STATISTICS IN MEDICINE, VOL. 15, 619–629 (1996)

A LIKELIHOOD APPROACH TO META-ANALYSIS WITH RANDOM EFFECTS

REBECCA J. HARDY AND SIMON G. THOMPSON

Medical Statistics Unit, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, U.K.

SUMMARY

In a meta-analysis of a set of clinical trials, a crucial but problematic component is providing an estimate and confidence interval for the overall treatment effect θ. Since in the presence of heterogeneity a fixed effect approach yields an artificially narrow confidence interval for θ, the random effects method of DerSimonian and Laird, which incorporates a moment estimator of the between-trial components of variance σ_B^2, has been advocated. With the additional distributional assumptions of normality, a confidence interval for θ may be obtained. However, this method does not provide a confidence interval for σ_B^2, nor a confidence interval for θ which takes account of the fact that σ_B^2 has to be estimated from the data. We show how a likelihood based method can be used to overcome these problems, and use profile likelihoods to construct likelihood based confidence intervals. This approach yields an appropriately widened confidence interval compared with the standard random effects method. Examples of application to a published meta-analysis and a multicentre clinical trial are discussed. It is concluded that likelihood based methods are preferred to the standard method in undertaking random effects meta-analysis when the value of σ_B^2 has an important effect on the overall estimated treatment effect.

1. INTRODUCTION

Meta-analysis provides an objective way of combining information from separate studies looking at the same clinical question and has been applied most often to treatment effects in randomized clinical trials. For example, consider k randomized trials comparing a new medication with placebo, in which each trial's treatment effect is estimated in terms of a difference in means of a quantitative variable, or, for a binary outcome, in terms of the log odds ratio. Standard meta-analysis methods for providing an overall estimate of the treatment effect rely on certain assumptions.[1,2] The fixed effect model is based on there being homogeneity of treatment effects across all k studies included in the meta-analysis. In other words it must be assumed that for each study i the estimated treatment effect $\hat{\theta}_i$ has a distribution with common mean θ and individual variance v_i for $i = 1, \ldots, k$. The treatment effect θ can then simply be estimated as a weighted average of the individual study estimates, that is

$$\hat{\theta} = \frac{\sum_{i=1}^{k} w_i \hat{\theta}_i}{\sum_{i=1}^{k} w_i} \tag{1}$$

where w_i is the weight given to study i.[3] Any choice of weight will lead to an unbiased estimate of the common treatment effect, but w_i is generally taken to be the reciprocal of the variance v_i for

CCC 0277–6715/96/060619–11

Received September 1994
Revised June 1995

study i. These particular weights provide the most precise estimate of the treatment effect, that is they minimize the variance of $\hat{\theta}$. The fact that in practice v_i must be estimated is considered in Section 4, but for the moment we assume v_i known which implies that $\text{var}(\hat{\theta}) = 1/\sum_{i=1}^{k} w_i$. Provided $\hat{\theta}$ can be considered normally distributed, for example because the $\hat{\theta}_i$ are approximately normally distributed, a confidence interval for θ can then be calculated.

If heterogeneity of the treatment effects exists across studies, then the assumption underlying the fixed effect model does not hold and the confidence interval for the overall mean will be spuriously narrow as it does not take into account the additional between-study variation. In practical medical research, clinical homogeneity is rare owing to the nature of the studies and the many variables involved, and a degree of statistical heterogeneity might be anticipated.[4] A formal test of statistical homogeneity can be performed, using the test statistic

$$Q = \sum_{i=1}^{k} w_i(\hat{\theta}_i - \hat{\theta})^2 \tag{2}$$

which has approximately a χ^2_{k-1} distribution under H_0: $\theta_i = \theta$ for all i.

The random effects model proposed by DerSimonian and Laird[5] provides a way of incorporating heterogeneity into the estimate by including a between-study component of variance σ_B^2. A moment estimator of σ_B^2 may be derived by equating the test statistic for heterogeneity Q with its expectation. Whitehead and Whitehead[1] placed the model in the context of a general random effects parametric approach, where the heterogeneity is restricted to a particular form. The model is assumed to be

$$\hat{\theta}_i \sim \text{N}(\theta_i, v_i) \tag{3}$$

$$\theta_i \sim \text{N}(\theta, \sigma_B^2) \tag{4}$$

and this therefore allows a confidence interval to be calculated for θ. The overall estimate of treatment effect is once again obtained as a weighted average, but the weights are adjusted to incorporate σ_B^2. Hence, the DerSimonian and Laird random effects estimate is given by

$$\hat{\theta} = \frac{\sum_{i=1}^{k} w_i^* \hat{\theta}_i}{\sum_{i=1}^{k} w_i^*} \tag{5}$$

and $\text{var}(\hat{\theta}) = 1/\sum_{i=1}^{k} w_i^*$, where $w_i^* = 1/(v_i + \hat{\sigma}_B^2)$. To calculate the variance, the assumption that the weights, and therefore σ_B^2, are known is made.

In practice, the point estimates from these two methods usually vary only slightly from each other, but the random effects model leads to wider confidence intervals for the overall treatment effect. However, since σ_B^2 is assumed known in the calculation of $\text{var}(\hat{\theta})$ when in practice it must be estimated from the data, the confidence interval is still too narrow. It does not allow for the imprecision in the estimate of the between-study variance. Furthermore, there is no published method for calculating a confidence interval for σ_B^2. With a limited number of studies in a meta-analysis, σ_B^2 will not be precisely estimated and so the corresponding confidence interval will be wide.

These problems may be solved by using a likelihood approach. A likelihood based confidence interval may be obtained for the between-study variance σ_B^2. Also, a likelihood based confidence interval can be found for the estimate of the overall treatment effect θ which allows for the fact that σ_B^2 is estimated with error. An exposition of these methods is now given.

2. METHODS

2.1. The Likelihood model

The random effects model was set up as in (3) and (4) above with distributional assumptions of normality. The marginal distribution of each estimated treatment effect $\hat{\theta}_i$ is therefore normal with mean θ and variance $(v_i + \sigma_B^2)$.[1] Hence the contribution from study i to the likelihood for θ and σ_B^2 is given by:

$$L_i(\theta, \sigma_B^2) = \frac{1}{\sqrt{\{2\pi(v_i + \sigma_B^2)\}}} \exp\left\{\frac{-(\hat{\theta}_i - \theta)^2}{2(v_i + \sigma_B^2)}\right\}. \tag{6}$$

For a meta-analysis involving k independent studies, the likelihood is given by the product of the individual study likelihoods and so the log-likelihood[6,7] is

$$l(\theta, \sigma_B^2) = -\sum_{i=1}^{k} \frac{1}{2} \ln 2\pi(v_i + \sigma_B^2) - \sum_{i=1}^{k} \frac{(\hat{\theta}_i - \theta)^2}{2(v_i + \sigma_B^2)}. \tag{7}$$

The maximum likelihood estimates (MLEs) of the two parameters of interest, θ and σ_B^2, are derived by solving equations (8) and (9)[6,8] in an iterative manner, beginning by substituting an initial value of $\hat{\sigma}_B^2$ into equation (8). Equations (8) and (9) are obtained, in the usual way, by taking the partial derivatives of equation (7), setting to zero and then rearranging in a form which is convenient for the iteration process:

$$\hat{\theta} = \frac{\sum_{i=1}^{k} \dfrac{\hat{\theta}_i}{(v_i + \hat{\sigma}_B^2)}}{\sum_{i=1}^{k} \dfrac{1}{(v_i + \hat{\sigma}_B^2)}} \tag{8}$$

$$\hat{\sigma}_B^2 = \frac{\sum_{i=1}^{k} \dfrac{(\hat{\theta}_i - \hat{\theta})^2 - v_i}{(v_i + \hat{\sigma}_B^2)^2}}{\sum_{i=1}^{k} \dfrac{1}{(v_i + \hat{\sigma}_B^2)^2}}. \tag{9}$$

Alternatively, the MLEs may be obtained directly from the likelihood in (7) using, for example, S-plus.[9]

2.2. Confidence regions and profile likelihoods

The joint log-likelihood of θ and σ_B^2 in (7) can be calculated and three-dimensional plots obtained. This likelihood can also be displayed in the form of a contour plot where the contours join all points which have the same log-likelihood. The contours on the plots represent likelihood based confidence regions. Hence, a 95 per cent confidence region is given by all paris of θ and σ_B^2 which satisfy,

$$l(\theta, \sigma_B^2) > l(\hat{\theta}, \hat{\sigma}_B^2) - 5{\cdot}99/2 \tag{10}$$

where 5·99 is the 5 per cent point of the χ_2^2 distribution.

The profile log-likelihood can be used to obtain likelihood based confidence intervals for θ and σ_B^2. The profile log-likelihood is the log-likelihood for the parameter of interest which takes into account the fact that the other parameter is also unknown and must be estimated, that is $l_1^*(\theta) = l(\theta, \hat{\sigma}_B^2(\theta))$ and $l_2^*(\sigma_B^2) = l(\hat{\theta}(\sigma_B^2), \sigma_B^2)$, where $\hat{\sigma}_B^2(\theta)$ is the MLE of σ_B^2 as the value of θ varies

Table I. Meta-analysis of nine trials of effects of diuretics on pre-eclampsia[10]

(a) Results and odds ratios for individual trials

Study number	Cases of pre-eclampsia/total number of patients		Odds ratio	95% C.I.
	Treated	Control		
1	14/131 (10·7%)	14/136 (10·3%)	1·04	(0·48, 2·28)
2	21/385 (5·5%)	17/134 (12·7%)	0·40	(0·20, 0·78)
3	14/57 (24·6%)	24/48 (50·0%)	0·33	(0·14, 0·74)
4	6/38 (15·8%)	18/40 (45·0%)	0·23	(0·08, 0·67)
5	12/1011 (1·2%)	35/760 (4·6%)	0·25	(0·13, 0·48)
6	138/1370 (10·1%)	175/1336 (13·1%)	0·74	(0·59, 0·94)
7	15/506 (3·0%)	20/524 (3·8%)	0·77	(0·39, 1·52)
8	6/108 (5·6%)	2/103 (1·9%)	2·97	(0·59, 15·07)
9	65/153 (42·5%)	40/102 (39·2%)	1·14	(0·69, 1·91)

(b) Results from three methods of meta-analysis

Method	$\hat{\sigma}_B^2$ (95% C.I.)	$e^{\hat{\theta}}$ (95% C.I.)
Fixed effect	0·00 (–)	0·67 (0·56, 0·80)
Random effects:		
Moment estimator	0.23 (–)	0·60 (0·40, 0·89)
Likelihood estimator	0·24 (0·03, 1·13)	0·60 (0·37, 0·95)

and similarly $\hat{\theta}(\sigma_B^2)$ is the MLE of θ as σ_B^2 varies. The profile log-likelihoods for θ and σ_B^2 may be plotted and can be used directly to obtain the required confidence intervals. Hence 95 per cent confidence intervals for the two parameters are given by all values which satisfy:

$$l_1^*(\theta) > l_1^*(\hat{\theta}) - 3{\cdot}84/2 \tag{11}$$

and

$$l_2^*(\sigma_B^2) > l_2^*(\hat{\sigma}_B^2) - 3{\cdot}84/2 \tag{12}$$

where 3·84 is the 5 per cent point of the χ_1^2 distribution.

A test of homogeneity may also be derived form these profile likelihoods, since a null hypothesis of homogeneity is equivalent to stating that $\sigma_B^2 = 0$. The relevant likelihood ratio statistic to test this hypothesis is therefore $-2\{l_2^*(0) - l_2^*(\hat{\sigma}_B^2)\}$. Since σ_B cannot be negative, the square root of this likelihood ratio statistic is compared with a one-tailed standard normal distribution.

3. EXAMPLES

3.1. A meta-analysis of nine trials in pre-eclampsia

The first example is a meta-analysis of nine clinical trials investigating the effect of taking diuretics during pregnancy on the risk of pre-eclampsia.[10] The treatment effect is measured in terms of the log odds ratio θ, and so the results are reported here in terms of the odds ratio e^{θ}. The data are summarized in Table I(a).

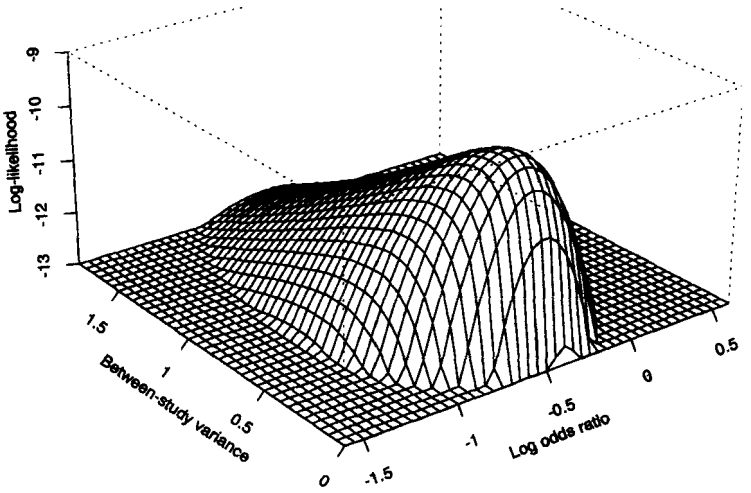

Figure 1. Joint log-likelihood of the overall log odds ratio and the between-study variance for the diuretics trials

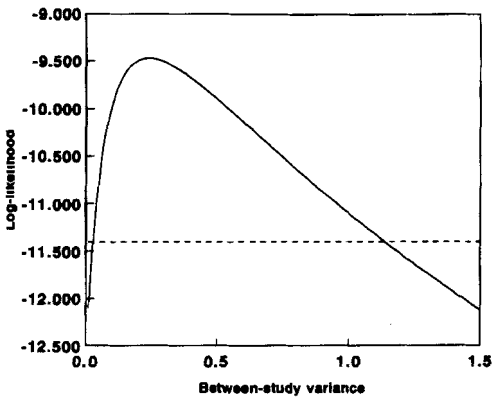

Figure 2. Profile log-likelihood for the between-study variance for the diuretics trials
The maximum likelihood estimate of σ_B^2 is 0·24. The dotted line represents the maximum log-likelihood minus 1·92

It can be seen from the three-dimensional likelihood plot (Figure 1) as well as from the profile log-likelihood (Figure 2) that the distribution of the between-study variance is, as expected, very asymmetric. The profile log-likelihood for the overall treatment effect is much more symmetric in shape (Figure 3) and so will produce a confidence interval which is almost symmetric. However, unlike the standard methods, using the profile log-likelihood does not force the interval to be symmetric.

The estimates and their corresponding confidence intervals for this particular meta-analysis based on the likelihood model can be compared with those obtained from both the fixed effect method and the standard random effects method using the moment estimator of σ_B^2 (Table I(b)). The maximum likelihood estimates agree well with the standard random effects estimates. The fact that a reasonably large estimate of the between-study variance is obtained indicates a lack of homogeneity in this set of studies. The likelihood ratio test for heterogeneity produces a highly significant result ($p = 0{\cdot}006$).

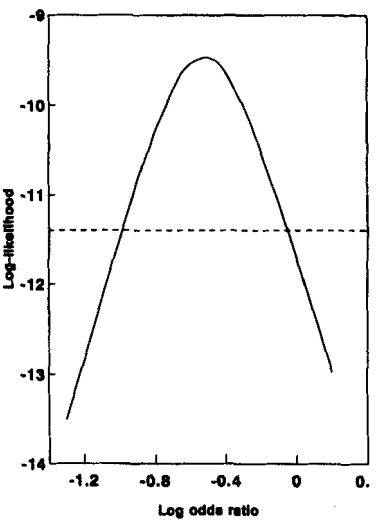

Figure 3. Profile log-likelihood for the overall log odds ratio for the diuretics trials
The maximum likelihood estimate of the overall log odds ratio is $-0{\cdot}51$. The dotted line represents the maximum log-likelihood minus 1·92

The overall estimate of treatment effect and the likelihood based confidence interval will obviously depend on the value of the between-study variance and the precision with which it is estimated. When $\sigma_B^2 = 0$, the estimate of treatment effect is simply that obtained from the fixed effect model. As σ_B^2 increases, the distribution of the weights between the studies becomes increasingly even. As σ_B^2 tends to infinity, the weights tend to equality and the overall estimate therefore tends to the simple unweighted average. A plot of $\hat{\theta}$ against σ_B^2 may be used as a form of sensitivity analysis to investigate the robustness of $\hat{\theta}$ to changes in the value of σ_B^2. Such plots provide a guide to the effect that an imprecise estimate of σ_B^2 may have on $\hat{\theta}$ and hence how much the likelihood based confidence interval will differ from the standard random effects interval. The important characteristic of the plot in this respect is what happens to the estimate of θ in the region around $\hat{\sigma}_B^2$. For example, if the estimate of θ remains almost constant across the values of σ_B^2 contained in its 95 per cent confidence interval, then the likelihood based confidence interval for θ will be little different from that obtained in the standard random effects method. However, the greater the variation in $\hat{\theta}$ over the region of interest, the greater the increase in width of the interval for θ.

In the context of this example, the MLE of σ_B^2 is 0·24. The sensitivity plot of the estimated overall odds ratio e^{θ} is plotted against $x = \sigma_B^2/(0{\cdot}24 + \sigma_B^2)$ in Figure 4, so that $x = 0$ corresponds to $\sigma_B^2 = 0$, $x = 0{\cdot}5$ to $\sigma_B^2 = 0{\cdot}24$ (the MLE) and $x = 1$ to $\sigma_B^2 = \infty$, and the whole range of σ_B^2 is reduced to a finite scale. Figure 4 shows that the overall odds ratio estimate changes by only 0·03 in the region covered by the 95 per cent confidence interval for σ_B^2. Therefore the estimation of the between-study variance does not have much influence on $\hat{\theta}$ in this example, and hence we may expect only a moderate increase in the width of the confidence interval for θ when using the likelihood method as compared with the standard random effects method. This is indeed the case, as shown in Table I(b).

In moving from the invalid fixed effect model to the likelihood model, which allows both for the heterogeneity and the estimation of the between-study variance, the certainty with which conclusions may be drawn from this meta-analysis changes. In particular, the confidence interval

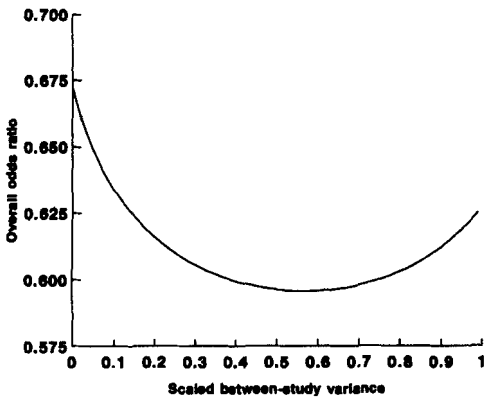

Figure 4. Relationship between the estimated overall odds ratio and the between-study variance for the diuretics trials Scaled between-study variance = $\sigma_B^2/(0{\cdot}24 + \sigma_B^2)$ where 0·24 is the MLE of the between-study variance from the random effects model

given by the likelihood method is somewhat wider than that from the standard random effects analysis, and the treatment benefit of diuretics is less significant, with the upper limit of the confidence interval being only slightly below unity.

3.2. A multicentre trial in mild hypertension

The second example uses data from the Medical Research Council's multicentre trial of the treatment of mild hypertension.[11] Each centre is taken to be equivalent to a 'trial' or 'study' in meta-analysis terminology. This means that there are a large number of 'trials', 189 in total, to be combined in a meta-analysis. The outcome considered here is the reduction in diastolic blood pressure (in mmHg) between entry to the trial and one year after entry. The outcome measure is a difference between two means, that is the difference between the mean reduction in blood pressure in the treatment group and the mean reduction in the control group. Hence, by contrast to the first example, the outcome is a continuous variable, but the methodology carries through in an exactly similar way as for the log odds ratio.

There is considerable evidence of heterogeneity between centres but there are almost no differences between the standard random effects results and the likelihood results (Table II). The confidence interval for the between-study variance is fairly wide, even though $\hat{\sigma}_B^2$ is based on a large number of centres. However, changes in the value of σ_B^2 do not affect the estimate of θ to any great extent (Figure 5), so that the profile likelihood based confidence interval for θ is apparently identical to the confidence interval derived from the standard random effects method. The interval for θ is approximately symmetric and that for σ_B^2 is also more symmetric than in the previous example, resulting from the large number of centres involved.

4. DISCUSSION

We have shown how likelihood methods may be employed in a random effects meta-analysis to provide a confidence interval for the overall treatment effect θ which allows for the fact that the between-study variance σ_B^2 has to be estimated. Such likelihood support intervals may be interpreted for practical purposes as being approximate confidence intervals.[12] The resulting

Table II. Results from three methods of meta-analysis for the mild hypertension trial data

Method	$\hat{\sigma}_B^2$ (95% C.I.)	$\hat{\theta}$ (95% C.I.)
Fixed effect	0·00 (–)	5·31 (5·03, 5·59)
Random effects:		
Moment estimator	1.81 (–)	5·29 (4·94, 5·63)
Likelihood estimator	1·78 (0·83, 3·05)	5·29 (4·94, 5·63)

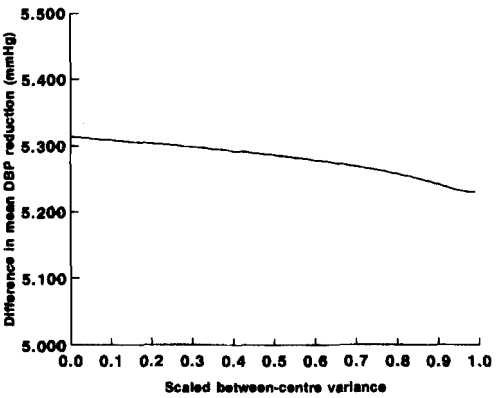

Figure 5. Relationship between the estimated difference in the mean reductions in diastolic blood pressure (DBP) and the between-centre variance for the MRC multicentre trial for the treatment of mild hypertension
Scaled between-centre variance = $\sigma_B^2/(1{\cdot}78 + \sigma_B^2)$ where 1·78 is the MLE of the between-centre variance from the random effects model

interval for θ is wider than that obtained from the standard random effects analysis which takes no account of the uncertainty in the estimate of σ_B^2.

The likelihood method presented yields a confidence interval for σ_B^2, so the precision of $\hat{\sigma}_B^2$ can be directly judged. Obviously, the fewer trials involved, the less precise will be the estimate of σ_B^2. However, even in the second example with 189 'trials', the confidence interval covered a wide range. Hence, in any meta-analysis in practice there will be considerable imprecision in estimating σ_B^2. Whether the value of σ_B^2 used substantially affects the overall estimate of treatment effect is a separate issue, and can easily be investigated using the sensitivity plot of $\hat{\theta}$ against σ_B^2 discussed (for example, Figure 4). In situations where there is little effect of σ_B^2 on $\hat{\theta}$ (as in the second example, Figure 5) the likelihood method yields similar results to the standard method and the latter can be considered adequate.

The question remains as to the circumstances in which the likelihood approach will yield an important widening of the confidence interval for θ compared with the standard random effects method. We have carried out the likelihood method proposed here for some other examples of meta-analyses with heterogeneity.[13,14] The results were quite similar to the first example in this paper, with a moderate increase in the width of the confidence interval but to a degree that would not substantially change the practical conclusions. An extreme and perhaps rather artificial

Table III. Meta-analysis of two trials of aspirin in prevention of non-fatal myocardial infarction

(a) Results and odds ratios for individual trials

Study	Number of non-fatal MIs/total number of patients		Odds ratio	95% C.I.
	Treated	Control		
U.K.[15]	87/3429 (2·5%)	38/1710 (2·2%)	1·15	(0·78, 1·68)
U.S.A[16]	129/11037 (1·2%)	211/11034 (1·9%)	0·61	(0·49, 0·76)

(b) Results from three methods of meta-analysis

Method	$\hat{\sigma}_B^2$ (95% C.I.)	$e^{\hat{\theta}}$ (95% C.I.)
Fixed effect	0·00 (−)	0·71 (0·59, 0·86)
Random effects:		
Moment estimator	0.18 (−)	0·82 (0·44, 1·52)
Likelihood estimator	0·07 (0·00, 1·73)	0·80 (0·39, 1·78)

example was also investigated, that of the only two low risk (primary prevention) trials of aspirin in the prevention of non-fatal myocardial infarction (MI)[15,16] which were part of a published meta-analysis.[17] The data are shown in Table III(a) and the meta-analysis results in Table III(b) in terms of the log odds ratio scale θ. Here the confidence interval for the overall treatment effect from the likelihood method is considerably wider than that from the standard random effects model. This is not only because the 95 per cent confidence interval for σ_B^2, being based on only two trials, is extremely wide, but also because within this interval there is a large range of overall odds ratio estimates (from 0·71 to 0·83). Hence we conclude that it is of practical importance to use the likelihood methods when σ_B^2 is imprecisely estimated and when the value of σ_B^2 affects the overall estimate of the treatment effect in the region around the MLE of σ_B^2. It will be a matter of practical experience as to whether this is of concern in less pathological examples than that of the two aspirin trials above.

Although the likelihood method proposed allows for the estimation of the between-study variance, it still assumes that the individual study variances are known, when in practice they too must be estimated. The full likelihood, in the case of binomial data, includes the conditional distribution of each 2×2 frequency table given its margins.[18] If a full likelihood method were pursued, the confidence intervals for the overall treatment effect would be expected to be even wider. However, it is in the smallest studies that the variances are most imprecisely estimated. Such studies not only take least weight in a meta-analysis but also have their relative weight determined more by the value of σ_B^2 than by v_i. Except when all the trials are small, the additional uncertainty would not therefore be expected to have a great impact on the results and so pursuing a full likelihood approach is unnecessarily sophisticated for most practical purposes. For instance, for the nine diuretics trials in the first example, the full likelihood method gives a 95 per cent confidence interval for the odds ratio of 0·37 to 0·97, rather than 0·37 to 0·95 (Table I(b)).

The proposed method is simple to program in S-plus[9] (details available from the authors), and can be applied to the analysis of continuous, ordinal and survival outcome measures[1] as well as binary outcome measures. An alternative approach to the analysis of continuous outcome

measures is to use a linear mixed model analysis. This can be easily done using PROC MIXED in SAS[19] if individual patient data are available and a common variance is assumed. It is, however, not straightforward if only summary statistics such as means and standard errors are given. For binary outcomes, the full likelihood method[18] requires the individual 2×2 contingency tables to be available, whereas the likelihood method proposed here only requires each trial's treatment effect and variance. This has the advantage that the method can be applied when the raw data are not available, or when the published odds ratios have been adjusted for prognostic factors.

The proposed likelihood model also makes the distributional assumption of normality, which may or may not hold for any particular set of data. This normality assumption may be checked using normal probability plots of the standardized residuals, $\sqrt{w_i^*}(\hat{\theta}_i - \hat{\theta})$. Under the normal distribution random effects model the distribution of these residuals will be approximately standard normal. Hence, normal probability plots which deviate substantially from a straight line of unit gradient through the origin indicate that the model does not fit the data. In such cases, either the likelihood method presented here should be interpreted as approximate, or alternative methods for example using a non-parametric distribution for the random effects should be pursued.[18]

Approximations to the true profile likelihoods may be obtained using quadratic curves derived from the asymptotic variance–covariance matrix for (θ, σ_B^2). Tests of the treatment effect may also be derived as shown for the case of continuous outcome measures by Rosner.[6] Using such a quadratic approximation forces the profile likelihood of both θ and σ_B^2 to be symmetric in shape. Obviously this approximation is poor for σ_B^2 and hence to obtain a confidence interval for σ_B^2 it would be preferable to work in terms of $\log(\sigma_B^2)$. There is, however, no advantage to these approximations, as the calculations involved are no simpler than those involved in working with the true likelihood. To take some account of the fact that σ_B^2 is estimated from a finite number of trials, Rosner[6] suggested that the test statistic may be better approximated by a t-distribution than by the N(0, 1) distribution. However, as we have shown, the widening of the confidence interval depends more on the strength of the relationship between σ_B^2 and $\hat{\theta}$ than it does on simply the number of trials involved and the precision of $\hat{\sigma}_B^2$.

Likelihood based methods for meta-analysis have been advocated before.[5,7,18] Although the normally distributed random effects model can of course be criticized for being unrealistic,[2] the widening in the confidence interval that it introduces compared with the fixed effect approach is intuitively sensible in terms of caution in the extrapolation of results to future trials or future patients. The assumptions underlying a random effects model may be more acceptable in analysing a multicentre trial as in the second example, where variations between centres can be more realistically considered as random. Although allowing explicitly for between-centre variability in treatment effects in a multicentre trial by using a random effects model is not what is usually done in practice, there are indeed arguments in favour of such an approach.[20,21]

The examples in this paper show that caution is required when interpreting results from the standard meta-analysis methods. Certainly, the confidence intervals from a fixed effect model will be too small when heterogeneity exists. However, even the confidence interval for the overall treatment effect from the usual random effects model may be too narrow when the overall estimated treatment effect varies substantially according to the value of the between-study variance. The use of a sensitivity plot, such as Figures 4 and 5, will therefore provide insight into whether the likelihood method is required or whether the simpler standard random effects analysis using a moment estimator of the between-study variance is adequate. This issue is of practical importance since the increased width of the confidence intervals can severely limit the conclusions that can be drawn from a meta-analysis.

ACKNOWLEDGEMENTS

We thank Michael Hills, Richard Hayes, Stephen Pyke and two referees for helpful comments on this paper, and Hans Van Houwelingen for providing a computer program to carry out the full likelihood analysis given in his paper.[18] We are also indebted to the MRC Working Party on Mild to Moderate Hypertension for allowing us to use data from the MRC trial, and thank Professor T. W. Meade and P. J. Brennan for their assistance in this.

REFERENCES

1. Whitehead, A. and Whitehead, J. 'A general parametric approach to the meta-analysis of randomized clinical trials', *Statistics in Medicine*, **10**, 1665–1677 (1991).
2. Thompson, S. G. 'Controversies in meta-analysis: The case of trials of serum cholesterol reduction', *Statistical Methods in Medical Research*, **2**, 173–192 (1993).
3. Woolf, B. 'On estimating the relation between blood group and disease', *Annals of Human Genetics*, **19**, 251–253 (1955).
4. Thompson, S. G. and Pocock, S. J. 'Can meta-analysis be trusted?', *Lancet*, **338**, 1127–1130 (1991).
5. DerSimonian, R. and Laird, N. M. 'Meta-analysis in clinical trials', *Controlled Clinical Trials*, **7**, 177–188 (1986).
6. Rosner, B. 'A generalization of the paired t-test', *Applied Statistics*, **31**, 9–13 (1982).
7. Goodman, S. N. 'Meta-analysis and evidence', *Controlled Clinical Trials*, **10**, 188–204 (1989).
8. Maritz, J. S. and Lwin, T. *Empirical Bayes Methods*, Chapman and Hall, London, New York, 1989.
9. Becker, R. A., Chambers, J. M. and Wilks, A. R. *The New S Language. A Programming Environment for Data Analysis and Graphics*, Wadsworth & Brooks/Cole, Pacific Grove, California, 1988.
10. Collins, R., Yusuf, S. and Peto, R. 'Overview of randomised trials of diuretics in pregnancy', *British Medical Journal*, **290**, 17–23 (1985).
11. Medical Research Council Working Party. 'MRC trial of treatment of mild hypertension: Principal results', *British Medical Journal*, **291**, 97–104 (1985).
12. Clayton, D. and Hills, M. *Statistical Models in Epidemiology*, Oxford University Press, Oxford, 1994.
13. Wood, D. A., Kinmonth, A. L., Pyke, S. D. M. and Thompson, S. G. 'Randomized controlled trial evaluating cardiovascular screening and intervention in general practice: Principal results of the British Family Heart Study', *British Medical Journal*, **308**, 313–320 (1994).
14. Thompson, S. G. 'Why sources of heterogeneity in meta-analysis should be investigated', *British Medical Journal*, **309**, 1351–1355 (1994).
15. Antiplatelets Trialists' Collaboration. 'Secondary prevention of vascular disease by prolonged antiplatelet treatment', *British Medical Journal*, **296**, 320–331 (1988).
16. Steering Committee of the Physicians' Health Study Research Group. 'Final report of the aspirin component of the ongoing Physicians' Health Study', *New England Journal of Medicine*, **321**, 129–135 (1989).
17. Antiplatelet Trialists Collaboration. 'Collaborative overview of randomised trials of antiplatelet therapy — I: Prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients', *British Medical Journal*, **308**, 81–106 (1994).
18. Van Houwelingen, H. C., Zwinderman, K. H. and Stijnen, T. 'A bivariate approach to meta-analysis', *Statistics in Medicine*, **12**, 2273–2284 (1993).
19. SAS Institute Inc. *SAS Technical Report P-229, Release 6.07, The MIXED Procedure*, SAS Institute Inc., Cary, NC, pp. 287–366 (1992).
20. Chakravarthi, S. R. and Grizzle, J. E. 'Analysis of data from multiclinic experiments', *Biometrics*, **31**, 325–338 (1975).
21. Raghunathan, T. E. and Yoichi, II. 'Analysis of binary data from a multicentre clinical trial', *Biometrika*, **80**, 127–139 (1993).