STATISTICS IN MEDICINE, VOL. 11, 141–158 (1992)

META-ANALYSIS FOR 2 × 2 TABLES: A BAYESIAN APPROACH

JOHN B. CARLIN

Clinical Epidemiology and Biostatistics Unit, Royal Children's Hospital, Flemington Road, Parkville, Vic 3052, Australia

SUMMARY

This paper develops and implements a fully Bayesian approach to meta-analysis, in which uncertainty about effects in distinct but comparable studies is represented by an exchangeable prior distribution. Specifically, hierarchical normal models are used, along with a parametrization that allows a unified approach to deal easily with both clinical trial and case-control study data. Monte Carlo methods are used to obtain posterior distributions for parameters of interest, integrating out the unknown parameters of the exchangeable prior or 'random effects' distribution. The approach is illustrated with two examples, the first involving a data set on the effect of beta-blockers after myocardial infarction, and the second based on a classic data set comprising 14 case-control studies on the effects of smoking on lung cancer. In both examples, rather different conclusions from those previously published are obtained. In particular, it is claimed that widely used methods for meta-analysis, which involve complete pooling of 'O–E' values, lead to understatement of uncertainty in the estimation of overall or typical effect size.

1. INTRODUCTION

Meta-analysis has been defined as the 'statistical analysis of a collection of analytic results for the purpose of integrating the findings'.[1] The last five years have seen rapidly increasing interest in meta-analysis in the medical research literature, an important early work being the study of Yusuf *et al.*[2] When used for estimation rather than for significance testing, the methodology proposed in that paper (the 'Peto method') leads effectively to a complete pooling of effect-size estimates, assuming the target of analysis is a common, constant, 'true effect'. (This assumption is avoided in classical significance testing with the Peto method, since sampling distributions are then based on the assumption of a null effect throughout, with an explicit alternative hypothesis not required. The Bayesian perspective of this paper addresses estimation throughout.) An alternative approach, due to DerSimonian and Laird,[1] invokes a random effects model, thus allowing for variation among the true effects of the different studies. There has been considerable discussion in the literature on the relative merits of these two approaches, and on many other less directly statistical issues involved in the conduct of meta-analyses.[3] Berlin *et al.*[4] conducted an empirical comparison of the two methods; their paper is also a good source of references for a number of published meta-analyses.

This paper develops a conceptually new approach to meta-analysis, closely related to the random effects methodology of DerSimonian and Laird[1] but based on Bayesian principles and emphasizing the assessment of uncertainty in meta-analytic conclusions. After introducing two archetypal examples from the fields of clinical trials and of case-control studies (Section 2), and

0277–6715/92/020141–18$09.00

Received June 1990
Revised May 1991

then reviewing methods for single-study analysis in each of these contexts (Section 3), I discuss the conceptual underpinning of meta-analysis in Section 4, emphasizing the need for clarity in the stated aims of any effort to integrate study results. In the following section, the specific framework of hierarchical normal models is developed and in Section 6 this is used to provide detailed illustrations of the Bayesian approach based on the two examples.

2. EXAMPLES

The ideas of the paper will be developed in the context of two specific examples. Not only do the two examples illustrate rather divergent meta-analytic conclusions (Section 6), but they span two major fields of epidemiological research: case-control studies and clinical trials.

The first example is taken from Yusuf *et al.*[2] and is an analysis of 22 long-term clinical trials of the prophylactic use of beta-blockers after myocardial infarction. Principal interest lies in the comparison of mortality between treated and control groups. The data for analysis are shown in the third and fourth columns of Table I, and consist of a standard 2×2 table for each trial included in the meta-analysis. Mortality varies from 3 per cent to 21 per cent across all the studies, most of which show a modest, though not statistically significant, benefit from the use of beta-blockers. Yusuf *et al.*'s analysis concludes that on combining the study results, there is strong evidence of a reduction of approximately 20 per cent in mortality due to beta-broker treatment.

The second example involves a now classic data set in the epidemiological literature. Cornfield[5] analysed a collection of 14 case-control studies that address the question of association between smoking and lung cancer. See Table II for a summary of the data. The same data have subsequently been analysed by Gart[6] and Cox.[7] Each of these analyses addressed meta-analytic issues, that is, the extent to which the effect estimates in the different studies agree with each other and, to the extent that homogeneity is plausible, methods for pooling the study results to enable an overall estimate of lung cancer risk in smokers compared with non-smokers.

3. SINGLE-STUDY ANALYSIS

Each of the examples involves data in the form of several 2×2 tables, the simplest type of comparative epidemiological data. In the clinical trials, with n_0 subjects in the control group and n_1 in the treatment group, giving rise to x_0 and x_1 deaths in treatment and control groups, respectively, the usual sampling model involves two independent binomial distributions with risks of death π_0 and π_1, respectively. The target of inference may be one or other of the risk difference, $\pi_1 - \pi_0$, the risk ratio (or relative risk), π_1/π_0, or the odds ratio,

$$\rho = \pi_1(1 - \pi_0)/\pi_0(1 - \pi_1).$$

In a similar notation, for the case-control study we suppose that sampling results in n_1 cases and n_0 controls, of whom x_1 and x_0, respectively, have been exposed to the putative etiologic agent in question. Again, the binomial model is assumed, whereby the 'risk' of exposure is π_1 among cases and π_0 among controls. The difference in exposure risks, $\pi_1 - \pi_0$, is of no interest, but the odds ratio, ρ, defined as before, is of interest for well-known reasons related to Bayes' Theorem (see, for example, Greenland and Thomas[8] for a thorough review of this area). Because of its common interpretability in both prospective and case-control designs, I concentrate in the following on inference for the odds ratio, ρ, or more specifically, the log odds ratio, $\Delta = \log(\rho)$.

Before treating methods for combining study results, I briefly review two methods for asymptotic inference in single studies based on crude analysis of a 2×2 table. In the clinical trials

Table I. Results of 22 clinical trials of beta-blockers for reducing mortality after myocardial infarction, from Yusuf *et al.*,[2] with approximate posterior mean and standard deviation of the log odds ratio, under three methods of analysis

Study number*	Basic data (deaths/total) Treated	Control	Study weight	Crude Mean	SD	Effect estimates EB(ML)† Mean	SD	Bayes‡ Mean	SD
1 (5·1)	3/38	3/39	0·11	0·03	0·84	−0·24	0·13	−0·23	0·20
2 (5·2)	7/114	14/116	0·35	−0·71	0·46	−0·28	0·13	−0·31	0·19
3 (5·3)	5/69	11/93	0·27	−0·51	0·53	−0·26	0·13	−0·27	0·19
4 (5·4)	102/1533	127/1520	2·40	−0·24	0·14	−0·25	0·10	−0·25	0·11
5 (5·5)	28/355	27/365	0·85	0·07	0·28	−0·20	0·12	−0·16	0·16
6 (5·6)	4/59	6/52	0·18	−0·57	0·66	−0·26	0·13	−0·27	0·19
7 (5·7)	98/945	152/939	2·43	−0·50	0·14	−0·36	0·09	−0·40	0·12
8 (5·8)	60/632	48/471	1·41	−0·08	0·20	−0·21	0·11	−0·18	0·14
9 (5·9)	25/278	37/282	0·91	−0·42	0·27	−0·28	0·12	−0·29	0·15
10 (5·10)	138/1916	188/1921	2·89	−0·33	0·12	−0·29	0·09	−0·30	0·10
11 (5·11)	64/873	52/583	1·49	−0·22	0·20	−0·24	0·11	−0·23	0·14
12 (5·12)	45/263	47/266	1·19	−0·04	0·23	−0·20	0·12	−0·17	0·15
13 (5·13)	9/291	16/293	0·44	−0·58	0·41	−0·27	0·13	−0·30	0·18
14 (5·14)	57/858	45/883	1·42	0·28	0·20	−0·12	0·11	−0·03	0·17
15 (5·15)	25/154	31/147	0·78	−0·32	0·30	−0·26	0·12	−0·27	0·16
16 (2·1)	33/207	38/213	0·96	−0·14	0·26	−0·23	0·12	−0·21	0·16
17 (2·3)	28/251	12/122	0·56	0·14	0·36	−0·21	0·12	−0·17	0·17
18 (2·4)	8/151	6/154	0·25	0·32	0·55	−0·22	0·13	−0·18	0·19
19 (2·5)	6/174	3/134	0·16	0·42	0·68	−0·23	0·13	−0·20	0·19
20 (4·1)	32/209	40/218	0·98	−0·22	0·26	−0·24	0·12	−0·24	0·15
21 (4·2)	27/391	43/364	1·02	−0·58	0·25	−0·31	0·12	−0·35	0·16
22 (4·3)	22/680	39/674	0·95	−0·59	0·26	−0·30	0·12	−0·35	0·16

* In parentheses, number used by Yusuf *et al.*[2]

† Conditional posterior mean and SD based on the ML estimate $\sigma_\Delta^2 = 0{\cdot}0178$.

‡ Estimates based on Monte Carlo simulation using 5000 drawn values. This assures a standard error for the mean of, at most, about 0·003, and for the standard deviation of about 1 per cent.

literature, the approach to meta-analysis proposed by Yusuf *et al.*[2] is based on ideas that go back to the paper of Mantel and Haenszel.[9] It corresponds, in a single study, to estimating the log odds ratio as

$$\hat{\Delta}_S = \frac{O - E}{V}, \tag{1}$$

where $O = x_1$, $E = n_1(x_0 + x_1)/N$, and $V = n_1 n_0 (x_0 + x_1)(N - x_0 - x_1)/[N^2(N-1)]$, and approximate normal-based confidence intervals are obtained using the estimated standard error

$$\sigma_S(\hat{\Delta}_S) = V^{-1/2}. \tag{2}$$

The estimate (1) may be motivated as the first Newton–Raphson step from zero towards the maximum likelihood (ML) estimate.[2, 10] The asymptotic validity of the confidence interval depends on the approximate normality of the score statistic in a binomial or hypergeometric model under the null hypothesis that $\Delta = 0$. From a Bayesian point of view, the estimate and standard error can be regarded as approximating a posterior mean and standard deviation for the parameter Δ, as long as sample size is large and prior information is relatively weak, and assuming that Δ is reasonably small.

Table II. Results of 14 case-control studies of the association between smoking and lung cancer,[5, 7] with approximate posterior mean and standard deviation of the log odds ratio, under three methods of analysis

Study number	Basic data (number of smokers/total) Cases	Controls	Study weight	Crude Mean	SD	Effect estimates EB(ML)† Mean	SD	Bayes‡ Mean	SD
1	83/86	72/86	0·555	1·683	0·656	1·552	0·378	1·574	0·420
2	90/93	227/270	0·612	1·737	0·610	1·576	0·368	1·597	0·408
3	129/136	81/100	0·851	1·464	0·464	1·481	0·328	1·474	0·354
4	70/82	397/522	1·161	0·608	0·329	0·934	0·268	0·874	0·311
5	412/444	299/430	1·481	1·730	0·211	1·685	0·192	1·694	0·197
6	597/605	666/780	1·058	2·547	0·370	2·101	0·289	2·166	0·352
7	88/93	174/186	0·702	0·194	0·548	0·998	0·353	0·881	0·444
8	1350/1357	1296/1357	0·985	2·206	0·401	1·876	0·303	1·934	0·345
9	60/63	106/133	0·586	1·628	0·630	1·538	0·372	1·536	0·420
10	459/477	534/615	1·324	1·353	0·268	1·392	0·232	1·385	0·242
11	724/728	246/300	—	3·682	0·523	—	—	—	—
12	499/518	462/518	1·310	1·158	0·273	1·255	0·235	1·240	0·245
13	451/490	1729/2365	1·582	1·448	0·173	1·454	0·162	1·456	0·161
14	260/265	259/287	0·795	1·727	0·493	1·595	0·337	1·614	0·372

† Conditional posterior mean and SD based on the ML estimate $\sigma_\Delta^2 = 0{\cdot}186$
‡ Estimates based on Monte Carlo simulation using 5000 drawn values

In the case-control context, tradition favours estimates based on empirical logits. Thus we may estimate Δ as

$$\hat{\Delta}_L = \log\left(\frac{x_1}{n_1 - x_1}\right) - \log\left(\frac{x_0}{n_0 - x_0}\right), \tag{3}$$

with approximate standard error[11]

$$\sigma_L(\hat{\Delta}_L) = \left[\frac{1}{x_1} + \frac{1}{n_1 - x_1} + \frac{1}{x_0} + \frac{1}{n_0 - x_0}\right]^{1/2}. \tag{4}$$

Cox[7] discusses refinements of these estimates that improve the asymptotic normality of the sampling distributions involved (in particular, it is often recommended to add 1/2 to each of the four counts in the 2×2 table), but for practical purposes, where study-specific sample sizes are assumed moderately large, such details need not concern us. The estimate (3) should in general be preferred to (1), since it is asymptotically unbiased and consistent[7] for all values of Δ.

4. MODELS FOR META-ANALYSIS

In considering a formal meta-analysis of the type of data introduced above, it is important first to reconsider the underlying logic of the inferential process for the single study. This involves a belief, at least notionally, in a population from which the study sample was drawn, such that the effects of interest (risk, odds ratio, and so on) can be conceptualized as actual population values that could theoretically be measured by exhaustive enumeration. The sample provides limited information, the uncertainty in associated inferences being gauged by posterior measures of variance, as approximated by the standard error formulae above.

To perform a meta-analysis, we need to consider what it is that we wish to estimate. I approach this question within the usual meta-analysis context where we have a collection of (in some sense)

comparable studies whose results are to be combined. Many authors have discussed issues involved in determining whether studies should be regarded as worthy of inclusion in a formal meta-analysis.[3] Given, however, the basic assumption of comparability of the studies, there are three conceptual possibilities. The first is that we view the studies as identical repeats of each other, in the sense that they may be regarded as samples from the same population, with the same outcome measures and so on. The second possibility is that the studies are regarded as *exchangeable*, in other words we view them as each bearing on the same general question, with some differences from study to study, but such that the differences are not expected *a priori* to have predictable effects favouring one study over another. The third possibility is that the studies each bear on such unrelated questions that the notion of combining their results to estimate some common quantity is not reasonable. Clearly, the second possibility in fact represents a continuum between the two extremes, and it is this exchangeability model that I propose to pursue as the most reasonable for meta-analysis in the context of examples such as those introduced above.

Using de Finetti's Theorem,[12] the assumption of exchangeability translates in modelling terms to an exchangeable prior distribution for the effects in the different studies. Such a model assumes that the effects are independently and identically distributed conditional on the values of certain hyperparameters. Notice that this result emerges directly from the judgment of a symmetric or exchangeable relationship between the study effects (although the resulting model is formally equivalent to a 'random effects' model, which from a non-Bayesian point of view would perhaps require a narrower rationale, in terms of random sampling from a population).

Given this conceptual framework, we may seek to identify the object of estimation in meta-analysis. The first target is the location of the effect-size distribution, since this represents the overall 'average' effect across all studies that could be regarded as exchangeable with those examined. Other possible targets are the effect size in any of the specific studies and the effect size in another, comparable (exchangeable) study. As Rubin[13] points out, the real aim would in general to be to estimate a *response surface* such that we could predict an effect based on known characteristics of a population and its exposure to risk. In assuming exchangeability I have discounted the possibility of modelling and hence of estimating such a response surface, but this assumption appears close to the common meta-analysis approach, where studies are only included on the basis of their assumed comparability.

Exchangeability gives no guidance, unfortunately, on the form of the joint distribution of the study effects. In this paper I provide some illustrative analyses utilizing the convenient assumption of a normal distribution for the random effects, in conjunction with the approximate normal sampling distribution of the study-specific effect estimates given above. Further discussion of model assumptions and model checking is given in Sections 6 and 7.

5. NORMAL–NORMAL ANALYSIS

Let D_i represent generically the point estimate of the effect Δ_i in the ith study, obtained from either (1) or (3), where $i = 1, \ldots, k$. The first stage of a hierarchical normal model assumes that

$$D_i | \Delta_i, \sigma_i^2 \sim \mathrm{N}(\Delta_i, \sigma_i^2), \tag{5}$$

where σ_i represents the corresponding estimated standard error, from (2) or (4), and is assumed known without error in the following. The latter is a common assumption with underlying binomial sampling distributions. Our modelling effort concentrates on the Δ_i because these are the parameters of ultimate interest. Attaching a prior distribution to the σ_i^2 independently of the Δ_i or the underlying binomial parameters is not appropriate because of the functional relationship between the variance and mean of the binomial. In similar hierarchical normal models where

the σ_i^2 are free to vary independently of the Δ_i, inferences about the means are relatively insensitive to assumptions about the variances; but further investigation of the assumption of known variances in this context would be important (see further discussion of the normal approximation to the binomial likelihood in Section 7).

At the second stage, an exchangeable normal prior is introduced:

$$\Delta_i \mid \mu_\Delta, \sigma_\Delta^2 \sim \mathrm{N}(\mu_\Delta, \sigma_\Delta^2), \tag{6}$$

where μ_Δ and σ_Δ^2 are unknown hyperparameters.

For Bayesian analysis, the model requires a third stage where a prior distribution is specified for μ_Δ and σ_Δ^2. For practical purposes it seems reasonable to assume a non-informative or locally uniform prior for μ_Δ, that is, a prior density that is constant over the region where the likelihood function has appreciable magnitude (see Box and Tiao[14], Section 1.2.5). The rationale is that even with quite a small number of studies (say 10) the combined data become relatively informative about the location of the effect-size prior distribution (so that prior information would have to be quite strong to exert much influence). In the following I also assume a locally uniform prior for σ_Δ^2, essentially for convenience, although it is easy to modify the analysis to allow a more informative prior distribution.

With the framework of normal sampling distributions (with variances assumed known) and exchangeable normal prior, with non-informative prior on its mean, standard results readily yield closed form solutions for the posterior distributions of quantities of interest, *conditional* on the variance hyperparameter, σ_Δ^2. All of these distributions are normal, since both prior and likelihood are normal, and so they are fully characterized by their moments, as given below, where the notation $\mathbf{D}$ denotes the observed data, $D_1, \ldots, D_k$. The outline derivation given here follows the simple treatment given by Rubin;[15] see Berger[16] for a general account with many further references.

First, for the overall mean effect, μ_Δ, letting $\hat{\mu}_\Delta = E(\mu_\Delta \mid \mathbf{D}, \sigma_\Delta^2)$, we have:

$$\hat{\mu}_\Delta = \frac{\sum_{j=1}^k w_j D_j}{\sum_{j=1}^k w_j}, \tag{7}$$

and

$$\mathrm{var}(\mu_\Delta \mid \mathbf{D}, \sigma_\Delta^2) = \frac{\sigma_\Delta^2}{\sum_{j=1}^k w_j}, \tag{8}$$

where $w_j = (1 + \sigma_i^2/\sigma_\Delta^2)^{-1}$ (or σ_i^{-2} in the limiting case where $\sigma_\Delta^2 = 0$). These results are not difficult to obtain, by writing down the (approximate) normal likelihood and collecting terms appropriately.

For the individual study effect in the ith study, it is simplest to begin with the *independent* normal posteriors for the Δ_i that arise when we condition on μ_Δ as well as σ_Δ^2.(These are also used in the Monte Carlo computations below.) Because of the conditional independence of the D_i's, the posterior mean of Δ_i is a simple weighted combination of D_i and μ_Δ:

$$E(\Delta_i \mid \mathbf{D}, \mu_\Delta, \sigma_\Delta^2) = w_i D_i + (1 - w_i)\mu_\Delta, \tag{9}$$

with

$$\mathrm{var}(\Delta_i \mid \mathbf{D}, \mu_\Delta, \sigma_\Delta^2) = w_i \sigma_i^2. \tag{10}$$

The posterior distribution of the Δ_i conditional only on σ_Δ^2 may then be obtained by mixing the independent normals described by (9) and (10) over the posterior for μ_Δ, described by (7) and (8).

The resulting moments are:

$$E(\Delta_i|\mathbf{D}, \sigma_\Delta^2) = w_i D_i + (1 - w_i)\hat{\mu}_\Delta \tag{11}$$

and

$$\text{var}(\Delta_i|\mathbf{D}, \sigma_\Delta^2) = w_i \sigma_i^2 + (1 - w_i)^2 \frac{\sigma_\Delta^2}{\sum_{j=1}^{k} w_j}. \tag{12}$$

Formula (11) reflects the familiar shrinkage or regression to the mean of Bayesian estimation. An expression similar to the second term in (12) gives the posterior covariance between two study effects[15].

Finally, we consider inference for a 'predicted effect', denoted Δ_j, representing an effect that is exchangeable with those that have been studied in the meta-analysis: the effect that 'would have arisen' in a $(k+1)^{st}$ study. The (conditional) posterior mean of Δ_j is just $\hat{\mu}_\Delta$ from (7). The corresponding posterior variance is the variance in (8) with the addition of σ_Δ^2, reflecting the well-known difference between inference for the mean and prediction of a new value of the Δ_i's.

A simple empirical Bayes approach[1] would proceed by using the formulae above with the substitution of a point estimate of σ_Δ^2, typically obtained by maximum likelihood, as in the examples below. (Various alternative approaches to the estimation of σ_Δ^2 are of course possible; see, for example, DerSimonian and Laird.[1]) Such an approach entails the danger of underestimating uncertainty in resultant inferences since the prior variance can rarely be precisely estimated from the data. At the extremes, if σ_Δ^2 is assumed 0, we obtain complete pooling of effect estimates, which is essentially the approach adopted by the Peto method. On the other hand, as $\sigma_\Delta^2 \to \infty$ we reach a model that assumes all effects are unrelated and should best be estimated completely independently of each other. These correspond, respectively, to the first and third possibilities of the conceptual schema in the previous section. Both may be regarded as 'fixed effects' approaches in the sense that under the first approach all estimates are pooled to obtain a single common estimate (equivalent to a Bayesian model for a common effect with a non-informative prior) while under the second approach each effect is estimated by study-specific data alone (equivalent to a Bayesian model for k unrelated effects each with non-informative priors). Note, however, that the fixed/random distinction is not very meaningful within the Bayesian paradigm where all unknowns are represented by probability distributions.

A fully Bayesian analysis requires integration of each of the conditional posterior distributions summarized by (7), (8), (11) and (12) over the posterior distribution of σ_Δ^2. Unfortunately, such integrations cannot be performed in closed form. Several options exist for approximate solution, both asymptotic and numerical.[17–19] Here I adopt a Monte Carlo approach similar to that of Rubin,[15] since although it is computationally expensive relative to more efficient quadrature schemes, it is very simple to programme and also enables the immediate generation in graphical form of approximate posterior distributions for all quantities of interest. (A quadrature scheme requires that we choose in advance a limited range of parameters of the posterior distribution, for example particular moments and quantiles.) Computational efficiency is not an issue in the typical meta-analysis, which is concerned with only a small number of summary statistics, each arising from lengthy and costly data collection.

The approximate posterior density of σ_Δ^2 (based on the normal distribution assumptions) may be written

$$\pi(\sigma_\Delta^2|\mathbf{D}) = \pi(\sigma_\Delta^2)\mathrm{L}(\sigma_\Delta^2; \mathbf{D}) \tag{13}$$

where $\pi(\sigma_\Delta^2)$ is a prior density for σ_Δ^2 and $L(\sigma_\Delta^2; \mathbf{D})$ is the likelihood function:

$$L(\sigma_\Delta^2; \mathbf{D}) = \left(\frac{\prod_{i=1}^k w_i}{(\sigma_\Delta^2)^{k-1} \sum_{i=1}^k w_i} \right)^{1/2} \exp\left\{ -\frac{1}{2\sigma_\Delta^2} \left[\sum_{i=1}^k w_i D_i^2 - \frac{(\sum_{i=1}^k w_i D_i)^2}{\sum_{i=1}^k w_i} \right] \right\}, \tag{14}$$

with appropriate modification at $\sigma_\Delta^2 = 0$. Formula (14) may be obtained by integrating μ_Δ out of the joint marginal density of $\mathbf{D}$, or by the device of Rubin.[15]

The first step in the Monte Carlo-based analysis used below is to obtain the value of σ_Δ^2 with maximum posterior density: this may be achieved, for example, by the derivative-free method of Brent.[20] Next we obtain suitable lower and upper bounds on the parameter such that the posterior density at each extreme is 100 times lower than at the maximum (the lower bound may default to zero). The resulting range is discretized, in the present implementation to 100 equal-width intervals, and the posterior density approximated by the corresponding discrete distribution. The Monte Carlo proceeds by drawing from this distribution (using the alias method[21]). For each drawn value of σ_Δ^2, a value of μ_Δ is drawn from its normal posterior with mean and variance given by (7) and (8). For selected individual study effects, posterior sample values are then drawn using the conditional mean and variance in (9) and (10). For a new study effect, values are drawn similarly from the appropriate normal distribution. The resulting approximate posterior distributions may be displayed graphically and summarized as desired by descriptive statistical methods.

6. EXAMPLES

6.1. Beta-blocker clinical trials

In Figure 1 is shown the approximate likelihood function (14) for the parameter σ_Δ^2, based on the data in Table I, and using the Peto or score-type definition of D_i and σ_i. (In fact, almost identical results obtain if the empirical logit definition is used instead.) The likelihood peaks at a non-zero value of σ_Δ^2, 0·0178, although zero is clearly a plausible value. These studies are in fact relatively homogeneous. To assess the adequacy of the normal approximations used in the analysis, a weighted normal plot[22] of the estimated random effects, at the maximum likelihood estimate of the effect-size variance, was obtained: see Figure 2. Since patient numbers are large in most of the studies, thus ensuring a good approximation to normality for the sampling distribution of the D_i's, the diagnostic plot is primarily a test for normality of the random effect distribution. There is little evidence in the plot of substantial departures from normality, although there is clearly some skewness to the left.

Figure 1 also shows the dependence of the usual point estimate, or posterior mean, and posterior standard deviation, of μ_Δ (both conditional on σ_Δ^2) on the value of σ_Δ^2, superimposed on the likelihood function. The posterior mean of the mean effect size spans a very small range, from $-0{\cdot}26$ to just over $-0{\cdot}24$, but the posterior standard deviation changes by a factor of more than two across the range of plausible values of σ_Δ^2. This illustrates a common feature of inferences in this situation: posterior variance may be underestimated if effect variation is assumed zero, or even if it is assumed adequately described by the maximum likelihood value of the effect variance. (The posterior standard deviation of μ_Δ is 0·050 and 0·060, respectively at $\sigma_\Delta^2 = 0$ and 0·0178, while upon integration using a uniform prior distribution we find a value of 0·071).

Table I compares crude, empirical Bayes, and fully Bayesian estimates, with corresponding standard deviations, for each study effect. Note that the empirical Bayes estimates are equivalent to those that would be obtained by applying the method of DerSimonian and Laird[1] to the logistic difference parameter, rather than to the risk difference. Study weights w_j, as defined

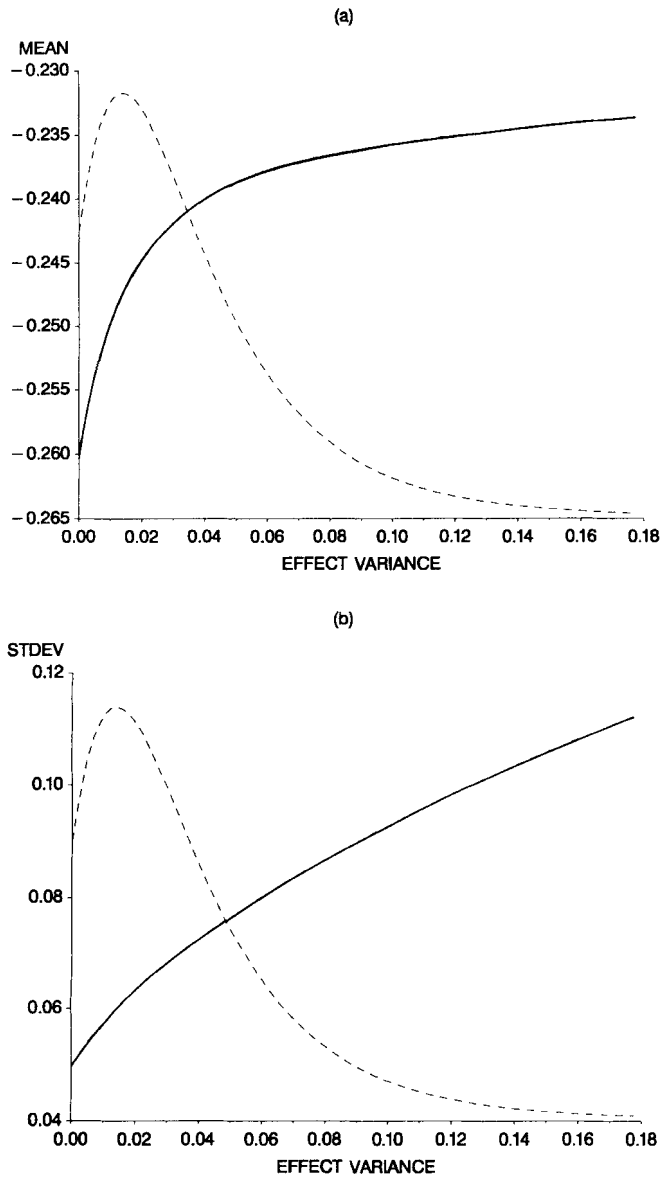


Figure 1. (*a*) Conditional posterior mean (solid line), from (7), of the mean treatment effect, μ_Δ, based on 22 beta-blocker trials in Table I, against the effect prior variance, σ_Δ^2, with the likelihood for σ_Δ^2, from (14), superimposed (dotted line). (*b*) As in (*a*) replacing the conditional posterior mean with the posterior standard deviation, from (8). n.b. The scale for the likelihood function is not shown, but the rescaling used ensures that the horizontal axis represents true zero for this function.

following (8) and using the ML value of σ_Δ^2, but normalized to average to unity, provide a quick indication of the precision of the individual study. Since the ML estimate of variance is quite small, considerable shrinkage is evident in the empirical Bayes estimates, especially for those studies with low internal precision (for example, studies, 1, 6, 18). Rather less shrinkage towards

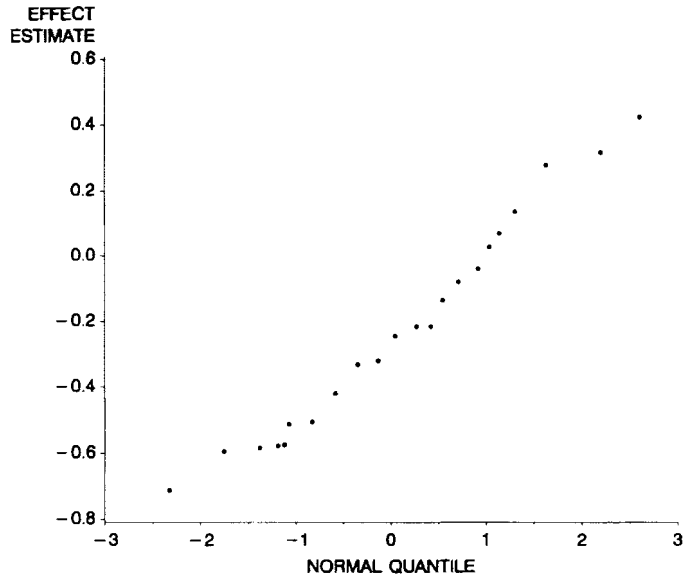


Figure 2. Weighted normal plot of empirical Bayes point estimates of effect size for 22 beta-blocker trials in Table I.

the overall mean is evident in the fully Bayesian estimates, an effect particularly pronounced for the strongest negative study, 14. The substantial degree of homogeneity between the studies is further reflected in the large reductions in posterior variance obtained upon going from the study-specific estimates to the Bayesian ones, which borrow strength from each other. Equally importantly, however, the naive empirical Bayes standard deviations are rather too small compared to the fully Bayesian ones.

Interpretation of the fully Bayes means and standard deviations is not straightforward because of the skewness in the corresponding posterior distributions. In Figure 3 are shown histograms of the simulated posteriors for four of the individual effects, as well as for the overall mean and for a predicted effect. Considerable skewness, away from the central value of the overall mean, is apparent for each of the individual effects, while the distributions of the overall mean and predicted effect show greater symmetry. The weaker studies, 2 and 18, exhibit longer tailed posterior distributions than the more precise ones, 7 and 14.

Table III gives a summary of posterior inferences for the overall mean and predicted effect. Several features are worth noting. First, an approximate 95 per cent highest posterior density interval for μ_A, when converted to the odds ratio scale gives the interval 0·68 to 0·91, which should be compared with the interval 0·70–0·85, give in Table 10 of Yusuf *et al.*[2] as a '95 per cent confidence interval for the true odds reduction [in mortality]'. The latter interval was obtained by complete pooling, equivalent to assuming $\sigma_A^2 = 0$ in the model used here. Thus, from this author's point of view, part of the explanation for the 'unusually narrow range of uncertainty' claimed by Yusuf *et al.* seems to be the use of an inappropriate model.

A related concern is that commonly used analyses tend to place undue emphasis on inference for the overall mean effect. Uncertainty about the probable treatment effect in a particular population where a study has not been performed (or indeed in a previously studied population but with a slightly modified treatment) might be more reasonably represented by inference for a new study effect, exchangeable with those for which studies have been performed, rather than for

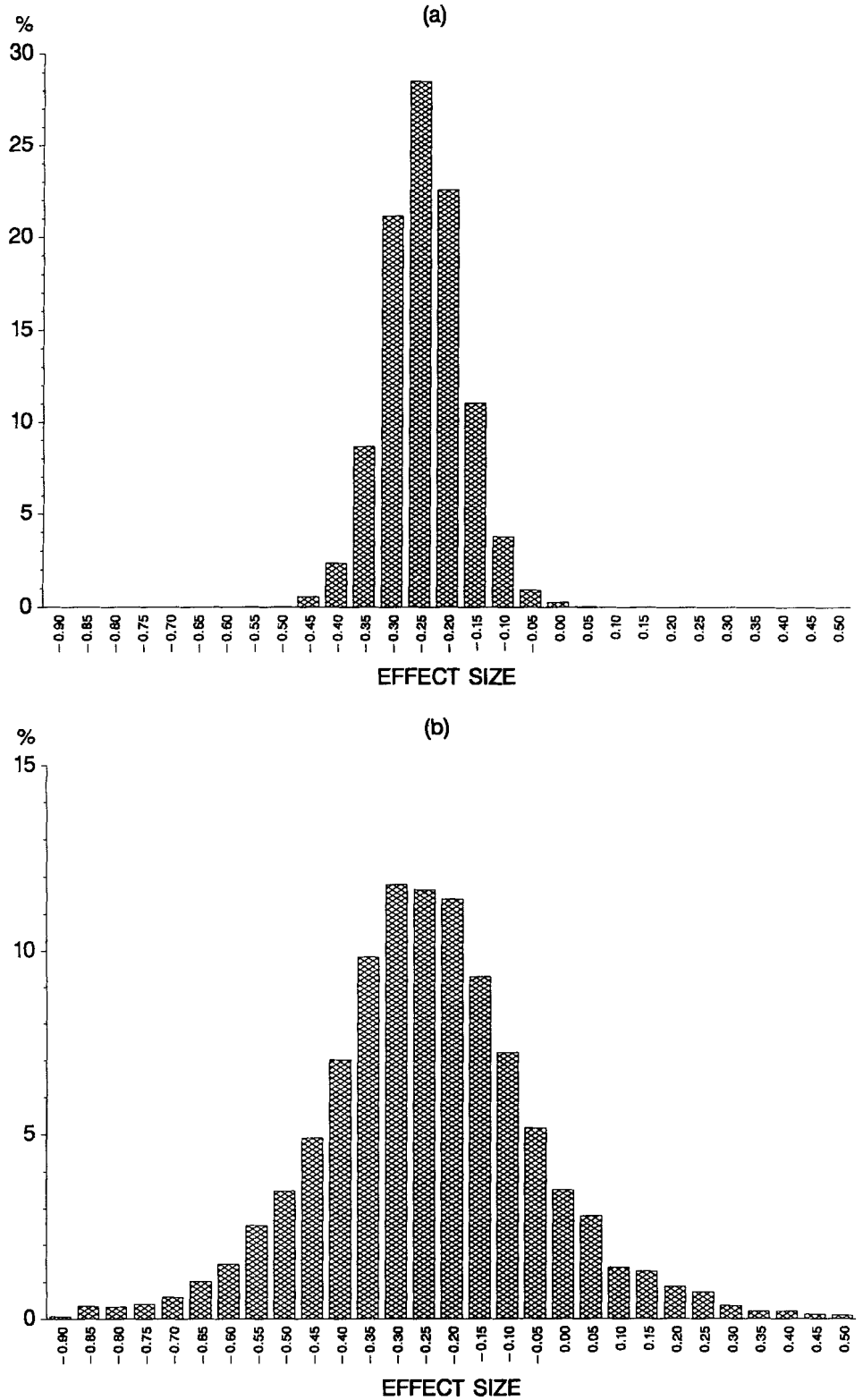
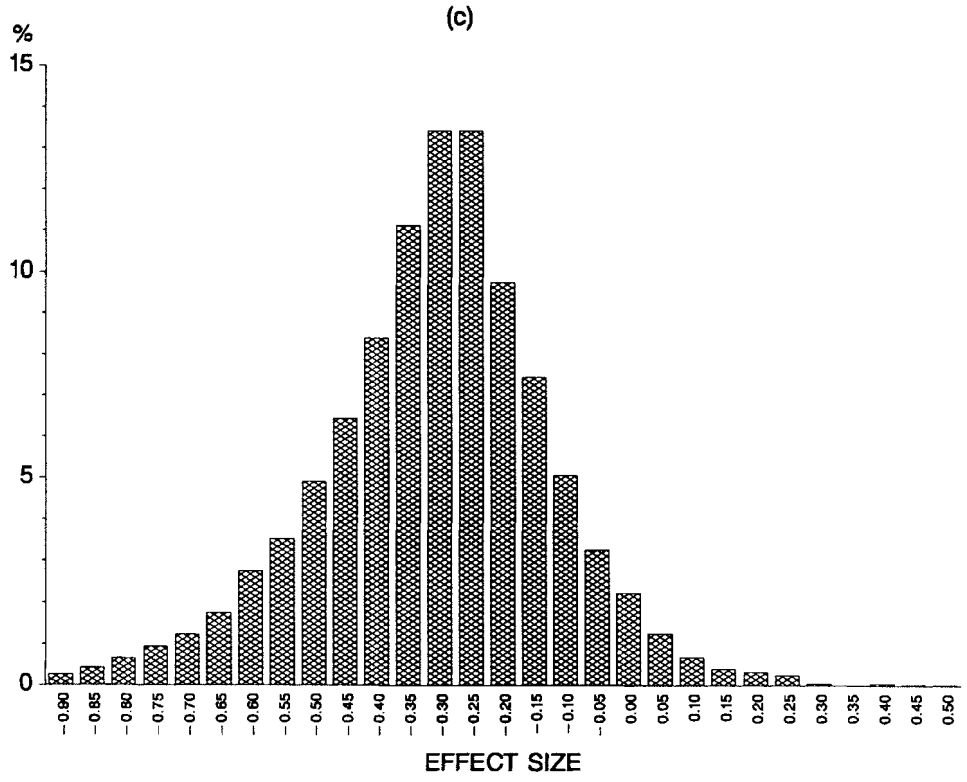


Figure 3. Histograms of simulated posterior distributions (based on 5000 drawn values) from beta-blocker trials meta-analysis: (*a*) mean of overall mean treatment effect, μ_Δ; (*b*) mean of new treatment effect; (*c*) mean of effect Δ_2; (*d*) mean of effect Δ_7; (*e*) mean of effect Δ_{14}; (*f*) mean of effect Δ_{18}.



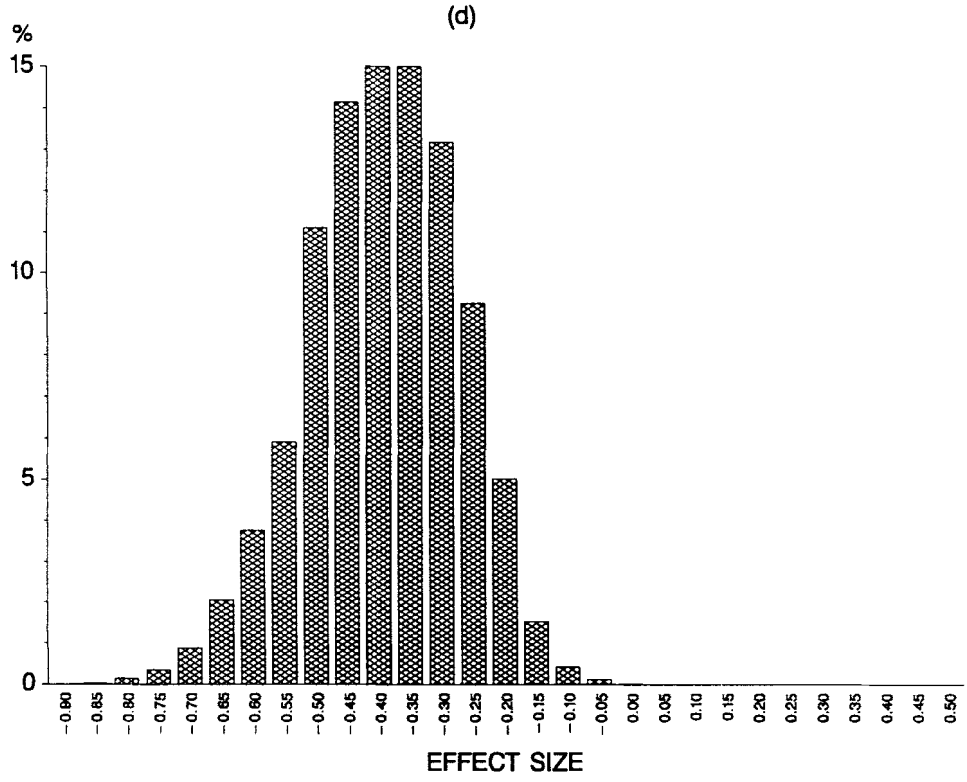
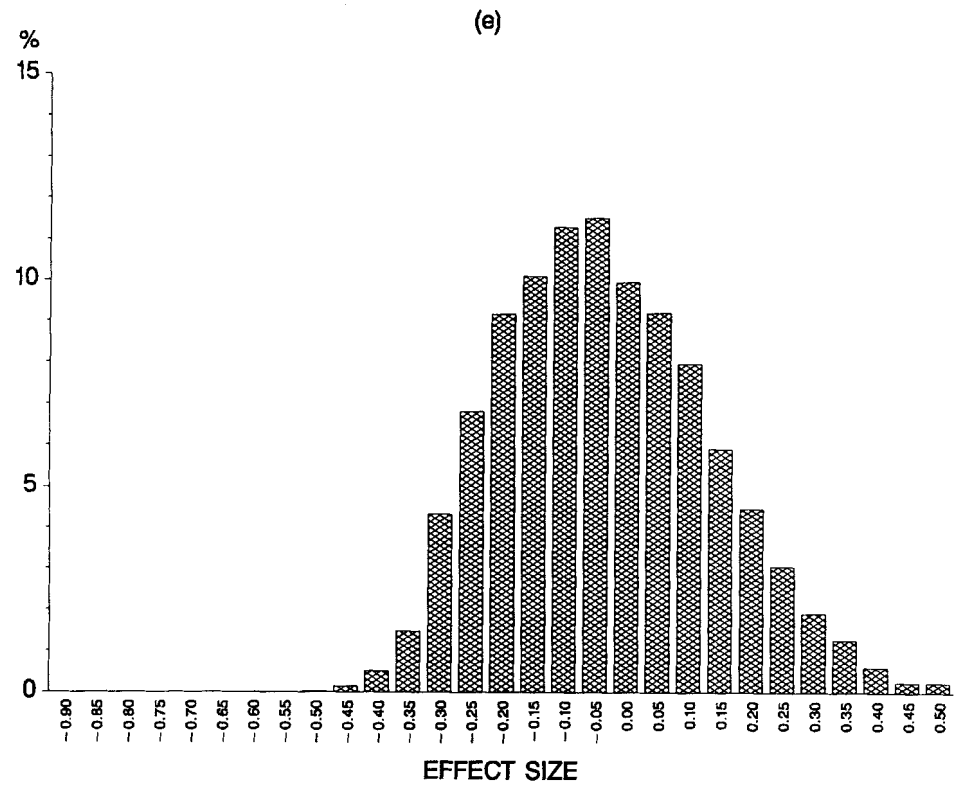


Figure 3. (Continued)



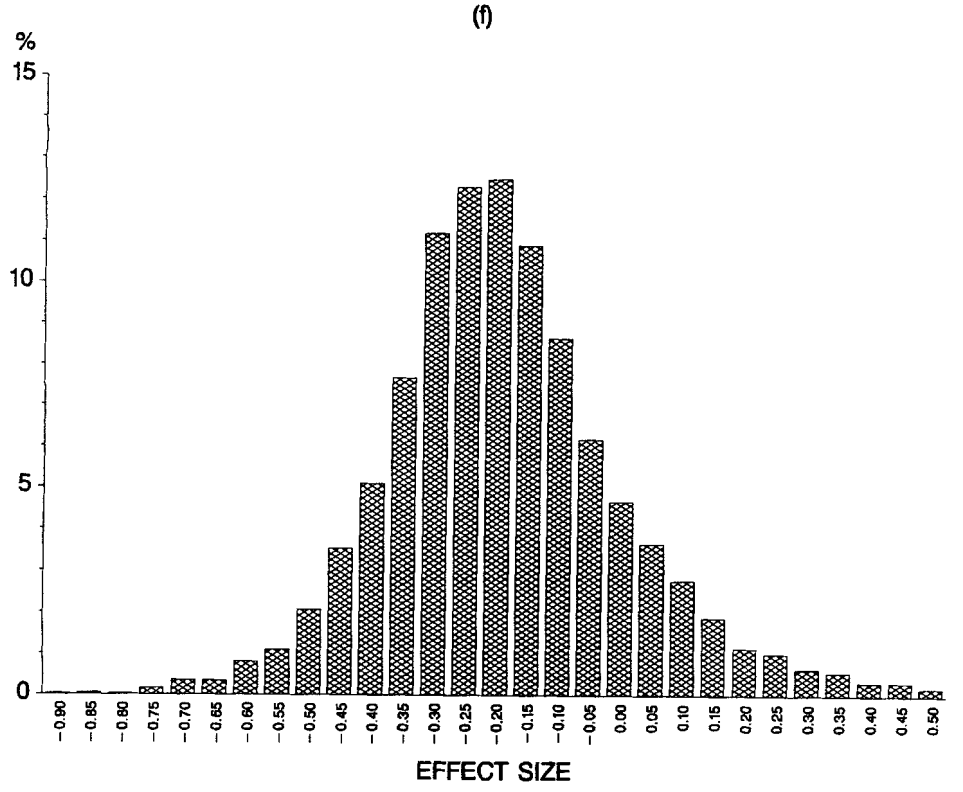


Figure 3. (Continued)

Table III. Summary of empirical Bayes and fully Bayesian inferences for the overall mean, and for the predicted effect in a new study, from meta-analysis of the beta-blocker trials in Table I

	EB(ML)		Fully Bayes			
	Mean	SD	Mean	SD	Lower 2½% (OR)	Upper 2½% (OR)
Mean, μ_Δ	−0·247	0·060	−0·243	0·071	−0·382 (0·68)	−0·096 (0·91)
Predicted effect, Δ_j	−0·247	0·132	−0·245	0·203	−0·653 (0·52)	0·185 (1·20)

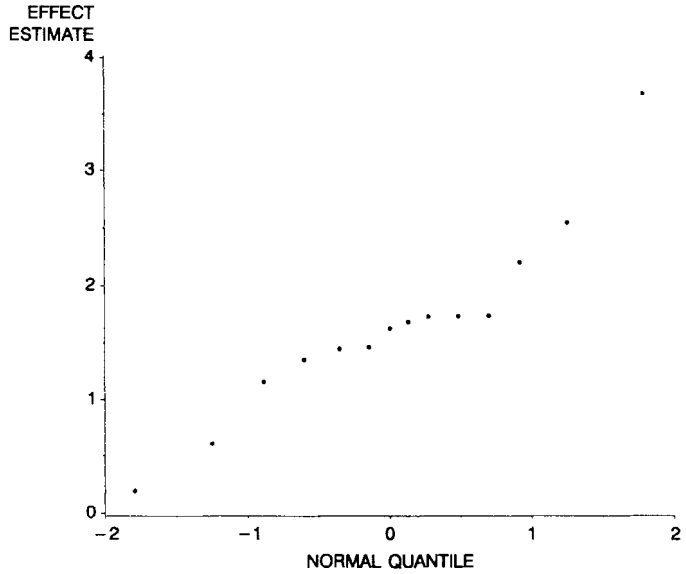


Figure 4. Weighted normal plot of empirical Bayes point estimates of effect size for all 14 lung cancer case-control studies in Table II.

the overall mean. This is a view also adopted recently by Skene and Wakefield.[23] In this case, uncertainty is of course much greater (see 'Predicted effect' in Table III; similarly, uncertainty for an individual patient would include yet another component of variation). In particular, with the above data, there is just over 10 per cent posterior probability that the effect in a new study would be negative.

6.2. Lung cancer case-control studies

Since the case-control study data in Table II come from a variety of investigations using differing protocols in a range of settings, it is not surprising that the results are less homogeneous than those of the randomized beta-blocker trials. Initial application of the hierarchical normal model, using the empirical logit definition of the D_i's, revealed a substantial lack of fit especially with regard to study number 11. This may be seen in Figure 4, which shows the weighted normal plot including all 14 studies, at the maximum likelihood estimate $\hat{\sigma}_\Delta^2 = 0{\cdot}42$.

The best approach in the face of this lack of fit might be to invoke a longer-tailed prior distribution than the normal. Further work is required to develop computational strategies for

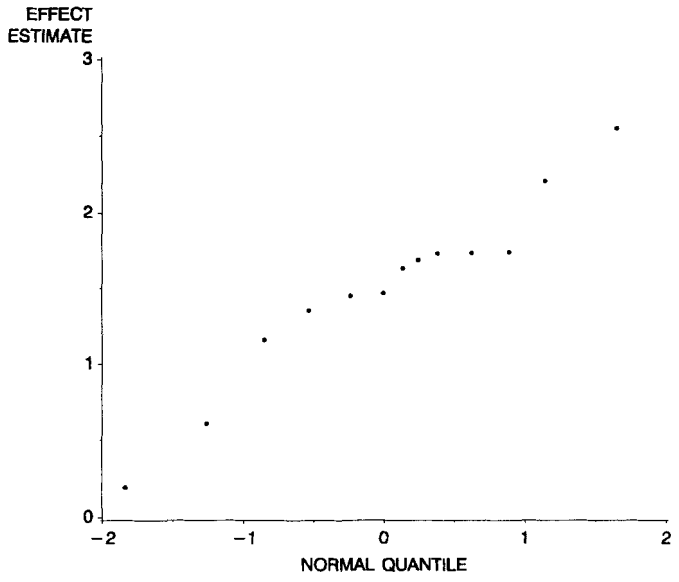


Figure 5. Weighted normal plot of empirical Bayes point estimates of effect size for 13 lung cancer case-control studies in Table II (study 1 excluded).

such an approach. Meanwhile, we adopt the more expedient method of omitting the discrepant study (a policy that might have more external support given particular knowledge of characteristics of the studies). In Figure 5, the weighted normal plot omitting study 11 and using the new ML estimate of $\hat{\sigma}_\Delta^2$, 0·186, suggests a substantially improved model fit. It should be noted that under the Bayesian or random effects approach, there is little ground for excluding more than the one study from the meta-analysis, in contrast with Cox,[7] who suggests excluding three studies (6, 8, 11), and Cornfield,[5] four studies (4, 6, 7, 11).

Continuing the analysis in the same fashion as for the previous example, we see in Figure 6 a more dispersed approximate likelihood (or posterior density) function. The value 0 for σ_Δ^2 would not be regarded as plausible, unless there were strong prior evidence to overrule that provided by the data. As usual (Figure 6(*a*)), the posterior mean for the mean effect size is insensitive to the unknown scale parameter, but the standard deviation varies substantially. Because of the greater heterogeneity when compared with the previous example, the empirical and fully Bayesian estimates in Table II are shrunk less towards the overall mean (1·50), and there is correspondingly less scope for the fully Bayesian analysis to show substantially different results from the naive analysis. As we should expect, conclusions in those studies with more extreme results (for example, study 6) and/or low internal precision (for example, study 7) are affected most by the Bayesian analysis. Table IV gives summary estimates for the lung cancer meta-analysis (based on 13 studies) along the same lines as in Table III for the beta-blocker trials.

7. DISCUSSION

This article has applied a Bayesian framework to the problem of meta-analysis, with the purpose both of clarifying the inferential aim of meta-analytic studies and of describing and implementing

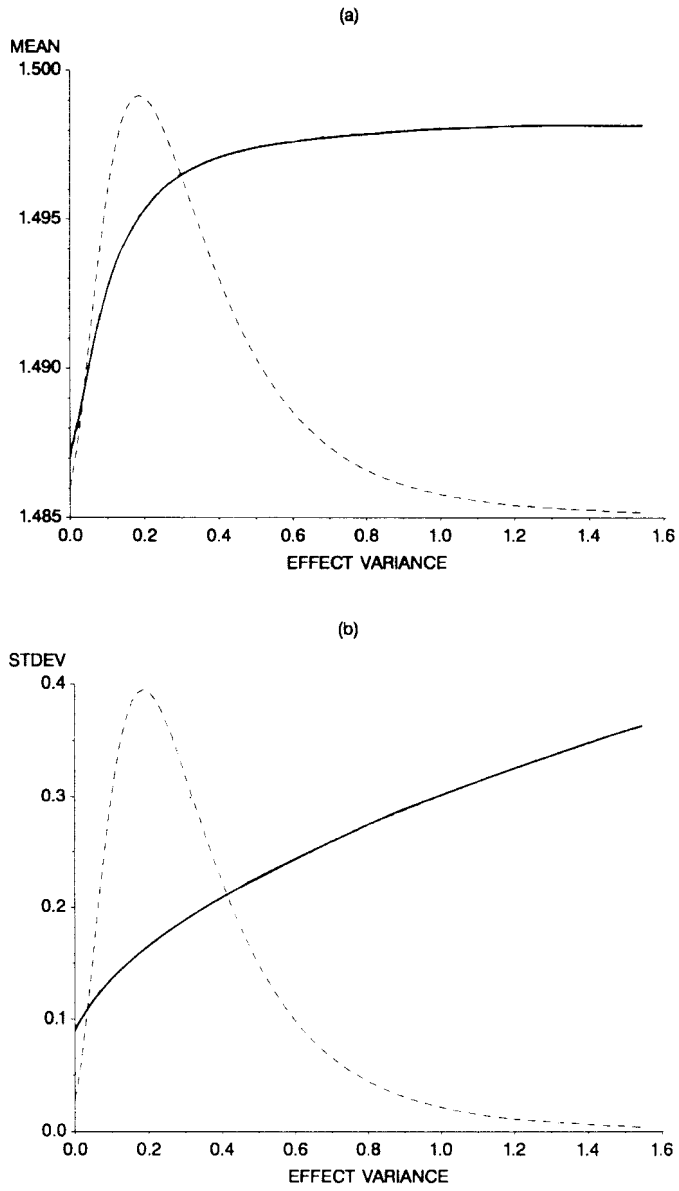


Figure 6. (*a*) Conditional posterior mean (solid line), from (7), of the mean treatment effect, μ_Δ, based on 13 case-control studies in Table II (study 11 excluded), against the effect prior variance, σ^2_Δ, with the likelihood for σ^2_Δ, from (14), superimposed (dotted line) (*b*) As in (a) replacing the conditional posterior mean with the posterior standard deviation, from (8). n.b. The scale for the likelihood function is not shown, but the rescaling used ensures that the horizontal axis represents true zero for this function

some appropriate computational tools. The analysis presented retains the usual assumption that studies to be combined are comparable, but this is held to translate naturally into an assumption of an exchangeable prior distribution for effects in different studies. It seems impossible, from the Bayesian point of view, to accept the notion that the effects being estimated in distinct studies are

Table IV. Summary of empirical Bayes and fully Bayesian inferences for the overall mean, and for the predicted effect in a new study, from meta-analysis of the lung cancer studies in Table II

	EB(ML)		Fully Bayes			
	Mean	SD	Mean	SD	Lower 2½% (OR)	Upper 2½% (OR)
Mean, μ_Δ	1·495	0·162	1·493	0·194	1·11 (3·02)	1·88 (6·53)
Predicted effect, Δ_j	1·495	0·461	1·489	0·617	0·25 (1·28)	2·74 (15·61)

identical quantities, an assumption that appears implicit in the Peto method[2] for meta-analysis, except when the latter is viewed as strictly a classical significance testing tool. The approach proposed has clear similarities to the random effects method of DerSimonian and Laird,[1] but it extends their method in allowing for a full accounting of uncertainty involved in estimating hyperparameters such as the (prior) variance of the effects. The examples illustrate that the Bayesian analysis may produce different conclusions, certainly in terms of the width of confidence intervals and even in some of the point estimates.

The methods described are based on the assumption of a hierarchical normal model, with either the empirical log odds ratio, or a score statistic-type estimate, used as the basic observation in each study. It is supposed that the sampling variance of each observation can be assumed known. The assumption of normally distributed observed values with known variance is likely to be reasonable in most situations, as long as the studies are large and observed counts are not too small. This approach ignores details associated with estimation of the usual nuisance parameter in the two-binomial model. A fully conjugate exact binomial approach, based on beta prior distributions, is possible (see, for example, Marshall,[24] who discusses the analysis of a single table), but where primary interest resides in the treatment effects, independent (beta) prior distributions for the two proportions involved in each study may not be appropriate. In fact it seems likely that little is lost in the present approach, where the nuisance parameters are ignored. Further study of this question should not be difficult in the present context of rapid advances in computational methods for Bayesian analysis: in particular, see Skene and Wakefield,[23] who discuss the same binomial-normal model as used here.

The validity of an assumption of a normal prior distribution for the true effects is more difficult to assess. In the examples, the use of a weighted normal plot for model checking was illustrated, but it is likely that alternative prior specifications would be difficult to distinguish on the basis of the data. A study of the sensitivity of conclusions to the choice of prior would be important. This study has concentrated on demonstrating the sensitivity of results to more basic aspects of the prior: whether or not its variance is zero and (if not) whether or not the variance can be precisely estimated from the data. The normal–normal analysis has the advantage of being relatively tractable. It would be possible to apply a similar analysis to the risk difference parameter, preferred by DerSimonian and Laird over the log odds ratio or logistic difference used here. The latter is preferred here because of its applicability to case-control studies as well as prospective designs; moreover, the assumption of a normal sampling distribution is more easily satisfied for estimators of the log odds ratio than for the risk difference.

The analysis presented utilized Monte Carlo computational methods for obtaining posterior distributions of quantities of interest. This approach to computation is attractive in its flexibility (in particular, the ability to produce full posterior distributions for many parameters simultaneously) and ease of programming, but for more extensive routine use it may be worth implementing more efficient algorithms.

REFERENCES

1. DerSimonian, R. and Laird, N. 'Meta-analysis in clinical trials', *Controlled Clinical Trials*, **7**, 177–188 (1986).
2. Yusuf, S., Peto, R., Lewis, J., Collins, R. and Sleight, P. 'Beta blockade during and after myocardial infarction: an overview of the randomized trials', *Progress in Cardiovascular Diseases*, **XXVII**, 335–371 (1985).
3. Yusuf, S., Simon, R. and Ellenberg, S. (eds.) 'Proceedings of the workshop on methodologic issues in overviews of randomized clinical trials, May 1986', *Statistics in Medicine*, **6**(3), entire issue (1987).
4. Berlin, J. A., Laird, N. M., Sacks, H. S. and Chalmers, T. C. 'A comparison of statistical methods for combining event rates from clinical trials', *Statistics in Medicine*, **8**, 141–151 (1989).
5. Cornfield, J. 'A statistical problem arising from retrospective studies', *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics*, **4**, 135–148 (1956).
6. Gart, J. J. 'On the combination of relative risks', *Biometrics*, **18**, 601–610 (1962).
7. Cox, D. R. *The Analysis of Binary Data*, Methuen, London, 1970.
8. Greenland, S. and Thomas, D. C. 'On the need for the rare disease assumption in case-control studies', *American Journal of Epidemiology*, **116**, 547–553 (1982).
9. Mantel, N. and Haenszel, W. 'Statistical aspects of the analysis of data from retrospective studies of disease', *Journal of the National Cancer Institute*, **22**, 719–748 (1959).
10. McCullagh, P. and Nelder, J. A. *Generalized Linear Models*, Chapman and Hall, London, 1983.
11. Woolf, B. 'On estimating the relationship between blood group and disease', *Annals of Human Genetics*, **19**, 251–253 (1955).
12. O'Hagan, A. *Probability: Methods and Measurement*, Chapman and Hall, London, 1988.
13. Rubin, D. B. 'A new perspective on meta-analysis', in Wachter, K. W. and Straf, Miron L. (eds.), *The Future of Meta-Analysis*, Russell Sage/NAS, 1989.
14. Box, G. E. P. and Tiao, G. C. *Bayseian Inference in Statistical Analysis*, Addison-Wesley, Reading, Massachusetts, 1973.
15. Rubin, D. B. 'Estimation in parallel randomized experiments', *Journal of Educational Statistics*, **6**, 377–401 (1981).
16. Berger, J. O. *Statistical Decision Theory and Bayesian Analysis*, 2nd edn. Springer-Verlag, New York, 1985.
17. Kass, R. E., Tierney, L. and Kadane, J. B. 'Asymptotics in Bayesian computation', in Bernardo, J. *et al.* (eds.), *Bayesian Statistics*, Oxford University Press, Oxford, 1988, pp. 261–278.
18. Smith, A. F. M., Skene, A. M., Shaw, J. E. H. and Naylor, J. C. 'Progress with numerical and graphical methods for practical Bayesian statistics', *The Statistician*, **36**, 75–82 (1987).
19. 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).
20. Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. *Numerical Recipes: the Art of Scientific Computing*, Cambridge University Press, Cambridge, 1986.
21. Kronmal, R. A. and Peterson, A. V. 'On the alias method for generating random variables from a discrete distribution', *The American Statistician*, **33**, 214–218 (1979).
22. Dempster, A. P. and Ryan, L. M. 'Weighted normal plots', *Journal of the American Statistical Association*, **80**, 845–850 (1985).
23. Skene, A. M. and Wakefield, J. C. 'Hierarchical models for multicentre binary response studies', *Statistics in Medicine*, **9**, 919–929 (1990).
24. Marshall, R. J., 'Bayesian analysis of case-control studies', *Statistics in Medicine*, **7**, 1223–1230 (1988).