STATISTICS IN MEDICINE, VOL. 12, 2273–2284 (1993)

A BIVARIATE APPROACH TO META-ANALYSIS

HANS C. VAN HOUWELINGEN AND KOOS H. ZWINDERMAN

Department of Medical Statistics, Faculty of Medicine, University of Leiden, P.O. Box 9604, 2300 RC Leiden, The Netherlands

AND

THEO STIJNEN

Department of Biostatistics and Epidemiology, Erasmus University, Rotterdam, The Netherlands

SUMMARY

The usual meta-analysis of a sequence of randomized clinical trials only considers the difference between two treatments and produces a point estimate and a confidence interval for a parameter that measures this difference. The usual parameter is the log(odds ratio) linked to Mantel–Haenszel methodology. Inference is made either under the assumption of homogeneity or in a random effects model that takes account of heterogeneity between trials. This paper has two goals. The first is to present a likelihood based method for the estimation of the parameters in the random effects model, which avoids the use of approximating Normal distributions. The second goal is to extend this method to a bivariate random effects model, in which the effects in both groups are supposed random. In this way inference can be made about the relationship between improvement and baseline effect. The method is demonstrated by a meta-analysis dataset of Collins and Langman.

1. INTRODUCTION

We consider a meta-analysis performed on the data from a number of related clinical trials in which treatment A is compared with treatment B. We restrict attention to clinical trials with a dichotomous outcome variable, but the method presented here can easily be generalized to different types of outcome. In each trial, we let n_A and n_B be the number of patients in the two treatment groups and X_A and X_B the number of patients for which a certain event is observed. The event may be beneficial (for example, remission of cancer), or adverse (recurrence, the occurrence of metastases). So the typical data we consider consist of a sequence of quadruples (n_A, n_B, X_A, X_B) for each trial, that can be represented by a 2×2 contingency table. We let π_A and π_B be the respective probabilities of the event of interest, and θ_A and θ_B the corresponding logits $(\theta = \log(\pi/(1-\pi)))$. The conventional meta-analysis uses the Mantel–Haenszel approach and produces a point estimate, a P-value and a confidence interval for the common odds ratio $(\pi_B/(1-\pi_B))/(\pi_A/(1-\pi_A))$ or its logarithm $\omega = \theta_B - \theta_A$. See, for example, Yusuf *et al.*[1]

This approach has been questioned by several authors and an alternative random effects model, based on a Normal approximation for the distribution of the estimated log(odds ratio) was introduced by Der Simonian and Laird.[2] A similar setup was used by Stijnen and Van Houwelingen[3] with focus on empirical Bayes estimation of the log(odds ratio) in each trial. In this paper we do not use Normal approximations, but base our analysis on the conditional non-central hypergeometric distribution. The parameters of the random effects model are estimated

0277–6715/93/242273–12$11.00

by the EM algorithm (Dempster *et al.*[4]), avoiding problems with the weights in the Der Simonian and Laird approach discussed by Mosteller and Chalmers.[5] Weighting of the trials is implicit in the maximum likelihood estimation procedure.

The Mantel–Haenszel procedure and its random effects extension focus on the estimation of the odds ratio or its logarithm ω. Information on the joint distribution of θ_A and θ_B is not used and there is a tacit assumption that ω is in some sense independent of θ_A (and θ_B). In this paper we show how insight into the relations between θ_A, θ_B (and ω) can be obtained by introducing a bivariate random effects model for (θ_A, θ_B). The bivariate approach is also feasible for parameterizations leading to either relative risk or risk difference as an effect measure. Conventional use of the log(odds ratio) as an effect measure is mainly for convenience.

Our approach is demonstrated by a meta-analysis of data from Collins and Langman[6] presented in Section 2. In Section 3 we review the likelihood version of the Mantel–Haenszel procedure and in Section 4 we discuss the random effects extension without the use of Normal approximations. Section 5 presents our bivariate approach and discusses the results, while Section 6 makes some concluding general remarks.

2. THE COLLINS AND LANGMAN DATASET

Collins and Langman[6] present an overview of 25 trials on the treatment of upper gastrointestinal bleeding by a histamine H_2 antagonist. In each trial treatment (T) is compared with a placebo control (C) and the numbers of patients with persistent or recurrent bleedings are recorded. The dataset is given in Table I.

The trials are ordered according to increasing standard error of the log(odds ratio). In Figures 1(a) and (b) a graphical presentation of the data is given. To avoid degeneracy when $X = 0$ or $X = n$, θ is estimated by $\log(X + 0{\cdot}5)/(n - X + 0{\cdot}5)$. The odds ratios in Table I are estimated similarly.

As effect measure we take the log(odds ratio); $\omega = \theta_C - \theta_T$. By definition $\omega > 0$ implies a positive treatment effect (reduction of the number of patients with recurrent or persistent bleedings). Collins and Langman[6] analysed the data using the method of Yusuf *et al.*[1] Where relevant, we compare these results with ours.

3. LIKELIHOOD BASED MANTEL–HAENSZEL-TYPE PROCEDURE

Using the conditional distribution of X_C given $X_T + X_C$ (this is the conditional non-central hypergeometric distribution of the 2×2 table given its marginals) leads to the likelihood (of each trial):

$$L(\omega) = \frac{\binom{n_C}{X_C}\binom{n_T}{X_T}e^{\omega X_C}}{\sum_y \binom{n_C}{y}\binom{n_T}{X_T + X_C - y}e^{\omega y}}.$$

Graphs of the scaled likelihoods $L(\omega)/\int_{-4}^{4} L(\omega)d\omega$ are present in Figure 2. From this figure we conclude that ω tends to be positive, because there is more of the likelihoods to the right half of the plot.

Under the assumption that all trials share a common odds ratio ω, the total log(likelihood) is given by

$$l(\omega) = \sum_{i=1}^{25} l_i(\omega),$$

Table I. Overview of 25 trials of a histamine H_2 antagonist (T) compared with placebo (C) in the treatment of upper gastrointestinal bleeding

Trial	Reference in Collins and Langman	Year of publication	n_T	X_T	n_C	X_C	Odds ratio	Log(odds ratio)
1	21	1983	259	50	260	51	1·02	0·02
2	15	1984	153	44	132	30	0·73	− 0·31
3	24	1984	106	16	107	15	0·92	− 0·08
4	19	1982	78	14	80	21	1·61	0·47
5	23	1984	51	16	54	15	0·84	− 0·17
6	4	1979	56	11	53	12	1·19	0·18
7	3	1979	50	8	50	16	2·39	0·87
8	9	1980	40	9	48	11	1·02	0·02
9	7	1979	33	12	36	10	0·68	− 0·38
10	11	1980	46	5	47	12	2·66	0·98
11	22	1984	31	6	29	12	2·80	1·03
12	25	1984	33	6	39	7	0·98	− 0·02
13	26	1981	36	11	26	5	0·57	− 0·57
14	14	1980	21	10	19	8	0·81	− 0·21
15	10	1980	20	7	20	8	1·22	0·20
16	13	1980	45	9	43	3	0·33	− 1·10
17	27	1982	34	3	31	13	6·57	1·88
18	18	1982	24	4	24	5	1·28	0·25
19	6	1979	14	2	15	6	3·42	1·23
20	28	1984	15	2	14	3	1·64	0·50
21	5	1979	18	5	12	1	0·32	− 1·14
22	17	1981	10	1	9	4	5·18	1·65
23	Brown*	?	10	3	11	1	0·31	− 1·18
24	12	1980	18	0	19	3	7·85	2·06
25	16	1981	14	0	14	0	1·00	0·00
Total			1215	254	1192	272		

* In the article of Collins and Langman no reference is given for this paper

where $l_i(\omega) = \log L_i(\omega)$ is the log-likelihood for the ith trial. Its graph is shown in Figure 3 (solid line). From this graph we conclude that the MLE $\hat{\omega} = 0{\cdot}12$, that $l(\hat{\omega}) = -53{\cdot}68$ and that the 95 per cent confidence interval under the assumption of homogeneity is $-0{\cdot}07 < \omega < 0{\cdot}31$ in perfect agreement with Collins and Langman.[6]

4. RANDOM EFFECTS EXTENSION OF THE MANTEL–HAENSZEL PROCEDURE

Relaxing the assumption of homogeneity, we consider ω random with distribution G, so that each trial has its own ω_i and $(\omega_1, \ldots, \omega_{25})$ is a random sample from G. The values of $\omega_1, \ldots, \omega_{25}$ are unobservable, like the ω-parameter of the previous section. We make the important assumption that ω is independent of the sample sizes (n_T, n_C). The likelihood for each trial is given by

$$L_i(G) = \int L_i(\omega)\, \mathrm{d}G(\omega)$$

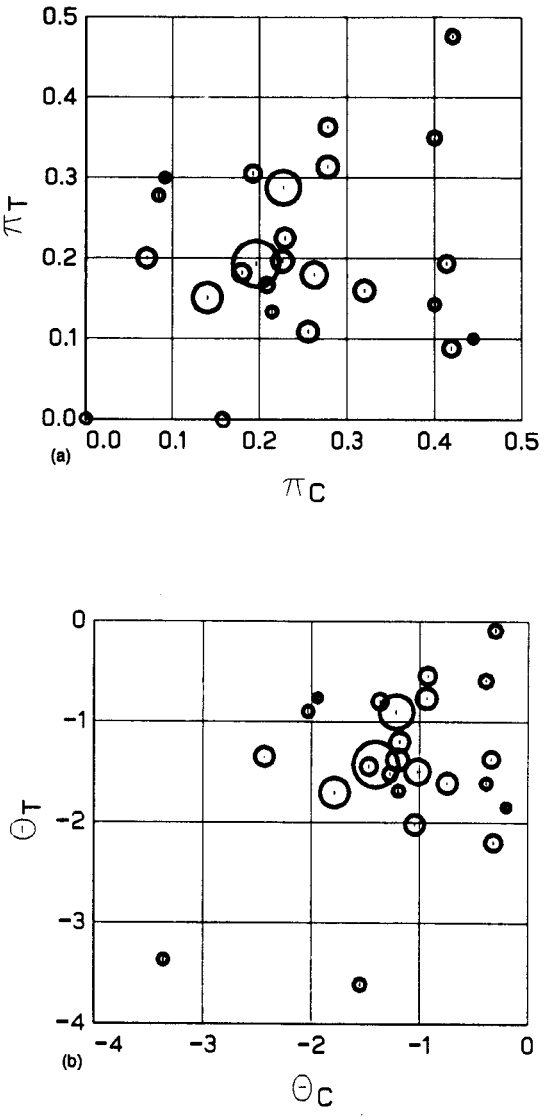

Figure 1. Graphical display of the Collins and Langman data set (Table I)
Figure 1(*a*) displays the estimated probabilities and in (*b*) the estimated log(odds). The area of each circle is proportional to the total sample size of the trial, so larger circles correspond to trials with larger weights

and G can be estimated by maximizing

$$l(G) = \sum_{1}^{25} l_i(G).$$

The *non-parametric* approach of Laird[7] estimates G by a discrete distribution via the EM-algorithm (Dempster *et al.*[4]). The reader is referred to their papers for details of the estimation procedure. For our dataset we obtain a two-point distribution for $\hat{G}$ with probability masses at $\omega = -0{\cdot}03$ and $\omega = 1{\cdot}10$ with respective probabilities 0·79 and 0·21. The corresponding

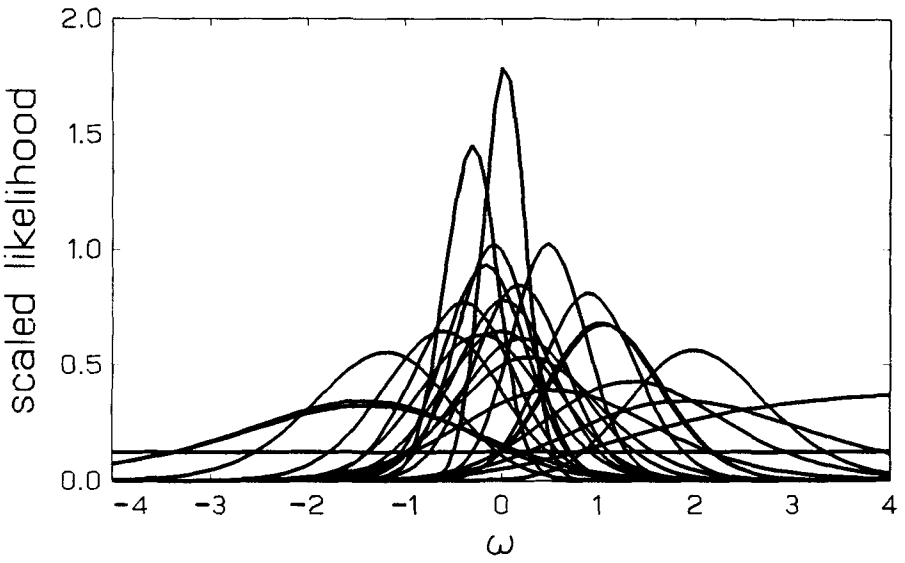

Figure 2. Scaled likelihoods for the individual trials
The sharply peaked curves correspond to the trials with large total sample size. The deviant trial on the right is no. 24 for which $L(\omega)$ is maximal at $\omega = \infty$

log(likelihood) is $l(\hat{G}) = -51{\cdot}76$. Comparing this with $l(\hat{\omega}) = -53{\cdot}68$, we see that the improvement of the likelihood is not substantial. On the other hand $\hat{G}$ gives some idea about the variability between trials. The moments of $\hat{G}$ are given by $\hat{\mu}_\omega = 0{\cdot}21$ and $\hat{\sigma}_\omega = 0{\cdot}47$.

A much smoother estimate of G is obtained from a parametric model. We consider the usual Normal model, that is $G = \mathrm{N}(\mu, \sigma^2)$, which leads to

$$L_i(\mu, \sigma) = \int L_i(\omega) \frac{1}{\sigma} \varphi\left(\frac{\omega - \mu}{\sigma}\right) \mathrm{d}\omega,$$

with $\varphi(x) = 1/\sqrt{2\pi}\mathrm{e}^{-1/2x^2}$, the standard Normal density. The parameters can be estimated by the EM algorithm. Details are given in the next section for the more general k-variate case. The results are

$$\hat{\mu} = 0{\cdot}17 \qquad \hat{\sigma} = 0{\cdot}35 \qquad l(\hat{\mu}, \hat{\sigma}) = -52{\cdot}99.$$

The method of Der Simonian and Laird[2] (omitting trial 25) gives $\hat{\mu} = 0{\cdot}18$ and $\hat{\sigma} = 0{\cdot}40$. The parametric model fits slightly worse than the non-parametric one, but it also has one degree of freedom less.

The difference $2(l(\hat{\mu}, \hat{\sigma}) - l(\hat{\omega})) = 2(-52{\cdot}99 + 53{\cdot}68) = 1{\cdot}38$ could be used for a formal likelihood ratio test of $\sigma = 0$. The corresponding P-value is 0·12, taking into account the one-sided nature of the test ($\sigma = 0$ versus $\sigma > 0$). For testing heterogeneity Collins and Langman[6] report a conventional $\chi^2_{[24]} = 38{\cdot}1$ with $P = 0{\cdot}02$. They concluded that there is some heterogeneity but ignored it in the computation of the confidence interval. In our view a formal test of heterogeneity is not relevant because the value $\sigma = 0$ is not very plausible.

A 95 per cent confidence interval for μ can be obtained via the profile log(likelihood):

$$pl(\mu) = \sup_{\sigma} l(\mu, \sigma) = l(\mu, \hat{\sigma}(\mu)).$$

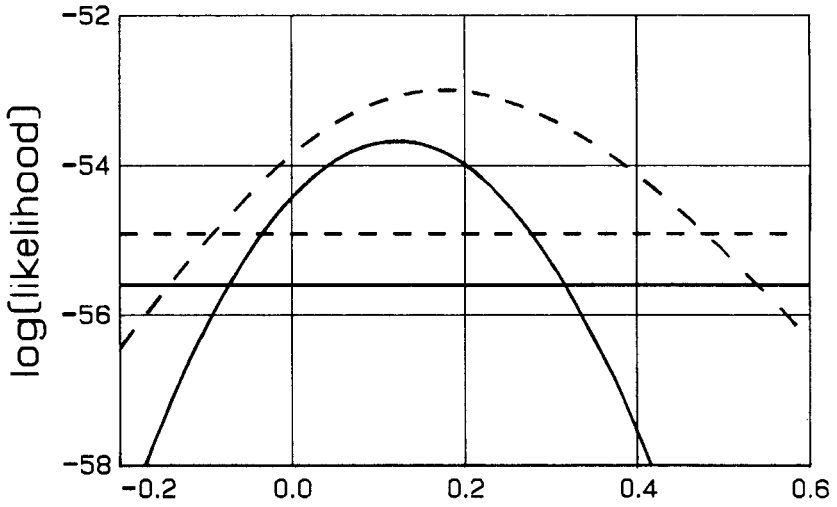

Figure 3. Graph of the log(likelihood) under the assumption of homogeneity (solid line) and the profile log(likelihood) under the random effects model (dashed line).
The horizontal lines are 1·92 = 1/2 × 3·84 below the dome of the curves. The intersection of each horizontal line with the corresponding curve gives the endpoints of the likelihood-based 95 per cent confidence interval

Again $\hat{\sigma}(\mu)$ can be estimated by the EM-algorithm and details are given in the next section. From the graph of $pl(\mu)$ in Figure 3 (dashed line) we see that the 95 per cent confidence interval for μ in this model is $-0{\cdot}09 < \mu < 0{\cdot}48$. The method of Der Simonian and Laird[2] gives a very similar confidence interval (-010, 0·45), which is wider than the one of Section 3. This is in line with the general objection that a fixed effect approach like the Mantel–Haenszel methodology produces confidence intervals that are too small because it ignores heterogeneity between trials. Figure 3 is typical for the comparison of homogeneous log(likelihood) and heterogeneous profile log(likelihood). The former is lower and more peaked than the latter.

We stress that the two lines in Figure 3 have a very different meaning. Under the homogeneity assumption, there is only a single parameter ω, for which a confidence interval can be obtained from the solid line in Figure 3. In the random effects model of this section, ω is a random variable with distribution G. The mean value μ of that distribution is certainly not the only parameter of interest; another is $Pr(\omega > 0) = \Phi(\mu/\sigma)$, estimated by $\Phi(0{\cdot}17/0{\cdot}35) = 0{\cdot}69$. It is also feasible to obtain a confidence interval for this parameter using the same methodology.

5. BIVARIATE RANDOM EFFECTS MODEL

5.1. Exact procedure

The full likelihood in each trial of the bivariate parameter $\theta = (\theta_T, \theta_C)'$ is given by

$$L(\theta) = \binom{n_T}{X_T}\binom{n_C}{X_C}\frac{e^{X_T\theta_T + X_C\theta_C}}{(1 + e^{\theta_T})^{n_T}(1 + e^{\theta_C})^{n_C}}.$$

It does not make sense to assume that θ is constant over trials, so we proceed with a random effects model in which θ has a bivariate distribution G. The non-parametric approach can be easily extended to the bivariate case, but we do not give details here, mainly because we think that

Table II. Bivariate random effect model with 4 mass points

θ_T	θ_C	$\omega = \theta_C - \theta_T$	Probability
− 1·52	− 1·49	0·03	0·43
− 0·90	− 1·10	− 0·20	0·26
− 1·72	− 0·79	0·93	0·26
− 0·30	− 0·43	− 0·12	0·05

$l(\hat{G}) = -1255{\cdot}15$

the discrete non-parametric estimate is not very realistic. The MLE $\hat{G}$ is not very stable. In Table II we give a stable solution with 4 mass points. Extension to more mass points gave no significant improvement of the log-likelihood.

The distribution is plotted in Figure 4 (triangles). The estimated distribution of ω is very much like the 2-point distribution of Section 4. The moments of $\hat{G}$ are

$$\hat{\mu} = (-1{\cdot}36, -1{\cdot}16) \qquad \text{yielding } \hat{\mu}_\omega = 0{\cdot}19$$

$$\hat{\Sigma}_\theta = \begin{pmatrix} 0{\cdot}150 & 0{\cdot}031 \\ 0{\cdot}031 & 0{\cdot}109 \end{pmatrix} \qquad \text{yielding } \hat{\sigma}_\omega = 0{\cdot}44.$$

Because of the instability of the non-parametric MLE and its very discrete nature, we prefer a parametric model for G. The obvious choice is $G = N(\mu, \Sigma)$. The likelihood per trial is given by

$$L_i(\mu, \Sigma) = \int L_i(\theta) f(\theta \mid \mu, \Sigma)\, d\theta$$

where $f(\theta \mid \mu, \Sigma)$ is the bivariate Normal density with parameters μ and Σ. The maximum likelihood estimates of μ and Σ are obtained by the EM algorithm. Consider the unobservable 2-vector $\theta_1, \ldots, \theta_{25}$ as missing observations. If $\theta_1, \ldots, \theta_{25}$ could be observed, the sufficient statistics are $\Sigma\theta_i$ and $\Sigma\theta_i\theta_i'$. Since

$$E(\theta_i) \mid \text{data, parameters}) = \tilde{\theta}_i = \frac{\int \theta L_i(\theta) f(\theta \mid \mu, \Sigma) d\theta}{\int L_i(\theta) f(\theta \mid \mu, \Sigma) d\theta}$$

and

$$E(\theta_i\theta_i' \mid \text{data, parameters}) = \tilde{\tau}_i = \frac{\int \theta\theta' L_i(\theta) f(\theta \mid \mu, \Sigma) d\theta}{\int L_i(\theta) f(\theta \mid \mu, \Sigma) d\theta},$$

the E-step consists of computing $\tilde{\theta}_i$ and $\tilde{\tau}_i$ and the M-step consists of

$$\hat{\mu} = \frac{1}{25} \Sigma \tilde{\theta}_i$$

$$\hat{\Sigma} = \frac{1}{25} \Sigma \tilde{\tau}_i - \hat{\mu}\hat{\mu}'.$$

Starting with an arbitrary μ and non-degenerate Σ, alternate iteration of E-step and M-step leads to the MLE of μ and Σ. Convergence is slow but sure. The algorithm is easy to program. The tedious part is the computation of all $\tilde{\theta}_i$ and $\tilde{\tau}_i$. In the application of the EM-algorithm in the previous section we used numerical integration to evaluate the integrals; this could be carried out

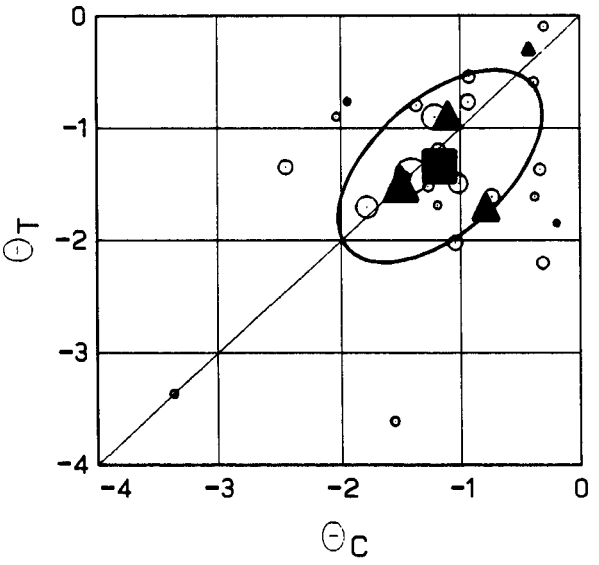

Figure 4. Datapoints and estimated mixing distributions
The datapoints are the same as in Figure 1(*b*). The triangles represent the non-parametric estimate of the mixing distribution. The size of the triangle is proportional to the probability of the point. The square represents the mean μ of the normal mixing distribution. The ellipse is the 95 per cent content ellipse $(\theta - \mu)'\Sigma^{-1}(\theta - \mu) = 6$

quite easily because the ω parameter is one-dimensional. In this section, where the parameter is two-dimensional, numerical integration is still feasible but very time consuming. This 'exact' method can be replaced by various approximations.

5.2. Approximate procedures

A first approximation is based on the MLE of $\hat{\theta}_i$ in each trial and its asymptotic covariance matrix

$$V_i = -l_i''(\theta)^{-1} \text{ evaluated at } \hat{\theta}_i.$$

Here $l_i(\theta) = \log(L_i(\theta))$ and $l_i''(\theta)$ is the matrix of second derivatives. Replacing $l_i(\theta)$ by $l_i(\theta) = l_i(\hat{\theta}_i) - (\theta - \hat{\theta}_i)'V_i^{-1}(\theta - \hat{\theta}_i)$ leads to

$$\tilde{\theta}_i = (V_i^{-1} + \Sigma^{-1})^{-1}(V_i^{-1}\hat{\theta}_i + \Sigma^{-1}\mu),$$

a weighted average of $\hat{\theta}_i$ and μ, and

$$\tilde{\tau}_i = \tilde{\theta}_i\tilde{\theta}_i' + (V_i^{-1} + \Sigma^{-1})^{-1}. \tag{1}$$

Iteration leads to the MLE of μ and Σ in the model where the conditional distribution of $\hat{\theta}_i$ is assumed to be $N(\theta_i, V_i)$. This is the kind of model used by Der Simonian and Laird,[2] and the asymptotic covariance matrix of $\hat{\mu}$ is given by

$$\text{covariance matrix of } \hat{\mu} = \left(\sum_i (V_i + \hat{\Sigma})^{-1}\right)^{-1}. \tag{2}$$

Table III. Bivariate random effects model: comparison of exact and approximate methods

Parameter	'Exact'	Approximate
$\hat{\mu}$	$(-1{\cdot}35, -1{\cdot}19)$	$(-1{\cdot}34, -1.17)$
$\hat{\mu}_\omega$	0·17	0·17
$\hat{\Sigma}$	$\begin{pmatrix} 0{\cdot}126 & 0{\cdot}067 \\ 0{\cdot}067 & 0{\cdot}129 \end{pmatrix}$	$\begin{pmatrix} 0{\cdot}122 & 0{\cdot}067 \\ 0{\cdot}067 & 0{\cdot}122 \end{pmatrix}$
$\hat{\sigma}_\omega$	0·35	0·33
$l(\hat{\mu}, \hat{\Sigma})$	−1257·53	−1257·62

The whole method is based on the assumption that $l_i''(\theta)$ is constant over the range of interest, which is generally not true. The method can only be applied if the sample sizes per trial are fairly large.

The method we advocate approximates $\tilde{\theta}_i$ by the value of θ that maximizes $L_i(\theta) f(\theta \mid \mu, \Sigma)$, that is by the solution of

$$l_i'(\theta) = \Sigma^{-1}(\theta - \mu)$$

($l_i'(\theta)$ is the vector of first derivatives).

The solution of this equation can be obtained by the Newton–Raphson method. The whole process can be accelerated by using the value of $\tilde{\theta}_i$ from the previous E-step as starting point and taking only the first Newton–Raphson step. The approximation of $\tilde{\tau}_i$ is obtained by Taylor expansion of the logarithm of the integrand about $\tilde{\theta}_i$. This leads to a value of $\tilde{\tau}_i$ that is given by (1) with V_i replaced by

$$\tilde{V}_i = -l_i''(\tilde{\theta}_i)^{-1}.$$

Similarly an approximate covariance matrix for $\hat{\mu}$ is given by (2) with V_i again replaced by $\tilde{V}_i$.

Approximations for the total log(likelihood) are obtained in a similar way. Convergence of the EM-algorithm is reached if the total log(likelihood) does not increase any more.

The approximate method can be easily extended to k-variate random components with $k > 2$. The 'exact' method gets very computer time intensive when $k > 2$.

5.3. Results from bivariate random effects model

These are presented in Table III. Both methods agree very well, but the approximate method is much faster.

The estimated mean and the 95 per cent content ellipse of the bivariate distribution are also plotted in Figure 4. The main axis of the ellipse happens to be parallel with the line $\theta_T = \theta_C$. So $\omega = \theta_C - \theta_T$ is uncorrelated with $(\theta_T + \theta_C)/2$. Both ω and $(\theta_T + \theta_C)/2$ show a substantial random fluctuation. The estimated probability that a trial with infinite sample size would result in a positive treatment effect is $Pr(\omega > 0) = \Phi(0{\cdot}17/0{\cdot}33) = 0{\cdot}70$.

The regression line of ω on θ_C has slope $(0{\cdot}122 - 0{\cdot}67)/0{\cdot}122 = 0{\cdot}48$. In our view this is not an indication that the log(odd ratio) model for the treatment is incorrect, but merely a manifestation of regression to the mean. The main axis of the ellipse gives the best description of the relation between θ_T and θ_C as in orthogonal regression. The ellipse shows the true variation among trials. The larger scatter of the data points is mainly due to pure random fluctuation caused by small sample sizes.

A tilted main axis of the ellipse is an indication that ω is not independent of $(\theta_T + \theta_C)/2$. This occurs if $\text{var}(\theta_T) \neq \text{var}(\theta_C)$. An additive risk model might be more appropriate if $\text{var}(\pi_T) = \text{var}(\pi_C)$ and a multiplicative risk model if $\text{var}(\ln(\pi_T)) = \text{var}(\ln(\pi_C))$. Approximate expressions for these variances can be obtained by the δ-method, $\text{var}(g(\theta)) = (g'(\mu))^2\text{var}(\theta)$. A more refined method is to assume a bivariate Normal model for either (π_T, π_C) or $(\log(\pi_T), \log(\pi_C))$. The likelihoods of the different models can be compared to select the best fitting model. However, technical problems can occur due to the boundary conditions $0 \leq \pi \leq 1$ (additive model) or $\log(\pi) \leq 0$ (multiplicative model).

The vector $\tilde{\theta}_i$ obtained in the final stage of the EM-process are just the posterior means of θ_i given the data. They are the empirical Bayes estimates of the θ_i and can be used to obtain adjusted estimates of the effects in each trial, adjusted for random measurement error. However, empirical Bayes methodology is beyond the scope of this paper and we will not go into graphical presentation of the $\tilde{\theta}_i$.

5.4. Confidence regions for means

Finally, we discuss the construction of a confidence region for μ. The approximate covariance matrix of $\hat{\mu}$, based on (2) with V_i replaced by $\tilde{V}_i$ equals

$$\begin{pmatrix} 0{\cdot}0118 & 0{\cdot}0032 \\ 0{\cdot}0032 & 0{\cdot}0114 \end{pmatrix}$$

yielding standard error of $\hat{\mu}_\omega$ of 0·13 which is in good agreement with the results of the previous section, where the method of Der Simonian and Laird gave a standard error of 0·14. The approximate covariance matrix can be used to compute a 95 per cent confidence region for the vector μ. Since the second derivative of the total log-likelihood is not constant, we prefer to construct a 95 per cent confidence region based on the likelihood ratio test. The EM-algorithm can easily be adapted to estimate Σ if μ is given. The value of μ does not change during the process and the estimate of Σ is now given by $\hat{\Sigma} + (\mu - \hat{\mu})(\mu - \hat{\mu})'$ with $\hat{\Sigma}$ and $\hat{\mu}$ as defined before. This can be used to compute the profile log-likelihood

$$pl(\mu) = \sup_{\Sigma} l(\mu, \Sigma)$$

Contour plots of the deviance, $2(pl(\hat{\mu}) - pl(\mu))$, are given in Figure 5, from which it can be seen that the 95 per cent confidence interval for ω coincides with the one of Section 4.

6. CONCLUDING REMARKS

First of all it is surprising how easily a multivariate random components model can be fitted by means of the EM-algorithm and some approximation of the integral of smooth functions with respect to the multivariate Normal density. We have work in progress in which the same methodology is applied to multicentre multitreatment studies where the treatment effect θ can vary randomly from centre to centre. By this approach we can quantify how likely it is that the treatments effects are in a specified order (analogous to $\text{Prob}(\omega > 0)$ in this paper).

Secondly, we think that the bivariate approach presented here has the great advantage of describing the relation between the effects of both treatments. It gives the information needed to make inferences about the nature of the relationship and to choose between additive risk, additive logits or multiplicative relative risks models. Moreover, it gives the opportunity to compute other parameters of interest. For instance, in the model of Section 5, the mean risk difference can be

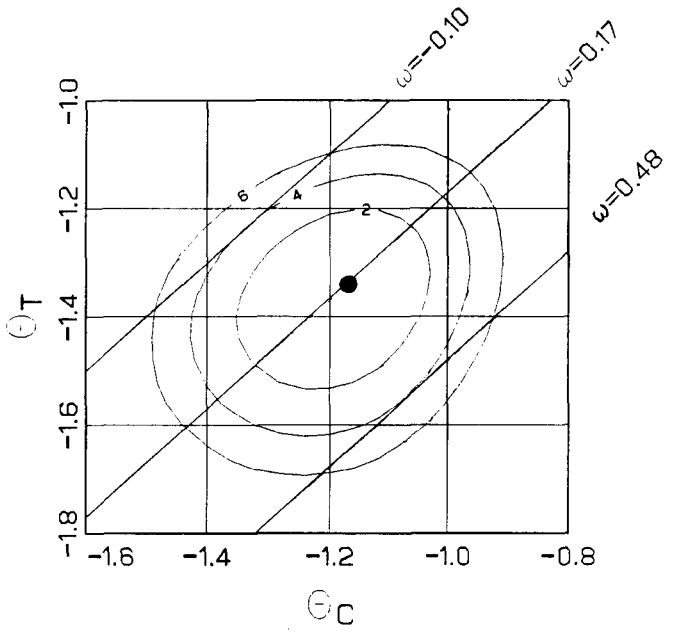

Figure 5. Deviance plot of the profile likelihood
Contours are plotted of $2(pl(\hat{\mu}) - pl(\mu))$. Deviance = 6 gives the bivariate 95 per cent confidence region. Deviance = 4 ($\simeq$ 3·84) gives the 95 per cent confidence interval for any linear combination, such as μ_{θ_1}, μ_{θ_2} and μ_ω. The tangents with slope = 1 to the deviance = 4 contour are plotted, from which the 95 per cent confidence interval for ω can be obtained. The solid circle corresponds with the MLE $\hat{\mu}$. Observe the egg shape of the confidence region

computed although the model is based on odds ratios. The analysis uses all available information and does not rely on conditioning, use of which can be questioned in many situations.

Thirdly, we have restricted attention to a simple bivariate model for Θ. The next step is to incorporate covariates in the model that might explain the variation in the Θ's. A first candidate would be the sample sizes (n_T, n_C), which would allow a check of our implicit assumption that θ is independent of (n_T, n_C). More work has to be done in building models that use all data and take into account more random variation than simple patient-to-patient variation.

Finally, a program written in Gauss is available from the first author. This program requires a set of 2×2 tables as input and performs all parametric analyses given in the paper. The discrete non-parametric procedure is not included because of its numeric instability.

ACKNOWLEDGEMENTS

Both referees are thanked for their valuable suggestions that improved the presentation of the paper.

REFERENCES

1. 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*, **27**, 335–371 (1985).
2. Der Simonian, R. and Laird, N. M. 'Meta-analysis in clinical trials', *Controlled Clinical Trials*, **7**, 177–188 (1986).
3. Stijnen, T. and Van Houwelingen, J. C. 'Empirical Bayes methods in clinical trials meta-analysis', *Biometrical Journal*, **32**, 335–346 (1990).

4. Dempster, A. P., Laird, N. M. and Rubin, D. B. 'Maximum likelihood from incomplete data via the EM algorithm', *Journal of the Royal Statistical Society, Series B*, **39**, 1–38 (1977).
5. Mosteller, F. and Chalmers, T. C. 'Some progress and problems in meta-analysis of clinical trials', *Statistical Science*, **7**, 227–236 (1992).
6. Collins, R. and Langman, M. 'Treatment with histamine H, antagonists in acute upper gastrointestinal hemorrhage', *New England Journal of Medicine*, **313**, 660–666 (1985).
7. Laird, N. M. 'Nonparametric maximum likelihood estimation of a mixing distribution', *Journal of the American Statistical Association*, **73**, 805–811 (1978).