200

15

Bayesian Meta-Analysis of Randomized Trials Using Graphical Models and BUGS

Teresa C. Smith and David J. Spiegelhalter Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge, England
Mahesh K. B. Parmar Medical Research Council Cancer Trials Office, Cambridge, England

ABSTRACT

In this chapter we describe how hierarchical random-effects models can be applied to meta-analysis using a fully Bayesian approach. Using a meta-analysis of randomized trials of selective decontamination of the digestive tract as an example, inferences are made using Gibbs sampling via BUGS, a freely available software package. We illustrate the usefulness of graphical modeling techniques for expressing the conditional independence assumptions of the parameters in the model and show how specification of the model in BUGS leads naturally from the graph formulation. Problems with using a standard noninformative prior distribution for the population variance are discussed and suitable alternative prior distributions are derived and compared.

1 INTRODUCTION

Hierarchical random-effects models are becoming increasingly popular for tackling problems involving complex structured data. In this chapter we describe how meta-analysis fits naturally into the hierarchical framework. Using an example of a meta-analysis of randomized trials of selective decontamination of the digestive tract [1], we illustrate how a Bayesian analysis can be carried out. Due to recent developments in Markov

chain Monte Carlo [2], implementing such models within a fully Bayesian framework is now a viable option, without recourse to approximations in which uncertainty about certain parameters is ignored.

Inferences on the parameters of interest are based on the posterior marginal distributions, which are obtained using the simulation technique known as Gibbs sampling. This is implemented using BUGS, a software package available free of charge, which provides a means of analyzing complex models essentially by describing their structure and automatically deriving the expressions necessary for the Gibbs sampling. In addition, we show how graphical modeling techniques may be used to express the conditional independence assumptions of the parameters in the model: not only can this provide valuable insight, but BUGS also exploits the resulting factorization of the full distribution of all the parameters and data as a product of simple conditional distributions.

Focus then centers on the choice of prior distribution and most importantly on the problems with using a standard noninformative prior for the population (or between-study) variance in a random-effects meta-analysis. Techniques for deriving suitable alternative priors based on three different approaches are also explored and their consequences compared.

2 THE PROBLEM

Preventing infections in intensive care units is a major area of concern. However, controversy still surrounds the best way to avoid them. One suggested strategy involves selectively decontaminating the digestive tract to prevent carriage of potentially pathogenic microorganisms from the oropharynx, stomach, and gut. An international collaborative group investigated the clinical benefits of selective decontamination of the digestive tract by carrying out a meta-analysis of 22 randomized trials [1]. In each trial, patients in intensive care units were randomized to either a treatment or a control group, where treatment consisted of different combinations of oral nonabsorbable antibiotics, with some studies in addition including a systemic component of the treatment. Patients in the control groups were given no treatment. For each trial the number who developed respiratory tract infections in the treatment and the control groups were recorded (Table 1).

The collaborative group analyzed the data using the classical Mantel-Haenszel-Peto method [3] to obtain estimates of both the individual treatment effects in each study and the pooled effect. The pooled-effect estimate is based on assuming a common effect across all studies.

202

able 1 Respiratory Tract Infections in Control nd Treatment Groups of 22 Trials, with Individual nd Pooled Estimates of Odds Ratios (95% onfidence Intervals) Obtained Using Mantel-laenszel-Peto Method

Study	Infections/total		Odds ratio
	Control	Treated	(95% confidence interval)
1	25/54	7/47	0.24 (0.10, 0.55)
2	24/41	4/38	0.13 (0.05, 0.32)
3	37/95	20/96	0.42 (0.23, 0.78)
4	11/17	1/14	0.10 (0.02, 0.40)
5	26/49	10/48	0.25 (0.11, 0.58)
6	13/84	2/101	0.17 (0.06, 0.48)
7	38/170	12/161	0.31 (0.17, 0.57)
8	29/60	1/28	0.14 (0.05, 0.36)
9	9/20	1/19	0.13 (0.03, 0.54)
10	44/47	22/49	0.11 (0.04, 0.25)
11	30/160	25/162	0.79 (0.44, 1.41)
12	40/185	31/200	0.67 (0.40, 1.12)
13	10/41	9/39	0.93 (0.33, 2.59)
14	40/185	22/193	0.48 (0.28, 0.82)
15	4/46	0/45	0.13 (0.02, 0.95)
16	60/140	31/131	0.42 (0.26, 0.70)
17	12/75	4/75	0.33 (0.12, 0.92)
18	42/225	31/220	0.72 (0.43, 1.18)
19	26/57	7/55	0.21 (0.09, 0.47)
20	17/92	3/91	0.21 (0.08, 0.54)
21	23/23	14/25	0.09 (0.02, 0.33)
22	6/68	3/65	0.52 (0.13, 1.99)
Pooled			0.36 (0.31, 0.43)

A FULL BAYESIAN RANDOM-EFFECTS MODEL

A random-effects model, unlike the fixed-effect method, adopts a probability model for individual study effects whose joint distribution is assumed not to depend on the order in which the studies are placed. In other words, we have no prior reason for thinking any particular study is different from another, and hence we are formally expressing a belief in *similarity*, as opposed to the extremes for *equivalence* or *complete independence*. This formal assumption, known as exchangeability, is

mathematically equivalent to assuming these effects are randomly drawn from some population.

Let r_i^C denote the number of patients in the control group with infections in the ith study, arising from n_i^C patients randomized to the control group, each assumed to have probability of p_i^C of developing an infection. Adopting equivalent notation for the treatment group, the full model can be written

$$\begin{aligned} r_i^C &\sim \text{binomial}(p_i^C, n_i^C) \\ r_i^T &\sim \text{binomial}(p_i^T, n_i^T) \\ \text{logit}(p_i^C) &= \mu_i - \delta_i/2 \\ \text{logit}(p_i^T) &= \mu_i + \delta_i/2 \\ \delta_i &\sim \text{normal}(d, \sigma^2) \end{aligned}$$

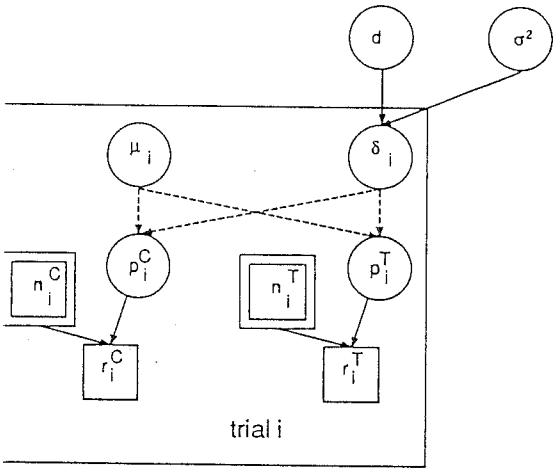
where $\text{logit}(p) = \log_e[p/(1-p)]$. The estimates of primary interest are the study-level treatment effects $\delta_i = \text{logit}(p_i^T) - \text{logit}(p_i^C)$, which is the log(odds ratio) in the ith study, and the population, or pooled, treatment effect d. $\mu_i = [\text{logit}(p_i^T) + \text{logit}(p_i^C)]/2$ may be considered the "average" (on the logit scale) infection rate in the ith trial.

Standard "empirical Bayes" methods [4] make inferences conditional on estimates of d and σ^2 that are obtained using a moment-matching procedure. This, however, ignores the uncertainty in the estimates of σ^2 when making inferences about d, and in σ^2 and d when estimating the precision of estimates of the individual trial effects, and this neglect may have a considerable impact if there are only a few studies.

A fully Bayesian analysis allows for this uncertainty by placing prior distributions on the unknown parameters, μ's, σ^2, and d. These priors will, however, generally be noninformative. Inferences about the parameters of interest (d, the population effect, and δ_i's, the individual trial effects) can then be made from the joint distribution formed by the prior and the likelihood by integrating out the unknown parameters (see Section 4). We may also be interested in predicting the true effect in a new study, as this in effect produces a prior distribution that may be used, for example, in assessing sample size for a further confirmatory trial.

The model can be expressed in the form of a graph [5] in which the nodes in the graph denote the data and parameters of the model (Figure 1). The idea of such a representation is to display qualitative aspects of the model without requiring algebraic formulas and hence to call attention to the essential assumptions. Constants fixed by the design of the study are denoted as double-edged rectangles (n_i^C, n_i^T), observed variables as

204



igure 1 Graphical model for random-effects meta-analysis.

ingle-edged rectangles (r_i^C, r_i^T), and unobserved variables as circles (p_i^C, $_i^T$, μ_i, δ_i, d, σ^2).

The graph is directed and acylic, and the arrows drawn between nodes ıdicate the conditional independence assumptions of the model. Direct ıfluences on a node are known as its "parents", using an obvious genetic nalogy, so, for example, d and σ^2 are the parents of each δ_i. The directed nks express the assumption that given its parent nodes, denoted pa(v), ach node v is independent of all other nodes except descendants of v. or example, from Figure 1, it can be seen that conditional on knowing $_i^T$ (the true treatment response rate in the ith trial), our beliefs in r_i^T ould be independent of the population parameters (d and σ^2) and data or the other trials. The links may either indicate a stochastic dependence solid arrow), for example, $r_i^C \sim$ binomial(p_i^C, n_i^C), or a logical function dashed arrow) such as logit(p_i^C) = $\mu_i - \delta_i/2$. Due to the conditional ıdependence assumptions as expressed in the graph, the full joint proba- ility distribution of all the quantities V can be specified as a simple factori- ation of the conditional parent-child distributions $p(v \mid \text{pa}(v))$ [6]. So

$$p(V) = \prod_{v \in V} p(v \mid \text{pa}(v)) \tag{1}$$

or the model described above, this joint distribution (ignoring the con-

stants $\underline{n}^C$, $\underline{n}^T$, and using the fact that p_i^C, p_i^T can be expressed in terms of μ_i, δ_i) takes the form

$$p(\underline{r}^C, \underline{r}^T, \underline{\mu}, \underline{\delta}, d, \sigma^2) \propto \prod_i [p(r_i^C \mid \mu_i, \delta_i)\, p(r_i^T \mid \mu_i, \delta_i)\, p(\mu_i) \\ p(\delta_i \mid d, \sigma^2)]\, p(d)\, p(\sigma^2) \tag{2}$$

A technical problem is to obtain the appropriate posterior marginal distributions for parameters of interest, conditional on having observed the data $\underline{r}^C$, $\underline{r}^T$. For example, inferences on the population effect d should be based on $p(d \mid \underline{r}^C, \underline{r}^T)$, which by Bayes' theorem is proportional to $p(\underline{r}^C, \underline{r}^T \mid d)\, p(d)$, so that

$$p(d \mid \underline{r}^C, \underline{r}^T) \propto \int p(\underline{r}^C, \underline{r}^T, \underline{\mu}, \underline{\delta}, d, \sigma^2)\, d\underline{\mu}\, d\underline{\delta}\, d\sigma^2$$

We thus need to integrate the joint distribution (2) over $\underline{\mu}$, $\underline{\delta}$, and σ^2. The integrand does not have a closed-form solution, even though it is composed of a product of terms each of which has a simple form. Some form of approximation or simulation technique is necessary for this and other inferences based on integrating out parameters from the full joint distribution. Developments in computer-intensive methodology have established what are known as Markov chain Monte Carlo methods as a practical proposition, a particular form of which is known as Gibbs sampling [7].

The graph is thus translated into a full probability model, and in the next section we now show how this model description directly forms the basis of the computational method.

4 INFERENCE USING GIBBS SAMPLING AND BUGS

A simple, although computationally demanding procedure for the numerical integration of complex functions, Gibbs sampling has come from its origins in statistical mechanics through image processing to play a major role in modern statistics. Initial values are given to all unknown quantities, which include all parameters, missing data, latent variables, and so on. Samples are then successively drawn from the conditional distribution of each variable in turn, given the current value of all the other variables, both observed data and unknown parameters set at their temporary values. It can be shown that under broad conditions *eventually* one will be sampling from the correct posterior distributions of the unknown parameters. There is a large literature on this topic: both methodological [2] and applications [8].

For any node v it is therefore necessary to sample from $p(v \mid V\backslash v)$, the full conditional distribution given all other nodes $V\backslash v$. However, the factorization (1) of the joint distribution expressed by the graphical model can be exploited to obtain,

$$p(v \mid V\backslash v) \propto \text{terms in } p(V) \text{ containing } v$$

$$= p(v \mid \text{pa}(v)) \prod_{v \in \text{pa}(w)} p(w \mid \text{pa}(w))$$

The full conditional distribution of any node v therefore depends only on a prior component $p(v \mid \text{pa}(v))$ and likelihood components arising from each of its children w.

The Gibbs sampling was implemented using the BUGS software [9, 10]. The BUGS language allows the model to be specified in much the same way as it was represented by the graphical model. The specification file has two sections. The first contains a declaration of all the nodes in the graph and the names of the files containing the data and the initial values, and the second is a list of the dependence relations expressed by the graph. This is shown below. The "average" infection rate for the ith trial, μ_i, and the overall treatment effect, d, were assigned normal(0, 4) and normal(0, 10) prior distributions, respectively. We note that BUGS parameterizes the normal distribution in terms of the precision rather than the variance in order to provide a conjugate prior analysis, such that "tau" represents $\tau = 1/\sigma^2$. The gamma(0.001, 0.001) prior distribution given to τ is a noninformative prior that is approximately equivalent to $p(\sigma^2) \propto 1/\sigma^2$. Suitable priors for τ are discussed in more detail in Sections 5 and 6.

```
for (i in 1:Num) {
        rt[i] ~ dbin(pt[i], nt[i]);
        rc[i] ~ dbin(pc[i], nc[i]);
  logit(pc[i]) <- mu[i] - (delta[i]/2);
  logit(pt[i]) <- mu[i] + (delta[i]/2);
    delta[i] ~ dnorm(d, tau);
      mu[i] ~ dnorm(0.0, 0.25);
}
        d ~ dnorm(0.0, 0.1);
      tau ~ dgamma(1.0E-3, 1.0E-3);
    sigma <- 1/sqrt(tau);
delta.new ~ dnorm(d, tau);
```

The syntax of the language should be largely self-explanatory and essentially consists of using the graph to express the joint distribution (2) as concisely as possible. Two forms of relation are shown: ~ translates to "is distributed as" and <- represents "is equal to." It is important

to understand that the language is *declarative* (describing a model), rather than *procedural* (specifying a sequence of steps as in a standard computer program). For example, the lines may be given in any order. (Details of how to obtain the program are given at the end of this chapter.)

From this model specification BUGS works out the parents and children of each node, constructs an internal representation of the graph, derives the necessary full conditional distributions, and carries out the Gibbs sampling. In fairly simple problems of this type a "burn-in" of, say, 500 iterations is generally sufficient to reach convergence and then summary statistics such as means and standard deviations of generated parameters values are monitored over, say, 1000 further iterations: formal techniques are available for checking convergence [11] and here a method described by Geweke [12] was adopted.

The full Bayesian random-effects model specified above gives a mean estimate for d of -1.39 (95% probability interval -1.82, -1.01). Figure 2 shows the individual study estimates for the random-effects model are drawn toward a central overall effect and have smaller intervals than the fixed-effects (Mantel-Haenszel-Peto) estimates. The pooled effect, however, has a wider interval. This is the expected pattern, due to some of the *within*-study variability in the fixed-effects analysis being accounted for as *between*-study variability in the random-effects model.

To obtain an estimate of the overall or pooled effect in a meta-analysis, the individual study estimates are combined in a way that enables some studies to contribute more to the pooled estimate than others. The amount that a study contributes to the pooled estimate is determined by its *weight*. In a fixed-effects analysis the studies are given weights that are inversely proportional to the variance of the estimated study effect, so

$$\hat{d} = \frac{\sum_{j=1}^{k} \hat{\delta}_j w_j}{\sum_{j=1}^{k} w_j}, \qquad w_j = \frac{1}{\mathrm{Var}(\hat{\delta}_j)}$$

and hence each weight w_j is proportional to study size.

In a random-effects analysis, however, the between-study variability σ^2, as well as the within-study variability, is taken into account in the weighting. If there is little heterogeneity between studies, then σ^2 will be small and the fixed-effects and random-effects estimates will be very similar. However, as the heterogeneity increases we will find that the weighting in the random-effects analysis will be less influenced by the within-study variance and become dominated by the between-study variance.

In the preceding example we see that random-effects analysis estimates a stronger relationship between infections and selective decontamination upon pooling. The random-effects analysis is placing less weight

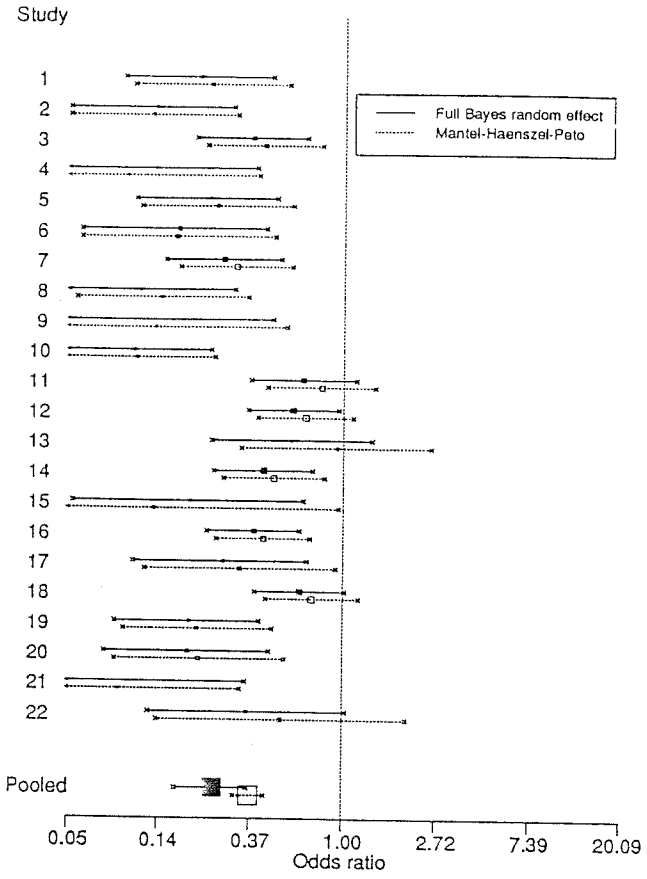


Figure 2 Estimated odd ratios for Bayesian random-effects model and Mantel-Haenszel-Peto method (area of mark is proportional to the sample size of the study).

on larger studies (with small variances) than the fixed-effect method due to the presence of heterogeneity, and hence this shift in estimated pooled log odds ratios is a result of the larger studies tending to have less strong treatment effects.

5 ISSUES WITH IMPROPER PRIORS IN HIERARCHICAL MODELS

It is important that appropriate consideration is given to the choice of prior distributions and that the effect they have on estimates of the param-

209

eters of interest is investigated. In this section we highlight the problems of priors for the population variance σ^2 and discuss the motivation behind the use of the gamma(0.001, 0.001) distribution for $\tau = 1/\sigma^2$ in Section 4. Alternative methods for deriving prior distributions are then explained.

The standard noninformative Jeffreys prior [13] for the variance, σ^2, of a normal distribution is of the form

$$p(\sigma^2) \propto \frac{1}{\sigma^2}$$

or equivalently $p(\tau) \propto 1/\tau$, where $\tau = \frac{1}{\sigma^2}$

which leads, when combined with an improper uniform prior for the mean, to the classical t-distribution for the posterior distribution of the mean.

However, as DuMouchel [14, 15] points out, when σ^2 is the variance of a random effect in a hierarchical model, the boundary value $\sigma^2 = 0$ is supported by a nonnegligible likelihood because it is theoretically possible that there are no trial-specific random effects. The asymptote at zero of $p(\sigma^2) \propto 1/\sigma^2$ is then sufficient to lead to an *improper* posterior distribution, and so suitable alternative priors should be used.

6 ALTERNATIVE PROPER PRIORS FOR POPULATION VARIANCE

6.1 Prior 1: "Just" Proper Prior

The standard noninformative prior for the population variance, $p(\sigma^2) \propto 1/\sigma^2$, is formally equivalent to an inverse gamma(0, 0) distribution. A "just" proper approximation to this is an inverse gamma(0.001, 0.001) distribution for σ^2 (Figure 3), or equivalently a gamma(0.001, 0.001) distribution for $\tau = 1/\sigma^2$. This was the prior used in Section 4.

6.2 Prior 2: Proper Prior by Introspection

By using knowledge of the particular context of the problem being analyzed a reasonable prior may be derived based on judgments about the likely size of between-study variability.

First, suppose that before looking at the data we consider it is plausible to observe one order of magnitude spread in odds ratios between the studies, so that the ratio of the maximum odds ratio to the minimum odds ratio could be 10. Converting to a log scale, this can be interpreted as having a prior belief that 95% of studies (contained in a range $\pm 1.96\sigma$)

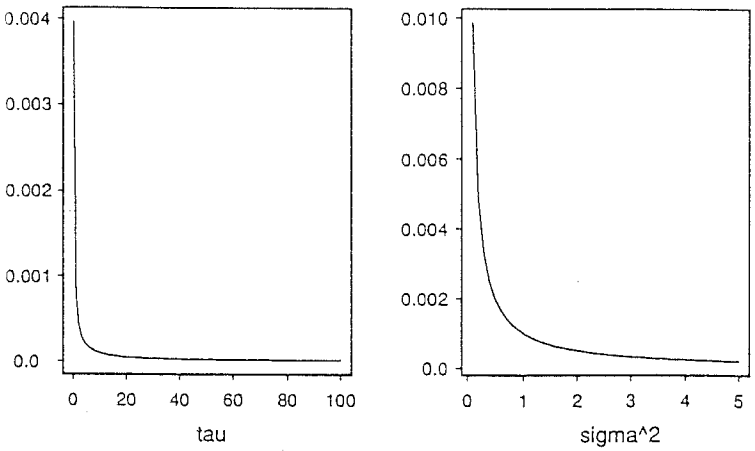


Figure 3 Inverse gamma (0.001, 0.001) distribution for σ^2.

cover log odds ratios in a range of log 10 = 2.3, and hence that a reasonable estimate of σ^2 is $(2.3/(2*1.96))^2 = 0.34$. Suppose in addition that we believe it would be very unlikely to observe two orders of magnitude difference (i.e., a hundredfold difference) between odds ratios in a meta-analysis. The ratio of the maximum odds ratio to the minimum odds ratio would then be 100 and this would lead to a "high" value of σ^2 of 1.37. A gamma(3, 1)

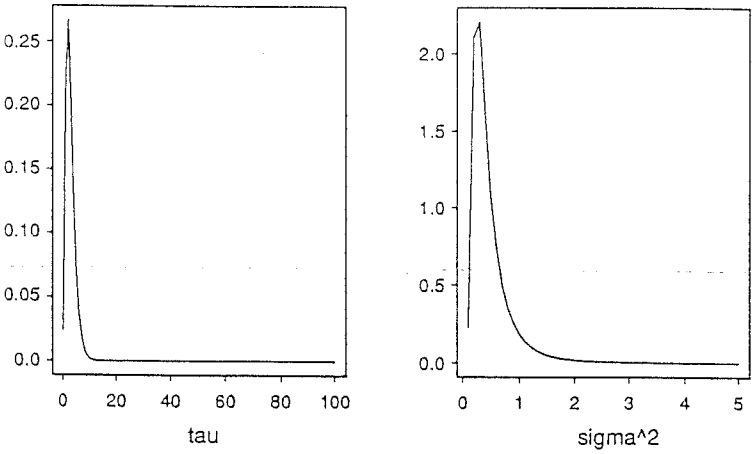


Figure 4 Inverse gamma (3, 1) distribution for σ^2.

distribution has a mean of 1/0.33 and a 96% probability of exceeding 1/1.37, and hence an inverse gamma(3, 1) distribution for σ^2 could represent these beliefs (Figure 4).

6.3 Prior 3: Proper Prior by Empirical Methods

From a large number of meta-analyses, each providing an estimate of the between-study variability, a prior for σ^2 can be obtained empirically. Parmar *et al.* [16] reviewed 30 meta-analyses of trials in a number of areas (Table 2) and the resulting estimates of σ^2 were used to find an inverse

Table 2 Estimates of $\hat{\sigma}^2$ from 30 Meta-analyses [16]

	No. of trials	$\hat{\sigma}^2$
1	21	0.044
2	5	0.055
3	21	0.000
4	64	0.145
5	7	0.918
6	8	0.005
7	21	0.000
8	26	0.000
9	4	0.000
10	6	0.063
11	13	0.111
12	13	0.012
13	7	0.023
14	10	0.109
15	7	0.011
16	11	0.477
17	6	1.297
18	19	0.056
19	11	1.352
20	21	0.000
21	31	0.000
22	28	0.000
23	31	0.016
24	9	0.000
25	23	0.004
26	20	0.076
27	8	0.000
28	10	0.064
29	14	0.327
30	8	0.533

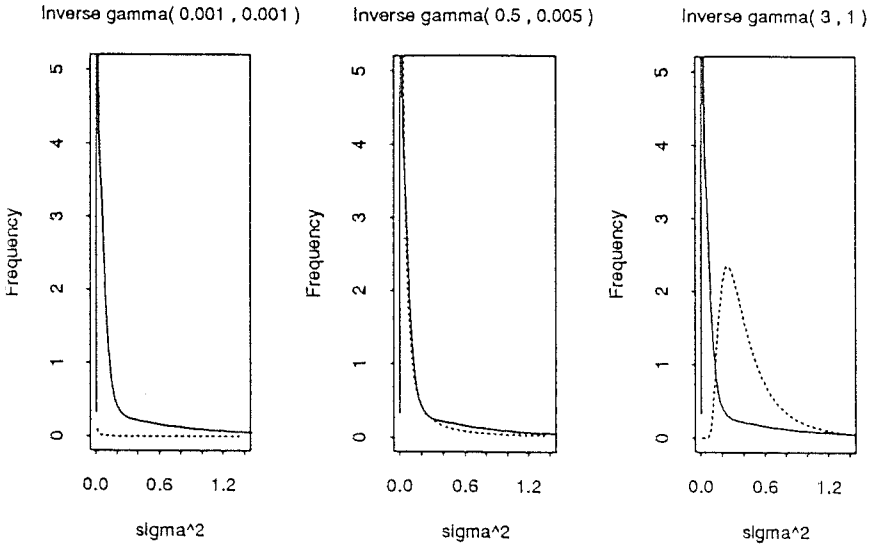


Figure 5 Kernel estimate of density for $\hat{\sigma}^2$ from meta-analyses of cancer trials (solid line = kernel estimate, dotted line = inverse gamma density).

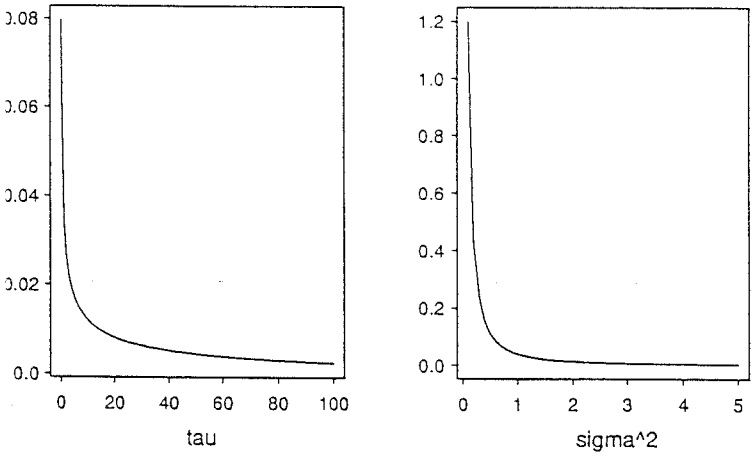


Figure 6 Inverse gamma (0.5, 0.005) distribution for σ^2.

gamma distribution that "best" described the observed between-study variability. We note that this empirical prior is derived from a number of different medical areas, but it should give an indication of plausible values of population variances in random-effects analyses.

The kernel estimate of the density of the sample was considered (Figure 5), and it suggested that a reasonable prior would be an inverse gamma (0.5, 0.005), shown in Figure 6. The kernel estimate of the density for the sample of $\hat{\sigma}^2$s differs quite considerably from the priors derived in Section 6.1 and 6.2 (Figure 5).

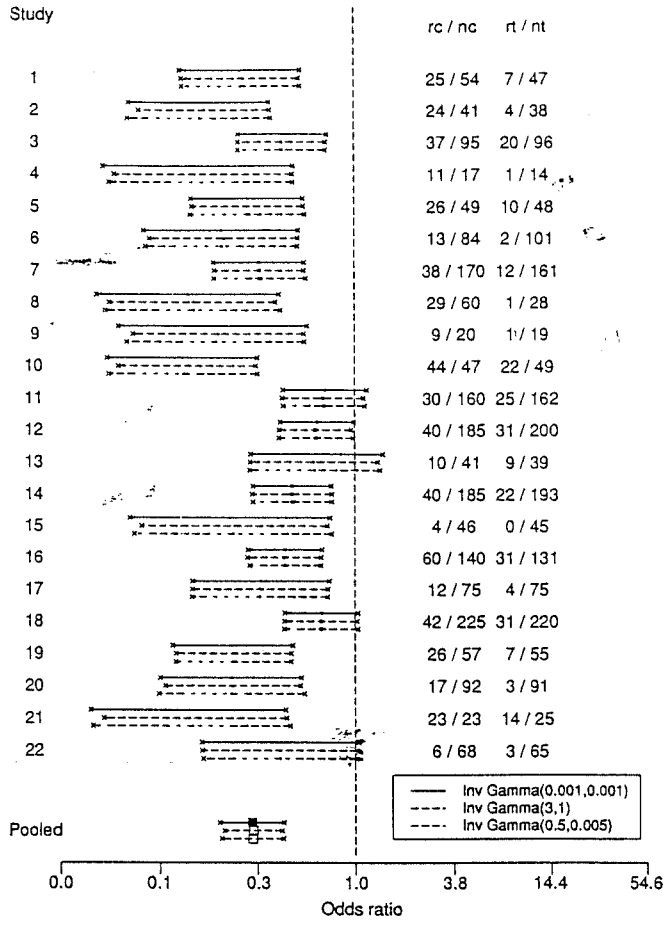


Figure 7 Meta-analysis of the decontamination example using three different priors for σ^2. Posterior mean and 95% credible intervals obtained from Gibbs sampling.

COMPARISON OF ALTERNATIVE PRIORS

ne three different priors gave almost identical results when applied to e meta-analysis of the selective decontamination example (Figure 7). ne only difference was that the estimated credible intervals from the trospective inverse gamma(3, 1) were slightly narrower compared with ose from the other two priors (particularly for the smaller study esti-ates). This is due to the just proper prior and the data-driven prior both voring small values of between-study variability σ^2, and therefore more the residual variability is accounted for as within-study variability.

Although the three approaches led to different inverse gamma priors, ere was actually little practical difference between the distributions. nis suggests that the just proper prior may well be a reasonable prior r σ^2, as it provides reasonable support to a wide range of plausible values σ^2.

DISCUSSION

he selective decontamination of the digestive tract meta-analysis was sed here to highlight a number of problems that frequently occur in meta-nalysis. In particular, any variability between studies should be taken to account in the analysis. The fixed-effect analysis estimated a 64% duction in infections with selective decontamination compared with the ontrols (odds ratio 0.36, 95% confidence interval 0.31, 0.43). When the ndom-effects model is used, however, a stronger relationship is found ith an estimated 75% reduction in infections (odds ratio 0.25, 95% proba-lity interval 0.16, 0.36). Because the random-effects analysis acknowl-lges the presence of heterogeneity, the uncertainty of estimating a pooled fect when individual study estimates vary greatly is reflected in its much ider 95% interval.

When heterogeneity among the studies in a meta-analysis is apparent, ossible differences between the studies that could be instrumental in ausing the heterogeneity could be adjusted for. In the decontamination xample the impression gained from Figure 2 is that the larger studies ave the less extreme results. In the Bayesian random-effects model we ould easily include the log of the study size ($n_i = n_i^T + n_i^C$) as a covariate o that

$$\text{logit}(p_i^T) - \text{logit}(p_i^C) = \delta_i + \beta(\log n_i - \log n.)$$

here $\log n. = (1/I) \sum_i \log n_i$. The estimated overall treatment effect ien has the slightly odd interpretation as the effect expected in an "aver-ge" size trial.

215

Other factors that could explain the difference in the effects of the studies include the variations in patient mix both within and between the studies. For example, the percentages of trauma, surgical, and medical patients varied widely. Differences in study design (whether they were double blind or not and variations in diagnostic measures) and treatment regimens could also have been important contributors. Heterogeneity should be explained, if possible, and not simply accommodated within a random-effects analysis.

Although the fully Bayesian model described here enables adjustments that can account for some of the between-study variability, interpretation requires care. In the example above it may be felt that the larger studies carry more credibility and we might expect carefully controlled studies to tend to show smaller effects. The larger centers are also more likely to stay up with recent developments and use more standard definitions of disease and treatment. The more extreme effects being observed in smaller studies might also point to publication bias, as small studies with small observed effects may not have been significant enough to be published.

The three population variance priors compared in this chapter, although themselves specific to the particular model, were derived using approaches that can be used to develop priors in many other areas.

This chapter illustrates just one application of BUGS. The software is capable of handling a wide range of problems, including hierarchical random effects and measurement error in generalized linear models, latent variable and mixture models, and various forms of missing and censored data. It also provides a mechanism for attaching models of different types within a single structure: essentially the graphical formalism permits models of arbitrary complexity. Naturally, for complex models issues of convergence of the sampled values becomes crucial, and a range of diagnostics are possible. BUGS is freely available, with full documentation, by anonymous ftp from the second author (e-mail: bugs@mrc-bsu.cam.ac.uk).

ACKNOWLEDGMENT

The BUGS project is supported by a grant from the UK Economic and Social Science Research Council's initiative for the Analysis of Large and Complex Datasets.

REFERENCES

1. Selective Decontamination of the Digestive Tract Trialists' Collaborative Group. (1993). Meta-analysis of randomised controlled trials of selective decontamination of the digestive tract. *British Medical Journal* 307: 525–532.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85: 398–409.

Antiplatelet Trialists' Collaboration. (1988). Secondary prevention of vascular disease by prolonged antiplatelet treatments. *British Medical Journal* 296: 320–331.

DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* 7: 177–188.

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Analysis*. John Wiley & Sons, Chichester, UK.

Lauritzen, S. L., Dawid, A. P., Larsen, B. N., and Leimer, H.-G. (1990). Independence properties of directed Markov fields. *Networks* 20: 491–505.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6: 721–741.

Gilks, W. R., Clayton, D. G., Spiegelhalter, D. J., Best, N. G., McNeil, A. J., Sharples, L. D., and Kirby, A. J. (1993). Modelling complexity: applications of Gibbs sampling in medicine (with discussion). *Journal of the Royal Statistical Society Series B* 55: 39–102.

Thomas, A., Spiegelhalter, D. J., and Gilks, W. R. (1992). BUGS: A program to perform Bayesian inference using Gibbs sampling. In *Bayesian Statistics 4*, ed. Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M. pp. 837–842. Clarendon Press, Oxford.

Gilks, W. R., Thomas, A., and Spiegelhalter, D. J. (1994). A language and program for complex Bayesian modelling. *The Statistician*, 43: 169–178.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* 7: 457–472.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian Statistics 4*, ed. Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., pp. 169–194. Clarendon Press, Oxford.

Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, MA.

DuMouchel, W. H. (1990). Bayesian meta-analysis. In *Statistical Methodology in the Pharmaceutical Sciences*, ed. Berry, D. A., 509–529. Marcel Dekker, New York.

DuMouchel, W. and Waternaux, C. (1992). Hierarchical models for combining information and for meta-analyses (discussion). In *Bayesian Statistics 4*, ed. Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., pp. 338–341. Clarendon Press, Oxford.

Parmar, M. H., Hughes, M., and Freedman, L. S. (1995). Heterogeneity and small trials in meta-analysis. In preparation.