STATISTICS IN MEDICINE
Statist. Med. 2005; **24**:2401–2428
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.2112

How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS

Paul C. Lambert*,†, Alex J. Sutton, Paul R. Burton, Keith R. Abrams and David R. Jones

Centre for Biostatistics and Genetic Epidemiology, Department of Health Sciences, University of Leicester, U.K.

SUMMARY

There has been a recent growth in the use of Bayesian methods in medical research. The main reasons for this are the development of computer intensive simulation based methods such as Markov chain Monte Carlo (MCMC), increases in computing power and the introduction of powerful software such as WinBUGS. This has enabled increasingly complex models to be fitted. The ability to fit these complex models has led to MCMC methods being used as a convenient tool by frequentists, who may have no desire to be fully Bayesian.

Often researchers want 'the data to dominate' when there is no prior information and thus attempt to use vague prior distributions. However, with small amounts of data the use of vague priors can be problematic. The results are potentially sensitive to the choice of prior distribution. In general there are fewer problems with location parameters. The main problem is with scale parameters. With scale parameters, not only does one have to decide the distributional form of the prior distribution, but also whether to put the prior distribution on the variance, standard deviation or precision.

We have conducted a simulation study comparing the effects of 13 different prior distributions for the scale parameter on simulated random effects meta-analysis data. We varied the number of studies (5, 10 and 30) and compared three different between-study variances to give nine different simulation scenarios. One thousand data sets were generated for each scenario and each data set was analysed using the 13 different prior distributions. The frequentist properties of bias and coverage were investigated for the between-study variance and the effect size.

The choice of prior distribution was crucial when there were just five studies. There was a large variation in the estimates of the between-study variance for the 13 different prior distributions. With a large number of studies the choice of prior distribution was less important. The effect size estimated was not biased, but the precision with which it was estimated varied with the choice of prior distribution leading to varying coverage intervals and, potentially, to different statistical inferences. Again there was less of a problem with a larger number of studies. There is a particular problem if the between-study variance is close to the boundary at zero, as MCMC results tend to produce upwardly biased estimates of the between-study variance, particularly if inferences are based on the posterior mean.

The choice of 'vague' prior distribution can lead to a marked variation in results, particularly in small studies. Sensitivity to the choice of prior distribution should always be assessed. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: Bayesian methods; Markov chain Monte Carlo; prior distributions; simulation study

*Correspondence to: Paul C. Lambert, Centre for Biostatistics and Genetic Epidemiology, Department of Health Sciences, University of Leicester, 22–28 Princess Road West, Leicester LE1 6TP, U.K.
†E-mail: pl4@le.ac.uk

Received April 2003
Accepted October 2004

1. INTRODUCTION

There has been a recent growth in the use of Bayesian methods in medical research and other areas [1–3]. Within a Bayesian analysis prior distributions for the unknown parameters need to be specified. In many situations vague prior distributions are chosen with the intention that they should have little or no impact in the inferences. However, naïve use of vague prior distributions may lead to them having an influence on any inference made. Here we assess the choice of prior distribution for the between-unit variance in a random effects meta-analysis using a simulation study. Although we consider one of the simplest random effects models, our findings are generalizable to more complex cases where random effects models are fitted.

One of the main reasons for the growth in Bayesian methods is the increase in computing power and the development of simulation based approaches such as Markov chain Monte Carlo (MCMC) methods [4]. This has led to specialist software being developed, in particular the BUGS software and the Windows implementation WinBUGS [5, 6]. In addition to the philosophical advantages of the Bayesian approach, the use of these methods has led to increasingly complex, but realistic, models being fitted [7]. Many of these models include hierarchical data structures where between-unit variation is modelled using random effects. Examples can be found in meta-analysis and generalized synthesis models [8], cluster randomized trials [9, 10], genetic epidemiology [11], institutional ranking [12] and subgroup analysis [13]. The use of hierarchical models is not unique to medicine and they are often applied in other areas such as education [14]. An advantage of the Bayesian approach is that the uncertainty in all parameter estimates is taken into account. This is particularly important if data are sparse.

When analysing data from a Bayesian perspective it is necessary to specify prior distributions for all unknown parameters. This can be a potential advantage, but in many situations there is a desire for the 'data to dominate' when no prior information is available (or when MCMC methods are being used for computational convenience and the researcher does not want to include prior information), which has led to the use of *vague* or *reference* priors [15]. We do not advocate the use of the term *non-informative prior distribution* as we consider all priors to contribute some information [16–18]. If data is sparse then even prior distributions that are intended to be vague may exert an unintentionally large degree of influence on any inferences. This may be a particular problem in random effects models as even though the total amount of data may be large, the number of units contributing to the estimation of the between-unit variation may be small. Therefore, with random effects models there is usually more concern regarding the influence of prior distributions on scale parameters rather than location parameters. In fact when the number of units contributing to the estimation of the between-unit variation is small it could be argued informative prior distributions are necessary.

The purpose of a reference prior is to be uniform over the range of interest and thus considers the possible values of the unknown parameters to be equally likely. However, a problem with the use of such prior distributions is the fact that uniformity is sensitive to transformation [19]. For example a prior distribution that is uniform on the variance scale will not be uniform on the standard deviation, precision or log variance scales. When using a reference prior one would hope that any parameter estimate would be unbiased and that the credible intervals would have coverage close to the nominal level. Although these are frequentist properties, their investigation is important as there is increasing use of MCMC methods as a convenient tool for fitting complex models rather than a desire to be fully Bayesian.

Previous work has shown that with a small number of units contributing to the estimation of the between-unit variation, inferences may be sensitive to the choice of prior distribution for this parameter [9, 10]. Browne and Draper [20] have recently investigated the use of two prior distributions for use in hierarchical models, namely a Gamma distribution on the precision and a uniform distribution on the variance. These two prior distributions are implemented in the hierarchical modelling software MLWin [21]. They found that when the number of units was small, the use of either of these prior distributions could lead to substantial bias and poor coverage. The same authors also found problems with a small number of units when using a Wishart prior distribution for correlated random effects [22].

The WinBUGS software has enabled complex models, that would be difficult or impossible to fit classically, to be fitted relatively straightforwardly using MCMC methods. With many of these models WinBUGS is used as a tool for fitting the models rather than for the desire to use informative prior distributions, so vague priors distributions are generally used in these situations. In addition to the problem of using vague prior distributions when using MCMC, there is the problem of whether the chains have converged. This can be difficult to assess in complex models.

In this paper we assess the performance of various prior distributions as implemented in the WinBUGS software [6] and thus our results may be sensitive to how this software implements the MCMC methods. We first consider a simple example of a random effects meta-analysis from the Cochrane Library, comparing the results of using different prior distributions. We then consider simulated data sets in the context of a random effects meta-analysis. We investigate the sensitivity of inferences to the choice of vague prior distribution for the between-study standard deviation by simulating meta-analysis data sets for nine different scenarios where the number of studies and size of the between-study standard deviation are varied.

In Section 2 we illustrate the sensitivity to the choice of prior distribution using an example from the Cochrane Library. In Section 3 we outline the procedure used to simulate the data for the nine scenarios. Section 4 presents the results of the simulations and Section 5 highlights the main findings and discusses issues for future research.

2. DEMONSTRATION OF PROBLEMS

Table I shows the odds ratios from a meta-analysis of short course (less than 7 days) vs long course (greater than 7 days) antibiotics for treatment of acute otitis media obtained from the Cochrane Library [23]. The outcome is treatment failure at 8–19 days. The original meta-analysis used a fixed-effects model even though there was strong evidence of heterogeneity of study effects using Cochran's test [24].

Table I. Odds ratios from five studies comparing the effects of short course (less than 7 days) vs long course (>7 days) antibiotics for acute otitis media.

Study	OR	95 per cent CI	Log(OR) (SE)
1	0.95	(0.39,2.28)	−0.05 (0.45)
2	0.80	(0.46,1.41)	−0.22 (0.29)
3	2.76	(1.00,7.63)	1.02 (0.52)
4	2.61	(1.54,4.43)	0.96 (0.27)
5	1.52	(0.95,2.42)	0.42 (0.24)

2.1. Bayesian hierarchical model

A Bayesian random effects model was fitted to the data presented in Table I. This is the same model that is explored further in the simulation study outlined in Section 3. Let y_i be the log-odds ratio in the ith study and s_i its associated standard error. A simple two-level hierarchical model can be fitted [25].

$$\begin{aligned} y_i &\sim \mathrm{N}(\mu_i, s_i^2) \\ \mu_i &\sim \mathrm{N}(\theta, \tau^2) \end{aligned} \tag{1}$$

This formulation of the model makes use of hierarchical centring, which can improve convergence [26]. Prior distributions need to be specified for the unknown parameters, i.e. the pooled log-odds ratio, θ, and the between-study standard deviation τ. For the pooled odds ratio, θ, a diffuse Normal distribution was used, i.e.

$$\theta \sim \mathrm{N}(0, 10000).$$

2.2. Prior distributions for variance components

For the between-study standard deviation τ, 13 different prior distributions were specified either for τ or some function of τ. However, it should be realized that specification of a prior distribution on, for example, the standard deviation scale, implies a distribution on the variance and precision scales. This is discussed with examples given below. The parameterisations for the different prior distributions are the same as those described in the WinBUGS manual [6]. The 13 prior distributions are as follows:

Prior 1

$$\frac{1}{\tau^2} \sim \mathrm{Gamma}(0.001, 0.001)$$

This is probably the most common used prior distribution for variance parameters, not least because it is used in many of the examples provided with the WinBUGS software [27, 28] This prior distribution is approximately uniform for most of the range, but has a 'spike' of probability mass close to zero.

Prior 2

$$\frac{1}{\tau^2} \sim \mathrm{Gamma}(0.1, 0.1)$$

This is of the same distributional form as prior 1, but with the two parameters set to 0.1, and thus provides a simple assessment of the sensitivity to the choice of these parameter values.

Prior 3

$$\log(\tau^2) \sim \mathrm{Uniform}(-10, 10)$$

This prior distribution is uniform on the log variance scale between two specified parameters. This has been used by Spiegelhalter [10] in the analysis of cluster randomized trials.

Prior 4

$$\log(\tau^2) \sim \mathrm{Uniform}(-10, 1.386)$$

As above, but only goes up to maximum of 1.386, the log of 4.0. This is because it seems implausible that the between-study variance could be larger than four (or equivalently the standard deviation to be greater than 2.0). In fact this value is probably still conservative, but it will demonstrate how estimates change when implausibly large values cannot be sampled. This is the first of a number of *weakly informative prior distributions* as it gives zero density to implausibly large values. The choice of what the upper bound should be is somewhat subjective and will vary between analysis situations.

Prior 5

$$\tau^2 \sim \text{Uniform}(1/1000, 1000)$$

Spiegelhalter [10] investigates uniform prior distributions on the variance. This is an option for specifying prior distributions for between level random effect variances in the MLWin software [21].

Prior 6

$$\tau^2 \sim \text{Uniform}(1/1000, 4)$$

This is the weakly informative version of prior 5. As for prior 4, the maximum value the between-study variance can be is 4.

Prior 7

$$\frac{1}{\tau^2} \sim \text{Pareto}(1, 0.001)$$

For a Pareto distribution with parameters α and c a uniform prior distribution for τ^k on the range $(0, r)$ can be expressed by setting $\alpha = k/2$ and $c = r^{-2}/k$ [5]. Hence values of $k = 2, 1$ and -2 give uniform prior distributions on the variance, standard deviation and precision scales respectively. Prior 7 is equivalent to a uniform distribution (0,1000) on the variance scale. These prior distributions have been used for variance components in genetic epidemiology models [11, 29].

Prior 8

$$\frac{1}{\tau^2} \sim \text{Pareto}(1, 0.25)$$

This is the weakly informative version of prior 7 and is equivalent to a uniform prior distribution for the variance in the range $(0, 4)$.

Prior 9

$$\tau \sim \text{Uniform}(0, 100)$$

This is a uniform prior distribution on the standard deviation scale in the range (0,100) and is a prior distribution recommended by Spiegelhalter *et al.* in a recent book [30].

Prior 10

$$\tau \sim \text{Uniform}(0, 2)$$

This is the weakly informative version of prior 9 and is thus is a uniform distribution on the standard deviation in the range (0,2).

Prior 11

$$\tau \sim \mathrm{N}(0, 100) \quad \text{for } \tau > 0$$

Prior 11 specifies a half normal distribution truncated at zero placed on the standard deviation scale. Such a distribution has been used previously in meta-analysis applications [31].

Prior 12

$$\tau \sim \mathrm{N}(0, 1) \quad \text{for } \tau > 0$$

This is the weakly informative version of prior 11, giving a smaller variance and thus giving a low probability to values greater than 4 for the between-study variance.

Prior 13

$$\frac{1}{\tau^2} \sim \mathrm{Logistic}(S_0) \quad S_0 = \sqrt{\frac{K}{\sum s_i^{-2}}} \quad K \text{ is the number of studies}$$

This prior distribution has been advocated by DuMouchel and Normand [32]. It is not strictly a 'vague' prior as it uses the observed variation to estimate the parameters for the prior distribution. It has a maximum at zero and is a decreasing function of τ.

Figure 1 shows the densities for five of the prior distributions on the variance, standard deviation and precision scales. It can be seen that both the choice of distribution and scale lead to different shaped distributions. For example, a prior that is uniform on the variance scale (Figure 1(c)) gives a triangular distribution on the standard deviation scale and a distribution with a spike near zero on the precision scale. This shows the importance of investigating the shape of prior distributions on different scales and demonstrates that the naïve belief that a 'flat' prior is uninformative is not necessarily correct.

The model (1) was fitted using WinBUGS using 5000 samples after a 'burn-in' of 1000 samples. Figure 2(a) shows the point estimates (medians) and 95 per cent credible intervals for the estimates of the pooled log-odds ratio and the between-study standard deviation from the meta-analysis of short versus long course antibiotic use for acute otitis media. It can be seen that although the point estimate of the pooled log-odds ratio is similar for the thirteen different prior distributions, there is considerable variability in the width of the 95 per cent credible interval. This is due to variability in both the point estimate and the width of the 95 per cent credible interval for the between-study standard deviation (Figure 2(b)).

The lack of the agreement in inferences when using the various prior distributions is worrying. For this reason we have conducted a simulation study which explores the sensitivity to the choice of prior distribution when varying the number of studies and the size of the between-study standard deviation. This is outlined in the next section.

3. SIMULATION STUDY

In a random effects meta-analysis the precision of the estimated between-study standard deviation depends upon the number of studies included in the meta-analysis and the actual magnitude of the between-study standard deviation. Therefore, data representing meta-analyses of size 5, 10 and 30 studies were generated. Each meta-analysis consisted of a number of hypothetical clinical trials comparing a standard treatment with a new treatment. The number

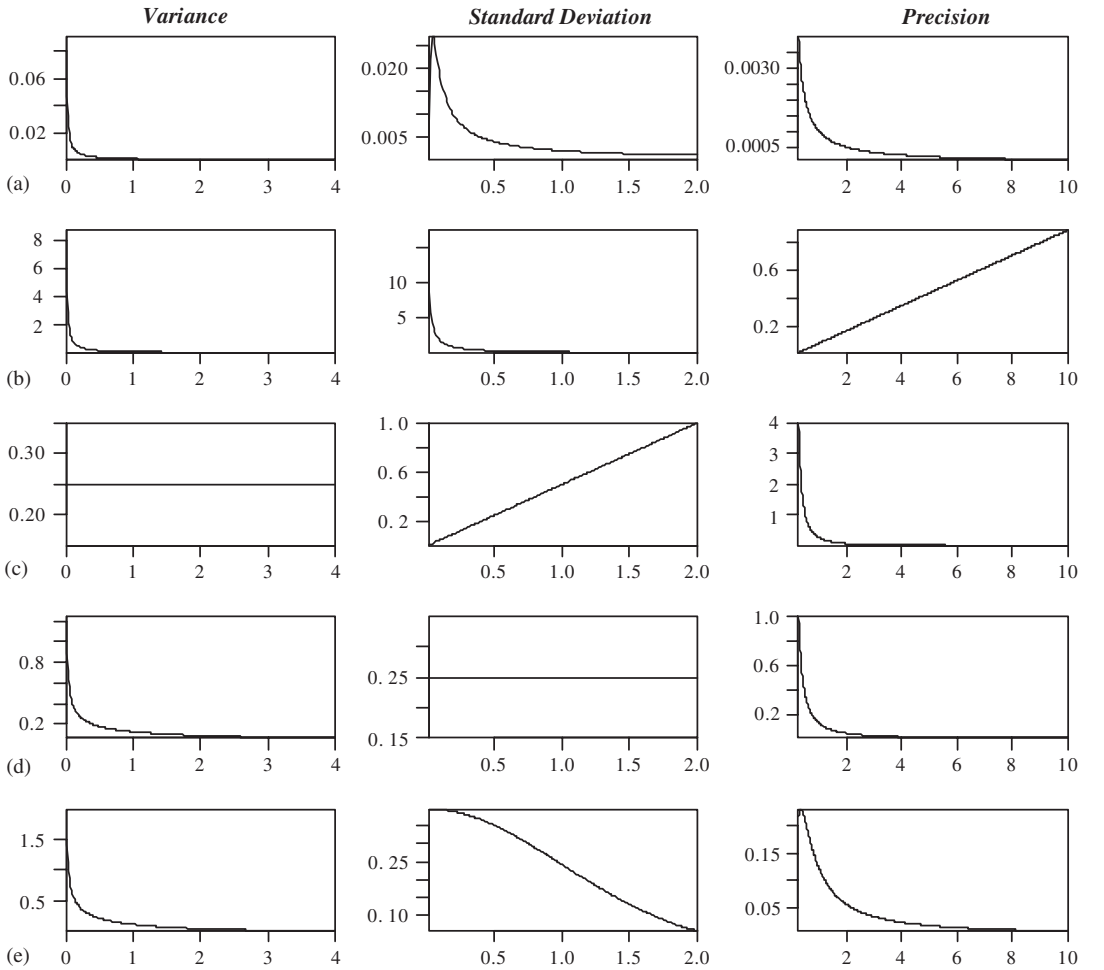

Figure 1. Density plots for priors: (a) $\frac{1}{\tau^2} \sim \text{Gamma}(0.001, 0.001)$ (Prior 1); (b) $\log(\tau^2) \sim \text{Uniform}(-10, 1.386)$ (Prior 4); (c) $\tau^2 \sim \text{Uniform}(1/1000, 4)$ (Prior 6); (d) $\tau \sim \text{Uniform}(0, 2)$ (Prior 10); and (e) $\tau \sim \text{N}(0, 1)I[0,]$ (Prior 12).

of studies in a meta-analysis of randomized controlled trials in medicine tends to be small and it is common to see meta-analysis performed on five or fewer studies. The outcome was defined as a dichotomous variable indicating the occurrence or not of the event of interest. Half of the patients on standard treatment had the event of interest. The underlying treatment effect in each meta-analysis was assumed to be an odds-ratio of 1.38 as outlined in Table II. Three different between-study standard deviations were investigated. These were 0.001 (effectively zero), 0.3 and 0.8. This leads to a different distribution of the underlying odds ratio across studies. A standard deviation of 0.001 (effectively zero) would indicate that a there is not true heterogeneity and a fixed effects model may be appropriate. For a standard

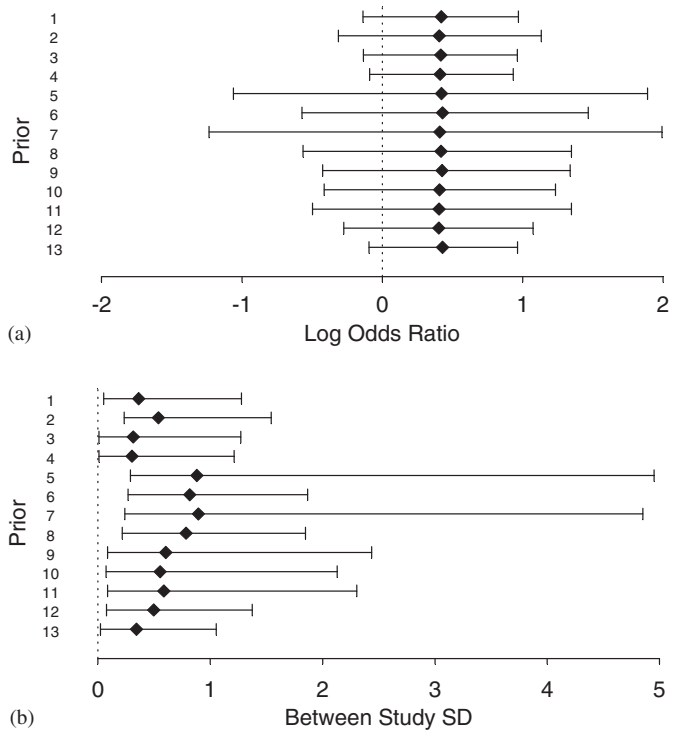

Figure 2. Point estimates and 95 per cent credible intervals for: (a) the pooled log-odds ratio; and (b) the between-study standard deviation from the meta-analysis of short vs long course antibiotic use for acute otitis media.

Table II. Assumed event rates and corresponding odds ratio for comparison of standard and new treatment groups.

Treatment	Outcome	
	−ve	+ve
Standard treatment	0.50	0.50
New treatment	0.42	0.58

True odds ratio is 1.381 (log-odds ratio 0.323).

deviation of 0.3 one would expect 95 per cent of the underlying treatment effects (odds ratios) to vary between 0.75 and 2.52. For a between-study standard deviation 0.8 one would expect 95 per cent of the underlying treatment effects to vary between 0.28 and 6.84. Note that a standard deviation of 0.8 may be unusual in the meta-analysis setting, but could be applicable in other areas where hierarchical models are used.

In the meta-analyses of five trials, the number of subjects in the five individual trials were 100, 200, 300, 400 and 500. For the meta-analyses of 10 and 30 trials, the same range of study sizes were simulated. There were two trials of each size in the 10 trial simulations and six trials of each size in the 30 trial simulations. Each meta-analysis data set was generated by

the following model:

$$\delta_i \sim \mathrm{N}(0, \tau^2)$$

$$\mathrm{logit}(p_{0i}) = \alpha$$

$$\mathrm{logit}(p_{1i}) = \alpha + \theta + \delta_i$$

$$r_{0i} \sim \mathrm{Binomial}(n_{0i}, p_{0i})$$
$$r_{1i} \sim \mathrm{Binomial}(n_{1i}, p_{1i})$$

where τ is the between-study standard deviation (0.001, 0.3 or 0.8), α is the log-odds for the standard therapy group (0), θ is the underlying log-odds ratio (0.323), n_{0i}, p_{0i} and r_{0i} are the number of subjects, the probability of an event and the number of events in the ith study for the standard treatment group respectively, with n_{1i}, p_{1i} and r_{1i} being the corresponding values for the new treatment group. For each of the nine scenarios, 1000 data sets were generated. For each data set the log-odds ratio and its associated standard error were calculated.

The Bayesian hierarchical model (1) was fitted to each generated dataset using the 13 different prior distributions. These models were fitted in WinBUGS 1.4. All 13000 models for a particular scenario were fitted simultaneously by looping over data sets and prior distributions. An example of the WinBugs code can be seen in Appendix I. For each dataset a burn-in of 1000 iterations was used, with sampling from a further 5000 iterations. In practice, when using MCMC methods for a single model, more iterations would be preferable. However, for the 117 000 models fitted in this paper it was not practical to run the chains for longer.

The 13 different vague prior distributions for the nine scenarios are evaluated using the frequentist criteria of bias and coverage under repeated sampling. With the increasing use of MCMC methods as a tool for fitting complex models, it is desirable for a Bayesian analysis with vague prior distributions to satisfy these frequentist criteria.

4. RESULTS

Figure 3 shows the point estimates (medians) and 95 per cent credible intervals for the first eight simulated data sets for the three scenarios (5, 10 and 30 studies) when the between-study standard deviation is 0.001 using each of the 13 prior distributions. Within each data set the point estimates of the log-odds ratios are broadly similar. However, the credible intervals vary and potentially could lead to different inferences. For example, in the first data set when there are five studies, four of the credible intervals exclude zero while the remaining nine include zero. As expected the width of the credible intervals narrows as the number of studies increases (note the three scenarios are plotted on different scales). Generally, there is more disagreement between the credible intervals when there are fewer studies included in the meta-analysis. However, even with 30 studies there is still some disagreement in the width of the credible interval.

Figures 4 and 5 are in the same format as Figure 3 and show the point estimates (medians) and the 95 per cent credible intervals for the first eight simulated meta-analysis data sets for 5, 10 and 30 studies when the underlying between-study standard deviation is 0.3 and 0.8, respectively. As expected, comparison of Figures 3–5 shows that the width of the credible interval increases. With 10 or 30 studies agreement in both the point estimates and

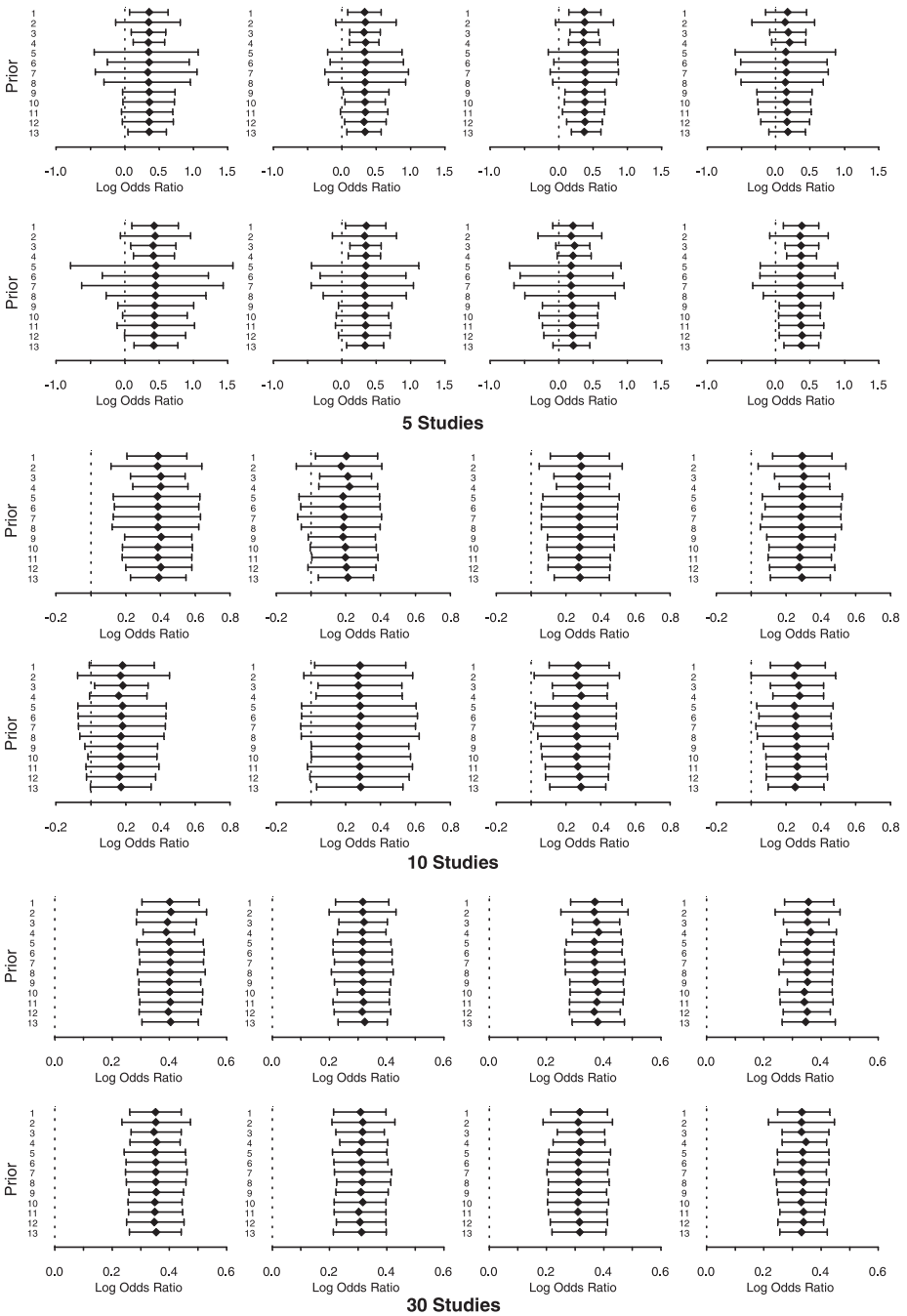

Figure 3. Point estimates and 95 per cent credible intervals for first eight simulated data sets when between-study S.D. = 0.001 for five studies, 10 studies and 30 studies.

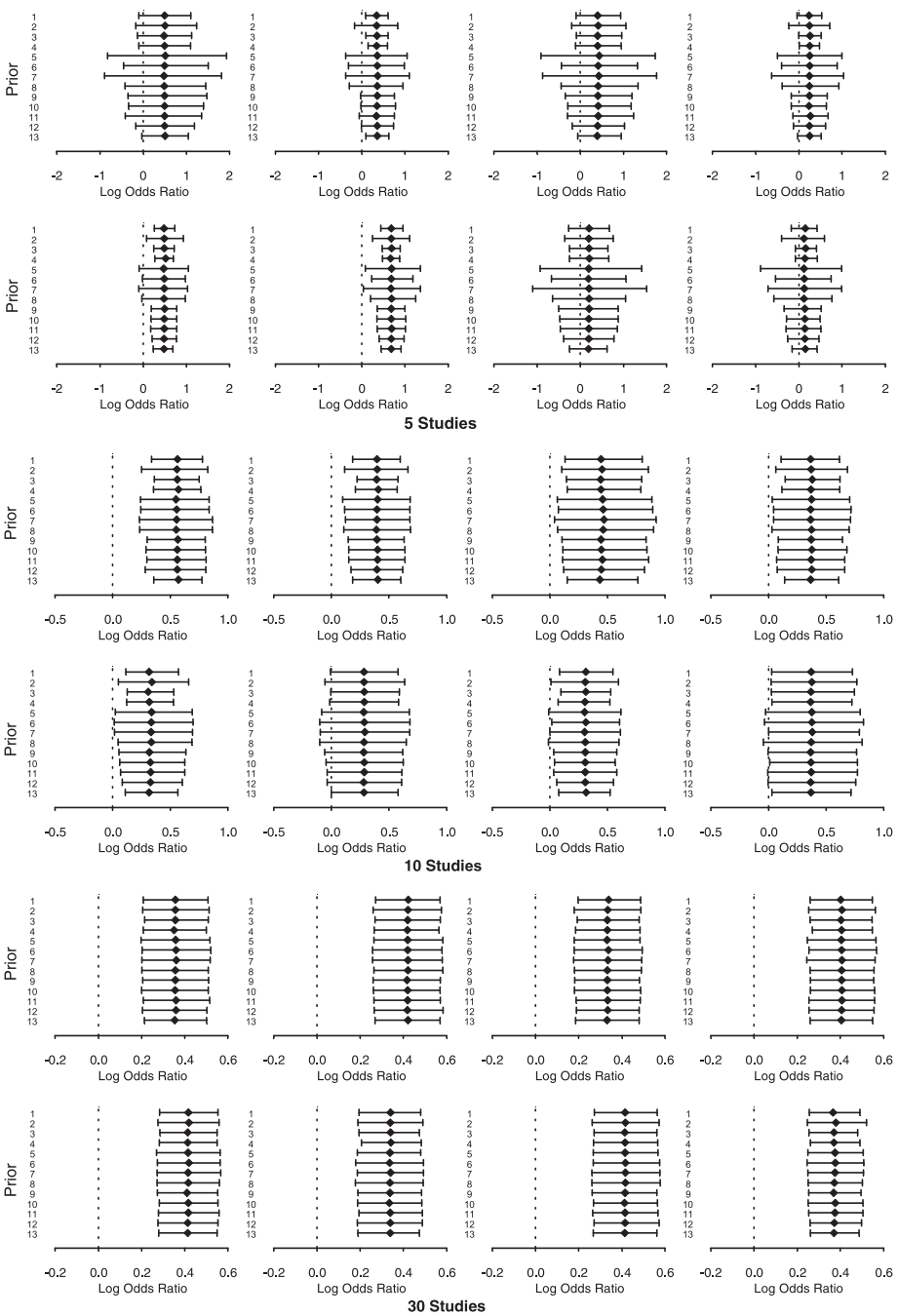

Figure 4. Point estimates and 95 per cent credible intervals for first eight simulated data sets when between-study S.D. = 0.3 for five studies, 10 studies and 30 studies.

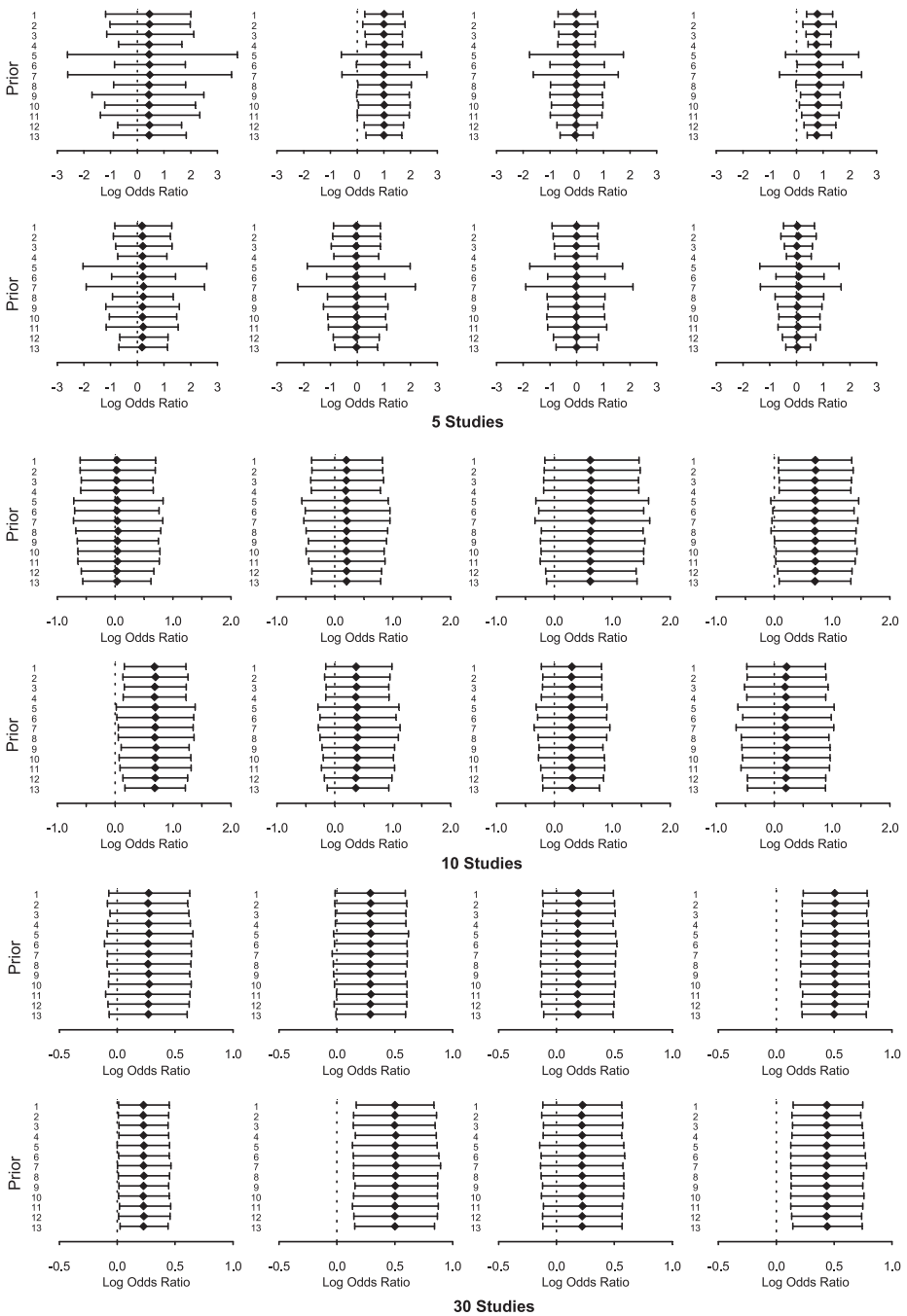

Figure 5. Point estimates and 95 per cent credible intervals for first eight simulated data sets when between-study S.D. = 0.8 for five studies, 10 studies and 30 studies.

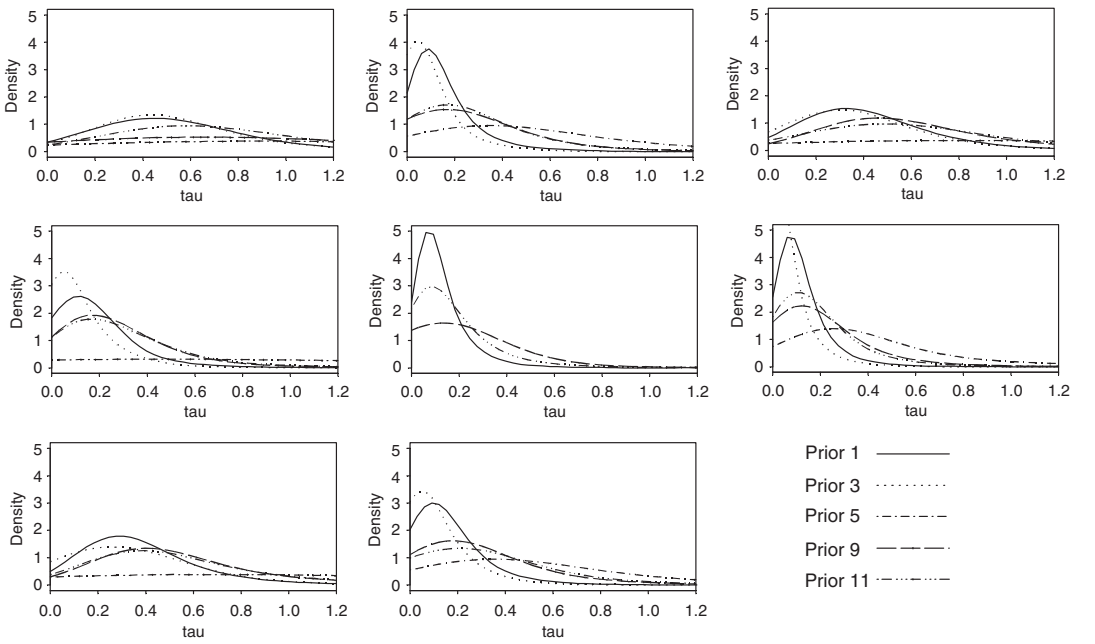

Figure 6. Posterior distributions for the between-study standard deviation for the first eight data sets simulated with five studies and a between-study standard deviation of 0.3 (only prior distributions 1, 3, 5, 9 and 11 are shown).

credible intervals is high compared to when there are five studies. This indicates that the prior distributions are exerting less influence relative to the data compared to the five study scenario.

Figure 6 shows the posterior densities for the first eight simulated data sets of the between-study standard deviation for five selected prior distributions for meta-analyses with five studies and a simulated between-study standard deviation of 0.3. These plots show how the different prior distributions result in considerably different shaped posterior distributions for the same data set.

Figure 7 shows for meta-analyses with five studies and a simulated between-study standard deviation of 0.001: (a) scatter plots of the agreement between the point estimates (medians) of the between-study standard deviations; and (b) scatter plots of the agreement between the estimated standard deviations of the pooled treatment effect for five selected prior distributions. It can be seen that there is disagreement between both the estimated between-study standard deviation and the standard deviation of the pooled treatment effect when using the five different prior distributions. Prior 5 (uniform on the variance) appears particularly discordant with the other prior distributions. Furthermore, in some instances a particular prior distribution consistently gives higher estimates than other prior distributions, e.g. Prior 11 (Normal on the standard deviation) gives higher estimates than Prior 1 (Gamma on the precision). In addition the plots show the relationship between the point estimate of the between-study standard deviation and the standard deviation of the pooled treatment effect, in that higher estimates

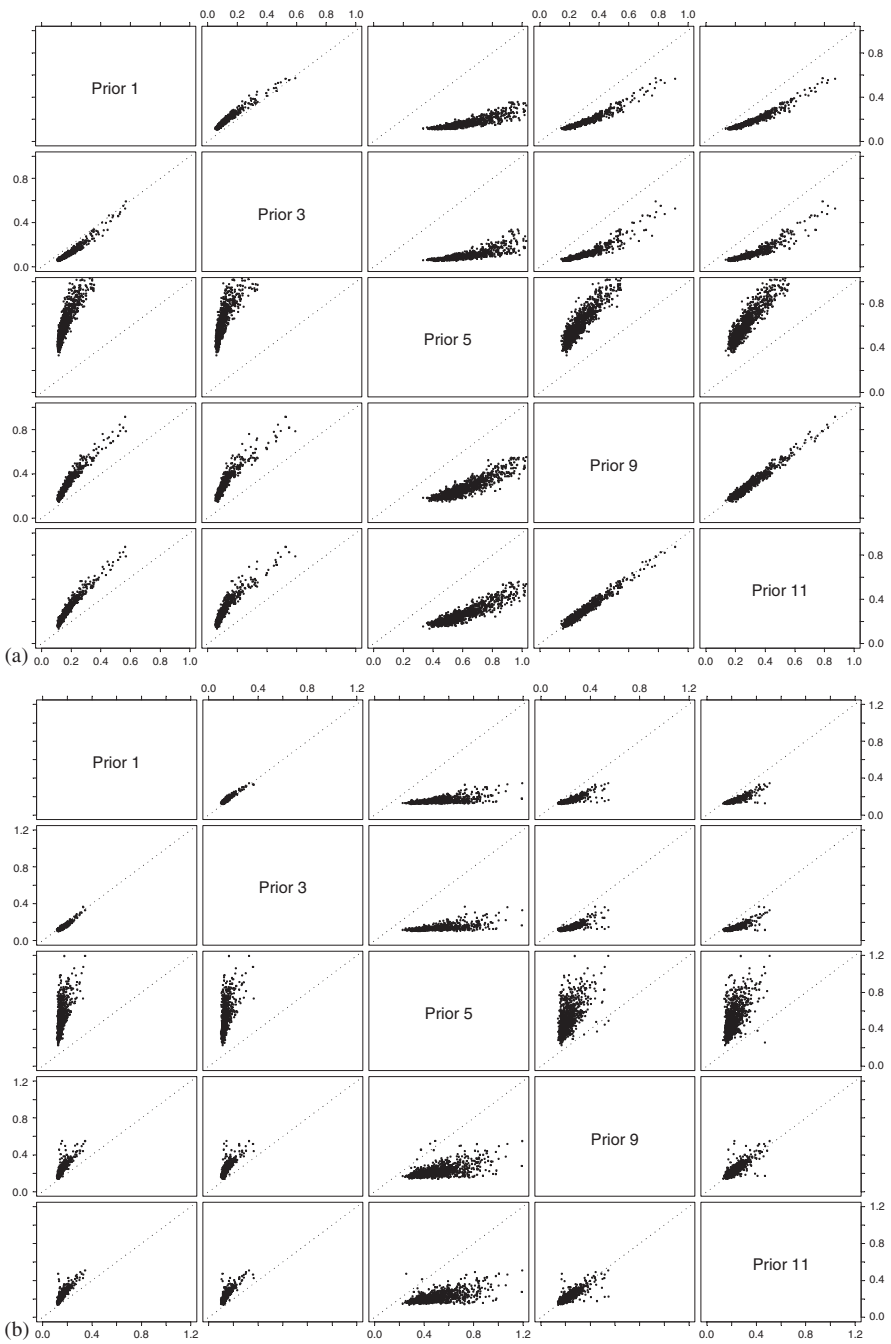

Figure 7. Scatter plot matrix for selected prior distributions for the five study scenario with a between-study standard deviation of 0.001, displaying: (a) the between-study standard deviation (median); and (b) the standard deviation of pooled log-odds ratio.

of the between-study standard deviation lead to higher estimates of the standard deviation of the pooled treatment effect (comparison of panels a and b).

Figure 8 shows similar agreement plots, but for meta-analyses with 10 studies and a between-study standard deviation of 0.8. These show that the agreement between estimates using different prior distributions for both the between-study standard deviation and the standard deviation of the pooled treatment effect is markedly much improved over the plots shown in Figure 7. This indicates that as the study size and the between-study deviation increases the influence of the prior distributions is reduced.

Figure 9 shows further agreement plots for meta-analyses with 30 studies and a simulated between-study standard deviation of 0.001. The plots show that, even when the number of studies is large, agreement is poor when the between-study standard deviation is small. However, a degree of caution is required when interpreting these plots due to issues of convergence as discussed below.

In general it is recommended to assess the convergence and mixing of MCMC chains on which inferences are based [33]. However, with 117 000 models fitted this is unrealistic for this simulation study. We therefore investigated informally (by visual inspection of the trace plots) the first 50 data sets in each scenario for each of the 13 prior distributions. The main problems we found were that the MCMC chain for the estimated between-study standard deviation occasionally got 'stuck' close to zero for some of the prior distributions and were slow mixing. This problem appeared to get more severe as the number of studies increased. Thus, for the data sets generated with 30 studies and a simulated between-study standard deviation of 0.001 there were a number of data sets where if a definitive analysis was being performed then in practice one would probably want to run the chains for longer. Two examples of such trace plots can be seen in Figure 10, where the trace plots are also shown on the log scale to aid interpretation. The prior distributions that visual inspection revealed to be particularly poor were Priors 3 and 4 (uniform on the log variance scale), Priors 11 and 12 (Gaussian on the standard deviation scale) and Prior 13 (DuMouchel). The convergence problems were less severe for data sets generated with 10 and 5 studies and a between-study standard deviation of 0.001. This problem getting 'stuck' close to zero has been recognized elsewhere [11, 34] Hence, non-convergence is clearly a possibility, and this emphasises the need for comprehensive diagnostic assessment to be used in routine application of even simple models [33].

Table III shows the mean values of the median pooled effect size, the mean of the standard deviation of the effect size and the coverage of the 95 per cent credible interval of the pooled effect size. Coverage was assessed by calculating 95 per cent credible intervals using the 2.5th and 97.5th percentiles of the distribution and evaluating the number of credible intervals that contained the known estimate. The table shows that the estimates of the pooled effect size appear to be unbiased. However, there is large variation in the standard deviation of the pooled effect size when the analysis consists of five studies. This variation is reduced, but still present, for analysis of 10 studies and reduced still further for 30 studies. The coverage of the pooled effect size tends to be too high when the between-study standard deviation is 0.001. This can be explained by the fact that when the true standard deviation is close to zero it will be upwardly biased as the MCMC sampler must always sample a positive value [11, 34] Coverage tends to improve as the number of studies analysed increases.

Table IV shows the mean values of the median between-study standard deviation, the mean of the standard deviation of the between-study standard deviation and the coverage of the 95

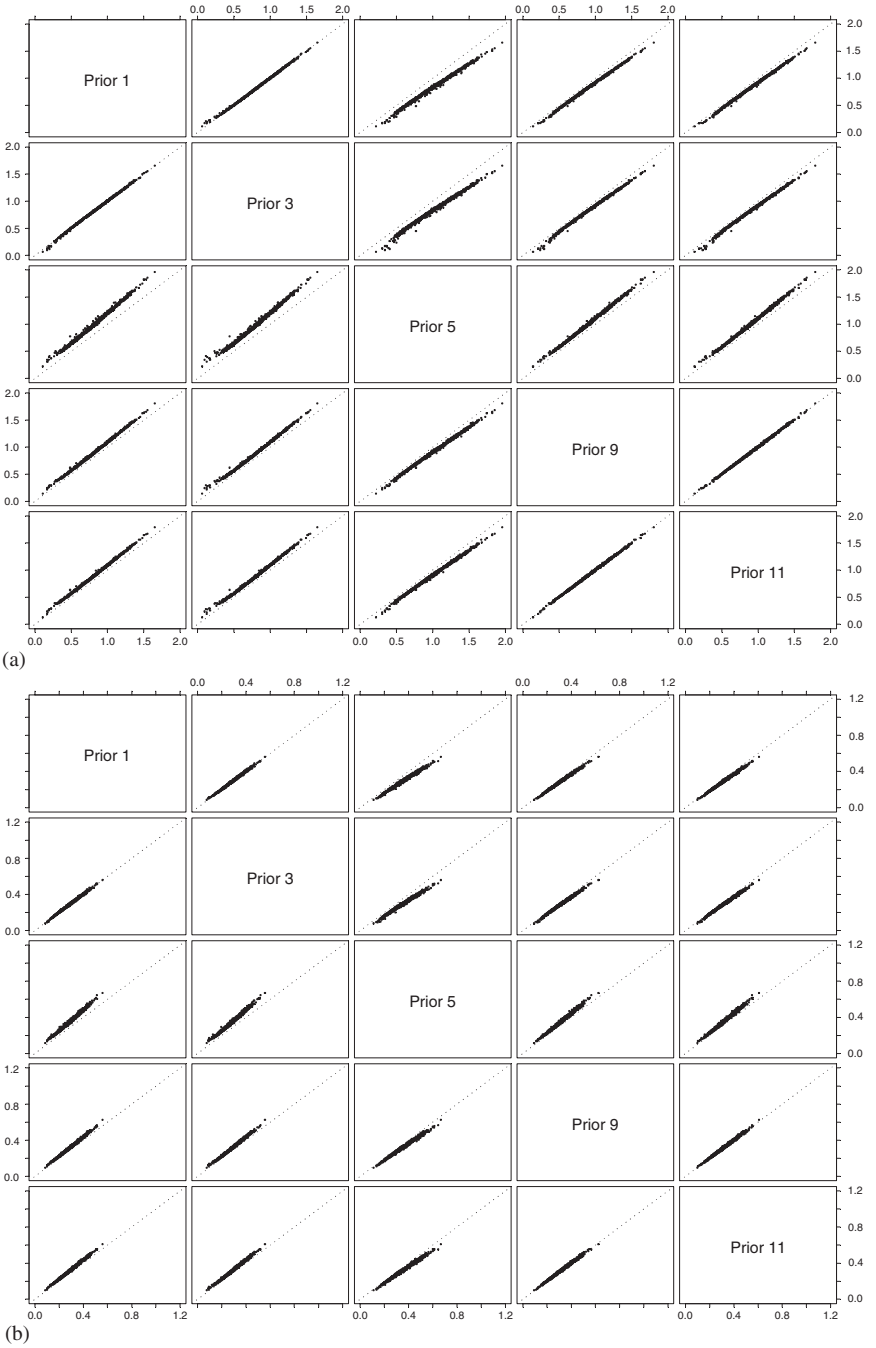

Figure 8. Scatter plot matrix for selected prior distributions for the 10 study scenario with a between-study standard deviation of 0.8, displaying: (a) the between-study standard deviation (median); and (b) the standard deviation of pooled log-odds ratio.

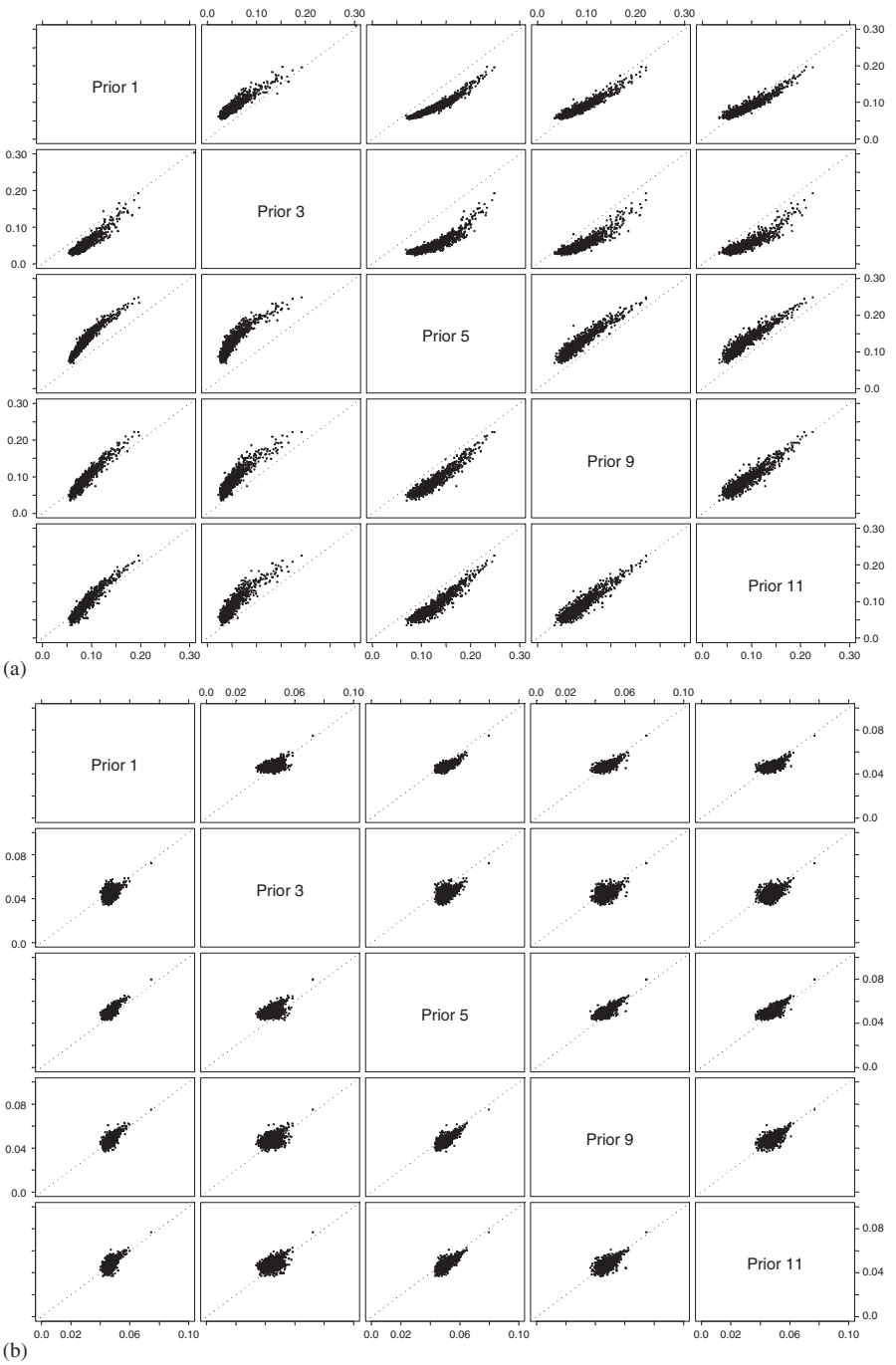

Figure 9. Scatter plot matrix for selected prior distributions for the 30 study scenario with a between-study standard deviation of 0.001, displaying: (a) the between-study standard deviation (median); and (b) the standard deviation of pooled log-odds ratio.

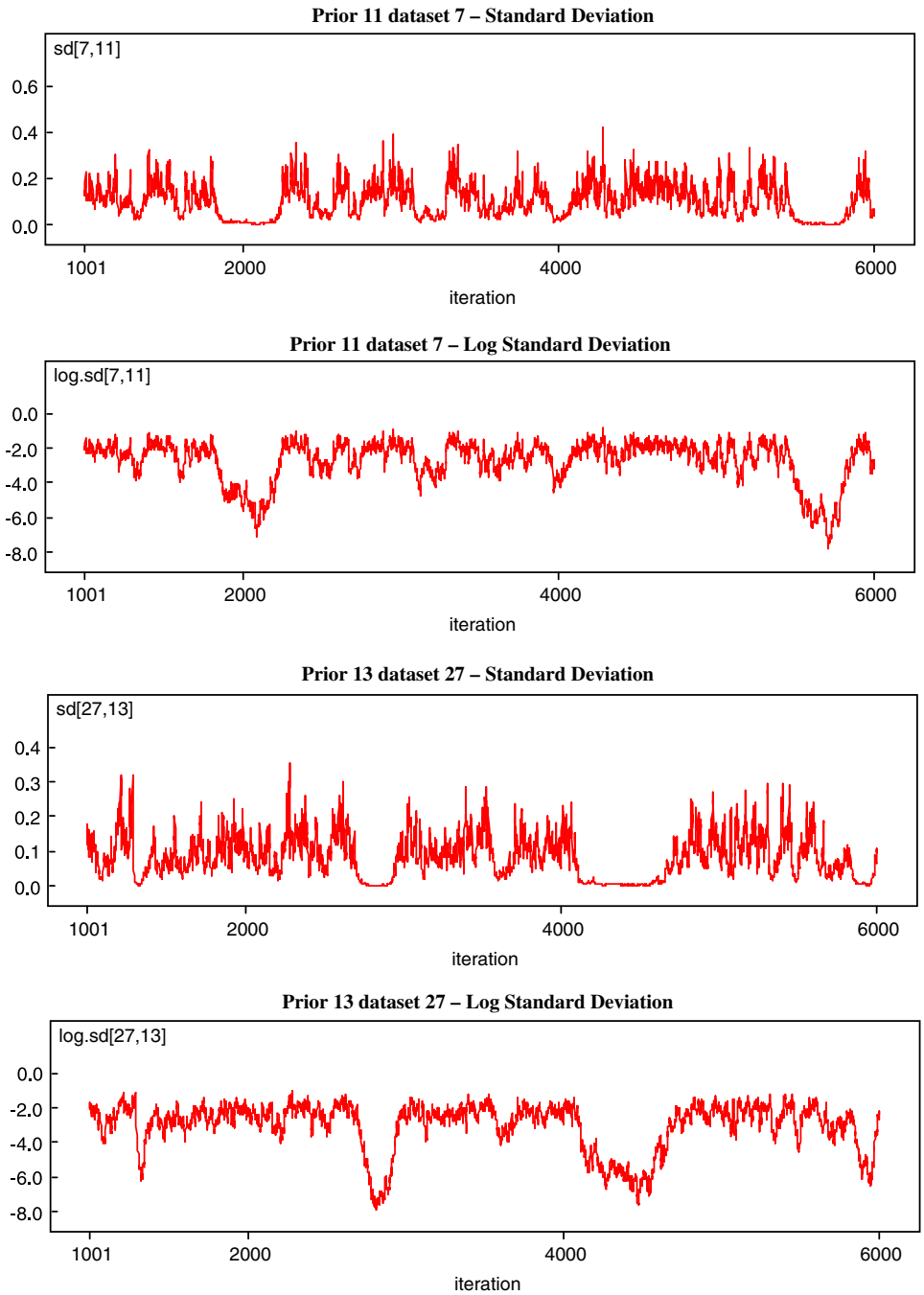

Figure 10. Example trace plots for four simulated data sets for the between-study standard deviation when convergence appeared to be a problem with 30 studies and a between-study standard deviation of 0.001.

per cent credible interval of the pooled effect size. Coverage was assessed as above. When the between-study standard deviation is near zero the MCMC estimate will always be upwardly biased as sampled numbers must always be positive, as explained above. With large samples, i.e. 30 studies, the between-study standard deviation is close to its nominal value when it is some distance from zero. However, there are still some problems with bias with 10 studies and particularly with five studies.

When the between-study standard deviation is close to zero the coverage is extremely poor due to the upward bias issue discussed above. For values away from zero coverage is improved, although one must take into account the complex relationship between bias and coverage.

5. DISCUSSION

We have performed a simulation study that demonstrates the potential influence of using prior distributions believed to be vague. We have generated data from a meta-analysis context, but the problems we have identified are likely to be generalizable to most areas in which hierarchical /variance component modelling is undertaken. Thirteen different prior distributions are assessed under nine different scenarios in which the size of the between unit variation and the number of units are varied. For each scenario we have assessed the frequentist properties of bias and coverage of the parameter estimates.

The results can be broadly summarised as follows. The point estimate of fixed effect (i.e. the log-odds ratio) has little bias. However, the use of the different prior distributions lead to problems with coverage of the fixed effect estimate. This is because of the variation in the estimates of the between unit standard deviation. Thus, even though all of these prior distributions were intended to be vague, their use could lead to different inferences. As the number of units increases the influence of the prior distribution is reduced, with the data truly dominating. However, there are potential problems when the true between unit standard deviation is close to zero. This is because the MCMC sampler is 'forced' to sample a positive value at every iteration of the sampler. A further problem is that when estimates are close to zero, poor mixing of the sampler can occur. This appeared to be more problematic as the number of units increased. There is no one prior distribution which performed best for all scenarios, but certain prior distributions performed particularly poorly in terms of the frequentist properties of bias and coverage. The priors that were uniform on the variance scale $(5, 6, 7, 8)$ were particularly poor with a small number of units and if a vague prior is desired there is little reason for their use.

There are a number of implications for practice resulting from this work. Firstly, when using priors distributions that are intended to be vague for the between unit variance, a sensitivity analysis is crucial. This is particularly important when the number of units is small and/or the estimated variance is close to zero. Although we have simulated data from a meta-analysis context, we feel that our results should apply to any area where random effects models are used, which display such characteristics. It seems unlikely that if these problems exist for simplistic models then they would disappear in more complex scenarios. This has been demonstrated by Browne and Draper [22] who observed bias and incorrect coverage when the number of units was small when using a random effects model with a Wishart prior distribution on correlated random effects for the intercept and slope in a hierarchical linear

Table III. Results of simulation study.

Prior distribution for variance	5 Studies			10 Studies			30 Studies		
	S.D. = 0	S.D. = 0.3	S.D. = 0.8	S.D. = 0	S.D. = 0.3	S.D. = 0.8	S.D. = 0	S.D. = 0.3	S.D. = 0.8
(1) $\frac{1}{\tau^2} \sim$ Gamma(0.001, 0.001)	**0.324**	**0.325**	**0.333**	**0.321**	**0.321**	**0.316**	**0.325**	**0.323**	**0.320**
	0.152	0.209	0.488	0.089	0.129	0.286	0.047	0.072	0.157
	99.4%	*94.4%*	*94.4%*	*97.3%*	*94.1%*	*93.1%*	*96.9%*	*94.1%*	*94.7%*
(2) $\frac{1}{\tau^2} \sim$ Gamma(0.1, 0.1)	**0.326**	**0.327**	**0.335**	**0.322**	**0.322**	**0.317**	**0.326**	**0.323**	**0.320**
	0.243	0.283	0.500	0.127	0.156	0.289	0.059	0.077	0.157
	100%	*99.5%*	*96.1%*	*99.7%*	*98.8%*	*94.0%*	*99.2%*	*96.0%*	*94.6%*
(3) $\log(\tau^2) \sim$ Uniform(−10.10)	**0.325**	**0.324**	**0.332**	**0.321**	**0.320**	**0.316**	**0.325**	**0.323**	**0.320**
	0.134	0.194	0.484	0.083	0.124	0.285	0.045	0.071	0.157
	99.0%	*91.5%*	*93.3%*	*96.3%*	*92.7%*	*93.3%*	*95.6%*	*93.6%*	*95.0%*
(4) $\log(\tau^2) \sim$ Uniform(−10, 1.386)	**0.324**	**0.324**	**0.332**	**0.321**	**0.320**	**0.316**	**0.326**	**0.323**	**0.320**
	0.133	0.187	0.409	0.083	0.124	0.281	0.044	0.071	0.157
	98.0%	*92.1%*	*92.6%*	*96.3%*	*92.2%*	*92.9%*	*95.4%*	*94.1%*	*94.6%*
(5) $\tau^2 \sim$ Uniform(1/1000, 1000)	**0.326**	**0.328**	**0.337**	**0.322**	**0.322**	**0.318**	**0.325**	**0.323**	**0.320**
	0.501	0.654	1.242	0.116	0.168	0.345	0.051	0.077	0.164
	100%	*100%*	*99.9%*	*99.3%*	*99.0%*	*96.8%*	*98.0%*	*95.8%*	*95.7%*
(6) $\tau^2 \sim$ Uniform(1/1000, 4)	**0.326**	**0.327**	**0.337**	**0.322**	**0.322**	**0.317**	**0.326**	**0.323**	**0.321**
	0.300	0.362	0.537	0.116	0.167	0.327	0.051	0.077	0.164
	100%	*100%*	*99.1%*	*99.3%*	*99.1%*	*96.6%*	*98.0%*	*95.7%*	*95.6%*
(7) $\frac{1}{\tau^2} \sim$ Pareto(1, 0.001)	**0.326**	**0.328**	**0.337**	**0.322**	**0.322**	**0.318**	**0.325**	**0.323**	**0.320**
	0.503	0.663	1.237	0.116	0.168	0.345	0.051	0.077	0.164
	100%	*100%*	*100%*	*99.3%*	*98.8%*	*96.6%*	*98.4%*	*95.5%*	*95.6%*

(8) $\frac{1}{\tau^2} \sim$ Pareto$(1, 0.25)$	**0.326**	**0.327**	**0.336**	**0.322**	**0.322**	**0.317**	**0.325**	**0.323**	**0.320**
	0.300	0.363	0.537	0.116	0.167	0.327	0.051	0.077	0.164
	100%	*100%*	*99.0%*	*99.3%*	*98.9%*	*96.5%*	*98.1%*	*95.8%*	*95.9%*
(9) $\tau \sim$ Uniform$(0, 100)$	**0.325**	**0.326**	**0.334**	**0.322**	**0.321**	**0.317**	**0.325**	**0.323**	**0.320**
	0.219	0.313	0.699	0.098	0.145	0.312	0.048	0.074	0.160
	100%	*98.0%*	*97.7%*	*98.3%*	*97.2%*	*94.5%*	*96.4%*	*94.4%*	*95.2%*
(10) $\frac{1}{\tau^2} \sim$ Pareto$(0.5, 0.0625)$	**0.325**	**0.326**	**0.334**	**0.321**	**0.321**	**0.317**	**0.326**	**0.323**	**0.320**
	0.211	0.266	0.472	0.098	0.145	0.303	0.047	0.074	0.160
	100%	*98.2%*	*96.8%*	*98.3%*	*97.5%*	*94.5%*	*97.0%*	*94.8%*	*94.9%*
(11) $\tau \sim$ N$(0, 100)I[0,]$	**0.325**	**0.327**	**0.335**	**0.321**	**0.321**	**0.317**	**0.325**	**0.323**	**0.320**
	0.216	0.304	0.652	0.097	0.145	0.312	0.047	0.074	0.160
	100%	*98.3%*	*97.4%*	*98.4%*	*97.0%*	*95.0%*	*96.6%*	*94.7%*	*95.2%*
(12) $\tau \sim$ N$(0, 1)I[0,]$	**0.325**	**0.326**	**0.334**	**0.321**	**0.321**	**0.316**	**0.325**	**0.323**	**0.320**
	0.182	0.238	0.429	0.097	0.142	0.284	0.047	0.074	0.158
	99.9%	*97.0%*	*95.1%*	*98.3%*	*97.4%*	*93.8%*	*96.9%*	*94.6%*	*94.9%*
(13) $\frac{1}{\tau^2} \sim$ Logistic(S_0)	**0.324**	**0.324**	**0.332**	**0.321**	**0.320**	**0.316**	**0.325**	**0.323**	**0.320**
	0.141	0.190	0.422	0.087	0.126	0.273	0.046	0.071	0.155
	99.0%	*92.7%*	*92.5%*	*96.8%*	*93.8%*	*92.6%*	*96.1%*	*93.6%*	*94.4%*

Note: Mean values of pooled median effect size (bold font)—true effect is 0.323, mean of the standard deviation of the pooled effect size (normal font) and coverage for 95% credible interval for pooled effect size (italic font).

Table IV. Results of simulation study.

Prior distribution for variance	5 Studies			10 Studies			30 Studies		
	S.D. = 0	S.D. = 0.3	S.D. = 0.8	S.D. = 0	S.D. = 0.3	S.D. = 0.8	S.D. = 0	S.D. = 0.3	S.D. = 0.8
(1) $\frac{1}{\tau^2} \sim$ Gamma(0.001, 0.001)	**0.124**	**0.242**	**0.780**	**0.099**	**0.258**	**0.772**	**0.077**	**0.283**	**0.788**
	0.156	0.235	0.524	0.082	0.134	0.253	0.046	0.071	0.125
	0.0%	*98.7%*	*91.7%*	*0.0%*	*95.0%*	*93.4%*	*0.0%*	*94.1%*	*94.6%*
(2) $\frac{1}{\tau^2} \sim$ Gamma(0.1, 0.1)	**0.361**	**0.434**	**0.831**	**0.276**	**0.371**	**0.787**	**0.200**	**0.321**	**0.791**
	0.232	0.279	0.508	0.101	0.134	0.249	0.045	0.066	0.124
	0.0%	*96.4%*	*97.4%*	*0.0%*	*95.6%*	*95.2%*	*0.0%*	*97.4%*	*94.7%*
(3) $\log(\tau^2) \sim$ Uniform(−10, 10)	**0.061**	**0.188**	**0.769**	**0.049**	**0.227**	**0.771**	**0.039**	**0.279**	**0.788**
	0.137	0.225	0.523	0.076	0.137	0.254	0.045	0.073	0.125
	0.0%	*97.0%*	*90.1%*	*0.0%*	*89.7%*	*93.3%*	*0.0%*	*93.0%*	*94.4%*
(4) $\log(\tau^2) \sim$ Uniform(−10, 1.386)	**0.060**	**0.187**	**0.735**	**0.049**	**0.226**	**0.768**	**0.038**	**0.279**	**0.788**
	0.131	0.203	0.320	0.075	0.137	0.234	0.045	0.073	0.125
	0.0%	*96.4%*	*91.2%*	*0.0%*	*89.3%*	*93.6%*	*0.0%*	*92.9%*	*94.5%*
(5) $\tau^2 \sim$ Uniform(1/1000, 1000)	**0.413**	**0.600**	**1.341**	**0.208**	**0.385**	**0.923**	**0.121**	**0.318**	**0.827**
	0.892	0.131	1.996	0.134	0.186	0.350	0.058	0.075	0.135
	0.0%	*88.8%*	*88.7%*	*0.0%*	*93.1%*	*92.9%*	*0.0%*	*94.8%*	*94.8%*
(6) $\tau^2 \sim$ Uniform(1/1000, 4)	**0.395**	**0.557**	**1.048**	**0.208**	**0.385**	**0.908**	**0.121**	**0.318**	**0.827**
	0.365	0.391	0.386	0.134	0.183	0.280	0.058	0.074	0.135
	0.0%	*91.0%*	*93.8%*	*0.0%*	*93.1%*	*93.6%*	*0.0%*	*94.8%*	*95.4%*
(7) $\frac{1}{\tau^2} \sim$ Pareto(1, 0.001)	**0.413**	**0.600**	**0.341**	**0.208**	**0.385**	**0.922**	**0.121**	**0.318**	**0.827**
	0.904	0.157	1.994	0.134	0.186	0.350	0.058	0.075	0.135
	0.0%	*89.1%*	*88.8%*	*0.0%*	*93.1%*	*93.2%*	*0.0%*	*94.9%*	*95.2%*

(8) $\frac{1}{\tau^2} \sim$ Pareto$(1, 0.25)$	**0.395**	**0.556**	**1.048**	**0.208**	**0.385**	**0.909**	**0.121**	**0.318**	**0.827**
	0.365	0.391	0.386	0.134	0.183	0.281	0.058	0.075	0.135
	0.0%	*91.0%*	*93.7%*	*0.0%*	*93.0%*	*93.4%*	*0.1%*	*94.7%*	*95.3%*
(9) $\tau \sim$ Uniform$(0, 100)$	**0.287**	**0.367**	**0.981**	**0.128**	**0.314**	**0.840**	**0.081**	**0.300**	**0.807**
	0.306	0.436	0.936	0.109	0.159	0.293	0.056	0.073	0.130
	0.8%	*96.6%*	*93.5%*	*3.7%*	*96.7%*	*94.5%*	*7.6%*	*94.2%*	*95.0%*
(10) $\tau^2 \sim$ Uniform$(0, 2)$	**0.207**	**0.367**	**0.963**	**0.129**	**0.314**	**0.840**	**0.081**	**0.300**	**0.807**
	0.279	0.366	0.581	0.110	0.159	0.292	0.056	0.073	0.130
	0.2%	*96.5%*	*94.4%*	*3.7%*	*96.6%*	*94.7%*	*9.2%*	*94.6%*	*94.9%*
(11) $\tau \sim$ N$(0, 100)I[0,]$	**0.207**	**0.367**	**0.974**	**0.129**	**0.314**	**0.840**	**0.080**	**0.300**	**0.807**
	0.296	0.410	0.796	0.109	0.159	0.293	0.056	0.073	0.130
	0.5%	*96.7%*	*94.5%*	*3.5%*	*96.6%*	*94.4%*	*10.4%*	*94.1%*	*95.0%*
(12) $\tau \sim$ N$(0, 1)I[0,]$	**0.192**	**0.330**	**0.784**	**0.126**	**0.306**	**0.784**	**0.080**	**0.298**	**0.794**
	0.207	0.251	0.334	0.106	0.148	0.230	0.056	0.072	0.124
	0.5%	*97.2%*	*96.3%*	*3.5%*	*97.0%*	*95.5%*	*9.2%*	*94.3%*	*95.0%*
(13) $\frac{1}{\tau^2} \sim$ Logistic(S_0)	**0.104**	**0.218**	**0.709**	**0.083**	**0.247**	**0.742**	**0.061**	**0.281**	**0.779**
	0.139	0.198	0.398	0.084	0.130	0.232	0.050	0.071	0.122
	4.1%	*99.0%*	*90.3%*	*8.8%*	*94.1%*	*92.7%*	*15.0%*	*94.0%*	*94.6%*

Note: Mean values of median between-study standard deviation (bold font)—true effect is provided in column headings, mean of the standard deviation of the between-study standard deviation (normal font) and coverage for 95% credible interval for the between-study standard deviation (italic font).

model. We welcome the new addition to the WinBUGS examples that demonstrates how a number of different prior distributions can be fitted to the same model and their results plotted in a similar format to those of Figure 3, allowing immediate comparison with relative ease [35]. When there is only a sparse number of units it may be valuable to consider true prior information as with such sparse data, the between unit variance is never going to be estimated well using no prior information. We observed that convergence was potentially problematical when the estimated between unit standard deviation was close to zero and this specific problem appeared to get worse as the number of units increased. This highlights the need to check for convergence routinely, even in models that may be perceived as elementary, such as those used here.

Some of the prior distributions used here are clearly unrealistic in that they give support to unfeasibly large values for the between unit standard deviations. The use of such priors has been criticized [36]. We would therefore recommend investigation of prior distributions that are vague within a realistic range for the data set under consideration within a sensitivity analysis. This approach has been considered previously in Bayesian meta-analysis [37]. A related method is to use previous empirical observations to derive reasonable prior distribution. This has been considered in a meta-analysis context where the prior distribution for the between-study variance has been derived from investigation of the observed heterogeneity from previous meta-analyses in the same clinical area [38]. A further approach to the choice of prior distribution is to use uniform shrinkage priors [10, 39, 40]. These are similar to the approach of DuMouchel used here (prior 13). Whichever prior distributions are used for the main and sensitivity analyses, on the grounds of transparency and following previous recommendations [41], we strongly advocate the reporting of all prior distributions considered, their impact on results and an assessment of their convergence.

All analyses were performed using WinBUGS (version 1.4), and hence it should be realized that results may not just be theoretical differences but may also reflect how the software implements the MCMC methods. Use of alternative methods of estimation or software could potentially lead to different results. However, with an increasing number of analysts using MCMC methods and WinBUGS in particular for a wide range of models, we feel that it is an important message that the use of vague prior distributions should be treated with a degree of caution and that sensitivity analysis to the choice of vague prior distributions, particularly in small samples, is crucial to any analysis.

APPENDIX I: SIMULATION IN BUGS

```
model hiersim {

# nsim=number of simulations
# nstud=number of studies
# nprior=number of different priors for variance

# create replicates of datasets

  for(i in 1:nsim){
    for(j in 1:nstud){
      for(k in 1:nprior){
```



```
        y[i,j,k] <- y.dat[i,j]
        }
    }
  }

# loop over datasets i
  for(i in 1:nsim){

# loop over number of priors k
      for(k in 1:nprior){

# loop over number of studies j
        for(j in 1:nstud){
          y[i,j,k]   dnorm(mu[i,j,k],prec[i,j])
          mu[i,j,k]   dnorm(theta[i,k],tau[i,k])
        }

# prior for pooled effect
    theta[i,k]   dnorm(0,0.0001)
    }

# priors for variances

# prior 1 - Gamma(0.001,0.001) on precision
  tau[i,1]   dgamma(0.001,0.001)
  var[i,1] <- 1/tau[i,1]
  sd[i,1] <- sqrt(var[i,1])

# prior 2 - Gamma(0.1,0.1) on precision
  tau[i,2]   dgamma(0.1,0.1)
  var[i,2] <- 1/tau[i,2]
  sd[i,2] <- sqrt(var[i,2])

# prior 3 - Uniform [-10,10] on log variance
  lv[i,3]   dunif(-10,10)
  log(var[i,3]) <- lv[i,3]
  tau[i,3] <- 1/var[i,3]
  sd[i,3] <- sqrt(var[i,3])

# prior 6 - Uniform [0,4] on variance
  var[i,6]   dunif(0,4)
  tau[i,6] <- 1/var[i,6]
  sd[i,6] <- sqrt(var[i,6])

# prior 4 - Uniform [-10,4] on log variance
  lv[i,4]   dunif(-10,1.386)
  log(var[i,4]) <- lv[i,4]
  tau[i,4] <- 1/var[i,4]
  sd[i,4] <- sqrt(var[i,4])

# prior 5 - Uniform [0,1000] on variance
```



```
    var[i,5]   dunif(0,1000)
    tau[i,5] <- 1/var[i,5]
    sd[i,5] <- sqrt(var[i,5])

  # prior 7 - Pareto(1,0.001) (equiv to unif(0,1000) on variance)
    tau[i,7]   dpar(1,0.001)
    var[i,7] <- 1/tau[i,7]
    sd[i,7] <- sqrt(var[i,7])

  # prior 8 - Pareto(1,0.25) (equiv to unif(0,2) on sd)
    tau[i,8]   dpar(1,0.25)
    var[i,8] <- 1/tau[i,8]
    sd[i,8] <- sqrt(var[i,8])

  # prior 9 - Uniform(0,100) on sd
    tau[i,9] <- 1/var[i,9]
    var[i,9] <- pow(sd[i,9],2)
    sd[i,9]   dunif(0,100)

  # prior 10 - Uniform(0,2) on sd
    tau[i,10] <- 1/var[i,10]
    var[i,10] <- pow(sd[i,10],2)
    sd[i,10]   dunif(0,2)

  # prior 11 - half-normal on sd var=100
    tau[i,11] <- 1/var[i,11]
    var[i,11] <- pow(sd[i,11],2)
    sd[i,11]   dnorm(0,0.01)I(0,)

  # prior 12 - half-normal on sd - var=1
    tau[i,12] <- 1/var[i,12]
    var[i,12] <- pow(sd[i,12],2)
    sd[i,12]   dnorm(0,1)I(0,)

  # prior 13 - log-logistic on sd (from DuMouchel)
    p[i]   dunif(0,1)
    sd[i,13] <- p[i] *s0[i]/(1-p[i])
    tau[i,13] <- 1/var[i,13]
    var[i,13] <- pow(sd[i,13],2)

    s0[i] <- sqrt(nstud/sum(prec[i,]))
    }
  }
```

REFERENCES

1. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. An introduction to Bayesian methods in health technology assessment. *British Medical Journal* 1999; **319**:508–512.
2. Congdon P. *Bayesian Statistical Modelling*. Wiley: New York, 2001.
3. Berry DA, Stangl DK. *Bayesian Biostatistics*. Marcel Dekker: New York, 1996.
4. Brooks SP. Markov chain Monte Carlo method and its application. *The Statistician* 1998; **47**:69–100.

5. Spiegelhalter DJ, Thomas A, Best NG, Gilks WR. *BUGS*: *Bayesian Inference Using Gibbs Sampling, Version 0.50*. MRC Biostatistics Unit: Cambridge, 1996.
6. Spiegelhalter DJ, Thomas A, Best NG, Lunn D. *WinBUGS*, *Version 1.4*, *User Manual*. MRC Biostatistics Unit: Cambridge, 2001.
7. Best NG, Spiegelhalter DJ, Thomas A, Brayne CEG. Bayesian-analysis of realistically complex-models. *Journal of the Royal Statistical Society, Series A, Statistics in Society* 1996; **159**:323–342.
8. Sutton AJ, Abrams KA, Jones DR, Sheldon TA, Song F. *Methods for Meta-analysis in Medical Research*. Wiley: Chichester, 2000.
9. Turner RM, Omar RZ, Thompson SG. Bayesian methods of analysis for cluster randomized trials with binary outcome data. *Statistics in Medicine* 2001; **20**:453–472.
10. Spiegelhalter DJ. Bayesian methods for cluster randomized trials with continuous responses. *Statistics in Medicine* 2001; **20**:435–452.
11. Burton PR, Tiller KJ, Gurrin LC, Cookson W, Musk AW, Palmer LJ. Genetic variance components analysis for binary phenotypes using generalized linear mixed models (GLMMS) and Gibbs sampling. *Genetic Epidemiology* 1999; **17**:118–140.
12. Goldstein H, Spiegelhalter DJ. League tables and their limitations—statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society, Series A, Statistics in Society* 1996; **159**:385–409.
13. Dixon DO, Simon R. Bayesian subset analysis in a colorectal cancer clinical trial. *Statistics in Medicine* 1992; **11**:13–22.
14. Goldstein H. *Multilevel Statistical Models*. Edward Arnold: London, 1995.
15. Kass RE, Wasserman L. The selection of prior distributions by formal rules. *Journal of the American Statistical Association* 1996; **91**:1343–1370.
16. Walley P, Gurrin LC, Burton PR. Analysis of clinical data using imprecise prior probabilities. *The Statistician* 1996; **45**:457–486.
17. Fisher LD. Comments on Bayesian and frequentist analysis and interpretation of clinical trials. *Control Clinical Trials* 1996; **17**:423–434.
18. Irony TZ, Singpurwalla ND. Noninformative priors do not exist: a discussion with Jose M. Bernardo. *Journal of Statistical Infererence and Planning* 1997; **65**:159–189.
19. Hughes MD. Reporting Bayesian analysis of clinical trials. *Statistics in Medicine* 1993; **12**:1651–1653.
20. Browne WJ, Draper D. A comparison of Bayesian and likelihood methods for fitting multilevel models. *Nottingham Statistics Research Report 04-01*, University of Nottingham, U.K., 2004.
21. Rasbash J, Browne WJ, Goldstein H, Yang M, Plewis IF, Healy MJR, Woodhouse G, Draper D, Langford IH, Lewis T. *A User's Guide to MLwiN, version 2.1d*. Institute of Education, University of London, 2002.
22. Browne WJ, Draper D. Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models. *Computational Statistics* 2000; **15**:391–420.
23. Glasziou PP, Del Mar CB, Sanders SL, Hayem M. Antibiotics for acute otitis media in children (Cochrane Review). *The Cochrane Library*, vol. 4. Wiley: New York, 2003.
24. Cochran WG. The combination of estimates from different experiments. *Biometrics* 1954; **10**:101–129.
25. Sutton AJ, Abrams KA. Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research* 2001; **10**:277–303.
26. Gelfand AE, Sahu SK, Carlin BP. Efficient parametrizations for normal linear mixed models. *Biometrika* 1995; **82**:479–488.
27. Spiegelhalter DJ, Thomas A, Best NG, Gilks WR. *BUGS Examples vol. 1*, *Version 0.5* (*version ii*). MRC Biostatistics Unit: Cambridge, 1996.
28. Spiegelhalter DJ, Thomas A, Best NG, Gilks WR. *BUGS Examples vol. 2*, *Version 0.5* (*version ii*). MRC Biostatistics Unit: Cambridge, 1996.
29. Scurrah KJ, Palmer LJ, Burton PR. Variance components analysis for pedigree-based censored survival data using generalized linear mixed models (GLMMs) and Gibbs sampling in BUGS. *Genetic Epidemiology* 2000; **19**:127–148.
30. Spiegelhalter DJ, Abrams KR, Myles J. *Bayesian Approaches to Clinical Trials and Health-care Evaluation*. Wiley: London.
31. Thompson SG, Smith TC, Sharp SJ. Investigating underlying risk as a source of heterogeneity in meta-analysis. *Statistics in Medicine* 1997; **16**:2741–2758.
32. DuMouchel W, Normand SL. Computer-modelling and graphical strategies for meta-analysis. In *Meta-analysis in Medicine and Health Policy*, Stangl DK, Berry DA (eds). Marcel Dekker: New York, 2000; 127–178.
33. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 1998; **7**:434–455.
34. Zeger SL, Karim MR. Generalized linear-models with random effects—a Gibbs sampling approach. *Journal of the American Statistical Association* 1991; **86**:79–86.
35. Spiegelhalter DJ, Thomas A, Best NG. Sensitivity to prior distributions: application to Magnesium meta-analysis. *WinBUGS*, *Version 1.4*, *User Manual*. MRC Biostatistics Unit: Cambridge, 2001.

36. Greenland S. Probability logic and probabilistic induction. *Epidemiology* 1998; **9**:322–332.
37. Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random effects meta-analysis: a comparative study. *Statistics in Medicine* 1995; **14**:2685–2699.
38. Higgins JPT, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine* 1996; **15**:2733–2749.
39. Daniels MJ. A prior for the variance in hierarchical models. *The Canadian Journal of Statistics* 1999; **27**:567–578.
40. Natarajan R, Kass RE. Reference Bayesian methods for generalised linear mixed models. *Journal of the American Statistical Association* 2000; **95**:227–237.
41. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. Bayesian methods in health technology assessment: a review. *Health Technology Assessment* 2000; **4**(38):1–130.

