 0.76, 0.93, and 0.96 with HN (upper batch) and HC (lower batch) 50%-RLMC-adjusted heterogeneity priors. See Table 3 for the corresponding base-scale parameter values of HN and HC heterogeneity priors

	RLMC = 0.76		RLMC = 0.93		RLMC = 0.96	
Par	S_P	I_L	S_P	I_L	S_P	I_L
$\mu^{(\text{HN})}$	0.1661	1.3805	0.2102	6.7392	0.2486	24.8017
$\tau^{(\text{HN})}$	3.4663	470.1564	2.3836	3393.3009	70.0537	7766.1207
$\theta_{\text{new}}^{(\text{HN})}$	0.1855	0.2790	0.2458	1.5349	0.2940	5.8388
$\mu^{(\text{HC})}$	0.0039	1.2579	0.0800	7.6894	0.3580	28.1756
$\tau^{(\text{HC})}$	0.5671	307.8816	0.1989	1099.9935	5.0586	2276.8277
$\theta_{\text{new}}^{(\text{HC})}$	0.0043	0.3043	0.0903	1.8186	0.4139	6.5216

`bayesmeta` format. Source code to reproduce the results is available as Supporting Information on the journal's web page. (`http://onlinelibrary.wiley.com/doi/xxx/suppinfo`)

4 | RESULTS

In this section, we provide posterior inference, *S-I* values, and relative *qS-qI* estimates for parameters $\{\mu, \tau, \theta_{\text{new}}\}$ for MPR (Table 1) and for both DM24 and DM23 in Table 2. Here, we consider six scenarios that arise for three RLMC targets induced by HN(0.5), HN(1), and HC(1) heterogeneity priors crossed with two 50%-RLMC-tail adjusted heterogeneity priors (HN, HC). The Supporting Information extends the results of the main manuscript and considers several additional topics, which are reviewed in Section 4.3.

4.1 | Mean pocket reduction: Comparison of HN and HC heterogeneity priors

For historical MPR data ($k = 13$, Table 1), HN(0.5), HN(1), and HC(1) heterogeneity priors, which are recommended in the literature, induce the RLMC equal to 0.76, 0.93, and 0.96, respectively. Table 3 focuses on these target RLMC values and reports the base-scale parameter values of the 50%-RLMC-adjusted HN and HC heterogeneity priors. The densities of these HN and HC priors for τ are shown in the top row of Figure 1. It is evident that the spread of both HN and HC increases with increasing base-scale parameter values and target RLMC values. The resulting posterior densities of μ (middle row of Figure 1) and τ (bottom row of Figure 1) induced by the HN and HC heterogeneity priors agree well across target 0.76, 0.93, and 0.96 RLMC values.

Table 4 complements these visual findings and provides posterior mean and standard deviation estimates for parameters $\{\mu, \tau, \theta_{\text{new}}\}$ together with the shortest 95% credible interval and the length we obtained for it in the six different scenarios. For both HN and HC, the length of 95%CIs for the main effect μ, the heterogeneity τ, and the predicted true effect in a new

TABLE 6 Mean pocket reduction $qS^{HN/HC}$-$qI^{HN/HC}$: quotients of sensitivity and identification estimates for HN with respect to HC defined in Equation (15) for parameters $\{\mu, \tau, \theta_{new}\}$ in Bayesian NNHMs across three RLMC targets 0.76, 0.93, and 0.96 with 50%-RLMC-adjusted HN and HC heterogeneity priors. See Table 3 for the corresponding base-scale parameter values of HN and HC heterogeneity priors

	RLMC = 0.76		RLMC = 0.93		RLMC = 0.96	
Par	$qS^{HN/HC}$	$qI^{HN/HC}$	$qS^{HN/HC}$	$qI^{HN/HC}$	$qS^{HN/HC}$	$qI^{HN/HC}$
μ	42.51	1.10	2.63	0.88	0.69	0.88
τ	6.11	1.53	11.99	3.08	13.85	3.41
θ_{new}	43.08	0.92	2.72	0.84	0.71	0.90

study θ_{new} increase with increasing RLMC. The 95%CIs are longer for HC than for HN for RLMC = 0.76 but are shorter across the 0.93 and 0.96 RLMC targets.

For the main effect μ and both the 0.76 and 0.93 RLMC targets, HC leads to low S values in Table 5. This property is reversed when RLMC = 0.96. The marginal posterior for τ is more affected by HN than by HC heterogeneity prior perturbations.

Note that for all 0.76, 0.93, and 0.96 RLMC targets, the marginal posterior for τ is highly affected by the σ_i-uncertainty for both HN and HC heterogeneity priors (Table 5). Moreover, the posteriors of all parameters $\{\mu, \tau, \theta_{new}\}$ are increasingly affected by σ_i-uncertainty as RLMC values increase.

To assess how the impact of the HN and HC heterogeneity priors differs, Equation (15) has been applied to S-I values of Table 5, leading to relative $qS^{HN/HC}$-$qI^{HN/HC}$ estimates in Table 6. For RLMC = 0.76, $qS^{HN/HC}$-$qI^{HN/HC}$ estimates are greater than 1, revealing that the impact of both the heterogeneity prior and the within-study σ_i standard deviation uncertainty on parameters $\{\mu, \tau, \theta_{new}\}$ is less pronounced for HC than for the HN heterogeneity prior. This relation for μ and θ_{new} is reversed for RLMC = 0.96 with $qS^{HN/HC}$-$qI^{HN/HC}$ estimates less than 1. For RLMC = 0.93, $qS^{HN/HC}$-$qI^{HN/HC}$ estimates show that posterior estimates of μ and θ_{new} obtained from HN are 2.63 and 2.72 times more sensitive than HC but they are less affected (0.88, 0.84) by σ_i standard deviation uncertainty. This indicates that for HN(1) and HC(0.67) corresponding to RLMC = 0.93, the advantages of the HN and the HC heterogeneity priors balance out and the decision as to which of the heterogeneity priors should be recommended for analysis of MPR is less clear-cut.

The S-I approach can be used to explore other RLMC targets such as, for example, 0.25, 0.5, and 0.75 in Section 4.1 of the Supporting Information. Note that Table 5 of the main manuscript and Table S9 of the Supporting Information together cover the range of RLMC targets between 0.25 and 0.96 and the range of base-scale parameters from 0.16 to 1.48 for HN and from 0.11 to 1 for HC. They show that HC is preferable for MPR data for a wide range of RLMC targets.

One advantage of the unified S-I approach is that it clearly demonstrates that depending on the RLMC value, the posterior estimates are affected to different extent by the heterogeneity prior uncertainty and the within-study standard deviation uncertainty. Moreover, it provides some guidance as to which Bayesian model specifications are beneficial. This can be particularly useful in cases where substantial information about the between-study heterogeneity (τ) is missing. For example, a practitioner concerned with the small sample sizes of individual studies and the validity of the corresponding within-study standard deviation estimates of historical MPR data (Table 1) could choose an HC(0.19) (RLMC = 0.5) heterogeneity prior for Bayesian analysis, because it renders the lowest I estimate for μ in Table S9 of the Supporting Information. On the other hand, when a practitioner prefers to use the HC(1) heterogeneity prior recommended in the literature (RLMC = 0.96 in Table 3), he/she should be aware that the validity of the within-study standard deviations in Table 1 is very important for this choice, because the posterior of μ is highly affected by the uncertainty in within-study standard deviation assumptions (Table 5). Moreover, he/she can now be aware that the use of an HN(1.48) rather than HC(1) heterogeneity prior would be recommended for this RLMC = 0.96 target (Table 5).

4.2 | Diabetes mellitus: Comparison of 24 and 23 studies

HN(0.5), HN(1), and HC(1) applied to DM24 data in Table 2 induced median RLMC values equal to 0.33, 0.67, and 0.82 (Table S16 of the Supporting Information). Note that these RLMC values differ from the RLMC values obtained for the same HN(0.5), HN(1), and HC(1) heterogeneity priors for MPR data analyzed in Section 4.1. For DM24 and DM23, we fixed target RLMC values at 0.33, 0.67, and 0.82 and adjusted the base-scale A_b parameters of HN and HC heterogeneity priors with the 50%-RLMC-based adjustment (Table S16 of the Supporting Information). There is a one-to-one correspondence

between base A_b scale parameters of HN and HC heterogeneity priors and RLMC targets. Larger A_b values are linked to larger RLMC and lead to prior density distributions for τ that demonstrate larger spread in Figures 2 and 3 (top row) of the Supporting Information. The resulting posterior distributions for μ (middle row) and τ (bottom row) agree well across 0.33, 0.67, and 0.82 RLMC targets. This visual impression is confirmed by stable posterior estimates reported in Table S17 of the Supporting Information.

Table S18 of the Supporting Information shows *S*-*I* estimates for both DM24 (upper batch) and DM23 (lower batch) data sets. Although both sample sizes $k = 23$ and 24 are large, the posterior for τ is sensitive to the within-study standard deviation uncertainty for both HN and HC heterogeneity priors. Removal of the large Study 20 leads to larger *S*-*I* estimates for DM23 than those for DM24, as indicated by $qS^{23/24}$-$qI^{23/24}$ estimates in Table S19 of the Supporting Information. This removal affects both *S* and *I* dimensions for the overall mean parameter μ but the *I* dimension is affected more. Posteriors of all parameters for DM23 are more affected by within-study standard deviation uncertainty than the parameters for DM24. In particular, the removal of Study 20 leads to posteriors for μ that are more than eight times more affected by within-study standard deviation uncertainty than posteriors for DM24. The sensitivity to heterogeneity prior uncertainty of posteriors induced by HC is more affected by the removal of Study 20 than the sensitivity of posteriors induced by HN.

For DM24, HC heterogeneity prior is of advantage only for the RLMC target equal to 0.33 (see $qS^{HN/HC}$-$qI^{HN/HC}$ estimates in Table S20 of the Supporting Information). Otherwise, the use of the HN heterogeneity prior induces posteriors that are less affected by the heterogeneity prior uncertainty and the within-study standard deviation uncertainty for both DM24 and DM23. Note that Tables S23– S25 of the Supporting Information based on 0.25, 0.5, and 0.75 RLMC targets provide similar results.

These results complement the analysis by Kumar et al. (2020) for both DM24 and DM23 data sets and equip them with *S*-*I* estimates. One advantage of the unified *S*-*I* approach is that we can explicitly quantify how the change of information allocation across individual studies after removal of one large study impacts the sensitivity of posteriors to both the heterogeneity prior uncertainty and the within-study standard deviation uncertainty on an equal footing. Moreover, if no substantial information about the between-study heterogeneity (τ) is available, we can recommend the use of HN(1) (RLMC = 0.67) and HN(1.48) (RLMC = 0.82) for the analysis of both DM24 and DM23 data sets.

4.3 | Summary of results provided in the Supporting Information

The Supporting Information extends the results of the main manuscript in several ways. First, it focuses on the grid of 0.25, 0.5, and 0.75 RLMC targets and demonstrates that the *S*-*I* approach can be anchored either on heterogeneity prior values known from the literature (main manuscript) or directly on RLMC targets. Numerical issues related to the choice of the RLMC perturbation *h* and the stability of estimates are addressed in Section 6 of the Supporting Information.

Second, Section 2.7 of the Supporting Information compares *S*-*I* with two other approaches to sensitivity assessment: the informal approach (Gustafson, 1996) and the epsilon-local sensitivity qS^{ϵ} measure (Roos et al., 2015). While the informal approach has no ability to quantify and compare sensitivity between different specifications of Bayesian meta-analyses, the epsilon-local sensitivity qS^{ϵ} measure focuses on data-to-no-data comparison and is therefore restricted to two parameters (μ and τ). In contrast, *S*-*I* quantifies the impact of base-scale parameter uncertainty of HN and HC heterogeneity priors and within-study standard deviation uncertainty on posterior estimates of all parameters $\{\mu, \tau, \theta_1, \ldots, \theta_k, \theta_{new}\}$ in Bayesian meta-analysis.

Third, Section 4.4 of the Supporting Information provides *S*-*I* estimates for one subgroup AGR2 based on $k = 2$ randomized studies and another subgroup AGR4 based on $k = 4$ nonrandomized studies. These *S*-*I* estimates warn that properties of randomized AGR2 and nonrandomized AGR4 data sets differ. This indicates that a more sophisticated model than the simple Bayesian NNHM would be better suited for the combined analysis of both disparate types of evidence. For example, the Bayesian generalized evidence synthesis (GES) based on a three-level hierarchical model, which is designed to synthesize disparate types of evidence, would be preferable for the analysis of the combined data set based on six observations (Prevost et al., 2000; Sutton & Abrams, 2001; Verde & Thomas, 2015; Verde, 2021).

Fourth, Sections 2.4 and 2.5 of the Supporting Information extend the *S*-*I* approach to Bayesian NtHM, which is useful when the assumption of the normality of random effects is violated. Section 4.1 of the Supporting Information extends the analysis of MPR data from Section 4.1 of the main manuscript and investigates whether using Bayesian NtHM has advantages over using Bayesian NNHM for this data set. Table S12 of the Supporting Information demonstrates that applying Bayesian NtHM to MPR data produces posterior estimates that are less affected by both the heterogeneity prior uncertainty and the within-study σ_i uncertainty than posteriors produced by Bayesian NNHM across 0.25, 0.5, and 0.75 RLMC

targets. Table S14 (lower part) confirms that for 0.25, 0.5, and 0.75 RLMC targets, the advantage of using HC rather than HN also applies to the Bayesian NtHM.

Finally, Section 5 of the Supporting Information extends one simulation design suggested by Hardy and Thompson (1998) to cover the range of cases, when the assumption of the normality of random effects is violated. This simulation study based on 1000 simulations for each scenario shows that *S-I* can correctly demonstrate that the Bayesian NtHM leads to lower *S-I* estimates than the Bayesian NNHM for both HN and HC heterogeneity priors.

5 | DISCUSSION

We showed that HC was preferable when summarizing the historical MPR data for the conventional periodontal therapy. These results demonstrate that historical data carry more information than what can be summarized by plain estimates, knowledge that provides some guidance on the choice of the heterogeneity prior for further analysis. The extensive use of all of the information provided by historical data is particularly relevant for areas of research for which obtaining patients is costly and where meta-analyses commonly contain small sample sizes. Thus, *S-I* estimates could support the application of common methods of learning from historical data (Neuenschwander et al., 2010; Schmidli et al., 2014; Neuenschwander & Schmidli, 2020), in evidence-based dentistry and beyond. More work is, however, necessary to determine if *S-I* estimates could also help judge whether mean difference or standardized mean difference is better suited to analyze continuous outcomes in applications (Hedges, 1981; Sedgwick & Marston, 2013; White & Thomas, 2005).

The proposed two-dimensional *S-I* system for sensitivity comparisons is designed for the conventional NNHM model for aggregated data, which is widely used in applications (Normand, 1999; Gelman et al., 2014; Hoaglin, 2015; Jackson et al., 2018). Note that the accuracy of *S-I* estimates depends on how well the NNHM assumptions match the data at hand, and this match is expected to be better for continuous outcomes than for noncontinuous ones. Although Spiegelhalter et al. (2004) argue that the normal approximation to log-OR scale is reasonable, it should be noted that *S-I* values for binary DM and AGR data are, in fact, an approximation.

We considered the DM24 data set dealing with the association of diabetes mellitus and the severe course of COVID-19, explicitly quantifying the impact on *S-I* estimates of omitting one large study in DM23. Moreover, we showed that applying an HN heterogeneity prior for the analysis of DM24 and DM23 has the advantage of producing posterior estimates that are less affected both by the heterogeneity prior and by the within-study σ_i uncertainty. Although the advantage of using HN rather than HC was clear for DM24 and DM23, HC was preferable when analyzing MPR. These results show that the HC heterogeneity prior (Gelman, 2006; Polson & Scott, 2012; Gelman et al., 2017) is a valid alternative to HN, which is often recommended in the literature (Bender et al., 2018; Friede et al., 2017a, 2017b). Future work aims to determine under which specific conditions HC and HN are beneficial in applications.

At first glance, the method for taking derivatives of the BC with respect to RLMC perturbations, resulting in a unified, two-dimensional *S-I* measure for the Bayesian NNHM, may seem complicated. However, we are aiming for a broadly applicable, unified, two-dimensional method, independent of the scale of the primary outcome, and achieving such general applicability necessitates the use of these mathematical tools. Taking derivatives of a distance between two posterior distributions is a well-known tool used in Bayesian robustness (Dey & Birmiwal, 1994; Gustafson, 2000; McCulloch, 1989). The particular choice of the BC distance is motivated by its boundedness, which renders better numerical properties than those of other known distances, such as the unbounded Kullback–Leibler distance (Jeffreys, 1961; Roos & Held, 2011; Roos et al., 2015). RLMC transforms the relation between the between-study heterogeneity and the within-study standard deviations into a dimensionless bounded value, independent of the scale of the primary outcome. Moreover, RLMC controls the flow of mass between the between-study heterogeneity and the within-study standard deviations. These properties make RLMC a valid basis for the unified, two-dimensional *S-I* measure that assesses the impact of both the heterogeneity prior and the within-study standard deviations on the posterior results.

The unified *S-I* approach proposed is similar in spirit to the approach taken by Hardy and Thompson (1998) and Gustafson et al. (2006), who show that the between-study heterogeneity τ can only be well understood when the within-study standard deviation σ_i values are taken into account. In our opinion, the mutual interplay between the heterogeneity τ and the within-study standard deviation σ_i cannot be easily disentangled, so we choose to simultaneously quantify the impact of τ and σ_i. However, the nature of the sensitivity analysis (based on prior perturbations) S is very different from that of the identification analysis (based on perturbation of within-study standard deviation σ_i values) I. While the sensitivity analysis can guide in specifying the more appropriate prior on the heterogeneity parameter, the within-study standard deviation σ_i values are provided by the data and cannot be modified. Nonetheless, the identification analysis is

important, because it serves as a reminder that the within-study standard deviation σ_i values are not true values but rather uncertain estimates, and this uncertainty can also affect posteriors.

One advantage of the unified *S-I* approach is that *S-I* estimates facilitate direct comparisons of sensitivity estimates induced by different model specifications on an equal footing. *S-I* estimates can be easily transformed into relative *qS-qI* estimates based on quotients. These quotients compare two specifications of the Bayesian meta-analysis and determine which of them leads to posteriors that are less affected by uncertainty in the prior choice and the uncertainty of the within-study standard deviation values. These comparisons may include different heterogeneity prior distributions (HN and HC), data with different allocation of information across individual studies after removal of one large study (k and $k-1$), different latent model assumptions (NNHM and NtHM), and subgroups of randomized and nonrandomized studies. Other comparisons can be designed depending on the scientific question of interest.

Although NNHM is popular among researchers, the assumption of the normality of random effects is sometimes violated in practice. Therefore, we extended the unified *S-I* approach to cover Bayesian NtHM fitted by MCMC in `JAGS`. To compute the corresponding *S-I* estimates, `JAGS` runs models several times based on the number of chains, the length of burn-in, and thinning prespecified by the user and does not conduct any convergence analysis. Because these computations can be time-consuming, more work is necessary to speed up procedures based on MCMC samples. Another approach to reduce the time necessary for computation of *S-I* estimates in Bayesian NtHM would be the implementation of the Bayesian NtHM in `INLA` (Martins & Rue, 2014; Martins et al., 2013; Rue et al., 2009, 2017). This approach is promising because procedures based on Bayesian numerical approximation, such as, for example, `bayesmeta` for Bayesian NNHM, require less computational time and postprocessing, and fewer specifications to fit the model.

The unified *S-I* approach could be further extended in at least two ways. One promising extension would be the application of the *S-I* approach to the Bayesian GES based on a three-level hierarchical model fitted by MCMC (Prevost et al., 2000; Sutton & Abrams, 2001). This Bayesian GES considers one between-study heterogeneity, two type-specific between-study heterogeneities, and three heterogeneity priors. Although a generalized RLMC can be derived and systematic perturbations of data (*I*-dimension) and three heterogeneity priors (*S*-dimension) can be designed for this Bayesian GES, more work is necessary to implement these systematic perturbations in `R` and to provide the full *S-I* functionality for users of this Bayesian GES. A successful implementation would provide a basis for further extensions to the modern Bayesian bias-corrected meta-analysis (Verde, 2021).

Another promising extension would be the use of study-specific factors for within-study standard deviation perturbation. Currently, likelihood perturbations are tuned by a factor f, which produces the same percent change on each within-study standard deviation. In applications, different within-study standard deviation values are quite common, such as, for example, when evidence of many small studies is combined with a few large studies (Hardy & Thompson, 1998). To better reflect such cases, our method could be extended to allow more complex perturbations that tune various within-study standard deviation values differently. However, allocation of different tuning factors to each study would be very complicated, because endless different combinations of tuning could be considered (Hardy & Thompson, 1998). In particular, general rules should be developed for choosing studies that require such specific tuning. This could be done based on a threshold of, for example, 50% contribution of a specific study to the total information (Hardy & Thompson, 1998) or based on the type of evidence, such as randomized or nonrandomized study (Prevost et al., 2000; Sutton & Abrams, 2001). More work is necessary to implement this new idea in its full generality in `R`.

The method proposed is provided as an `R` package `si4bayesmeta` openly accessible on R-Forge. We believe that this software will facilitate the application of the unified approach to sensitivity and identification quantification in a Bayesian meta-analysis in a wide range of contexts and in many research fields.

ACKNOWLEDGMENT

We thank the Editor, the Associate Editor, and two anonymous reviewers for their constructive comments and suggestions, which considerably extended the original focus of the manuscript. We also thank Kimberly Lewis for English proofreading the manuscript. Support by the Swiss National Science Foundation (no. 175933) granted to Małgorzata Roos is gratefully acknowledged.

CONFLICT OF INTEREST

The authors have declared no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data shown in Tables 1–2 and used to demonstrate the applicability of the *S-I* framework have already been analyzed by Zaugg et al. (2014) and Kumar et al. (2020). The software for *S-I* computation is bundled in an `R` package called `si4bayesmeta` and is freely accessible on GitHub (`https://github.com/hunansona/si4bayesmeta`). Source code to reproduce the results is available as Supporting Information on the journal's web page (`http://onlinelibrary.wiley.com/doi/10.1002/bimj.202000193/suppinfo`).

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section.

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

ORCID

Małgorzata Roos https://orcid.org/0000-0001-9878-6969
Haakon Bakka https://orcid.org/0000-0001-8272-865X
Håvard Rue https://orcid.org/0000-0002-0222-1881

REFERENCES

Bender, R., Friede, T., Koch, A., Kuss, O., Schlattmann, P., Schwarzer, G., & Skipka, G. (2018). Methods for evidence synthesis in the case of very few studies. *Research Synthesis Methods*, *9*, 382–392.

Berger, J., & Berliner, L. M. (1986). Robust Bayes and empirical Bayes analysis with ϵ-contaminated priors. *The Annals of Statistics*, *14*, 461–486.

Berger, J. O., Ríos Insua, D., & Ruggeri, F. (2000). Bayesian robustness. In D. R. Insua & F. Ruggeri (Eds.), *Robust Bayesian analysis* (pp. 1–32). Springer-Verlag.

Bjordal, J. M., Ljunggren, A. E., Klovning, A., & Slørdal, L. (2004). Non-steroidal anti-inflammatory drugs, including cyclo-oxygenase-2 inhibitors, in osteoarthritic knee pain: Meta-analysis of randomised placebo controlled trials. *British Medical Journal*, *329*, 1317–1320.

de Carvalho, M., Page, G. L., & Barney, B. J. (2019). On the geometry of Bayesian inference. *Bayesian Analysis*, *14*, 1013–1036.

Crins, N. D, Röver, C., Goralczyk, A. D., & Friede, T. (2014). Interleukin-2 receptor antagonists for pediatric liver transplant recipients: A systematic review and meta-analysis of controlled studies. *Pediatric Transplantation*, *18*, 839–850.

Davey, J., Turner, R. M., Clarke, M. J., & Higgins, J. P. (2011). Characteristics of meta-analyses and their component studies in the Cochrane Database of Systematic Reviews: A cross-sectional, descriptive analysis. *BMC Medical Research Methodology*, *11*, 1–11.

Dey, D. K., & Birmiwal, L. R. (1994). Robust Bayesian analysis using divergence measures. *Statistics & Probability Letters*, *20*, 287–294.

Efron, B., & Hastie, T. (2016). *Computer age statistical inference*. Cambridge University Press.

Friede, T., Röver, C., Wandel, S., & Neuenschwander, B. (2017a). Meta-analysis of few small studies in orphan diseases. *Research Synthesis Methods*, *8*, 79–91.

Friede, T., Röver, C., Wandel, S., & Neuenschwander, B. (2017b). Meta-analysis of two studies in the presence of heterogeneity with applications in rare diseases. *Biometrical Journal*, *59*, 658–671.

GBD. (2016). Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: A systematic analysis for the Global Burden of Disease Study 2015. *Lancet*, *388*, 1545–1602.

Gelfand, A. E., & Sahu, S. K. (1999). Identifiability, improper priors, and Gibbs sampling for generalized linear models. *Journal of the American Statistical Association*, *94*, 247–253.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (Comment on article by Browne and Draper). *Bayesian Analysis*, *1*, 515–534.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). Chapman & Hall/CRC Press.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gelman, A., Simpson, D., & Betancourt, M. (2017). The prior can often only be understood in the context of the likelihood. *Entropy*, *19*, 555.

Günhan, B. K., Röver, C., & Friede, T. (2020). Random-effects meta-analysis of few studies involving rare events. *Research Synthesis Methods*, *11*, 74–90.

Gustafson, P. (1996). Robustness considerations in Bayesian analysis. *Statistical Methods in Medical Research*, *5*, 357–373.

Gustafson, P. (2000). Local robustness in Bayesian analysis. In D. R. Insua & F. Ruggeri (Eds.), *Robust Bayesian analysis* (pp. 71–88). Springer-Verlag.

Gustafson, P. (2015). *Bayesian inference for partially identified models. exploring limits of limited data*. Chapman & Hall/CRC Press.

Gustafson, P., Hossain, S., & MacNab, Y. C. (2006). Conservative prior distributions for variance parameters in hierarchical models. *The Canadian Journal of Statistics*, *34*, 377–390.

Hardy, R. J., & Thompson, S. G. (1998). Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine*, *17*, 841–856.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*, 107–128.
Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*, 1539–1558.
Hjort, N. L., & Glad, I. K. (1995). Nonparametric density estimation with a parametric start. *The Annals of Statistics*, *23*, 882–904.
Hoaglin, D. C. (2015). We know less than we should about methods of meta-analysis. *Research Synthesis Methods*, *6*, 287–289.
IntHout, J., Ioannidis, J. P., Borm, G. F., & Goeman, J. J. (2015). Small studies are more heterogeneous than large ones: A meta-meta-analysis. *Journal of Clinical Epidemiology*, *68*, 860–869.
Jackson, D., Law, M., Stijnen, T., Viechtbauer, W., & White, I. R. (2018). A comparison of seven random-effects models for meta-analyses that estimate the summary odds ratio. *Statistics in Medicine*, *37*, 1059–1085.
Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.
Jiang, J. (2013). The subset argument and consistency of MLE in GLMM: Answer to an open problem and beyond. *The Annals of Statistics*, *41*, 177–195.
Johnson, N. L., Kotz, S., & Balkrishnan, N. (1994). *Continuous univariate distributions. Volume 1* (2nd ed.). John Wiley & Sons.
Kass, R. E., & Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, *90*, 928–934.
Kumar, A., Arora, A., Sharma, P., Anikhindi, S. A., Bansal, N., Singla, V., Khare, S., & Srivastava, A. (2020). Is diabetes mellitus associated with mortality and severity of COVID-19? A meta-analysis. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, *14*, 535–545.
Lele, S. R., Nadeem, K., & Schmuland, B. (2010). Estimability and likelihood inference for generalized linear mixed models using data cloning. *Journal of the American Statistical Association*, *105*, 1617–1625.
Mardia, K. V, Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. Academic Press.
Martins, T. G., & Rue, H. (2014). Extending integrated nested Laplace approximation to a class of near Gaussian latent models. *Scandinavian Journal of Statistics*, *41*, 893–912.
Martins, T. G., Simpson, D., Lindgren, F., & Rue, H. (2013). Bayesian computing with INLA: New features. *Computational Statistics & Data Analysis*, *67*, 68–83.
McCulloch, R. E. (1989). Local model influence. *Journal of the American Statistical Association*, *84*, 473–478.
Neuenschwander, B., Capkun-Niggli, G., Branson, M., & Spiegelhalter, D. J. (2010). Summarizing historical information on controls in clinical trials. *Clinical Trials*, *7*, 5–18.
Neuenschwander, B., & Schmidli, H. (2020). Use of histarical data. In E. Lesaffre, G. Baio, & B. Boulanger (Eds.), *Bayesian methods in pharmaceutical research* (pp. 111–137). Chapman & Hall/CRC Press.
Normand, S. T. (1999). Meta-analysis: Formulating, evaluating, combining, and reporting. *Statistics in Medicine*, *18*, 321–359.
Ott, M., Hunanyan, S., Held, L. & Roos, M. (2021). Sensitivity quantification in Bayesian meta-analysis. Statistical Methods in Medical Research, in preparation.
Plummer, M. (2016). JAGS: Just another Gibbs sampler. http://mcmc-jags.sourceforge.net.
Polson, N. G., & Scott, J. G. (2012). On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis*, *7*, 887–902.
Prevost, T. C., Abrams, K. R., & Jones, D. R. (2000). Hierarchical models in generalized synthesis of evidence: An example based on studies of breast cancer screening. *Statistics in Medicine*, *19*, 3359–3376.
Ríos Insua, D., Ruggeri, F., & Martín, J. (2000). Bayesian sensitivity analysis. In A. Saltelli, K. Chan, & E.M. Scott (Eds.), *Sensitivity Analysis* (pp. 225–244). John Wiley & Sons.
Roos, M., & Held, L. (2011). Sensitivity analysis in Bayesian generalized linear mixed models for binary data. *Bayesian Analysis*, *6*, 259–278.
Roos, M., Martins, T., Held, L., & Rue, H. (2015). Sensitivity analysis for Bayesian hierarchical models. *Bayesian Analysis*, *10*, 321–349.
Röver, C. (2020). Bayesian random-effects meta-analysis using the bayesmeta R package. *Journal of Statistical Software*, *93*, 1–51.
Röver, C., & Friede, T. (2020). Dynamically borrowing strength from another study through shrinkage estimation. *Statistical Methods in Medical Research*, *29*, 293–308.
Rue, H., Martino, S., & 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*, 319–392.
Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., & Lindgren, F. K. (2017). Bayesian computing with INLA: A review. *Annual Review of Statistics and Its Applications*, *4*, 395–421.
Schmidli, H., Gsteiger, S., Roychoudhury, S., O'Hagan, A., Spiegelhalter, D. J., & Neuenschwander, B. (2014). Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*, *70*, 1023–1032.
Sedgwick, P., & Marston, L. (2013). Meta-analyses: Standardised mean differences. *British Medical Journal*, *347*, f7257.
Smith, T. C., Spiegelhalter, D. J., & Thomas, A. (1995). Bayesian approaches to random-effects meta-analysis: A comparative study. *Statistics in Medicine*, *14*, 2685–2699.
Spiegelhalter, D. J., Abrams, K. R., & Myles, J. P. (2004). *Bayesian approaches to clinical trials and health-care evaluation*. John Wiley & Sons.
Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, *64*, 583–616.
Sutton, A. J., & Abrams, K. R. (2001). Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research*, *10*, 277–303.
Turner, R. M., Davey, J., Clarke, M. J., Thompson, S. G., & Higgins, J. P. (2012). Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *International Journal of Epidemiology*, *41*, 818–827.
Verde, P. E. (2021). A bias-corrected meta-analysis model for combining, studies of different types and quality. *Biometrical Journal*, *63*, 406–422.

Verde, P. E., & Ohmann, C. (2015). Combining randomized and non-randomized evidence in clinical research: A review of methods and applications. *Research Synthesis Methods*, *6*, 45–62.

White, I. R., & Thomas, J. (2005). Standardized mean differences in individually-randomized and cluster-randomized trials, with applications to meta-analysis. *Clinical Trials*, *2*, 141–151.

Xie, Y., & Carlin, B. P. (2006). Measures of Bayesian learning and identifiability in hierarchical models. *Journal of Statistical Planning and Inference*, *136*, 3458–3477.

Zaugg, B., Sahrmann, P., Roos, M., Attin, T., & Schmidlin, P. R. (2014). Improving scaling and root planing over the past 40 years: A meta-analysis. *Dentistry*, *4*, 1–5.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Roos, M., Hunanyan, S., Bakka, H., & Rue, H. (2021). Sensitivity and identification quantification by a relative latent model complexity perturbation in Bayesian meta-analysis. *Biometrical Journal*, *63,* 1555–1574. https://doi.org/10.1002/bimj.202000193