STATISTICS IN MEDICINE, VOL. 17, 201–218 (1998)

APPROXIMATE BAYESIAN INFERENCE FOR RANDOM EFFECTS META-ANALYSIS

KEITH ABRAMS[1]* AND BRUNO SANSÓ[2]

[1]*Department of Epidemiology and Public Health, University of Leicester, U.K.*
[2]*Centro de Estadísticas y Software Matemático and Departamento de Cómputo Científico y Estadística Universidad Simon Bolívar, Caracas, Venezuela*

SUMMARY

Whilst meta-analysis is becoming a more commonplace statistical technique, Bayesian inference in meta-analysis requires complex computational techniques to be routinely applied. We consider simple approximations for the first and second moments of the parameters of a Bayesian random effects model for meta-analysis. These computationally inexpensive methods are based on simple analytical formulae that provide an efficient tool for a qualitative analysis and a quick numerical estimation of posterior quantities. They are shown to lead to sensible approximations in two examples of meta-analyses and to be in broad agreement with the more computationally intensive Gibbs sampling. © 1998 John Wiley & Sons, Ltd.

Statist. Med., **17**, 201–218 (1998)

1. INTRODUCTION

Over the last decade meta-analysis has become an accepted part of medical research.[1] In wanting to summarize evidence from a number of studies of the same type of design, that is, randomized controlled trials, cohort studies, case-control studies, a variety of statistical methods have been proposed.[1] Whilst not all of the methods proposed have involved the actual combination of the size of effects from the individual studies, the most widely used methods have. Of these methods the so-called *fixed effect model* is the most common. In such a model it is assumed that there is a true population effect, and that all the individual studies are estimating it.[2] A number of estimators for the population effect have been proposed, with many adopting a weighted estimator, the weights being a function of the precision of individual studies.[2]

If the heterogeneity, in terms of the effect sizes, between the studies is 'great' then a number of authors have proposed a *random effects model*.[3–6] In such models each individual study is assumed to be estimating its own, unknown, true effect which in turn is a perturbation about an overall population effect. Whilst intuitively appealing, some have argued that random effects models may mask the true reasons for the underlying heterogeneity, and that possible explanations should be investigated.[7] Perhaps pragmatically, generalized linear mixed models[8] offer the greatest flexibility in allowing both fixed covariates, thought to explain some of the heterogeneity, together with a random component to accomodate unexplained heterogeneity.[1,6]

* Correspondence to: Keith Abrams, Department of Epidemiology and Public Health, University of Leicester, 22–28 Princess Road West, Leicester LE1 6TP, U.K., E-mail: kral@le.ac.uk

CCC 0277–6715/98/020201–18$17.50
© 1998 John Wiley & Sons, Ltd.

Received December 1994
Revised March 1997

Whilst it would appear that such a range of modelling strategies would be able to accommodate a range of scenarios, a number of issues remain. Discrimination between fixed and random effect models in a particular situation has been advocated using a χ^2 test for heterogeneity, which it is accepted has low statistical power.[2,6,7] When a random effects model has been chosen, estimation of the variability of the population effect is sometimes problematic, and a number of authors report obtaining negative estimates in various circumstances.[4–6]

Whilst not pretending to be a panacea, the Bayesian approach would appear to offer the meta-analyst a more flexible (and robust) modelling strategy. In particular, there is a number of advantages conferred by adopting such an approach. The first is the ability to include 'background' information that may be thought to be pertinent to the clinical question being addressed[1], whilst the second is the fact that in estimating the true effects of individual studies they can in some way 'borrow strength' from other 'similar' studies.[9] In a similar manner, and more pragmatically, the fully Bayesian approach takes account of all parameter uncertainty,[10,11] a feature which is particularly important when considering the estimation of study-specific effects or the prediction of an effect likely to be observed in a future study.[12] Though empirical Bayes (EB) methods have been suggested as a means of allowing for *all* parameter uncertainty, they still require the use of Monte Carlo simulation or numerical integration methods.[10,11] More recently, methods based on the use of profile likelihood have also been advocated for estimating overall population effect whilst simultaneously allowing for the fact that the between-study heterogeneity had also been estimated.[13] Finally, a third advantage of the Bayesian approach is that it formalizes the methods used in cumulative meta-analyses[14] in which information/beliefs regarding treatment effects is sequentially updated as more trials become available.

A number of authors have advocated a variety of Bayesian approaches to meta-analysis. Some have adopted a hierarchical Gaussian modelling approach[9,11,15–18] in which the observed study effects are assumed to be Normally distributed about their true, but unknown, effects and that at the next level these unknown true effects are themselves assumed to be Normally distributed about an overall population effect, with suitable prior distributions assumed for the unknown variances and the overall population mean. Such a modelling strategy parallels the two-stage linear model of Lindley and Smith.[19] Other approaches have concentrated on assuming that within an individual study the event rates in the two comparison groups each follow a binomial distribution, and that only after taking some transformation of these rates, frequently logistic in nature, are the individual study effects related to an overall population effect.[20–22]

The models we adopt here are of a hierarchical Gaussian nature. The observed study effects are assumed to follow Normal distributions in which the variances are assumed unknown, but with non-informative improper prior distributions. The means of these distributions are themselves assumed to arise from a Normal distribution, in which the unknown mean represents the overall population effect, and has a Uniform prior distribution, and the variance, representing the between-study heterogeneity, is also assumed unknown and has an inverse gamma prior distribution. Such a prior specification, especially for the between-study heterogeneity, enables a variety of situations to be accommodated, from the case when there is a relatively little prior information available to the case when there may be substantial information *a priori* either in the form of subjective beliefs or in the form of quantitative evidence from other sources.

Implementing a Bayesian model such as the one that we propose in this paper presents the difficulty of having to evaluate integrals over several dimensions, in order to investigate the properties of the joint posterior distribution for the parameters, a task that is not easily achieved. Hence, approximations that produce simple formulae, and that are accurate for moderate to large

 © 1998 John Wiley & Sons, Ltd.

sample sizes, provide a very efficient tool to estimate certain posterior quantities of interest. In particular, the use of such approximations also enables an assessment of the sensitivity of posterior estimates to the choice of prior distributions to be made easily. As opposed to Laplace approximation,[23] modal approximations[24] and Markov chain Monte Carlo (MCMC) approximation methods,[25] the approximations proposed in this paper provide simple closed-form formulae, that facilitate a qualitative analysis of the behaviour of posterior quantities.

The rest of the paper is organized with Section 2 describing briefly two motivating examples of meta-analyses in which there are varying amounts of prior information and varying degrees of between-study heterogeneity. Section 3 outlines a general Bayesian random effects model for meta-analysis, and briefly describes parameter estimation using closed-form and Gibbs sampling approximation methods. Section 4 presents the results of applying the two methods of parameter estimation discussed to the examples in Section 2. Finally, Section 5 discusses the approximations presented here and the role Bayesian inference has to play in meta-analysis generally.

2. EXAMPLES

2.1. Infection Example

A previously conducted meta-analysis of randomized trials investigated the evidence of clinical benefits for the selective decontamination of the digestive tract for patients in intensive care units.[26] The data, presented in Table I, consist of 22 randomized trials intended to investigate the clinical benefits of selective decontamination of the digestive tract. In each trial patients in intensive care units were randomized to receive either a combination of non-absorbable antibiotics (treatment group) or no treatment (control group). For each trial, the number of respiratory tract infections in the treatment and control group were then recorded, and the odds ratio for developing an infection in the treatment group compared to the control, together with its associated 95 per cent confidence interval (CI) were calculated.

The data had originally been analysed using a fixed effect model,[26] but as Smith *et al.*[22,27] have remarked, there appears to be a considerable degree of heterogeneity present in the data, and a random effects to model would appear to be more appropriate. In fact, a χ^2 test for heterogeneity reveals that there is substantial evidence for the use of a random effect model; $\chi^2_{\mathrm{HET}} = 58{\cdot}0$, d.f. $= 21$ and $P = 0{\cdot}00001$.

Whilst there is no explicit prior information available for this example, previous Bayesian analyses[22,27] have used a range of different hypothesized prior distributions for the between-study variability. The one used here is derived subjectively by considering the likely between-study variability that would be observed[27] and is described in further detail in Section 4.1.

2.2. Dentifrice Example

The second example concerns a previously published meta-analysis which was conducted of all randomized controlled trials comparing sodium monofluorophosphate (SMFP) to sodium fluoride (NaF) dentifrices (toothpastes) in the prevention of caries development.[28] The outcome in each trial was the change, from baseline, in the decayed missing (due to caries) filled surface (DMFS) dental index[29] at three years follow-up. Of 12 studies identified as meeting the inclusion criteria, 9 considered a straight comparison of NaF and SMFP. Table II displays the data from these 9 studies, in terms of mean change in DMFS index for each treatment and the difference in mean change in DMFS index between treatments, SMFP – NaF, together with associated 95 per

© 1998 John Wiley & Sons, Ltd.

Table I. Randomized evidence, in terms of respiratory tract infections, regarding the use of selective decontamination of the digestive tract for patients in intensive care units (Digestive Tract Trialists' Collaborative Group)[26]

Study	Treated Infections	Treated Total	Control Infections	Control Total	Odds Ratios T/C	95% CI
1	7	47	25	54	0·21	(0·08, 0·55)
2	4	38	24	41	0·09	(0·03, 0·30)
3	20	96	37	95	0·42	(0·22, 0·79)
4	1	14	11	17	0·06	(0·01, 0·44)
5	10	48	26	49	0·24	(0·10, 0·58)
6	2	101	13	84	0·13	(003, 0·53)
7	12	161	38	170	0·28	(0·15, 0·57)
8	1	28	29	60	0·06	(0·01, 0·33)
9	1	19	9	20	0·10	(0·02, 0·64)
10	22	49	44	47	0·06	(0·02, 0·22)
11	25	162	30	160	0·79	(0·45, 1·41)
12	31	200	40	185	0·67	(0·40, 1·12)
13	9	39	10	41	0·93	(0·34, 2·56)
14	22	193	40	185	0·47	(0·27, 0·83)
15	0	45	4	46	0·10	(0·01, 1·99)
16	31	131	60	140	0·42	(0·25, 0·70)
17	4	75	12	75	0·32	(0·10, 0·99)
18	31	220	42	225	0·72	(0·43, 1·19)
19	7	55	26	57	0·18	(0·07, 0·46)
20	3	91	17	92	0·17	(0·05, 0·56)
21	14	25	23	23	0·03	(0·00, 0·49)
22	3	65	6	68	0·54	(0·14, 2·07)

Table II. Randomized evidence comparing sodium fluoride (NaF) with sodium monofluorophosphate (SMFP) dentrifrices in terms of differences from baseline in DMFS dental index (Johnson)[28]

Study	NaF N	NaF Mean	NaF SD	SMFP N	SMFP Mean	SMFP SD	SMFP–NaF	95% CI
1	134	5·96	4·24	113	6·82	4·72	+ 0·86	(−0·26, + 1·98)
2	175	4·74	4·64	151	5·07	5·38	+ 0·33	(−0·76, + 1·42)
3	137	2·04	2·59	140	2·51	3·22	+ 0·47	(−0·22, + 1·16)
4	184	2·70	2·32	179	3·20	2·46	+ 0·50	(+0·01, + 0·99)
5	174	6·09	4·86	169	5·81	5·14	− 0·28	(−1·34, + 0·78)
6	754	4·72	5·33	736	4·76	5·29	+ 0·04	(−0·50, + 0·58)
7	209	10·10	8·10	209	10·90	7·90	+ 0·80	(−0·73, + 2·33)
8	1151	2·82	3·05	1122	3·01	3·32	+ 0·19	(−0·07, + 0·45)
9	679	3·88	4·85	673	4·37	5·37	+ 0·49	(−0·06, + 1·04)

cent CIs. The original analysis used a fixed effect model, against which a χ^2 test for heterogeneity provided no substantial evidence; $\chi^2_{\mathrm{HET}} = 5{\cdot}4$, d.f. $= 8$ and $P = 0{\cdot}71$.

The three remaining studies out of the 12 considered the comparison of combination treatment, NaF together with SMFP, with the use of SMFP alone. Table III displays the data for these three

 © 1998 John Wiley & Sons, Ltd.

Table III. Randomized evidence comparing combined sodium fluoride (NaF) with sodium monofluorophosphate (SMFP) with sodium monofluorophosphate (SMFP) dentrifices in terms of differences from baseline in DMFS dental index (Johnson)[28]

Study	NaF + SMFP			SMFP			NaF + SMFP − SMFP			
	N	Mean	SD	N	Mean	SD	Difference	SD	N	95% CI
1	228	8·46	6·19	230	9·30	6·67	−0·84	0·60	458	(−2·02, +0·34)
2	858	3·67	4·59	827	3·74	4·84	−0·07	0·23	1685	(−0·52, +0·38)
3	512	11·27	7·47	515	11·16	7·94	+0·11	0·48	1027	(−0·83, +1·05)

studies in terms of mean change in DMFS index for each treatment and the difference in mean change in DMFS index between treatments, together with associated 95 per cent CIs. Without making the assumption that there is no form of interaction between NaF and SMFP this data cannot be combined directly with that of the direct comparison trials. However, it does contain some information about the between-study variability that is likely to be observed, and therefore could be used to derive a prior distribution regarding the likely between-study heterogeneity. The derivation of such a prior distribution is considered in more detail in Section 4.2.

3. METHODS

3.1. A random effects model

When all the available heterogeneity between studies has been explained in terms of explanatory factors, both at the study level and the patient level, any estimate of the overall treatment effect, and individual study effects must take account of the remaining heterogeneity. This is most appropriately accomplished by using a random effects model. A number of random effects models have been proposed, most notably DerSimonian and Laird.[5] By assuming a Gaussian error structure the following random effect model is obtained:[14]

$$y_i \sim \mathrm{N}[\theta_i, \sigma_i^2/n_i]$$

$$\theta_i \sim \mathrm{N}[\mu, \tau^2] \quad i = 1, \ldots, k. \tag{1}$$

Under the assumption that the data from each study are Normally distributed, as in the dentifrice example, y_i is the difference in mean responses for the ith study, σ_i^2 is the within-study variance, and $n_i = n_{1i}n_{2i}/(n_{1i} + n_{2i})$, where n_{1i} and n_{2i} are the number of patients in the treatment and control groups, respectively, in the ith study. If the original patient data correspond to binary outcomes, as in the infection example, then y_i is the logarithm of the odds ratio and model (1) is based on a Normal approximation to the logarithm of the odds of an event. Therefore, letting y_i be the logarithm of the odds ratio and following Cox and Snell,[30] σ_i^2 becomes $1/p_{1i}(1 - p_{1i}) + 1/p_{2i}(1 - p_{2i})$ where p_{1i} and p_{2i} are the probabilities of events in the treatment and control, respectively, in the ith study. Finally, in this case n_i is the number of patients in the treatment or control group, which are assumed to be equal, that is, $n_i = n_{1i} = n_{2i}$. For the case when $n_{1i} \neq n_{2i}$ there are a number of possible alternatives for n_i, including the minimum, maximum or average of n_{1i} and n_{2i}. Choice between alternatives will depend upon the specific context, but ideally a sensitivity analysis should be performed.

© 1998 John Wiley & Sons, Ltd.

In a model such as (1), θ_i is the true, but unknown, effect in the ith study, and μ is the unknown population effect, and it is this quantity that is often of most interest, since it represents the pooled effect indicated by the studies. Finally, τ^2 is the population variance, or the between-study variability, and is also of interest since it is a measure of how variable the effect is within a population. In the case when $\tau^2 = 0$ a fixed effect model is obtained.

The structure of (1) is similar to the hierarchical Bayesian models proposed by a number of authors.[24,31–33] In a Bayesian setting we would wish to put prior distributions on the unknown parameters of the model, that is, σ_i^2s, μ and τ^2. In practice prior information is only likely to be available possibly for μ, the population treatment effect, and τ^2 the population variability of the effect over studies. Therefore for σ_i^2, the individual trial variances, we would wish to assume a vague non-informative set of prior distributions. Following the common practice of using a Jeffreys' prior (Box and Tiao, p. 426)[33] we shall assume that $P(\sigma_i^2) \propto 1/\sigma_i^2$. We have found that the use of such a prior distribution leads to results very similar to those obtained by the common assumption that the σ_i^2s are known and are replaced by the s_i^2s, the observed within-study variances. A detailed discussion of the prior distributions for the σ_i^2s together with recursive formulae for the approximation of the first posterior moment can be found in Lindley[24]. Smith *et al.*[27] also discuss the choice of such a prior distribution for σ_i^2 and argue against assuming θ_i independent of σ_i^2.

A prior distribution for τ^2 has to be flexible enough to be able to easily accommodate *a priori* information whilst at the same time being mathematically convenient. For this reason an inverse gamma distribution, denote IG(a, b), is used.[34] Finally, for the purposes of this paper we assume that there is only vague prior information regarding the population effect, μ, thus maintaining some element of objectivity with respect to the estimation of the overall effect. We therefore assume it follows a locally uniform prior distribution.

3.2. Approximate inference

An approximation for the first moment of the effect of the ith study as described in model (1) is given by

$$E(\theta_i|\boldsymbol{y}, \boldsymbol{s}^2, \boldsymbol{n}) \approx \bar{y} - \frac{(n_i - 1)s_i^2 b(2a + k - 1)}{2n_i(n_i - 3)(1 + b\mathrm{RSS_B}/2)}(y_i - \bar{y}) \tag{2}$$

where $\boldsymbol{y} = (y_1, \ldots, y_k)^{\mathrm{T}}$, the vector of effect sizes in the k studies, $\boldsymbol{s}^2 = (s_1^2, \ldots, s_k^2)^{\mathrm{T}}$, the vector of within-study variances, $\boldsymbol{n} = (n_1, \ldots, n_k)^{\mathrm{T}}$, the vector of study sizes and $\mathrm{RSS_B} = \sum y_i^2 - k\bar{y}^2$ is the residual sum of squares between studies, in the usual ANOVA notation. Also, $\bar{y}$ is the arithmetic average of the y_is, and a and b are the parameters of the inverse gamma prior distribution for τ^2. An approximation to the second moment is given by

$$V(\theta_i|\boldsymbol{y}, \boldsymbol{s}^2, \boldsymbol{n}) \approx \frac{(n_i - 1)s_i^2}{n_i(n_i - 3)}\left\{1 + \frac{(n_i - 1)s_i^2 b^2(2a + k - 1)(n_i - 4 + 2a + k)(y_i - \bar{y})^2}{2n_i(n_i - 3)(n_i - 5)(1 + b\mathrm{RSS_B}/2)^2} + \frac{(n_i - 1)s_i^2 b(2a + k - 1)}{2kn_i(n_i - 5)(1 + b\mathrm{RSS_B}/2)}\right\} \tag{3}$$

where $\mathrm{RSS_B} = \sum y_i^2 - k\bar{y}^2$ is the residual sum of squares between studies, in the usual ANOVA notation. Similarly, an approximation for the first moment of σ_i^2, the variability in the ith study, is

 © 1998 John Wiley & Sons, Ltd.

given by

$$E(\sigma_i^2|\boldsymbol{s}^2, \boldsymbol{n}) \approx \frac{n_i}{n_i - 2} s_i^2. \tag{4}$$

Considering the population effect, μ, initial approximations, based on first-order Taylor series expansions, for the first and second moments are given by

$$E(\mu|\boldsymbol{y}, \boldsymbol{s}^2, \boldsymbol{n}) \approx \bar{y} \quad \text{and} \quad V(\mu|\boldsymbol{y}, \boldsymbol{s}^2, \boldsymbol{n}) \approx \frac{2(1 + b\mathrm{RSS_B}/2)}{bk(2a + k - 3)}. \tag{5}$$

Finally, approximations for the first and second moments of the population variability, τ^2, are given by

$$E(\tau^2|\boldsymbol{y}, \boldsymbol{s}^2, \boldsymbol{n}) \approx \frac{2(1 + b\mathrm{RSS_B}/2)}{b(2a + k - 3)} \quad \text{and} \quad V(\tau^2|\boldsymbol{y}, \boldsymbol{s}^2, \boldsymbol{n}) \approx \frac{8(1 + b\mathrm{RSS_B}/2)^2}{b^2(2a + k - 3)^2(k + 2a - 5)}. \tag{6}$$

All the approximations reported above are obtained using a first-order Taylor expansion of the posterior density; the idea of the expansion is presented in Appendix I, a more detailed exposition can be seen in Sansó.[35]

Note that the approximation of the posterior expectation of θ_i, (2), is the average of the observations for the ith study shrunk towards the overall mean, $\bar{y}$. The approximation of the posterior variance of θ_i, (3), is an estimate of the variance within the ith study inflated by a factor that corresponds to the random effect. The posterior expectation of the overall treatment effect, μ, is approximated by $\bar{y}$. Whilst this is clearly a sensible approximation it is also a rather crude one, and so the following approximation, based on a second-order Taylor expansion, is considered instead:

$$E(\mu|\boldsymbol{y}, \boldsymbol{s}^2, \boldsymbol{n}) \approx \frac{\bar{y} - \dfrac{b(k + 2a - 1)}{2(1 + b\mathrm{RSS_B}/2)} \displaystyle\sum_{i=1}^{k} \frac{n_i s_i^2}{n_i - 3} \left(\frac{\bar{y}(k - 3) + y_i}{k} - \frac{(\bar{y}^2(\bar{y} - y_i) + \bar{y}y_i^2)(k + 2a + 1)b}{2(1 + b\mathrm{RSS_B}/2)} \right)}{1 - \dfrac{b(k + 2a - 1)}{2(1 + b\mathrm{RSS_B}/2)} \displaystyle\sum_{i=1}^{k} \frac{n_i s_i^2}{n_i - 3} \left(\frac{k - 1}{k} - \frac{(\bar{y}(\bar{y} - y_i) + y_i^2)(k + 2a + 1)b}{2(1 + b\mathrm{RSS_B}/2)} \right)}. \tag{7}$$

Formulae (6) are reported for completeness since, in a Bayesian set-up, knowledge of the posterior distribution of μ is enough to describe the overall treatment effect.

Of particular interest is the case when $a = 0$ and $b = 2$, which we assume to be a reference prior distribution that expresses relatively vague *a priori* knowledge in that it corresponds to an inverse χ^2 distribution with zero degrees of freedom. Note that choosing $a = b = 0$, which is frequently proposed as a default prior distribution, does not produce sensible approximations for $V(\mu|\boldsymbol{y}, \boldsymbol{s}^2, \boldsymbol{n})$, $E(\tau^2|\boldsymbol{y}, \boldsymbol{s}^2, \boldsymbol{n})$ and $V(\tau^2|\boldsymbol{y}, \boldsymbol{s}^2, \boldsymbol{n})$.

3.3. Gibbs Sampling

For the model (1) and the prior distributions described in Section 3.1 the conditional posterior distributions may be obtained in closed form and thus Gibbs sampling may be used to obtain posterior estimates of the model parameters.[25] The functional forms of the conditional distributions are presented in Appendix II together with details on the methods used for assessing convergence of the Gibbs sampler.

© 1998 John Wiley & Sons, Ltd.

4. APPLICATIONS

4.1. Infection Example

In the previous Bayesian analysis of this data[22,27] a prior distribution for τ^2, the between-study variability, is derived based on subjective beliefs regarding the plausible range of effect sizes. It is assumed that there is likely to be a one order of magnitude difference between the maximum and minimum odds ratio observed, that is, that they differ by a factor of 10, but that it is very unlikely that they differ by two orders of magnitude, that is, that they differ by a factor of 100. Thus, a likely value of τ^2 is given by $[\log_e(10)/(2 \times 1{\cdot}96)]^2$, that is, 0·33, and an extreme value of τ^2 is given by $[\log_e(100)/(2 \times 1{\cdot}96)]^2$, that is, 1·38. It transpires that an inverse gamma distribution with parameters 3 and 1 provides a suitable distribution for such beliefs. In terms of τ^2 this prior distribution has a mode at 0·25, and there is 8 per cent prior probability that τ^2 will be greater than one order of magnitude different to the average within-study variance, that is, 0·50.

From Table I we can see that in trial 15 none of the patients in the treatment arm develops an infection, whilst in trial 21 all 23 patients in the control arm develop infections. As a result calculation of the log odds ratio and variance is problematic. We adopt the solution advocated by Cox and Snell[30] (pp. 31–32), in which correction factors of 0·5 are introduced into the log odds ratio and corresponding corrections are made to the variance. This correction also has the effect of improving the distributional assumption of Normality, and reducing the bias.[30] Nevertheless due care has to be taken when there are a number of trials which have particularly low or high response rates in either or both groups.

In order to use the approximations of Section 3.2 we assume that the number of patients in both the treatment and control group is the same for each study. Whilst from Table I we can see that this assumption is approximately satisfied in this example, though there is an imbalance in trial 8, in general this might not be so. In such circumstances choosing n_i to be the minimum guarantees a conservative estimation approach, as the individual trial standard errors will be larger. For this example, we obtained similar results using the minimum, average and maximum, though the results presented here were obtained using the average.

Table IV shows the means and variances for μ and τ^2, the population parameters, using each of the three estimation methods: maximum likelihood; the approximations of Sections 3.2 and 3.3. For the Bayesian approaches two different prior distributions for τ^2 were considered; the first in the reference prior distribution discussed in Section 3.2 in which $a = 0$ and $b = 2$, whilst the second was that based on subjective beliefs[27] briefly outlined above.

We can see from Table IV that, in terms of the population parameters, posterior estimates based on the approximations of Section 3.2 and those obtained using Gibbs sampling are in broad agreement. The differences between the Bayesian estimates and those obtained using maximum likelihood methods are partly explained by the larger estimates for τ^2 under the Bayesian methods, which in turn ensures that the smaller studies, that is, trials 4, 15 and 21, receive greater relative weight. However, these smaller studies also have larger effect sizes, and thus have an effect on the estimate of the overall effect. A sensitivity analysis using a range of different values for b yielded similar results.

In terms of the individual trial effects, Figure 1 shows the approximated trial effects, those obtained using Gibbs sampling, together with approximate 95 per cent credibility intervals, and maximum likelihood estimates with 95 per cent confidence intervals. In general the Bayesian approaches shrink the estimates of the individual trial effects towards the respective overall population effect. This is more noticeable for the smaller trials than for the larger ones.

 © 1998 John Wiley & Sons, Ltd.

Table IV. Comparison of estimation methods for infection data, assuming reference prior distributions for both μ and τ^2, together with a prior distribution based on subjective beliefs for τ^2 (*denotes that the approximation is based on a second-order Taylor expansion)

Parameter	Maximum likelihood	Approximation	Gibbs sampling
Uniform prior distribution for μ and $IG(0, 2)$ for τ^2			
$E(\mu)$	− 1·245	− 1·477	− 1·488
$E(\mu)$*	–	− 1·498	–
Median(μ)	–		–
$V(\mu)$	0·028	0·044	0·053
SD(μ)	0·167	0·210	0·230
$E(\tau^2)$	0·345	0·962	1·090
Median(τ^2)	–	–	1·005
$V(\tau^2)$	–	0·109	0·141
SD(τ^2)	–	0·330	0·375
Uniform prior distribution for μ and $IG(3, 1)$ prior for τ^2			
$E(\mu)$	− 1·245	− 1·477	− 1·463
$E(\mu)$*	–	− 1·494	–
Median(μ)	–	–	− 1·456
$V(\mu)$	0·028	0·035	0·041
SD(μ)	0·167	0·187	0·202
$E(\tau^2)$	0·345	0·771	0·665
Median(τ^2)	–	–	0·626
$V(\tau^2)$	–	0·052	0·053
SD(τ^2)	–	0·228	0·231

Comparing the two Bayesian approaches we observe that for particularly small and unbalanced studies, that is, trials 4 and 21, the approximations over-shrink the study effects and under-shrink the estimates of the variances. Comparing the Bayesian approaches with maximum likelihood, the results are similar for the larger trials, but for the smaller trials the Bayesian approaches shrink the estimates of the individual study effects towards the overall population effect.

4.2. Dentifrice Example

Before considering the analysis of the dentifrice data for the direct comparison of NaF with SMFP we need to consider the use of the data from the indirect comparison of NaF + SMFP with SMFP in order to derive a prior distribution for τ^2. One possible method would be to use maximum likelihood methods[4, 36] to derive an estimate of τ^2 from the indirect evidence. However, in this case, when there are only three studies, such an approach does not appear appropriate. An alternative approach is to consider the prior distributions of the θ_is, which conditional upon μ and τ^2, are members of the Student family of distributions. Thus, the parameters of the inverse gamma distribution for τ^2 may be expressed in terms of a plausible range for the θ_is. Further details are given in Appendix III. As there are only three studies, use of the overall range, that is 0·11–0·84 = 0·95, rather than the interquartile range will ensure that the resulting prior distribution is sufficiently diffuse so as to reflect the small sample size, and thus reduce the weight of prior evidence. This approach yields an IG(0·5, 8·86) prior distribution for τ^2, which has a mode at 0·08,

© 1998 John Wiley & Sons, Ltd.

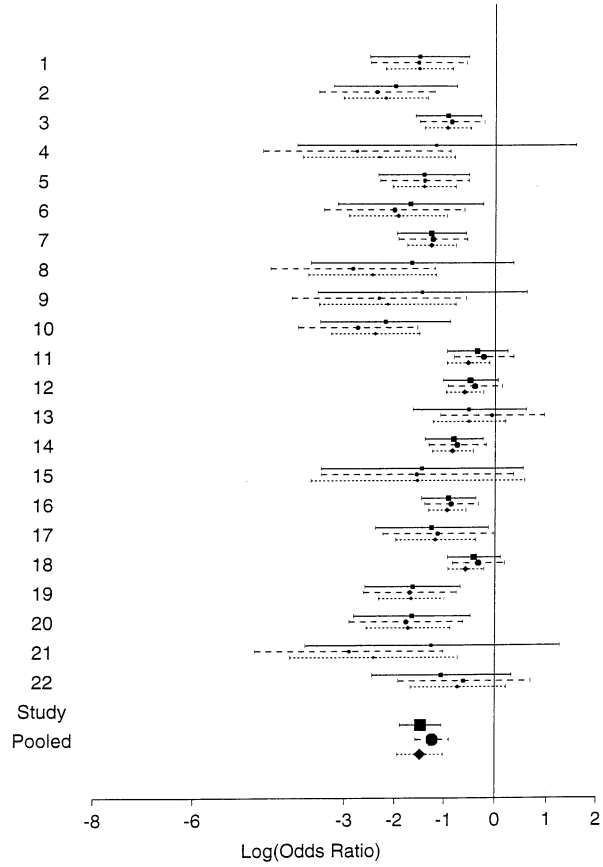

Figure 1. Individual study estimates of log odds ratios, for treated compared to control, and approximate 95 per cent credibility intervals for infection example using approximations and Gibbs sampling assuming $\tau^2 \sim \mathrm{IG}[0, 2]$ and maximum likelihood: —■— approximation; —●— maximum likelihood; ·♦· Gibbs sampling (size of symbols is proportional to overall size of study)

and prior probability that τ^2 is at least one order of magnitude greater than 0·42, the average between-study variability, of 54 per cent.

Table V shows the means and variances for μ and τ^2 using both a reference prior distribution for τ^2, that is, IG(0, 2), and the prior distribution derived from the indirect evidence briefly outlined above, that is, IG(0·5, 8·86). Maximum likelihood methods estimate τ^2 to be negative and therefore it is set to zero for this particular example. Whilst Bayesian methods also estimate τ^2 to be small, they certainly do not estimate it to be zero. This is as a consequence of the fact that both prior distributions spread their mass, to varying degrees, over the positive real line. Both the approximation methods and Gibbs sampling give estimates of the first and second moments of μ and τ^2 which are in broad agreement. In particular, using both the reference and informative prior distributions for τ^2, the uncertainty associated with μ is considerably greater than that estimated using maximum likelihood.

 © 1998 John Wiley & Sons, Ltd.

Table V. Comparison of estimation methods for dentifrice data, assuming reference prior distributions for both μ and τ^2, together with a prior distribution based on other randomized data for τ^2 (*denotes that the approximation is based on a second-order Taylor expansion)

Parameter	Maximum likelihood	Approximation	Gibbs sampling
Uniform prior distribution for μ and $IG(0, 2)$ for τ^2			
$E(\mu)$	0·283	0·378	0·352
$E(\mu)$*	–	0·434	–
Median(μ)	–	–	0·351
$V(\mu)$	0·009	0·038	0·052
SD(μ)	0·095	0·195	0·228
$E(\tau^2)$	0·0	0·339	0·345
Median(τ^2)	–	–	0·270
$V(\tau^2)$	–	0·057	0·081
SD(τ^2)	–	0·239	0·285
Uniform prior distribution for μ and $IG(0{\cdot}5, 8{\cdot}86)$ prior for τ^2			
$E(\mu)$	0·283	0·378	0·354
$E(\mu)$*	–	0·394	–
Median(μ)	–	–	0·343
$V(\mu)$	0·009	0·020	0·035
SD(μ)	0·095	0·141	0·187
$E(\tau^2)$	0·0	0·180	0·164
Median(τ^2)	–	–	0·080
$V(\tau^2)$	–	0·013	0·010
SD(τ^2)	–	0·114	0·100

In terms of the individual trial effects, Figure 2 shows the approximated trial effects, those obtained using Gibbs sampling, together with approximate 95 per cent credibility intervals, and maximum likelihood together with 95 per cent confidence intervals. We can see that for most of the trials the two Bayesian approaches yield similar results apart from those trials that have particularly large within-study variability, that is, trials 1, 2, 5, 7. In terms of comparing the Bayesian approaches with maximum likelihood, the results are similar for the larger trials, but for the smaller trials the Bayesian approaches shrink the estimates of the individual study effects towards the overall population effect.

5. DISCUSSION

We have shown that the relatively simple approximations that we have proposed in Section 3.2 lead to broadly similar results when compared to the more computationally intensive Gibbs sampling whilst still allowing for the inclusion of pertinent background information in a formal manner. The two examples we have considered highlight perhaps the extremes in the application of a meta-analytic techniques. The first example, on digestive tract infections, consisted of a relatively large number of trials of variable size, whilst the second example, on the comparison of dentifrices, though only containing a relatively small number of moderate size studies, did have

© 1998 John Wiley & Sons, Ltd.

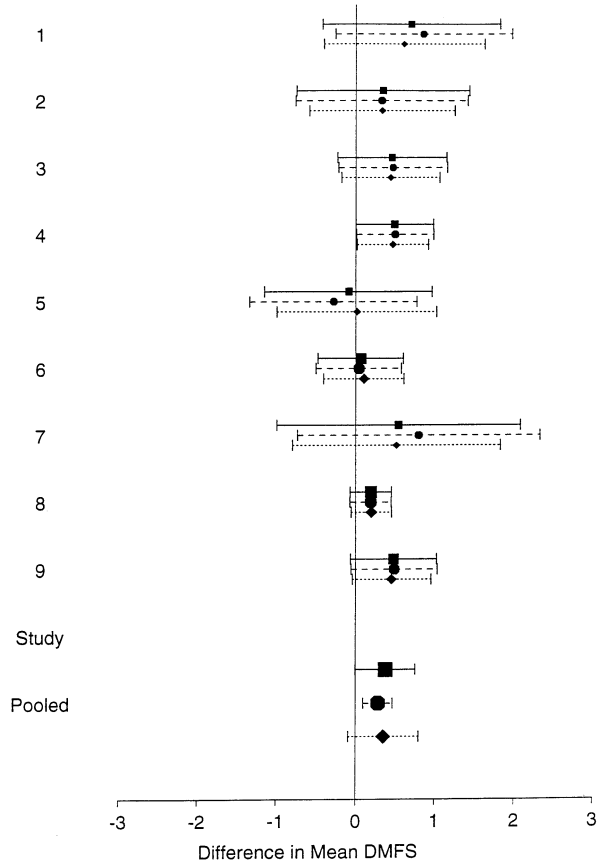

Figure 2. Individual study estimates of difference in mean DMFS index, SMFP – NaF, and approximate 95 per cent credibility intervals for dentifrice example using approximations and Gibbs sampling assuming $\tau^2 \sim \mathrm{IG}[0, 2]$ and maximum likelihood: —■— approximation; —●— maximum likelihood; ·♦· Gibbs sampling (size of symbols is proportional to overall size of study)

some studies which had relatively large within-study variances. In particular, we have demonstrated that is the case when a classical approach fails to estimate the between-study variability, a Bayesian approach, whether based on analytic approximations or Markov chain Monte Carlo approximation methods, can produce sensible estimates of the underlying between-study heterogeneity. This is as a result of choosing families of prior distributions for τ^2 which are only defined for positive real values.

We have also shown that although in general the approximations and Gibbs sampling are in broad agreement, what differences there are depend on either the sample sizes of the trials in the meta-analysis or the number of trials included. In the former case, differences at the individual study level parameters appear, whilst in the latter case, differences at the population level parameters emerge. In situations when there are a few trials, and each is relatively small, the rationale for using meta-analytic methods at all could be questioned. For practical purposes the approximations outlined here are not suitable for meta-analyses with fewer than six studies in

 © 1998 John Wiley & Sons, Ltd.

them. In terms of individual studies it is more difficult to give clear guidelines, but our experience would suggest extreme caution should be taken in using the approximations when there are studies that are either particularly small, that is, contain less than 30 patients, or when the within-study variances are relatively large, or when there is severe imbalance between the treatment groups within trials. This latter point is particularly relevant to the case when the outcome measure is binary, as the approximations assume the number of patients in the two treatment groups, within a trial, to be equal. Though a number of methods have been proposed to accommodate such imbalances, their exact effect upon the approximations ideally requires exploration via a sensitivity analysis.

Whilst the approximations in general yield broadly similar results to those obtained using Gibbs sampling, in poorly behaved examples, that is, a small number of small trials, it might be felt that the accuracy of the approximations is not sufficient to warrant formal inference, and we suggest that in such circumstances they could be used to provide appropriate starting values for Gibbs sampling, especially since the convergence of such methods has been shown to depend critically on starting values.[37] Though the approximations outlined here could be obtained to a higher degree of accuracy by considering higher order Taylor series expansions, a balance has to be maintained between accuracy and complexity. In fact, the simple formulae provided are particularly suited to performing sensitivity analyses in order to assess the influence of specific values of a and b, without being computationally expensive.

Model (1) has assumed that the σ_i^2s are unknown and have a non-informative improper prior distributions. In practice, even in random effects models, the σ_i^2s are often assumed to be known and take the values of the observed within-study variances. This could be incorporated into a model formulation such as (1) but is unlikely to lead to radically different approximations.

The approach taken here to the inclusion of background information thought pertinent to the issue in question has been to derive an appropriate prior distribution for the between-study heterogeneity. An alternative approach would have been to have included such information as another parallel level within a hierarchical modelling framework.[38,39] The choice of which approach is appropriate will often depend on the particular situation and the clinical question to be addressed.

So far we have not addressed issues regarding the more general role that Bayesian methods have to play in meta-analysis. Two areas deserve particular attention: cumulative meta-analysis, and the choice between competing models. The use of cumulative meta-analyses has been advocated as a means of establishing the stability of an intervention effect over time.[14,40] Though to date the majority of cumulative analyses have been reported from a classical perspective, the sequential updating of beliefs regarding an intervention is inherently Bayesian in nature and such methods offer a potentially useful method of analysis. The second area, that of model discrimination, and in particular discrimination between fixed and random effect models, is another area in which a Bayesian approach is of potential benefit. As noted by a number of authors the standard test for heterogeneity is of low statistical power[6,7,41] making practical interpretation of such a test difficult. Recently the use of Bayes factors[42–44] have been advocated as a means of discriminating between competing models. Though some would argue that a Bayesian approach to meta-analysis should take account of all uncertainty regarding model parameters, and therefore a random effects to model should always be adopted, Bayes factors do have the advantage of allowing a more judicious appraisal of the evidence for either model. They also allow for the possibility that model uncertainty may be incorporated formally into an analysis.[44]

© 1998 John Wiley & Sons, Ltd.

While the Bayesian approach offers an appealing methodology to the meta-analyst, its application in practice has been hampered by the lack of suitable user-friendly software. Recently this has begun to change[45] and the approximations outlined here add another complementary tool to the applied (Bayesian) statistician's toolkit.

APPENDIX I

The idea behind the approximations for the posterior first and second moments of the parameters in model (1) will be illustrated by considering the approximation for the marginal density of the data, $m(y)$

$$m(y) \propto \int_{\mathscr{R}^k} \int_0^{+\infty} \cdots \int_0^{+\infty} \int_0^{+\infty} \int_{-\infty}^{+\infty} \prod_{i=1}^{k} \left(\frac{1}{\sigma_i^2}\right)^{(n_i+2)/2} \left(\frac{1}{2\pi}\right)^{n_i/2}$$
$$\times \exp\left\{-\frac{n_i s_i^2}{2\sigma_i^2} - \frac{n_i(y_i-\theta_i)^2}{2\sigma_i^2}\right\} \mathrm{N}_k(\boldsymbol{\theta}|\mu, \tau^2 I)\, \mathrm{IG}(\tau^2|a, b)\mathrm{d}\mu\, \mathrm{d}\tau^2\, \mathrm{d}\boldsymbol{\sigma}^2\, \mathrm{d}\boldsymbol{\theta}$$

where s_i^2 is the observed variance of the ith study, $\mathrm{N}_k(\boldsymbol{x}|\boldsymbol{v}, W)$ denotes a k-variate Normal density evaluated at $\boldsymbol{x}$, with mean vector $\boldsymbol{v}$ and variance-covariance matrix W; $\mathrm{IG}(x|\alpha, \beta)$ denotes an inverted gamma density evaluated at x, with parameters α and β and I is a $k \times k$ identity matrix. Further, define $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_k^2)^{\mathrm{T}}$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^{\mathrm{T}}$ so that $\mathrm{d}\boldsymbol{\sigma}^2 = \mathrm{d}\sigma_1^2 \ldots \mathrm{d}\sigma_k^2$ and $\mathrm{d}\boldsymbol{\theta} = \mathrm{d}\theta_1 \ldots \mathrm{d}\theta_k$. Note that $m(y)$ appears in the denominator of the posterior moments of all parameters.

Changing the order of integration and integrating out $\boldsymbol{\theta}$, we obtain

$$m(y) = \int_0^{+\infty} \cdots \int_0^{+\infty} \int_0^{+\infty} \int_{-\infty}^{+\infty} \prod_{i=1}^{k} \left(\frac{2\pi}{n_i}\right)^{1/2} \prod_{i=1}^{k} \left(\frac{1}{\sigma_i^2}\right)^{(n_i+1)/2} \left(\frac{1}{2\pi}\right)^{n_i/2}$$
$$\times \mathrm{N}_k(\boldsymbol{y}|\mu, \mathrm{diag}(\sigma_i^2/n_i) + \tau^2 I)\, \mathrm{IG}(\tau^2|a, b) \exp\{-\textstyle\sum n_i s_i^2/2\sigma_i^2\}\mathrm{d}\mu\, \mathrm{d}\tau^2\, \mathrm{d}\boldsymbol{\sigma}^2$$

where $\boldsymbol{y} = (y_1, \ldots, y_k)^{\mathrm{T}}$. For large n_i it is sensible to think that $\mathrm{N}_k(\boldsymbol{y}|\mu, \mathrm{diag}(\sigma_i^2/n_i) + \tau^2 I)$ tends to $\mathrm{N}_k(\boldsymbol{y}|\mu, \tau^2 I)$. More precisely, consider the function

$$f(z_1, \ldots, z_k) = \left(\frac{1}{2\pi}\right)^{k/2} \prod_{i=1}^{k} \left(\frac{1}{z_i}\right)^{1/2} \exp\left\{-\sum \frac{(y_i\mu)^2}{2z_i}\right\}$$

then $f(\tau^2 + \sigma_1^2/n_1, \ldots, \tau^2 + \sigma_k^2/n_k) = \mathrm{N}_k(\boldsymbol{y}|\mu, \mathrm{diag}(\sigma_i^2/n_i) + \tau^2 I)$. A first-order Taylor expansion of f centred on $z^0 = (\tau^2, \ldots, \tau^2)^{\mathrm{T}}$ and evaluated at $z = (\tau^2 + \sigma_1^2, \ldots, \tau^2 + \sigma_k^2)^{\mathrm{T}}$ gives

$$\mathrm{N}_k(\boldsymbol{y}|\mu, \mathrm{diag}(\sigma_i^2/n_i) + \tau^2 I) \approx \mathrm{N}_k(\boldsymbol{y}|\mu, \tau^2 I)\left(1 + \sum -\frac{\sigma_i^2}{2n_i\tau^2} + \frac{(y_i-\mu)^2}{2n_i\tau^4}\right)$$

so that the previous approximation is equivalent to a one term Taylor expansion of $f(z_i, \ldots, z_k)$.

The former discussion yields an approximation for $m(y)$ such that

$$m(y) = \int_0^{+\infty} \cdots \int_0^{+\infty} \int_0^{+\infty} \int_{-\infty}^{+\infty} \prod_{i=1}^{k} \left(\frac{2\pi}{n_i}\right)^{1/2} \prod_{i=1}^{k} \left(\frac{1}{\sigma_i^2}\right)^{(n_i+1)/2} \left(\frac{1}{2\pi}\right)^{n_i/2}$$
$$\times \mathrm{N}_k(\boldsymbol{y}|\mu, \tau^2 I)\, \mathrm{IG}(\tau^2|a, b) \exp\{-\textstyle\sum n_i s_i^2/2\sigma_i^2\}\mathrm{d}\mu\, \mathrm{d}\tau^2\, \mathrm{d}\boldsymbol{\sigma}^2$$

and for this expression all integrals can be carried out in closed form.

 © 1998 John Wiley & Sons, Ltd.

A similar technique can be used to obtain approximations for the numerous in the formulae for $E(\theta_i|\boldsymbol{y}, \boldsymbol{s}^2, \boldsymbol{n})$, $E(\sigma_i^2|\boldsymbol{y}, \boldsymbol{s}^2, \boldsymbol{n})$, $E(\mu|\boldsymbol{y}, \boldsymbol{s}^2, \boldsymbol{n})$, $V(\theta_i|\boldsymbol{y}, \boldsymbol{s}^2, \boldsymbol{n})$, $V(\mu|\boldsymbol{y}, \boldsymbol{s}^2, \boldsymbol{n})$, $E(\tau^2|\boldsymbol{y}, \boldsymbol{s}^2, \boldsymbol{n})$ and $V(\tau^2|\boldsymbol{y}, \boldsymbol{s}^2, \boldsymbol{n})$ where $\boldsymbol{y} = (y_1, \ldots, y_k)^{\mathrm{T}}$, $\boldsymbol{s}^2 = (s_1^2, \ldots, s_k^2)^{\mathrm{T}}$ and $\boldsymbol{n} = (n_1, \ldots, n_k)^{\mathrm{T}}$. Taking the ratio of the approximations to the numerator and $m(y)$ in each case yields approximations to the first and second moments of the model parameters.

APPENDIX II

For the model (1) and the prior distributions described in Section 3.1 the conditional posterior distributions may be obtained in closed form and thus Gibbs sampling may be implemented without too much difficulty.[25] The full conditional distributions are given by

$$P(\theta_i|\mu, \boldsymbol{\sigma}^2, \tau^2, \boldsymbol{y}, \boldsymbol{s}^2, \boldsymbol{n}) = \mathrm{N}\left(\theta_i \left| \frac{y_i n_i/\sigma_i^2 + \mu/\tau^2}{n_i/\sigma_i^2 + 1/\tau^2}, \frac{1}{n_i/\sigma_i^2 + 1/\tau^2}\right.\right) \quad i = 1, \ldots, k$$

$$P(\mu|\boldsymbol{\theta}, \boldsymbol{\sigma}^2, \tau^2, \boldsymbol{y}, \boldsymbol{s}^2, \boldsymbol{n}) = \mathrm{N}\left(\mu|\sum \theta_i/k, \tau^2/k\right)$$

$$P(\sigma_i^2|\boldsymbol{\theta}, \mu, \tau^2, \boldsymbol{y}, \boldsymbol{s}^2, \boldsymbol{n}) = \mathrm{IG}(\sigma_i^2|n_i, 2/(n_i((y_i - \theta_i)^2 + s_i^2))) \quad i = 1, \ldots, k$$

and

$$P(\tau^2|\boldsymbol{\theta}, \mu, \boldsymbol{\sigma}^2, \boldsymbol{y}, \boldsymbol{s}^2, \boldsymbol{n}) = \mathrm{IG}\left(\tau^2|k/2 + a, 2/\sum(\theta_i - \mu)^2 + b\right).$$

Estimation of the marginal posterior densities for the model parameters is then achieved by sampling from each of the posterior conditional distributions in turn, conditional upon current estimates of other unknown model parameters, until 'convergence' is achieved.

Whilst in this case it would be relatively straightforward to implement such a sampler in any statistical programming environment, we have used the BUGS program.[45]

Several authors have discussed the convergence criteria of the Gibbs sampler,[46,47] the number of iterations required[48] and general sampling strategies that should be adopted.[49,50] We have adopted a mixture of the methods suggested, using multiple runs with different starting values and varying the number of iterations (both burn-in and sampled), and assessed convergence within an individual run by means of the diagnostic statistic proposed by Geweke[46] together with plots of the actual samples.

APPENDIX III

In order to elicit the values of the parameters a and b of the inverse gamma prior distribution for the population variance τ^2, consider model (1), then the prior distribution for the vector $\boldsymbol{\theta} = (\theta_1 \ldots \theta_k)^{\mathrm{T}}$ is given by

$$P(\boldsymbol{\theta}) = \int_0^\infty \mathrm{N}_k(\boldsymbol{\theta}|\mu, \tau^2 I)\mathrm{IG}(\tau^2|a, b)\,\mathrm{d}\tau^2$$

which is proportional to

$$\left(\frac{ba\sum_{i=1}^k(\theta_i - \mu)^2}{2a} + 1\right)^{-1/2(k+2a)}$$

a k-variate Student density with location μ, scale matrix I/ab and $2a$ degrees of freedom. The *a priori* marginal density of each θ_i is then a Student t-density with location μ, scale $1/ab$ and $2a$ degrees of freedom (note that this does not imply that the θ_is are independent).

© 1998 John Wiley & Sons, Ltd.

The choice of $a = 1/2$ produces a Cauchy distribution, the density with the flattest tails among the family of Student distributions with integer valued degrees of freedom. Since such a density does not have either first or second moments, the elicitation of b can be achieved using the quantiles. Consider a Cauchy distribution with location μ and scale λ then, if 2Δ is the interquartile range, due to the symmetry of the density

$$\int_{\mu-\Delta}^{\mu+\Delta} \frac{1}{1+(t-\mu)^2/\lambda^2} \frac{1}{\pi\lambda} \, \mathrm{d}t = \frac{1}{2}$$

and changing variables

$$\int_{-\Delta/\lambda}^{\Delta/\lambda} \frac{1}{1+u^2} \, \mathrm{d}t = \frac{\pi}{2},$$

and thus $2\tan^{-1}(\Delta/\lambda) = \pi/2$ so that $\Delta = \lambda$.

As a result of the former discussion the interquartile range for the θ_i equals $2\sqrt{(2/b)}$ and so, letting c_0 be a specified value for it, $b = 8/c_0^2$. The conclusion is that a prior distribution for τ^2, which is reasonably vague in that it does not have any finite moments and produces a robust inference in terms of the influence of large values of μ (see, for example, Pericchi and Sansó[51]), is given by an inverse gamma distribution with parameters 1/2 and $8/c_0^2$, where c_0 is a specified interquartile range for the θ_is. Note that the mode of such a distribution is attained at $c_0^2/12$. An analysis similar to the former one can be carried out for quantiles other than those corresponding to 25 per cent probability.

ACKNOWLEDGEMENTS

The authors would like to thank David R. Jones, Paul C. Lambert, Luis R. Pericchi and Derek Teather for useful discussions, and Teresa C. Smith for help with BUGS, and two referees whose comments on earlier versions of the paper greatly improved it. They would also like to thank the British Council (Venezuela) for providing financial support for travelling expenses. The second author was also partially supported by BID-CONICIT, Venezuela.

REFERENCES

1. Jones D. R. 'Meta-analysis: weighing the evidence', *Statistics in Medicine*, **14**, 137–149 (1995).
2. Fleiss, J. L. 'The statistical basis of meta-analysis', *Statistical Methods in Medical Research*, **2**, 121–145 (1993).
3. Cochran, W. G. 'Problems arising in the analysis of a series of similar experiments', *Journal of the Royal Statistical Society*, **4**, 102–118 supplement (1937).
4. Rao, P. S. R. S., Kaplan, J. and Cochran, W. G. 'Estimators for the one-way random effects model with unequal error variances', *Journal of the American Statistical Association*, **76**(373), 89–97 (1981).
5. DerSimonian, R. D. and Laird, N. 'Meta-analysis in clinical trials', *Controlled Clinical Trials*, **7**, 177–188 (1986).
6. Whitehead, A. and Whitehead, J. 'A general parametric approach to meta-analysis of randomised clinical trials', *Statistics in Medicine*, **10**, 1665–1677 (1991).
7. Thompson, S. G. 'Why sources of heterogeneity in meta-analysis should be investigated', *British Medical Journal*, **309**, 1351–1355 (1994).
8. Breslow, N. E. and Clayton, D. G. 'Approximate inference in generalized linear mixed models', *Journal of the American Statistical Association*, **88**, 9–25 (1993).
9. DuMouchel, W. 'Bayesian meta analysis, in Berry, D. (ed.), *Statistical Methodology in Pharmaceutical Sciences*, Marcel-Dekker, New York, 1990, pp. 509–529.

 © 1998 John Wiley & Sons, Ltd.

10. Morris, C. 'Parametric Empirical Bayes inference: theory and applications (with discussion)', *Journal of the American Statistical Association*, **78**, 47–65 (1983).
11. Carlin, J. B. 'Meta-analysis for 2×2 tables: a Bayesian approach', *Statistics in Medicine*, **11**, 141–158 (1992).
12. Lambert, P. C. and Abrams, K. R. 'Meta-analysis using multi-level models', *ML3 Newsletter*, **7**, 2, 17–19 (1995).
13. Hardy, R. J. and Thompson, S. G. 'A likelihood approach to meta-analysis with random effects', *Statistics in Medicine*, **15**(6), 619–629 (1996).
14. Lau, J., Antman, E. M., Jimenez-Silva, J., Kupelnick, B., Mosteller, F. and Chalmers, T. C. 'Cumulative meta-analysis of therapeutic trials for myocardial infarction', *New England Journal of Medicine*, **327**, 248–254 (1992).
15. DuMouchel, W. and Harris, J. E. 'Bayes methods for combining the results of cancer studies in humans and other species', *Journal of the American Statistical Association*, **78**, 293–315 (1983).
16. Pocock, S. J. and Hughes, M. D. 'Estimation issues in clinical trials and overviews', *Statistics in Medicine*, **9**, 657–671 (1990).
17. Morris, C. N. and Mormand, S. L. 'Hierarchical models for combining information and for meta-analysis', in Bernardo, J. M., Berger, J. O., Lindley, D. V. and Smith, A. F. M. (eds), *Bayesian Statistics 4*, Oxford University Press, Oxford, 1992.
18. Malec, D. and Sedransk, J. 'Bayesian methodology for combining the results from different experiments when the specifications for pooling are uncertain', *Biometrika*, **79**(3), 593–601 (1992).
19. Lindley, D. V. and Smith, A. F. M. 'Bayes estimates for the linear model (with discussion)', *Journal of the Royal Statistical Society, Series A*, **34**, 1–41 (1972).
20. Skene, A. M. and Wakefield, J. C. 'Hierarchical models for multicentre binary response studies', *Statistics in Medicine*, **9**, 919–929 (1990).
21. Rogatko, A. 'Bayesian approach for meta-analysis of controlled clinical trials', *Communications in Statistics – Theory and Methods*, **21**, 1441–1462 (1991).
22. Smith, T., Spiegelhalter, D. J. and Thomas, A. Bayesian graphical modelling for random-effects meta-analysis: a comparative study', *Statistics in Medicine*, **14**(24), 2685–2699 (1995).
23. Tierney, L. and Kadane, J. B. 'Accurate approximations for posterior moments and marginal densities', *Journal of the American Statistical Association*, **81**, 82–86 (1986).
24. Lindley, D. V. 'The estimation of many parameters', in Godamle, V. P. and Sprott, D. A. (eds), *Foundations of Statistical Inference*, 1971, pp. 435–455.
25. Gelfand, A. E. and Smith, A. F. M. 'Sampling-based approaches to calculating marginal densities', *Journal of the American Statistical Association*, **85**, 398–409 (1990).
26. Digestive Tract Trialists' Collaborative Group, 'Meta-analysis of randomised controlled trials of selective decontamination of the digestive tract', *British Medical Journal*, **307**, 525–532 (1993).
27. Smith, T. C., Spiegelhalter, D. J. and Parmar, M. K. B. 'Bayesian meta-analysis of randomised trials using graphical models and BUGS', in Berry, D. and Stangl, D. (eds), *Bayesian Biostatistics*, Marcel Dekker, Oxford, 1995, pp. 411–427.
28. Johnson, M. F. 'Comparative efficacy of NaF and SMFP dentifrices in caries prevention; a meta-analytic overview', *Caries Research*, **27**, 328–336 (1993).
29. Cohen, B. and Kramer, I. R. H. *Scientific Foundations of Dentistry*, William Heinemann, London, 1976.
30. Cox, D. R. and Snell, E. J. *Analysis of Binary Data*, 2nd edn, Chapman and Hall, London, 1989.
31. Raiffa, H. and Schlaifer, R. *Applied Statistical Decision Theory*, Havard Business School, Boston, 1961.
32. Box, G. E. P. and Tiao, G. C. 'Bayesian estimation of means for the random effects model', *Journal of the American Statistical Association*, **63**, 174–181 (1968).
33. Box, G. E. P. and Tiao, G. C. *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Massachusetts, 1973.
34. Bernardo, J. M. and Smith, A. F. M. *Bayesian Theory*, Wiley, Chichester, 1994.
35. Sansó, B. 'Approximations for the posterior moments of the location of a normal likelihood with unknown scale', Technical Report 94-01, CESMa, Universidad Simón Bolívar, 1994.
36. Harville, D. A. 'Maximum likelihood approaches to variance component estimation and to related problems', *Journal of the American Statistical Association*, **72**, 320–338 (1977).

© 1998 John Wiley & Sons, Ltd.

37. Smith, A. F. M. and Roberts, G. O. 'Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods', *Journal of the Royal Statistical Society*, *Series B*, **55**, 3–23 (1993).
38. Eddy, D. M., Hasselblad, V. and Shachter, R. 'A Bayesian method for synthesizing evidence: the Confidence Profile Method', *International Journal of Technology Assessment in Health Care*, **6**, 31–55 (1990).
39. Smith, T. C., Abrams, K. R. and Jones, D. R. 'Hierarchical models in generalised synthesis of evidence – a breast cancer screening example' Technical Report (Statistics) 95–02, Department of Epidemiology and Public Health, University of Leicester, 1995.
40. Davey-Smith, G., Song, F. and Sheldon, T. 'Cholesterol lowering and mortality: the importance of considering initial level of risk', *British Medical Journal*, **306**, 1367–1373 (1993).
41. Thompson, S. G. and Pocock, S. J. 'Can meta-analyses be trusted?', *Lancet*, **338**, 1127–1130 (1991).
42. Kass, R. R. 'Bayes factors in practice', *Statistician*, **42**, 551–560 (1993).
43. Berger, J. O. and Pericchi, L. R. 'The intrinsic Bayes factor for linear models', in Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. (eds), *Bayesian Statistics 5*, Oxford University Press, Oxford, 1996, pp. 25–44.
44. Abrams, K. R. and Sansó, B. 'Model discrimination in meta-analysis – a Bayesian perspective', Technical Report (Statistics) 95-03, Department of Epidemiology and Public Health, University of Leicester, 1995.
45. Gilks, W. R., Thomas, A. and Spiegelhalter, D. J. 'A language and program for complex Bayesian modelling', *Statistician*, **43**(1), 169–178 (1994).
46. Geweke, J. 'Evaluating the accuracy of sampling based approaches to the calculation of posterior moments', in Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. (eds), *Bayesian Statistics 4*, Oxford University Press, Oxford, 1992, pp. 169–193.
47. Roberts, G. O. 'Convergence diagnostics of the Gibbs sampler', in Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. (eds), *Bayesian Statistics 4*, Oxford University Press, Oxford, 1992, pp. 775–782.
48. Raftery, A. E. and Lewis, S. M. 'How many iterations in the Gibbs sampler?', in Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. (eds), *Bayesian Statistics 4*, Oxford University Press, Oxford, 1992, pp. 763–773.
49. Gelman, A. and Rubin, D. R. 'A single series from the Gibbs sampler provides a false sense of security', in Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. (eds), *Bayesian Statistics 4*, Oxford University Press, Oxford, 1992, pp. 625–631.
50. O'Hagan, A. *Bayesian Inference*, Volume 2b of *Kendall's Advanced Theory of Statistics*, Edward Arnold, London, 1994.
51. Pericchi, L. R. and Sansó, B. 'A note on bounded influence in Bayesian analysis', *Biometrika*, **82**, 1, 223–225 (1995).

 © 1998 John Wiley & Sons, Ltd.