Received 26 April 2016; | Revised 6 June 2016; | Accepted 6 June 2016

DOI: xxx/xxxx

ARTICLE TYPE

A Bayesian Fixed Effects Approach to Meta-Analysis: Decoupling Exchangeability and Hierarchical Modelling

Clara Domínguez Islas*[1] | Kenneth M. Rice[2]

[1]Fred Hutchinson Cancer Research Center, Vaccine and Infectious Disease Division, Seattle, WA, US
[2]Department of Biostatistics, University of Washington, WA, US

Present Address
*Clara Domínguez Islas, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N., M2-C200, PO Box 19024, Seattle, WA 98109-1024. Email: cdomingu@fredhutch.org

In meta-analysis, Bayesian methods seem a natural choice for the combination of sources of evidence. However, their sensitivity in practice to choice of prior distribution is much less attractive, particularly for parameters describing heterogeneity. A recent non-Bayesian approach to fixed-effects meta-analysis provides novel ways to think about estimation of both fixed-effect average effects and variability around this average. In this paper we describe the Bayesian analogs of those results, showing how Bayesian inference on fixed-effects parameters—properties of the study populations at hand—is more stable and less sensitive than standard methods. As well as these practical insights, our development also clarifies different ways in which prior beliefs like homogeneity and correlation can be reflected in prior distributions. We also carefully distinguish how random-effects models can be used to reflect sampling uncertainty versus their use reflecting *a priori* exchangability of study-specific effects, and how subsequent inference depends on which motivation is used. Throughout the paper, examples are drawn from real examples as well as simulation and theory, and the impact of the analysis choices is seen even when meta-analyzing small numbers of studies.

KEYWORDS:
meta-analysis, fixed effects, random effects, exchangeability, heterogeneity

1 | INTRODUCTION

In recent years, meta-analysis has proven to be an extremely useful tool for summarizing and synthesizing results from different studies. Two main approaches have been historically used to address the different questions posed in a meta-analysis: the fixed and random effects approaches. Extensive literature is available not only describing these approaches, but also attempting to provide guidelines for choosing one or the other.[1,2]

An important difference between fixed and random effect approaches is the type of statistical inference that they enable.[3,2] In the random effects approach the inference goal is usually to characterize a distribution of effect sizes —perhaps a hypothetical distribution—from which the effect sizes observed in the studies at hand are assumed to be sampled. From a Bayesian perspective, this approach can also be motivated from the belief that the effects (not the studies themselves) are exchangeable or that their magnitudes cannot be differentiated *a priori*.[4] However, targeting inference to hyper-parameters can be problematic, as we often have little information on them.[5] This is typical when the number of studies included is small: the classical frequentist estimates can behave poorly,[6,7] while Bayesian inference can be correspondingly very sensitive to the choice of prior distribution.[8,9]

In contrast, the fixed effects approach, based on the assumption that the effects estimated in the different studies are unknown but fixed quantities, allows for inference that is conditional on the studies at hand, with no random distribution of effect sizes.[10] Although some authors advocate for the use of fixed effects only when the effects are assumed to be the same, other authors advocate for its use more generally, providing inference that is conditional on the studies at hand.[11,2]

In recent work,[12] the fixed effects approach to meta-analysis was motivated as an optimal estimation problem: specifically, that the inverse variance weighted average of the study specific effects, in addition to being an interpretable parameter in its own right,[13] is the weighted average that can be most precisely estimated in frequentist analyses. In Bayesian analysis the interpretability of the fixed effects parameter carries over directly, and the precision property strongly suggests that any problems of its prior sensitivity should be smaller than for other potential target parameters. In this paper we show how this intuition does indeed apply, providing stable inference where more standard approaches are driven heavily by the prior. We show this not only for the fixed effects location parameter, but also for measures of heterogeneity.

This paper is structured as follows: in Section 2 we briefly review a data-adaptive fixed effects approach to meta-analysis. In Section 3 we propose a Bayesian analogue to this approach and discuss the use of a family of conjugate multivariate priors. We provide a closed-form expression for the posterior distribution of the parameters of interest, whose properties are also explored. In Section 4, we reconcile this approach with the multilevel hierarchical model used in classic Bayesian random effects approach, allowing simultaneous estimation parameters for both conditional and un-conditional inference. Lastly, in Section 5 we use a previously published meta-analysis as an example to evaluate the posterior distribution of the proposed parameters, comparing it to the distribution of the parameters typically targeted in a random effects Bayesian approach, illustrating and contrasting their properties.

2 | REVIEW OF FIXED EFFECTS META-ANALYSIS

We denote $\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k$ as point estimates of the true effect size parameters $\beta_1, \beta_2, ...\beta_k$ from k studies included in a meta-analysis, and let $\sigma_1^2, \sigma_2^2, ..., \sigma_k^2$ be the variances of these estimates. The aim of a fixed effects analysis is to estimate an average of the true effects, like the simple mean or the precision weighted average. In the specific case in which all the effect sizes are the same, any weighted average of the estimates would estimate that true common effect, with the precision weighted average being the most precise.

From all possible weighted averages to which we could target our inference, the inverse-variance weighted average is statistically optimal, if the variances are known, as it is the weighted average for which the corresponding estimator has the minimum variance[12]. Using the same notation, we denote the inverse-variance weighted average by β_F:

$$\beta_F = \frac{\sum_{i=1}^k \frac{1}{\sigma_i^2}\beta_i}{\sum_{i=1}^k \frac{1}{\sigma_i^2}}, \tag{1}$$

Complementing the summary of the overall location of the effects, as given by β_F, a parameter that quantifies the amount of heterogeneity among the effects sizes is the weighted average of the squared squared deviations of each effect from β_F, which we denote as ζ^2:

$$\zeta^2 = \frac{\sum_{i=1}^k \frac{1}{\sigma_i^2}(\beta_i - \beta_F)^2}{\sum_{i=1}^k \frac{1}{\sigma_i^2}} = \frac{\sum_{i=1}^k \frac{1}{\sigma_i^2}\beta_i^2}{\sum_{i=1}^k \frac{1}{\sigma_i^2}} - \beta_F^2. \tag{2}$$

Writing $\sigma_i^2 = (n_i\phi_i)^{-1}$, where ϕ_i is the amount of information about β_i that provided by each of the n_i observations in study i, (1) and (2) can be expressed as:

$$\begin{aligned}
\beta_F &= \frac{\sum_{i=1}^k n_i\phi_i\beta_i}{\sum_{i=1}^k n_i\phi_i} = \frac{\sum_{i=1}^k \eta_i\phi_i\beta_i}{\sum_{i=1}^k \eta_i\phi_i} \\
\zeta^2 &= \frac{\sum_{i=1}^k n_i\phi_i(\beta_i - \beta_F)^2}{\sum_{i=1}^k n_i\phi_i} = \frac{\sum_{i=1}^k \eta_i\phi_i(\beta_i - \beta_F)^2}{\sum_{i=1}^k \eta_i\phi_i},
\end{aligned} \tag{3}$$

where $\eta_i = n_i / \sum_{i'} n_{i'}$ is the proportion of observations in study i relative to the total number of observations from all studies in the meta-analysis.

Several frequentist estimation methods for β_F and ζ^2 have been proposed, under the standard large-sample assumption that the $\hat{\beta}_i$ are jointly Normal, i.e. that k-variate ($\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \boldsymbol{\Sigma})$), with a diagonal variance-covariance matrix $\boldsymbol{\Sigma}$ =diag$\{\sigma_i^2\}$. Table 1 summarizes estimators for both parameters in which the common assumption of known variances is made. We note that, using the same notation as above, the variance of $\hat{\beta}_F$ is inversely proportional to $\Phi = \sum_{i=1}^k n_i \phi_i = N \sum_{i=1}^k \eta_i \phi$, which we refer to as the *total amount of information*.

TABLE 1 Estimators for the fixed effects meta-analysis location and heterogeneity summary parameters

Parameter	Estimator	Confidence interval (1-α)
β_F	$\hat{\beta}_F = \frac{\sum_{i=1}^k \frac{1}{\sigma_i^2}\hat{\beta}_i}{\sum_{i=1}^k \frac{1}{\sigma_i^2}}$	Normal approximation:
.		$\hat{\beta}_F \pm z_{\alpha/2}\sqrt{\Phi^{-1}}$, with $\Phi = \sum_{i=1}^k \frac{1}{\sigma_i^2}$
ζ^2	$\hat{\zeta}^2 = \max\left\{0, \frac{Q-(k-1)}{\Phi}\right\}$,	Inverted non-central χ^2 probability interval:
	with $Q = \sum_{i=1}^k \frac{1}{\sigma_i^2}(\hat{\beta}_i - \hat{\beta}_F)^2$	$\{\zeta^2 : \chi^2_{k-1,\alpha/2}(\Phi\zeta^2) \le Q \le \chi^2_{k-1,1-\alpha/2}(\Phi\zeta^2)\}$

Methods that further relax the assumption of known σ_i^2 are proposed and discussed elsewhere .[12] These methods take into account the uncertainty in the estimation of the study variances (σ_i^2), which has a greater impact on the estimators from Table 1 when there is significant heterogeneity in the study effects.

3 | BAYESIAN FIXED EFFECTS META-ANALYSIS

With the standard Normality assumptions on the $\hat{\beta}_i$, the fixed effects setting also permits Bayesian methods when a prior is specified on the unknown parameters. Here we shall denote these in matrix notation, as parameter vector $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_k)$ and with

$$\begin{aligned} \beta_F &= \mathbf{1}_k^T \boldsymbol{W}\boldsymbol{\beta}, \\ \zeta^2 &= \boldsymbol{\beta}^T \left(\boldsymbol{W} - \boldsymbol{W}\mathbf{1}_k\mathbf{1}_k^T\boldsymbol{W}\right)\boldsymbol{\beta}, \end{aligned} \tag{4}$$

where $\mathbf{1}_k^T$ denotes the unit vector of length k and $\boldsymbol{W} = (\mathbf{1}_k^T\boldsymbol{\Sigma}^{-1}\mathbf{1}_k)^{-1}\boldsymbol{\Sigma}^{-1}$, with $\boldsymbol{\Sigma}$ =diag$\{\sigma_i^2\}$. Expressing the target parameters β_F and ζ^2 as a linear combination and a quadratic form in the β_i, emphasizes that, however it is specified, the joint prior on the β_i induces a prior on the target parameters, automatically.

Specifying a joint prior distribution on $\boldsymbol{\beta}$ not only allows the analysis to incorporate prior beliefs or information on each individual effect-size parameter, but also determines how much information each study's effect-size provides on the others. A multivariate Normal prior, for example, can be used to reflect beliefs on the particular location of each effect-size (the mean vector parameter), the uncertainty in these beliefs for each study (the variances) and on how related the effect-sizes from different studies are though to be (the pairwise correlation coefficients). In the simple case of a these correlations all being equal – a consequence of assuming exchangeability[14] – by varying the value of the correlation from 0 to 1, we can specify a range of scenarios, from believing that study effects are totally uninformative about each other (fixed effects) to believing that they each parameter tells us as much about any other parameter as about itself – for example when assuming that all study effect-sizes are exactly the same (common effect).

In what follows we will use both conjugate priors, that provide closed-form posteriors that can be inspected analytically, and also non-conjugate choices. Given modern computational methods, specifically Markov Chain Monte Carlo (MCMC) algorithms[15] and the availability of off-the-shelf software[16,17], the calculations for Bayesian meta-analyses are straightforward in practice, allowing us to focus directly on impact of different assumptions on subsequent inference.

3.1 | Multivariate Normal conjugate prior

We now present analytical results obtained using a conjugate prior distribution and discuss some properties of the corresponding posterior distribution.

Under a fixed effects approach with the standard large-sample assumption of Normality of the effects estimates $\hat{\beta}_i$, the multivariate Normal prior is conjugate for the parameter vector $\boldsymbol{\beta}$, with the posterior distribution of $\boldsymbol{\beta}$ conditional on the data $\hat{\boldsymbol{\beta}}$ also being a multivariate Normal [15]. Letting the prior for $\boldsymbol{\beta}$ be k-variate normal with mean vector $\boldsymbol{\nu}$ and variance-covariance matrix $\boldsymbol{\Upsilon}$, then:

$$\boldsymbol{\beta}|\hat{\boldsymbol{\beta}} \sim N_k\left((\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Upsilon}^{-1})^{-1}\left(\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\beta}}+\boldsymbol{\Upsilon}^{-1}\boldsymbol{\nu}\right),(\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Upsilon}^{-1})^{-1}\right). \tag{5}$$

As a linear combination of a k-variate Normal distribution, the posterior distribution of β_F is also Normal:

$$\beta_F|\hat{\boldsymbol{\beta}} \sim N\left(\mathbf{1}_k^T\boldsymbol{W}(\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Upsilon}^{-1})^{-1}\left(\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\beta}}+\boldsymbol{\Upsilon}^{-1}\boldsymbol{\nu}\right),\mathbf{1}_k^T\boldsymbol{W}(\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Upsilon}^{-1})^{-1}\boldsymbol{W}\mathbf{1}_k\right). \tag{6}$$

The posterior distribution of ζ^2 is obtained by noting that ζ^2 is a quadratic form of the k-variate Normal vector $\boldsymbol{\beta}$, which has a Normal posterior. Closed form expressions for the poster mean and variance of ζ^2 can therefore be obtained using known results for quadratic forms [18, Chapter 2], while quantiles of the distribution can be obtained by expressing the posterior of ζ^2 as a weighted sum of non-central chi-square random variables, as shown in **(author?)** [19]. The corresponding calculations can be implemented with the `R` package `CompQuadForm` [19]: details are given in Appendix A.

3.2 | Exchangeable multivariate conjugate prior expressed as a hierarchical structured distribution (incorporating information on μ)

For the special case of *a priori* exchangeability of the β_i, the prior is a k-variate Normal with every element of the mean vector $\boldsymbol{\nu}$ being equal, prior variance for each β_i also equal, and an exchangeable correlation matrix $\boldsymbol{\Upsilon}_k(\rho) = (1-\rho)\boldsymbol{I}_k + \rho\mathbf{1}_{kk}$, where ρ denotes the correlation coefficient between any two β_i. This multivariate prior can also be expressed as a hierarchical prior, in which the study effects β_i are i.i.d. Normal with mean μ (a hyperparmeter) which is assigned a Normal prior. These two priors are shown in Table 2 , together with the transformations used to switch between them.

TABLE 2 Exchangeable multivariate Normal prior for $\boldsymbol{\beta}$ and its equivalent parameterization as a hierarchical structured model. The relevant transformations to switch between the two parameterizations are also given.

	Multivariate	Hierarchical
Prior	$\boldsymbol{\beta} \sim N_k(\nu\mathbf{1}_k, \xi^2\boldsymbol{\Upsilon}_k(\rho))$	$\beta_i\|\mu$ iid $N(\mu,\tau^2)$
	$\boldsymbol{\Upsilon}_k(\rho) = (1-\rho)\boldsymbol{I}_k + \rho\mathbf{1}_k\mathbf{1}_k^T$	$\mu \sim N(\nu,\psi^2)$
	$0 \le \rho \le 1, \xi^2 \ge 0$	$\tau^2 \ge 0, \psi^2 \ge 0$
Reparametrization	$\xi^2 = \tau^2 + \psi^2$	$\tau^2 = (1-\rho)\xi^2$
	$\rho = \psi^2/(\tau^2+\psi^2)$	$\psi^2 = \rho\xi^2$

We notice that both priors reflect the same belief of the study effects being exchangeable: explicitly in the multivariate model and implicitly in the hierarchical model. A careful comparison of both parameterizations allows us to better understand how prior beliefs can be incorporated. For example, prior beliefs on the average location of the study effects are incorporated through the hyper-parameter ν in both models. Prior beliefs on how similar the study effects are (homogeneity) can be explicitly stated in the value the correlation parameter ρ, with 1 corresponding to a prior assumption of perfect homogeneity while values close to 0 correspond to a prior assumption of unrelated effects. The same correlation can be implicitly incorporated in the hierarchical model as $\psi^2/(\tau^2+\psi^2)$, so that high values of ψ^2 relative to τ^2 induce a high correlation of the study effects. Lastly, the parameter ξ^2 in the multivariate normal is equivalent to the sum of the parameters τ^2 and ψ^2 from the hierarchical model, so we can interpret ξ^2 as a total variance due both to the heterogeneity of the effects and the uncertainty on their overall location.

The posterior distribution of β_F can be obtained from (6). But furthermore, by noticing that the total information ($\Phi = \sum_i n_i\phi_i = \sum_i \sigma_i^{-2}$) can be expressed as $\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}$, so that $\boldsymbol{\Sigma}^{-1} = \Phi\boldsymbol{W}$, then the posterior mean and variance of $\beta_F|\hat{\boldsymbol{\beta}}$ can be written as:

$$\mathrm{Var}[\beta_F|\hat{\boldsymbol{\beta}}] = \mathbf{1}_k^T \boldsymbol{W} \left(\Phi \boldsymbol{W} + \xi^{-2}\boldsymbol{\Upsilon}_k(\rho)^{-1}\right)^{-1} \boldsymbol{W}\mathbf{1}_k \tag{7}$$

$$\mathbb{E}[\beta_F|\hat{\boldsymbol{\beta}}] = \mathbf{1}_k^T \boldsymbol{W} \left(\Phi \boldsymbol{W} + \xi^{-2}\boldsymbol{\Upsilon}_k(\rho)^{-1}\right)^{-1} \left(\Phi \boldsymbol{W}\hat{\boldsymbol{\beta}} + \xi^{-2}\boldsymbol{\Upsilon}_k(\rho)^{-1}\mathbf{1}_k \nu\right) \tag{8}$$

From these expressions we can see that the posterior distribution of β_F is a Normal distribution with mean and variance that approach $\mathbf{1}_k^T \boldsymbol{W}\hat{\boldsymbol{\beta}} = \hat{\beta}_F$ and Φ^{-1}, respectively, as the total information ($\Phi = \sum n_i \phi_i$) increases or the total amount of variance due to heterogeneity and uncertainty ($\xi^2 = \tau^2 + \psi^2$) decreases. This means that with large sample sizes or in the absence of strong beliefs on the location of the study effects or their homogeneity, the Bayesian posterior mean of the precision weighted average reduces to the frequentist estimator, as given in Table 1 .

It is of interest to compare the properties of the Bayesian estimator of β_F to those of the estimator of hyperparameter μ, as this is usually the target of Bayesian random effects meta-analyses. Using the equivalences noted in Table 2 , the posteriors for these two parameters are

$$\beta_F|\hat{\boldsymbol{\beta}} \sim N\left(\sum_{i=1}^{k}\left(\frac{\sigma_i^{-2}(1-\lambda_i)}{\Phi}\right)\hat{\beta}_i + \left(\sum_{i=1}^{k}\frac{\sigma_i^{-2}\lambda_i}{\Phi}\right)\nu,\ \frac{1}{\Phi^2}\sum_{i=1}^{k}\sigma_i^{-2}(1-\lambda_i)\right)$$

$$\text{with} \qquad \lambda_i = \left(\frac{1}{1+\psi^2\sum_i \frac{1}{\sigma_i^2+\tau^2}}\right)\left(\frac{\sigma_i^2}{\sigma_i^2+\tau^2}\right), \tag{9}$$

and

$$\mu|\hat{\boldsymbol{\beta}} \sim N\left(\left(\frac{1}{\psi^2}+\sum_{i=1}^{k}\frac{1}{\sigma_i^2+\tau^2}\right)^{-1}\left(\frac{\nu}{\psi^2}+\sum_{i=1}^{k}\frac{\hat{\beta}_i}{\sigma_i^2+\tau^2}\right),\ \left(\frac{1}{\psi^2}+\sum_{i=1}^{k}\frac{1}{\sigma_i^2+\tau^2}\right)^{-1}\right). \tag{10}$$

See Appendix B for a detailed derivation.

First we notice that the posterior mean of β_F can be expressed as a weighted average of the k effect estimates ($\hat{\beta}_1, ..., \hat{\beta}_k$) and the mean of the prior (μ), meaning that it will never fall outside of the range of these values. The same is true for the posterior mean of μ. Also, we notice that the posterior variance of β_F is always less than Φ^{-1}, that is, for any values of ψ^2 and τ^2 in our prior, the precision of the Bayesian estimate of β_F is at least the same precision we get for the frequentist estimator. This is not true for μ, for which large values of ψ^2 and τ^2 in the prior will produce a posterior variance that can be arbitrarily larger than Φ^{-1}.

Table 3 shows the limit values of the posterior mean and variance of β_F and μ for some extreme values of ψ^2 and τ^2. The only case when the limit value of mean and variance is the same for both estimators is when using a prior reflecting almost perfect homogeneity ($\tau^2 \to 0$). In this case, the posterior distribution of both β_F and μ is a weighted average of the inverse-variance weighted average of the estimates ($\hat{\beta}_F$) and the prior mean (μ), with the weights depending on the information provided by the data (Φ) and the precision of the prior (ψ^2). In contrast, when using a prior that reflects large heterogeneity ($\tau^2 \to \infty$), the posterior distribution of μ approaches its prior distribution, with little or no influence from the data, while the posterior distribution of β_F approaches that of the corresponding frequentist estimator.

When using a diffuse prior for the overall location of the effects ($\psi^2 \to \infty$) the posterior distribution of μ approaches that of the frequentist estimator of μ under a random effects model with known between-studies variance. Its variance is always greater than Φ^{-1} and increases with τ^2. For a very small value of τ^2 the posterior mean and variance of μ approximate $\hat{\beta}_F$ and Φ^{-1}, respectively, while for a value of τ^2 sufficiently large relative to the study variances (σ_i^2), the posterior mean and variance of μ are approximately equal to $\sum_i \hat{\beta}_i/k$ and τ^2/k, respectively. That is, a diffuse prior for μ along with moderate to large heterogeneity will produce a Bayesian estimate of μ that approaches the unweighted average of the effect estimates with its precision directly depending on the number of studies and completely independent of the size or precision of such studies. This is not the case for β_F, for which the posterior distribution under a vague prior reduces again to that of the frequentist estimator, with the precision increasing with the total amount of information ($\Phi = \sum_i \sigma_i^{-2} = N\sum_i \eta_i\phi_i$). In other words, in the absence of strong prior beliefs, more precise estimates of the individual effect-size parameters will produce a more precise Bayesian estimation of β_F, but not of μ.

Finally, when using a very precise or informative prior ($\psi^2 \to 0$), the posterior distribution of β_F approaches that of a precision weighted average of the effect-size parameters, after each one being 'corrected' or 'shrunk' towards a Normal prior with mean ν and variance τ^2. The correction depends on the precision of the corresponding estimate (σ_i^2) relative to the precision of that prior (τ^2). We notice that the gain in precision of this estimate relative to the frequentist estimator depends on τ^2, the degree of homogeneity induced by the prior; greater gains will be obtained when τ^2 is small, that is, when the effect sizes are thought to

TABLE 3 Limit values of the mean and variance for the posterior normal distributions of μ and β_F, when using a hierarchical normal prior distribution as in Table 2 .

Prior distribution			Posterior distribution	
Belief	Value of hyper-parameter	Estimator	Mean	Variance
Homogeneity of the study effects	$\tau^2 \to 0$	$\beta_F\|\hat{\boldsymbol{\beta}}$	$\frac{\Phi\hat{\beta}_F+\psi^{-2}\nu}{\Phi+\psi^{-2}}$	$\Phi^{-1}\left(\frac{\Phi}{\Phi+\psi^{-2}}\right)$
		$\mu\|\hat{\boldsymbol{\beta}}$	$\frac{\Phi\hat{\beta}_F+\psi^{-2}\nu}{\Phi+\psi^{-2}}$	$\Phi^{-1}\left(\frac{\Phi}{\Phi+\psi^{-2}}\right)$
Large heterogeneity of the study effects	$\tau^2 \to \infty$	$\beta_F\|\hat{\boldsymbol{\beta}}$	$\hat{\beta}_F$	Φ^{-1}
		$\mu\|\hat{\boldsymbol{\beta}}$	ν	ψ^2
Informative prior for the location of study effects	$\psi^2 \to 0$	$\beta_F\|\hat{\boldsymbol{\beta}}$	$\frac{1}{\Phi}\sum_i \sigma_i^{-2}\left(\frac{\tau^2\hat{\beta}_i+\sigma_i^2\nu}{\sigma_i^2+\tau^2}\right)$	$\frac{1}{\Phi^2}\sum \sigma_i^{-2}\left(\frac{\tau^2}{\sigma_i^2+\tau^2}\right)$
		$\mu\|\hat{\boldsymbol{\beta}}$	ν	0
Vague prior for the location of study effects	$\psi^2 \to \infty$	$\beta_F\|\hat{\boldsymbol{\beta}}$	$\hat{\beta}_F$	Φ^{-1}
		$\mu\|\hat{\boldsymbol{\beta}}$	$\frac{\sum(\sigma_i^2+\tau^2)^{-1}\hat{\beta}_i}{\sum(\sigma_i^2+\tau^2)^{-1}}$	$\frac{1}{\sum(\sigma_i^2+\tau^2)^{-1}}$

be similar and are then allowed to 'borrow strength' from each other. In contrast, the posterior distribution of μ will again be close to the prior distribution when $\psi^2 \to 0$, with little influence of the data.

To better understand the difference between the posterior distributions of β_F and μ, we look further at their means. From (10) we can see that the weight for ν in $\mathrm{E}(\mu|\hat{\boldsymbol{\beta}})$ is ψ^{-2} and the weight for each $\hat{\beta}_i$ is $(\sigma_i^2+\tau^2)^{-1}$. On the other hand, from (9), the weights for ν and each of the β_i in $\mathrm{E}(\beta_F|\hat{\boldsymbol{\beta}})$ can be expressed as being proportional to ψ^{-2} and $(\sigma_i^2+\tau^2)^{-1}\{1+\tau^2\sigma_i^{-2}[1+\psi^{-2}(\sum_i \frac{1}{\sigma_i^2+\tau^2})^{-1}]\}$, respectively. We notice then that more weight is given to the effect size estimates $\hat{\beta}_i$ in the posterior mean of β_F than in the posterior mean of μ. In this sense, we can say that the posterior for β_F is 'closer' to the data than the posterior for μ.

In summary, we can say that priors reflecting certainty and homogeneity about the β_i do influence the posterior distribution of β_F, giving a gain in precision relative to the estimation obtained without any prior information, i.e. frequentist estimator, while priors that reflect uncertainty or heterogeneity have little influence, reducing the posterior distribution to that of the frequentist estimator but with no harm in its precision. On the other hand, priors reflecting certainty and homogeneity will have a much greater influence in the posterior distribution of μ, allowing little contribution from the data, while priors reflecting uncertainty and heterogeneity can produce estimates with very poor precision.

3.3 | Prior beliefs on the between-studies variance

So far, we have considered hierarchical models with the hyper-parameter τ^2 fixed, which induce a k-variate normal prior distribution for the parameter vector $\boldsymbol{\beta}$. However, as it is the standard practice in Bayesian random effects meta-analysis, a prior distribution can be used for τ^2. A fixed effects approach to meta-analysis would not exclude the use of these types of hierarchical priors, whenever the effect-sizes are considered to be exchangeable. However, if we keep in mind that in the fixed effects approach the inference is targeted to functions of the vector $\boldsymbol{\beta}$, then it is evident that we need to understand and illustrate the prior (multivariate) distribution of $\boldsymbol{\beta}$ that is induced by the hierarchical structure that includes priors for both μ and τ^2. To do this, we consider the following hierarchical model:

$$\begin{aligned} \beta_1,...,\beta_k|\mu,\tau^2 &\quad \text{iid} \quad N(\mu,\tau^2) \\ \mu|\nu,\psi^2 &\quad \sim \quad N(\nu,\psi^2) \\ \tau^2|\alpha &\quad \sim \quad U(0,\alpha) \end{aligned} \tag{11}$$

Figure 1 shows the prior distribution of the parameter vector $\boldsymbol{\beta} = (\beta_1,...,\beta_k)^T$ that is induced by different combinations of Normal priors for μ and uniform priors for τ^2 as in (11). We observe that the marginal distribution of the individual β_i is bell-shaped, with variance $\psi^2 + E(\tau^2)$. However, as can be observed in the contour plots of the bivariate joint prior of any

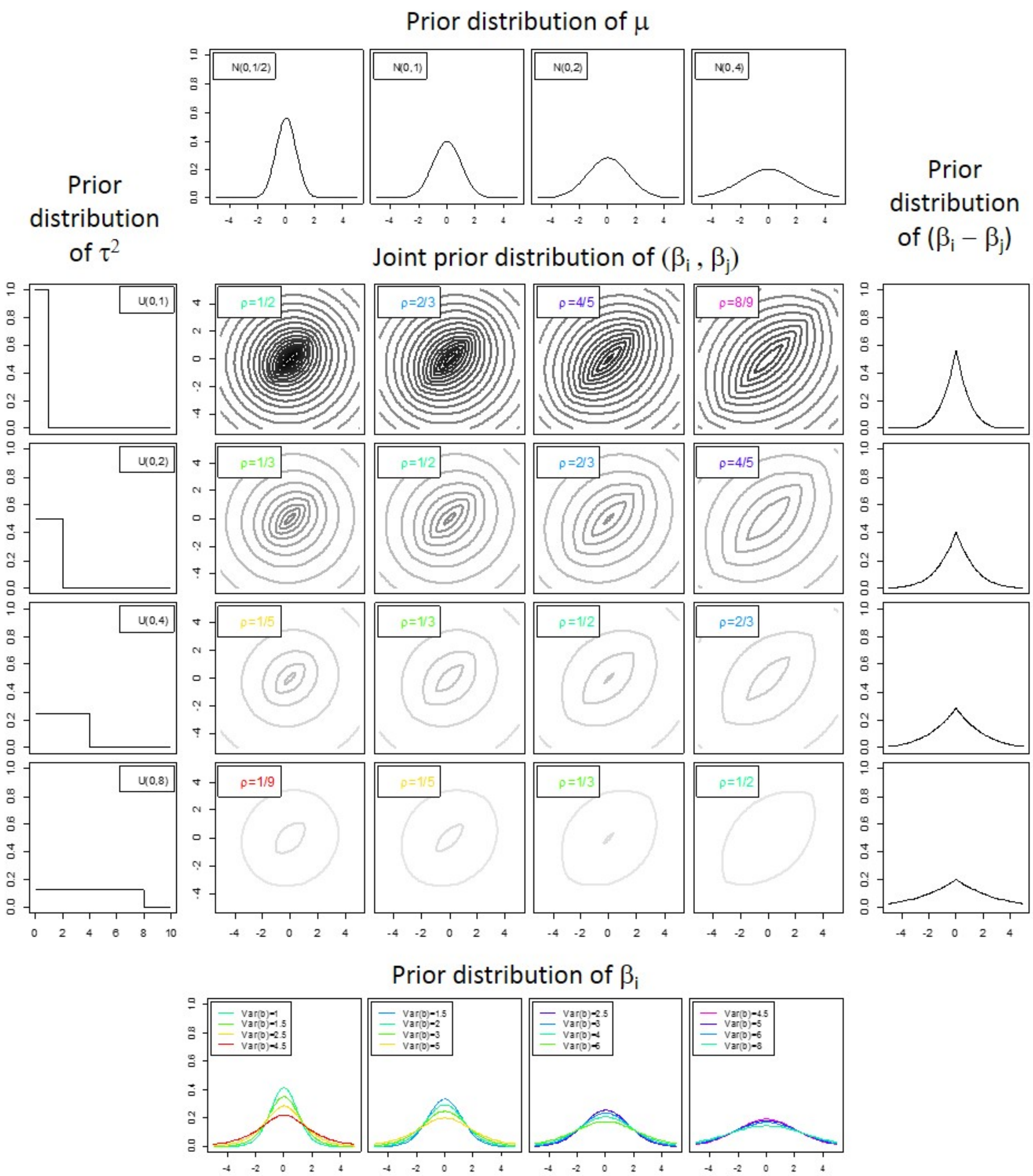

FIGURE 1 Marginal distributions of the effect size parameters $(\beta_1, ..., \beta_k)$ induced by different combinations of priors in μ (top) and τ^2 (left) in a hierarchical model of the form $\beta_i|\mu, \tau^2 \sim N(\mu, \tau^2)$ for $i = 1, ..., k$: the bivariate joint distribution of any pair (β_i, β_j) (middle), the marginal distribution of β_i (bottom) and the marginal distribution of $(\beta_i - \beta_j)$ (right).

pair (β_i, β_j), the distribution shows a 'ridge' along the $\beta_i = \beta_j$ line. This is also evident in the 'pointy' marginal distribution of the difference $(\beta_i - \beta_j)$. This reveals a prior that more strongly suggests homogeneity of the effect-size parameters than, say, a multivariate Normal prior with the same variance and correlation. Similar 'ridge-like' joint distributions of $\boldsymbol{\beta}$ are obtained when using families of distributions other than uniform as priors for τ^2 (results not shown). And even 'sharper' distributions are obtained when 'vague' or 'diffuse' priors are used for μ, as is commonly done, which induce almost flat priors with very little information on the location of the parameters, but with a very strong suggestion of homogeneity, which in turn produce the shrinkage observed in the updated distributions of the effect sizes.

The use of MCMC sampling methods for the estimation of the posterior distributions for all parameters is well known[15], and will not be discussed here. We just note that the posterior distribution of the parameters β_F and ζ^2, as functions of the parameter

Study	Zinc N	Zinc Mean	Zinc SD	Placebo N	Placebo Mean	Placebo SD	Mean Difference, 95% CI
Douglas 1987	33	12.1	9.8	30	7.7	9.8	4.40 [-0.44, 9.24]
Petrus 1998	52	4.4	1.4	49	5.1	2.8	-0.70 [-1.58, 0.18]
Prasad 2000	25	4.5	1.6	23	8.1	1.8	-3.60 [-4.48, -2.72]
Prasad 2008	25	4	1	25	7.1	1.3	-3.10 [-3.72, -2.48]
Turner 2000b	68	6.9	3.4	71	7.5	4	-0.70 [-1.94, 0.54]
Turner 2000c	72	7.9	4.2	71	7.5	4	0.40 [-0.99, 1.79]

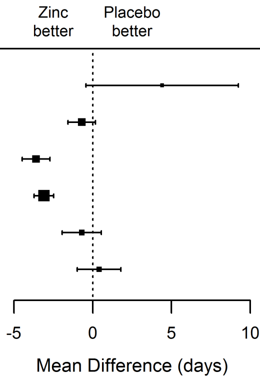

FIGURE 2 Meta-analysis on the efficacy of zinc acetate lozenges in reducing the duration of cold symptoms, published in[20].

vector $\boldsymbol{\beta}$, can be then easily obtained from a random sample of the joint posterior distribution of $\boldsymbol{\beta}$. Sample code (using `R` package `R2WinBUGS`) is shown in Appendix C.

A well known characteristic of hierarchical models like the one in (11) is the sensitivity of the resulting inference, both for the location and heterogeneity parameters μ and τ^2, to the choice of prior for τ^2 [8]. Because of this, sensitivity analyses are recommended as a routine practice[9]. In the following sections, we use an example meta-analysis to study the sensitivity of the proposed parameters β_F and ζ^2 to the choice of prior.

4 | EXAMPLE

In this section we illustrate the Bayesian estimation methods discussed in previous Sections of this Chapter, applied to the example meta-analysis on the efficacy of zinc lozenges in reducing the duration of common cold symptoms (see Figure 2).

In Figure 3 we present results from frequentist and Bayesian estimation of the proposed parameters β_F and ζ^2 from a fixed effects approach, as well as the parameters μ and τ^2 from a random effects approach. Results of Bayesian analyses include those from selected priors of the family of conjugate Normal distributions in Table 2 , as well hierarchically structured priors compiled in the paper by Lambert et al.[8] For the former, the mean and variance of the posterior Normal distribution of β_F and μ were obtained using equations (6) and (10), respectively, while the mean and variance of the posterior distribution of ζ^2 were obtained from equations in Appendix A, and selected percentiles were obtained using the `R` package `CompQuadForm`.

These results illustrate how the point estimate and precision of μ are very sensitive to the choice of prior distribution, and specifically to the level of heterogeneity in the prior. For example, when using a very flat prior for μ ($N(0, 1000)$) with a homogeneous prior for the study effects ($\beta_i \sim N(\mu, 0.25)$) then $[\mu|\hat{\boldsymbol{\beta}}] \sim N(-1.72, 0.31^2)$, while the same vague prior for μ with a more heterogeneous prior for the study effects ($\beta_i \sim N(\mu, 25)$) results in a very different posterior: $[\mu|\hat{\boldsymbol{\beta}}] \sim N(-0.72, 2.08^2)$. This is not the case for β_F, for which the posterior distribution, in both cases, closely approximates that of the standard frequentist estimator.

Further results of Bayesian analyses using a hierarchical model with Normal prior for μ and a fixed value of τ^2 (equivalent to a multivariate Normal prior for $\boldsymbol{\beta}$) are shown in Figure 4 , while results from hierarchical models with a Uniform prior distribution for τ^2 are shown in Figure 5 . In these plots we show the posterior mean and 95% credible interval of the location parameters μ and β_F, for a range of values of τ^2 (or the hyper-parameter θ that determines its distribution). These correspond to priors that go from close to homogeneous ($\tau^2 \to 0$) to more heterogeneous ($\tau^2 \to \infty$). We also use selected values of ψ^2, which correspond to priors that go from very precise ($\psi^2 = 0.01$) to very flat or vague ($\psi^2 = 1000$). It is evident again that the estimation of μ is more sensitive to the choice of prior than the estimation β_F. The example also illustrates the behavior of the estimates in the limit cases described in Table 3 . For example, the posterior distribution of μ approaches the distribution of β_F as the heterogeneity decreases, while it approaches either the prior distribution (when ψ^2 is small) or the distribution of the un-weighted average (when ψ^2 is large) as the heterogeneity increases. On the other hand, the posterior distribution of β_F is more stable, i.e. more robust to vagueness and/or heterogeneity in the prior, approaching in such cases the distribution of the frequentist estimator.

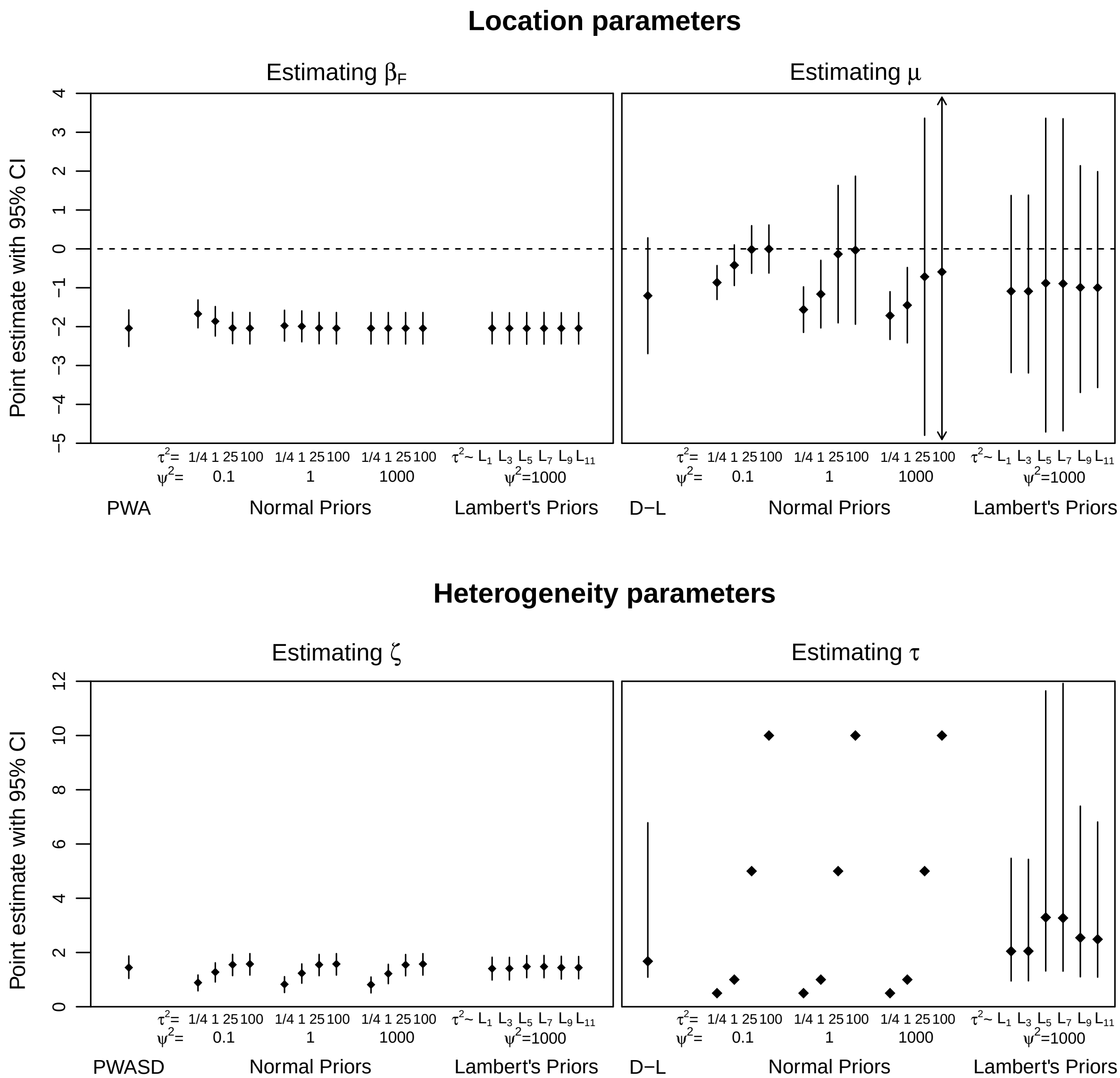

FIGURE 3 Posterior distribution (mean with 95% probability interval) of the location parameters μ and β_F (top) and posterior distribution (median with 95% probability interval) of the heterogeneity parameters $\tau = \sqrt{\tau^2}$ and $\zeta = \sqrt{\zeta^2}$ (bottom), along with frequentist estimates (the fixed effects precision weighted average (PWA, $\hat{\beta}_F$), the precision weighted average squared deviation (PWASD, $\hat{\zeta}^2$), and the random effects DerSimonian-Laird (D-L) estimators). Results from hierarchical Normal prior distributions ($\beta_i|\mu \sim N(\mu, \tau^2)$, for $i = 1, ..., k$; $\mu \sim N(0, \psi^2)$), with a fixed value for the between-study heterogeneity or a diffuse prior distribution taken from Lamber et al[8] (L_1: $\tau^{-2} \sim$ Gamma$(0.001, 0.001)$; L_3: $\log(\tau^2) \sim$ Uniform$(-10, 10)$; L_5: $\tau^{-2} \sim$ Uniform$(1/1000, 1000)$; L7: τ^{-2} ~Pareto$(1, 0.001)$; L_9: τ ~Uniform$(0, 100)$; $L_1 1$:$\tau \sim N(0, 100)$ for $\tau > 0$).

As for the quantification of heterogeneity, we can see that the posterior distribution of ζ^2 is, as expected, influenced by prior beliefs on the heterogeneity of the study effects. However, this influence is limited to a range of values τ^2, with the posterior distribution of ζ^2 'stabilizing', as consequence of the posterior distribution of the individual study effects stabilizing around their frequentist estimates. Although a similar 'stabilizing' behavior is observed in the median of the posterior distribution of τ^2, the precision (as reflected by the credible intervals) is importantly sensitive to the choice of prior (see Figure 5).

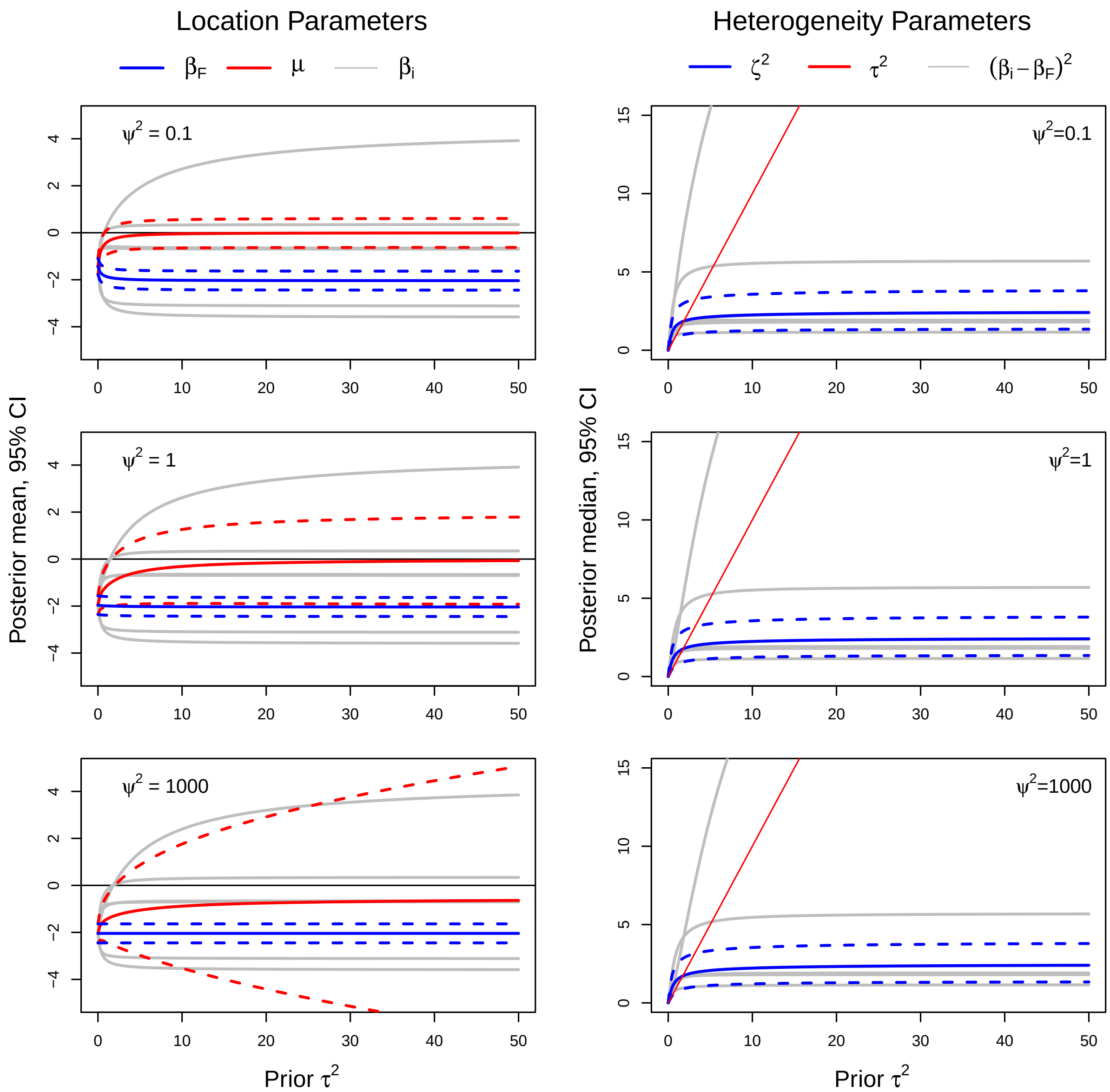

FIGURE 4 Mean and 95% credible interval from posterior distribution of the location parameters β_F and μ (left), and median with 95% credible interval from posterior distribution of heterogeneity parameter ζ^2 along with value of τ^2 (right). Results are presented for different values of hyper-parameters ψ^2 and τ^2 from the hierarchical Normal prior distribution ($\beta_i|\mu \sim N(\mu, \tau^2)$, for $i = 1, ..., k$; $\mu \sim N(0, \psi^2)$).

5 | CONCLUSIONS

We have proposed and implemented a fixed effects approach to meta-analysis within a Bayesian framework, while maintaining exchangeability of the study effects that is a key feature of many random effects analyses. Our approach is based on the estimation of a weighted average, β_F, which describes the overall location of the effect-size parameters, along with the estimation of a weighted average of their squared deviations ζ^2), which describes their heterogeneity.

In the frequentist framework for meta-analysis, the inverse variance weighted average is optimal among all affine combinations of the effect size parameters, in the sense that it is the one for which the corresponding estimator has the minimum variance. When a distributional assumption is made, this same optimality translates into it being the affine combination for which the

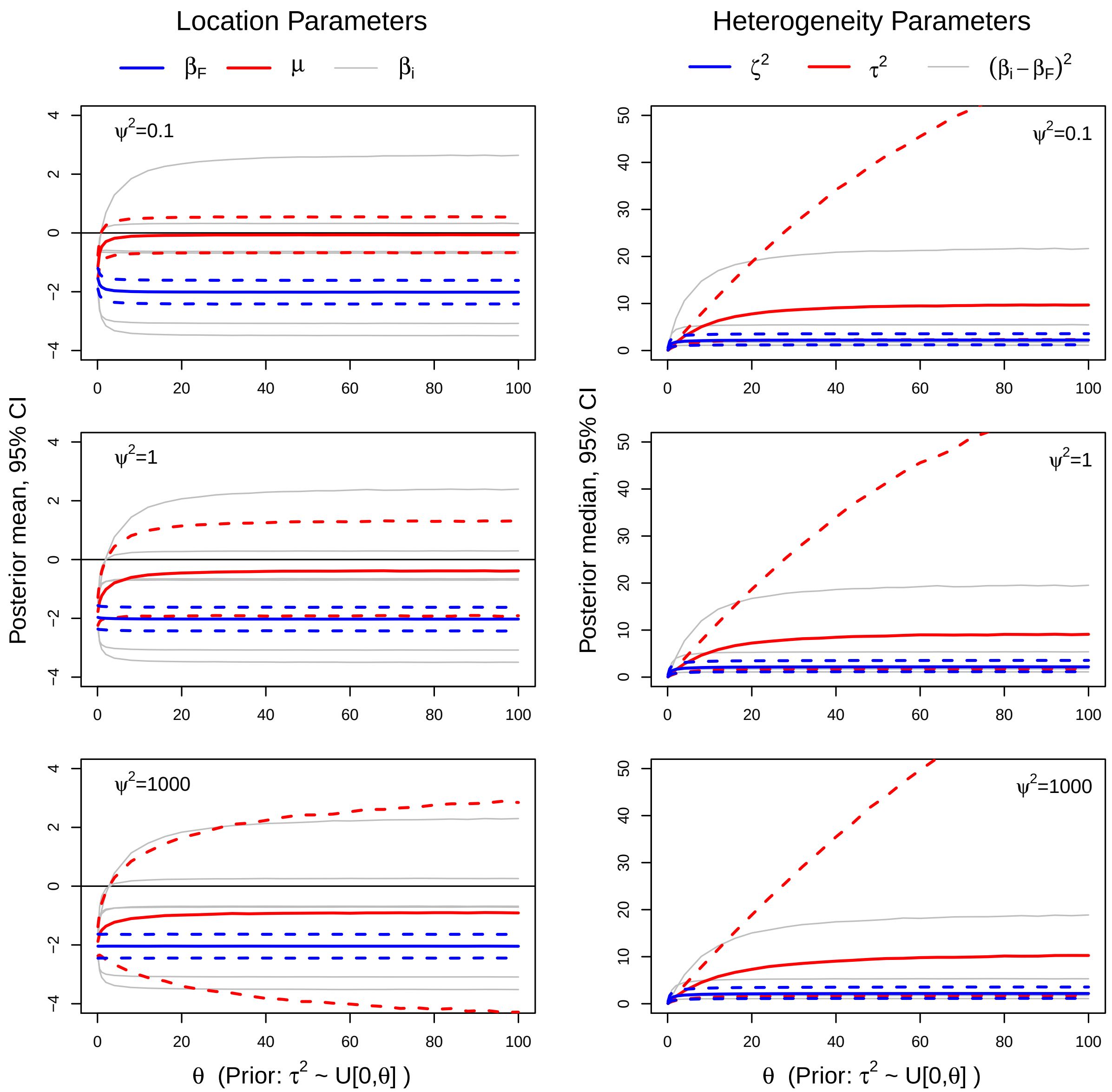

FIGURE 5 Mean and 95% credible interval from MCMC samples of posterior distributions of the location parameters β_F and μ (left), and median and 95% credible interval from posterior distribution of heterogeneity parameter ζ^2 along with value of τ^2 (right). Results are presented for different values of hyper-parameters ψ^2 and θ from the hierarchical model: $\beta_i|\mu \sim N(\mu, \tau^2)$, for $i = 1, ..., k$; $\mu \sim N(0, \psi^2)$; $\tau^2 \sim U(0, \theta)$.

corresponding estimator provides the maximum possible information. This maximum amount of information is actually the sum of the information provided by each estimator for its parameter. It should be noticed that, in practice, the true values of the variances $\sigma_1^2, ..., \sigma_k^2$ are unknown, so the analysis, as proposed here, actually estimates a weighted average where the weights are proportional to $\hat{\sigma}_i^{-2}$ rather than to σ_i^{-2}. This results in our proposed analysis estimating not exactly the optimal weighted average as described above, but one that is very close (as close as we can get) to that optimal parameter.

When considering a different approach for constructing a Bayesian analog to the optimal weighted average, we can think of targeting our inference to the affine combination for which the posterior variance is minimized. Thus, by using weights that are proportional to the inverse of the posterior variances, we would obtain the weighted average with the highest posterior precision. However, given that the posterior variances depend both on the prior and the data, these weights would also depend on the prior, meaning that the target parameter would be somewhat driven by the prior. As this seems undesirable, we instead propose an analysis in which the target parameter (an specific weighted average) is driven by the data, while the estimation is based on the data but is improved and strengthened by the information contained in the prior. The fact that the data drive the target parameter is what makes the the estimation of this parameter more robust and much less sensitive to the prior. As result, we have a desirable behavior of a Bayesian estimator defaulting to the frequentist estimate in the absence of strong beliefs, but gaining in precision when the such beliefs are reflected in the prior distribution.

In this paper, we have also discussed different prior distributions, either in the form of multivariate priors or hierarchical models with priors on hyperparameters (like the mean and variance of a normal random effects model). We have analyzed how these priors reflects beliefs of homogeneity and exchangeability. When using priors that involve hyperparameters, we have made a point that that inference is not, and should not, be restricted to these hyperparameters, but that results for parameters under the two models (fixed effects or random effects) can be obtained, thus providing information that is complementary. However, a cautionary note should be made related to the priors that are implied by a classic random effects model, and this is that the joint distribution that they induce on the effect-size parameters contains a "ridge", suggesting strong homogeneity of the effects. While this produces the shrinkage that is well known in the random effects analysis, in practice it might suggest homogeneity much more strongly that is actually intended. Therefore, we advise careful evaluation of such priors, to make sure that they truly reflect prior beliefs on how effect sizes might differ.

In summary, we have proposed a Bayesian fixed effects approach to meta-analysis, in which parameters describing location and heterogeneity correspond to data-driven weighted averages, which make them robust and stable to the choice of prior, even when the number of studies is small. We emphasize once more that such gains are not given by an estimation method but rather by targeting inference to a parameter for which the data provide more information.

References

1. Borenstein M, Hedges LV, Higgins J, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods.* 2010;1(2):97–111.

2. Hedges LV, Vevea JL. Fixed-and random-effects models in meta-analysis.. *Psychol Methods.* 1998;3(4):486.

3. Cooper H, Hedges LV, Valentine JC. *The handbook of research synthesis and meta-analysis*. New York: Russel Sage Foundation; 2009.

4. Higgins J, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc.* 2009;172(1):137–159.

5. Goel PK. Information measures and Bayesian hierarchical models. *J Am Stat Assoc.* 1983;:408–410.

6. Li Y, Shi L, RH Daniel. The bias of the commonly-used estimate of variance in meta-analysis. *Commun Stat Theory Methods.* 1994;23(4):1063-1085.

7. Viechtbauer W. Bias and efficiency of meta-analytic variance estimators in the random-effects model. *J Educ Behav Stat.* 2005;30(3):261–293.

8. Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Stat Med.* 2005;24(15):2401–2428.

9. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian approaches to clinical trials and health-care evaluation*. John Wiley & Sons; 2004.

10. Konstantopoulos S, Hedges LV. Analyzing effect sizes: fixed-effects models. In: Cooper H, Hedges LV, Valentine JC, eds. *The handbook of research synthesis and meta-analysis*, New York: Russel Sage Foundation 2009.

11. Peto R. Discussion. *Stat Med.* 1987;6(3):241–244.

12. Domínguez Islas Clara, Rice Kenneth M. Addressing the estimation of standard errors in fixed effects meta-analysis. *Statistics in medicine.* 2018;37(11):1788–1809.

13. Rice K, Higgins JPT, Lumley T. A re-evaluation of fixed effect(s) meta-analysis. *J R Stat Soc Ser A Stat Soc.* 2017;. http://dx.doi.org/10.1111/rssa.12275. Accessed Jan 2017.

14. Tong YL. The role of the covariance matrix in the least-squares estimation for a common mean. *Linear algebra and its applications.* 1997;264:313–323.

15. DeGroot Morris H. *Optimal statistical decisions*. John Wiley & Sons; 2005.

16. Lunn David J, Thomas Andrew, Best Nicky, Spiegelhalter David. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing.* 2000;10(4):325–337.

17. Martin Andrew D, Quinn Kevin M, Park Jong Hee. MCMCpack: Markov Chain Monte Carlo in R. *Journal of Statistical Software.* 2011;42(9):1–21.

18. Searle Shayle R, Gruber Marvin HJ. *Linear models*. John Wiley & Sons; 2016.

19. Duchesne P, Micheaux P. Computing the distribution of quadratic forms: Further comparisons between the Liu-Tang-Zhang approximation and exact methods. *Computational Statistics and Data Analysis.* 2010;54:858-862.

20. Singh M, Das RR. Zinc for the common cold. *Cochrane Database of Systematic Reviews.* 2013;6.

How to cite this article: C. Domínguez Islas, and K.M. Rice, (2018), A Bayesian Fixed Effects Approach to Meta-Analysis: Decoupling Exchangeability and Hierarchical Modelling, *Stat. Med.*, *2018;00:1-6*.

APPENDIX

A POSTERIOR DISTRIBUTION OF ζ^2 WHEN USING A CONJUGATE MULTIVARIATE NORMAL PRIOR

The following equations have been adapted from **(author?)** [19]. Let the matrix $\boldsymbol{C}$ be the Cholesky decomposition of $\text{Cov}[\beta_F|\hat{\boldsymbol{\beta}}]$, so that $\boldsymbol{C}^T\boldsymbol{C} = \text{Cov}[\beta_F|\hat{\boldsymbol{\beta}}]$, and let $\boldsymbol{P}$, with $\boldsymbol{P}\boldsymbol{P}^T$, such that it diagonalizes the the matrix $\boldsymbol{C}(\boldsymbol{W} - \boldsymbol{W}\mathbf{1}_{kk}\boldsymbol{W})\boldsymbol{C}^T$, so that

$$\boldsymbol{P}\boldsymbol{C}(\boldsymbol{W} - \boldsymbol{W}\mathbf{1}_{kk}\boldsymbol{W})\boldsymbol{C}^T\boldsymbol{P}^T = \boldsymbol{D} = \text{diag}\{\lambda_1, \lambda_2, ..., \lambda_k\}$$

Given that the rank of $(\boldsymbol{W} - \boldsymbol{W}\mathbf{1}_{kk}\boldsymbol{W})$ is $k-1$, we have that $\lambda_1 \geq ... \geq \lambda_{k-1} \geq 0$ and $\lambda_k = 0$. Then, letting $\boldsymbol{Y} = \boldsymbol{P}(\boldsymbol{C}^T)^{-1}(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}})$, with $\boldsymbol{Y} \sim N_k(\boldsymbol{P}(\boldsymbol{C}^T)^{-1}\mathbb{E}[\beta_F|\hat{\boldsymbol{\beta}}], \boldsymbol{I}_k)$, we get that

$$\begin{aligned}(\zeta^2|\hat{\boldsymbol{\beta}}) &= (\boldsymbol{\beta}|\hat{\boldsymbol{\beta}})^T\left(\boldsymbol{W} - \boldsymbol{W}1_{kk}\boldsymbol{W}\right)(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}}) \\ &= \boldsymbol{Y}^T\boldsymbol{D}\boldsymbol{Y} = \sum_{i=1}^{k-1}\lambda_i\chi^2(\delta_i)\end{aligned}$$

where δ_i equal to the square of the ith element of $\boldsymbol{P}(\boldsymbol{C}^T)^{-1}\mathbb{E}[\beta_F|\hat{\boldsymbol{\beta}}]$. From this equation, the expected value and variance of $(\zeta^2|\boldsymbol{\beta})$ can be easily evaluated:

$$\mathbb{E}[\zeta^2|\hat{\boldsymbol{\beta}}] = \sum_{i=1}^{k-1}\lambda_i(1+\delta_i)$$

$$\text{Var}[\zeta^2|\hat{\boldsymbol{\beta}}] = \sum_{i=1}^{k-1}\lambda_i^2(2+4\delta_i)$$

Also, [19] have implemented some algorithms to evaluate probabilities of a sum of chi-square random variables into the `R` package `CompQuadForm`, which can be used to obtain selected percentiles of this posterior distribution.

B POSTERIOR DISTRIBUTION OF β_F WHEN USING A HIERARCHICAL PRIOR

The posterior variance and mean of $\beta_F = \mathbf{1}_k^T \boldsymbol{W}\boldsymbol{\beta}$ are given by (8) and (7). Re-parametrizing in terms of τ^2 and ψ^2 we obtain the following identities:

$$\boldsymbol{\Upsilon}^{-1} = \left(\tau^2 \boldsymbol{I}_i + \psi^2 \mathbf{1}_{kk}\right)^{-1} = \tau^{-2}\boldsymbol{I}_k - \left(\frac{\psi^2/\tau^2}{\tau^2 + k\psi^2}\right)\mathbf{1}_{kk}$$

$$(\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Upsilon}^{-1})^{-1} = \tau^2 \mathrm{Diag}\left\{\frac{\sigma^2}{\sigma^2+\tau^2}\right\} + \left(\frac{\psi^2}{1+\psi^2\sum_i \frac{1}{\sigma_i^2+\tau^2}}\right)\mathrm{Diag}\left\{\frac{\sigma^2}{\sigma^2+\tau^2}\right\}\mathbf{1}_{kk}\mathrm{Diag}\left\{\frac{\sigma^2}{\sigma^2+\tau^2}\right\}$$

Then, for the posterior variance of β_F we obtain:

$$\begin{aligned}
\mathrm{Var}[\beta_F|\hat{\boldsymbol{\beta}}] =& \frac{1}{\Phi^2}\mathbf{1}_k^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Upsilon}^{-1})^{-1}\boldsymbol{\Sigma}^{-1}\mathbf{1}_k \\
=& \frac{1}{\Phi^2}\mathbf{1}_k^t\left[\tau^2\mathrm{Diag}\{\sigma_i^{-2}\}\mathrm{Diag}\left\{\frac{\sigma^2}{\sigma^2+\tau^2}\right\}\mathrm{Diag}\{\sigma_i^{-2}\}\right]\mathbf{1}_k \\
&+ \frac{1}{\Phi^2}\mathbf{1}_k^t\left[\left(\frac{\psi^2}{1+\psi^2\sum_i\frac{1}{\sigma_i^2+\tau^2}}\right)\mathrm{Diag}\left\{\frac{1}{\sigma^2+\tau^2}\right\}\mathbf{1}_k\mathbf{1}_k^t\mathrm{Diag}\left\{\frac{1}{\sigma^2+\tau^2}\right\}\right]\mathbf{1}_k \\
=& \frac{1}{\Phi^2}\left[\sum_{i=1}^k\left(\frac{1}{\sigma^2}\right)\left(\frac{\tau^2}{\sigma^2+\tau^2}\right) + \left(\frac{\psi^2}{1+\psi^2\sum_i\frac{1}{\sigma_i^2+\tau^2}}\right)\left(\sum_{i=1}^k\frac{1}{\sigma^2+\tau^2}\right)^2\right] \\
=& \frac{1}{\Phi^2}\sum_{i=1}^k\frac{1}{\sigma_i^2}\left[\left(1-\frac{\sigma_i^2}{\sigma_i^2+\tau^2}\right) + \left(\frac{\psi^2\sum_i\frac{1}{\sigma_i^2+\tau^2}}{1+\psi^2\sum_i\frac{1}{\sigma_i^2+\tau^2}}\right)\frac{\sigma_i^2}{\sigma_i^2+\tau^2}\right] \\
=& \frac{1}{\Phi^2}\sum_{i=1}^k\frac{1}{\sigma_i^2}\left[1-\left(\frac{1}{1+\psi^2\sum_i\frac{1}{\sigma_i^2+\tau^2}}\right)\frac{\sigma_i^2}{\sigma_i^2+\tau^2}\right].
\end{aligned}$$

And for the posterior mean:

$$\mathrm{E}[\beta_F|\hat{\boldsymbol{\beta}}] = \frac{1}{\Phi}\mathbf{1}_k^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Upsilon}^{-1})^{-1}\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\beta}} + \frac{1}{\Phi}\mathbf{1}_k^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Upsilon}^{-1})^{-1}\boldsymbol{\Upsilon}^{-1}\mathbf{1}_k\nu,$$

where the first term reduces to:

$$\begin{aligned}
&\frac{1}{\Phi^2}\mathbf{1}_k^t\left[\tau^2\mathrm{Diag}\{\sigma_i^{-2}\}\mathrm{Diag}\left\{\frac{\sigma^2}{\sigma^2+\tau^2}\right\}\mathrm{Diag}\{\sigma_i^{-2}\}\right]\hat{\boldsymbol{\beta}} \\
&+ \frac{1}{\Phi^2}\mathbf{1}_k^t\left[\left(\frac{\psi^2}{1+\psi^2\sum_i\frac{1}{\sigma_i^2+\tau^2}}\right)\mathrm{Diag}\left\{\frac{1}{\sigma^2+\tau^2}\right\}\mathbf{1}_k\mathbf{1}_k^t\mathrm{Diag}\left\{\frac{1}{\sigma^2+\tau^2}\right\}\right]\hat{\boldsymbol{\beta}} \\
=& \frac{1}{\Phi}\sum_{i=1}^k\left[\left(\frac{1}{\sigma_i^2}\right)\left(\frac{\tau^2}{\sigma_i^2+\tau^2}\right) + \left(\frac{\psi^2\sum_i\frac{1}{\sigma_i^2+\tau^2}}{1+\psi^2\sum_i\frac{1}{\sigma_i^2+\tau^2}}\right)\left(\frac{1}{\sigma_i^2+\tau^2}\right)\right]\hat{\beta}_i \\
=& \frac{1}{\Phi}\sum_{i=1}^k\left(\frac{1}{\sigma_i^2}\right)\left[\left(1-\frac{\sigma_i^2}{\sigma_i^2+\tau^2}\right) + \left(\frac{\psi^2\sum_i\frac{1}{\sigma_i^2+\tau^2}}{1+\psi^2\sum_i\frac{1}{\sigma_i^2+\tau^2}}\right)\left(\frac{\sigma_i^2}{\sigma_i^2+\tau^2}\right)\right]\hat{\beta}_i \\
=& \frac{1}{\Phi}\sum_{i=1}^k\left(\frac{1}{\sigma_i^2}\right)\left[1-\left(\frac{1}{1+\psi^2\sum_i\frac{1}{\sigma_i^2+\tau^2}}\right)\left(\frac{\sigma_i^2}{\sigma_i^2+\tau^2}\right)\right]\hat{\beta}_i,
\end{aligned}$$

and within the second term we have that:

$$
\begin{aligned}
&\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Upsilon}^{-1})^{-1}\boldsymbol{\Upsilon}^{-1} \\
&= \text{Vec}\left\{\frac{\sigma_i^2}{\sigma_i^2+\tau^2}\right\} + \left(\frac{\psi^2\tau^{-2}\sum_i \frac{\sigma_i^2}{\sigma_i^2+\tau^2}}{1+\psi^2\sum_i \frac{1}{\sigma_i^2+\tau^2}}\right)\text{Vec}\left\{\frac{\sigma_i^2}{\sigma_i^2+\tau^2}\right\} - \left(\frac{k\psi^2}{\tau^2+k\psi^2}\right)\text{Vec}\left\{\frac{\sigma_i^2}{\sigma_i^2+\tau^2}\right\} \\
&\quad -\left(\frac{k\psi^2}{1+\psi^2\sum_i \frac{1}{\sigma_i^2+\tau^2}}\right)\left(\frac{\psi^2\tau^{-2}}{\tau^2+k\psi^2}\right)\left(\sum_i \frac{\sigma_i^2}{\sigma_i^2+\tau^2}\right)\text{Vec}\left\{\frac{\sigma_i^2}{\sigma_i^2+\tau^2}\right\} \\
&= \left[1-\frac{k\psi^2}{\tau^2+k\psi^2}+\left(\frac{\psi^2\tau^{-2}}{1+\psi^2\sum_i \frac{1}{\sigma_i^2+\tau^2}}\right)\left(\sum_i \frac{\sigma_i^2}{\sigma_i^2+\tau^2}\right)\left(1-\frac{k\psi^2}{\tau^2+k\psi^2}\right)\right]\text{Vec}\left\{\frac{\sigma_i^2}{\sigma_i^2+\tau^2}\right\} \\
&= \left[\frac{\tau^2}{\tau^2+k\psi^2}\left(1+\frac{k\psi^2\tau^{-2}-\psi^2\sum_i \frac{1}{\sigma_i^2+\tau^2}}{1+\psi^2\sum_i \frac{1}{\sigma_i^2+\tau^2}}\right)\right]\text{Vec}\left\{\frac{\sigma_i^2}{\sigma_i^2+\tau^2}\right\} \\
&= \left(\frac{1}{1+\psi^2\sum_i \frac{1}{\sigma_i^2+\tau^2}}\right)\text{Vec}\left\{\frac{\sigma_i^2}{\sigma_i^2+\tau^2}\right\}.
\end{aligned}
$$

so that

$$
\mathrm{E}[\beta_F|\hat{\boldsymbol{\beta}}] = \frac{1}{\Phi}\sum_{i=1}^{k}\left(\frac{1}{\sigma_i^2}\right)\left\{\left[1-\left(\frac{1}{1+\psi^2\sum_i \frac{1}{\sigma_i^2+\tau^2}}\right)\left(\frac{\sigma_i^2}{\sigma_i^2+\tau^2}\right)\right]\hat{\beta}_i + \left[\left(\frac{1}{1+\psi^2\sum_i \frac{1}{\sigma_i^2+\tau^2}}\right)\left(\frac{\sigma_i^2}{\sigma_i^2+\tau^2}\right)\right]\nu\right\}
$$

C SAMPLE CODE FOR MCMC SAMPLING

The following is a sample code in **R** (as an interface to WinBUGS[16]) used for sampling of the posterior distribution of the parameters of interest, through MCMC methods.

```
library(R2WinBUGS)

## Setting hierarchical model
hmodel <- function()
{
for (i in 1:k)
{
w[i] <- pow(se[i], -2)
hatbeta[i] ~ dnorm(beta[i], w[i])
theta[i] ~ dnorm(mu,prec)
num[i] <- w[i]*beta[i]
qua[i] <- w[i]*pow(beta[i],2)
}
mu ~ dnorm(0,0.001)
tau ~ dunif(0,100)
tau2 <- pow(tau,2)
prec <- 1/tau2
betaF <- sum(num[])/sum(w[])
zeta2 <- sum(qua[])/sum(w[]) - pow(betaF,2)
}
```

```
## Saving to file
myfilename <- file.path(mydirectory, "hier_model.bug")
write.model(hmodel, myfilename)

## Initial Values (3 chains)
myinitial   <- list( list(beta=rep(0,k),mu=0, tau=0.1),
list(beta=rep(0,k),mu=0, tau=0.9),
list(beta=rep(0,k),mu=0, tau=0.4) )

## Data
mydata <- list(hatbeta=y,se=yse,k=6,psi2=100)

## Parameters of interest
myparameters <- c("beta","mu","tau2","betaF","zeta2")

## Running chains
output <- bugs(data=mydata, inits=initial, parameters=myparameters,
model.file=myfilename, n.chains=3, n.iter=200000,
n.thin=10, bugs.directory=bugs.dir, debug=TRUE)

## Displaying results
summary(output)
```