STATISTICS IN MEDICINE
Statist. Med. 2002; **21**:589–624 (DOI: 10.1002/sim.1040)

TUTORIAL IN BIOSTATISTICS
Advanced methods in meta-analysis: multivariate approach and meta-regression

Hans C. van Houwelingen[1,*,†], Lidia R. Arends[2] and Theo Stijnen[2]

[1]*Department of Medical Statistics, Leiden University Medical Center, P.O. Box 9604, 2300 RC Leiden, The Netherlands*
[2]*Department of Epidemiology & Biostatistics, Erasmus Medical Center Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands*

SUMMARY

This tutorial on advanced statistical methods for meta-analysis can be seen as a sequel to the recent Tutorial in Biostatistics on meta-analysis by Normand, which focused on elementary methods. Within the framework of the general linear mixed model using approximate likelihood, we discuss methods to analyse univariate as well as bivariate treatment effects in meta-analyses as well as meta-regression methods. Several extensions of the models are discussed, like exact likelihood, non-normal mixtures and multiple endpoints. We end with a discussion about the use of Bayesian methods in meta-analysis. All methods are illustrated by a meta-analysis concerning the efficacy of BCG vaccine against tuberculosis. All analyses that use approximate likelihood can be carried out by standard software. We demonstrate how the models can be fitted using SAS Proc Mixed. Copyright © 2002 John Wiley & Sons, Ltd.

KEY WORDS: meta-analysis; meta-regression; multivariate random effects models

1. INTRODUCTION

In this paper we review advanced statistical methods for meta-analysis as used in bivariate meta-analysis [1] (that is, two outcomes per study are modelled simultaneously) and meta-regression [2]. It can be seen as a sequel to the recent Tutorial in Biostatistics on meta-analysis by Normand [3]. Meta-analysis is put in the context of mixed models using (approximate) likelihood methods to estimate all relevant parameters. In the medical literature meta-analysis is usually applied to the results of clinical trials, but the application of the theory presented in this paper is not limited to clinical trials only. It is the aim of the paper not only to discuss the underlying theory but also to give practical guidelines how to carry out these analyses.

As the leading example we use the meta-analysis data set of Colditz *et al.* [4]. This data set is also discussed in Berkey *et al*. [2]. Wherever feasible, it is specified how the analysis can

*Correspondence to: Hans C. van Houwelingen, Department of Medical Statistics, Leiden University Medical Center, P.O. Box 9604, 2300 RC Leiden, The Netherlands
†E-mail: jcvanhouwelingen@lumc.nl

Received December 1999
Accepted June 2001

be performed by using the SAS procedure Proc Mixed. The paper is organized as follows. In Section 2 we review the concept of approximate likelihood that was introduced in the meta-analysis setting by DerSimonian and Laird [5]. In Section 3 we review the meta-analysis of one-dimensional treatment effect parameters. In Section 4 we discuss the bivariate approach [1] and its link with the concept of underlying risk as source of heterogeneity [6–10]. In Section 5 we discuss meta-regression within the mixed model setting. Covariates considered are aggregate measures on the study level. We do not go into meta-analysis with patient-specific covariates. In principle that is not different from analysing a multi-centre study [11]. In Section 6 several extensions are discussed: exact likelihood's based on conditioning; non-normal mixtures; multiple endpoints; other outcome measures, and other software. This is additional material that can be skipped at first reading. Section 7 is concerned with the use of Bayesian methods in meta-analysis. We argue that Bayesian methods can be useful if they are applied at the right level of the hierarchical model. The paper is concluded in Section 8.

2. APPROXIMATE LIKELIHOOD

The basic situation in meta-analysis is that we are dealing with n studies in which a parameter of interest ϑ_i $(i=1,\ldots,n)$ is estimated. In a meta-analysis of clinical trials the parameter of interest is some measure of the difference in efficacy between the two treatment arms. The most popular choice is the log-odds ratio, but this could also be the risk or rate difference or the risk or rate ratio for dichotomous outcome or similar measures for continuous outcomes or survival data. All studies report an estimate $\hat{\vartheta}_i$ of the true ϑ_i and the standard error s_i of the estimate. If the studies only report the estimate and the p-value or a confidence interval, we can derive the standard error from the p-value or the confidence interval. In the Sections 3 to 5, which give the main statistical tools, we act as if $\hat{\vartheta}_i$ has a normal distribution with unknown mean ϑ_i and known standard deviation s_i, that is

$$\hat{\vartheta}_i \sim \mathrm{N}(\vartheta_i, s_i^2) \tag{1}$$

Moreover, since the estimates are derived from different data sets, the $\hat{\vartheta}_i$ are conditionally independent given ϑ_i. This approximate likelihood approach goes back to the seminal paper by DerSimonian and Laird [5]. However, it should be stressed that it is *not* the normality of the frequency distribution of $\hat{\vartheta}_i$ that is employed in our analysis. Since our whole approach is likelihood based, we only use that the likelihood of the unknown parameter in each study looks like the likelihood of (1). Thus, if we denote the log-likelihood of the ith study by $\ell_i(\vartheta)$, the real approximation is

$$\ell_i(\vartheta) = -\frac{1}{2}(\vartheta - \hat{\vartheta}_i)^2/s_i^2 + c_i \tag{2}$$

where c_i is some constant that does not depend on the unknown parameter.

If in each study the unknown parameter is estimated by maximum likelihood, approximation (2) is just the second-order Taylor expansion of the (profile) log-likelihood around the MLE $\hat{\vartheta}_i$. The approximation (2) is usually quite good, even if the estimator $\hat{\vartheta}_i$ is discrete. Since most studies indeed use the maximum likelihood method to estimate the unknown parameter, we are confident that (2) can be used as an approximation. In Section 6 we will discuss

some refinements of this approximation. In manipulating the likelihoods we can safely act as if we assume that (1) is valid and use, for example, known results for mixtures of normal distributions. However, we want to stress that actually we only use assumption (2).

The approach of Yusuf *et al.* [12], popular in fixed effect meta-analysis, and of Whitehead and Whitehead [13] are based on a Taylor expansion of the log-likelihood around the value $\vartheta = 0$. This is valid if the effects in each study are relatively small. It gives an approximation in the line of (2) with different estimators and standard errors but a similar quadratic expression in the unknown parameter.

As we have already noted, the most popular outcome measure in meta-analysis is the log-odds ratio. Its estimated standard error is equal to ∞ if one of the frequencies in the 2×2 table is equal to zero. That is usually repaired by adding 0.5 to all cell frequencies. We will discuss more appropriate ways of handling this problem in Section 6.

3. ANALYSING ONE-DIMENSIONAL TREATMENT EFFECTS

The *analysis under homogeneity* makes the assumption that the unknown parameter is exactly the same in all studies, that is $\vartheta_1 = \vartheta_2 = \cdots = \vartheta_n = \vartheta$. The log-likelihood for ϑ is given by

$$\ell(\vartheta) = \sum_i \ell_i(\vartheta) = -\frac{1}{2} \sum_i [(\vartheta - \hat{\vartheta}_i)^2 / s_i^2 + \ln(s_i^2) + \ln(2\pi)] \tag{3}$$

Maximization is straightforward and results in the well-known estimator of the common effect

$$\hat{\vartheta}_{\text{hom}} = \left[\sum_i \hat{\vartheta}_i / s_i^2\right] \Big/ \left[\sum_i 1/s_i^2\right]$$

with standard error

$$\text{SE}(\hat{\vartheta}_{\text{hom}}) = 1 \Big/ \sqrt{\left(\sum_i 1/s_i^2\right)}$$

Confidence intervals for ϑ can be based on normal distributions, since the s_i^2 terms are assumed to be known. Assuming the s_i^2 terms to be known instead of to be estimated has little impact on the results [14]. This is the basis for the traditional meta-analysis.

The assumption of homogeneity is questionable even if it is hard to disprove for small meta-analyses [15]. That is, heterogeneity might be present and should be part of the analysis even if the test for heterogeneity is not significant. Heterogeneity is found in many meta-analyses and is likely to be present since the individual studies are never identical with respect to study populations and other factors that can cause differences between studies.

The popular model for the *analysis under heterogeneity* is the normal mixture model, introduced by DerSimonian and Laird [5], that considers the ϑ_i to be an independent random sample from a normal population

$$\vartheta_i \sim \text{N}(\vartheta, \sigma^2)$$

Normality of this mixture is a true assumption and not a simplifying approximation. We will further discuss it in Section 6. The resulting marginal distribution of ϑ_i is easily obtained as

$\hat{\vartheta}_i \sim \mathrm{N}(\vartheta, \sigma^2 + s_i^2)$ with corresponding log-likelihood

$$\ell(\vartheta, \sigma^2) = -\frac{1}{2}\sum_i [(\vartheta - \hat{\vartheta}_i)^2/(\sigma^2 + s_i^2) + \ln(\sigma^2 + s_i^2) + \ln(2\pi)] \tag{4}$$

Notice that (3) and (4) are identical if $\sigma^2 = 0$.

This log-likelihood is the basis for inference about *both* parameters ϑ and σ^2. Maximum likelihood estimates can be obtained by different algorithms. In the example below, it is shown how the estimates can be obtained by using the SAS procedure Proc Mixed. If σ^2 were known, the ML estimate for ϑ would be

$$\hat{\vartheta}_{\mathrm{het}} = \left[\sum_i (\hat{\vartheta}_i/(\sigma^2 + s_i^2)\right] \Big/ \left[\sum_i [1/(\sigma^2 + s_i^2)\right]$$

with standard error

$$\mathrm{SE}(\hat{\vartheta}_{\mathrm{het}}) = 1 \Big/ \sqrt{\left\{\sum_i 1/(\sigma^2 + s_i^2)\right\}}$$

The latter can also be used if σ^2 is estimated and the estimated value is plugged in, as is done in the standard DerSimonian and Laird approach.

The construction of confidence intervals for both parameters is more complicated than in the case of a simple sample from a normal distribution. Simple χ^2- and t-distributions with d.f. $= n-1$ are not appropriate. In this article all models are fitted using SAS Proc Mixed, which gives Satherthwaite approximation based confidence intervals. Another possibility is to base confidence intervals on the likelihood ratio test, using profile log-likelihoods. That is, the confidence interval consists of all parameter values that are not rejected by the likelihood ratio test. Such confidence intervals often have amazingly accurate coverage probabilities [16, 17]. Brockwell and Gordon [18] compared the commonly used DerSimonian and Laird method [5] with the profile likelihood method. Particularly when the number of studies is modest, the DerSimonian and Laird method had coverage probabilities considerably below 0.95 and the profile likelihood method achieved the best coverage probabilities.

The profile log-likelihoods are defined by

$$p\ell_1(\vartheta) = \max_{\sigma^2} \ell(\vartheta, \sigma^2) \quad \text{and} \quad p\ell_2(\sigma^2) = \max_{\vartheta} \ell(\vartheta, \sigma^2)$$

Based on the usual $\chi^2_{[1]}$-approximation for $2(p\ell_1(\hat{\vartheta}) - p\ell_1(\vartheta))$, the 95 per cent confidence interval for ϑ is obtained as all ϑ's satisfying $p\ell_1(\vartheta) > p\ell_1(\hat{\vartheta}) - 1.92$ (1.92 is the 95 per cent centile of the $\chi^2_{[1]}$ distribution 3.84 divided by 2) and similarly for σ^2. Unlike the usual confidence interval based on Wald's method, this confidence interval for ϑ implicitly accounts for the fact that σ^2 is estimated.

Testing for heterogeneity is equivalent to testing $H_0{:}\sigma^2 = 0$ against $H_1{:}\sigma^2 > 0$. The likelihood ratio test statistic is $T = 2(p\ell_2(\hat{\sigma}^2) - p\ell_2(0))$. Since $\sigma^2 = 0$ is on the boundary of the parameter space, T does not have a $\chi^2_{[1]}$-distribution, but its distribution is a mixture with probabilities half of the degenerate distribution in zero and the $\chi^2_{[1]}$-distribution [19]. That means that the p-value of the naive LR-test has to be halved. Once the mixed model has been fitted, the

following information is available at the overall level:

(i) $\hat{\vartheta}$ and its confidence interval, showing the existence or absence of an overall effect;
(ii) $\hat{\sigma}^2$ and its confidence interval (and the test for heterogeneity), showing the variation between studies;
(iii) approximate 95 per cent prediction interval for the *true* parameter $\hat{\vartheta}_{\text{new}}$ of a new unrelated study: $\hat{\vartheta} \pm 1.96\hat{\sigma}$ (approximate in the sense that it ignores the error in the estimation of ϑ and σ);
(iv) an estimate of the probability of a positive result of a new study:

$$P(\vartheta_{\text{new}} > 0) = \Phi(\hat{\vartheta}/\hat{\sigma})$$

(where Φ is the standard normal cumulative distribution function).

The following information is available at the individual study level:

(i) posterior confidence intervals for the true ϑ_i's of the studies in the meta-analysis based on the posterior distribution $\vartheta_i \mid \hat{\vartheta}_i \sim \text{N}(\hat{\vartheta} + B_i(\hat{\vartheta}_i - \hat{\vartheta}), B_i s_i^2)$ with $B_i = \hat{\sigma}^2/(\hat{\sigma}^2 + s_i^2)$. The posterior means or so-called empirical Bayes estimates give a more realistic view on the results of, especially, the small studies. See the meta-analysis tutorial of Normand [3] for more on this subject.

3.1. Example

To illustrate the above methods we make use of the meta-analysis data given by Colditz *et al.* [4]. Berkey *et al.* [2] also used this data set to illustrate their random-effects regression approach to meta-analysis. The meta-analysis concerns 13 trials on the efficacy of BCG vaccine against tuberculosis. In each trial a vaccinated group is compared with a non-vaccinated control group. The data consist of the sample size in each group and the number of cases of tuberculosis. Furthermore some covariates are available that might explain the heterogeneity among studies: geographic latitude of the place where the study was done; year of publication, and method of treatment allocation (random, alternate or systematic). The data are presented in Table I.

We stored the data in an SAS file called `BCG_data.sd2` (see Data step in SAS commands below). The treatment effect measure we have chosen is the log-odds ratio, but the analysis could be carried out in the same way for any other treatment effect measure.

3.1.1. Fixed effects model. The analysis under the assumption of homogeneity is easily performed by hand. Only for the sake of continuity and uniformity do we also show how the analysis can be carried out using SAS software.

The ML-estimate of the log-odds ratio for trial i is

$$\ln OR_i = \log\left(\frac{Y_{\text{A},i}/(n_{\text{A},i} - Y_{\text{A},i})}{Y_{\text{B},i}/(n_{\text{B},i} - Y_{\text{B},i})}\right)$$

where $Y_{\text{A},i}$ and $Y_{\text{B},i}$ are the number of disease cases in the vaccinated (A) and non-vaccinated group (B) in trial i, and $n_{\text{A},i}$ and $n_{\text{B},i}$ the sample sizes. The corresponding within-trial variance,

Table I. Example: data from clinical trials on efficacy of BCG vaccine in the prevention of tuberculosis [2, 4].

Trial	Vaccinated		Not vaccinated		ln(OR)	Latitude	Year	Allocation*
	Disease	No disease	Disease	No disease				
1	4	119	11	128	−0.93869	44	48	Random
2	6	300	29	274	−1.66619	55	49	Random
3	3	228	11	209	−1.38629	42	60	Random
4	62	13536	248	12619	−1.45644	52	77	Random
5	33	5036	47	5761	−0.21914	13	73	Alternate
6	180	1361	372	1079	−0.95812	44	53	Alternate
7	8	2537	10	619	−1.63378	19	73	Random
8	505	87886	499	87892	0.01202	13	80	Random
9	29	7470	45	7232	−0.47175	27*	68	Random
10	17	1699	65	1600	−1.40121	42	61	Systematic
11	186	50448	141	27197	−0.34085	18	74	Systematic
12	5	2493	3	2338	0.44663	33	69	Systematic
13	27	16886	29	17825	−0.01734	33	76	Systematic

*This was actually a negative number; we used the absolute value in the analysis.

computed from the inverse of the matrix of second derivatives of the log-likelihood, is

$$\mathrm{var}(\ln OR_i) = \frac{1}{Y_{\mathrm{A},i}} + \frac{1}{n_{\mathrm{A},i} - Y_{\mathrm{A},i}} + \frac{1}{Y_{\mathrm{B},i}} + \frac{1}{n_{\mathrm{B},i} - Y_{\mathrm{B},i}}$$

which is also known as Woolf's formula.

These within-trial variances were stored in the same SAS data file as above, called `BCG_data.sd2`. In the analysis, these variances are assumed to be known and fixed.

```
# THE DATA STEP;

data BCG_data;
input TRIAL VD VWD NVD NVWD LATITUDE YEAR ALLOC;
LN_OR=log((VD/VWD)/(NVD/NVWD));
EST=1/VD+1/VWD+1/NVD+1/NVWD;
datalines;
 1     4       119        11       128         44       48       1
 2     6       300        29       274         55       49       1
 3     3       228        11       209         42       60       1
 4    62     13536       248     12619         52       77       1
 5    33      5036        47      5761         13       73       2
 6   180      1361       372      1079         44       53       2
 7     8      2537        10       619         19       73       1
 8   505     87886       499     87892         13       80       1
 9    29      7470        45      7232         27       68       1
10    17      1699        65      1600         42       61       3
11   186     50448       141     27197         18       74       3
12     5      2493         3      2338         33       69       3
13    27     16886        29     17825         33       76       3
;
proc print;run;
```


Running these SAS commands gives the following output:

```
OBS  TRIAL    VD    VWD  NVD    NVWD    LATITUDE   YEAR   ALLOC  LN_OR      EST

 1     1      4     119   11     128       44       48      1    -0.93869   0.35712
 2     2      6     300   29     274       55       49      1    -1.66619   0.20813
 3     3      3     228   11     209       42       60      1    -1.38629   0.43341
 4     4     62   13536  248   12619       52       77      1    -1.45644   0.02031
 5     5     33    5036   47    5761       13       73      2    -0.21914   0.05195
 6     6    180    1361  372    1079       44       53      2    -0.95812   0.00991
 7     7      8    2537   10     619       19       73      1    -1.63378   0.22701
 8     8    505   87886  499   87892       13       80      1     0.01202   0.00401
 9     9     29    7470   45    7232       27       68      1    -0.47175   0.05698
10    10     17    1699   65    1600       42       61      3    -1.40121   0.07542
11    11    186   50448  141   27197       18       74      3    -0.34085   0.01253
12    12      5    2493    3    2338       33       69      3     0.44663   0.53416
13    13     27   16886   29   17825       33       76      3    -0.01734   0.07164
```

The list of variables matches that in Table I (VD = vaccinated and diseased, VWD = vaccinated and without disease, NVD = not vaccinated and diseased, $NVWD$ = not vaccinated and without disease). The variable `ln_or` contains the estimated log-odds ratio of each trial and the variable `est` contains its variance per trial. In the Proc Mixed commands below, SAS assumes that the within trial variances are stored in a variable with the name '`est`'.

```
# THE FIXED EFFECTS MODEL;
Proc mixed method =ml data=BCG_data;     #call SAS procedure;
class trial;                             #specifies 'trial' as classification variable;
model ln_or=/ s ;                        #an intercept only model; print the solution s;
repeated /group=trial;                   #each trial has its own within-trial variance;
parms / parmsdata=BCG_data               #the parmsdata-option reads in the variable EST
                                         (indicating the within-trial variances) from
                                         the dataset BCG_data.sd2;
eqcons=1 to 13;                          #the within trial variances are considered to
                                         be known and must be kept constant;
run;
```

Running this analysis gives the following output:

```
                          The MIXED Procedure
(...)
                       Solution for Fixed Effects

Effect          Estimate     Std Error    DF       t  Pr > |t|  Alpha     Lower   Upper
INTERCEPT   -0.43627138    0.04227521    12  -10.32    0.0001   0.05   -0.5284 -0.3442
```

The estimate of the common log-odds ratio is equal to -0.436 with standard error $= 0.042$ leading to a 95 per cent Wald based confidence interval of the log-odds ratio from -0.519 to -0.353. (Although it seems overly precise, we will present results to three decimals, since these are used in further calculations and to facilitate comparisons between results of different

models.) This corresponds to an estimate of 0.647 with a 95 per cent confidence interval from 0.595 to 0.703 for the odds ratio itself. Thus we can conclude that vaccination is beneficial.

The confidence intervals and p-values provided by SAS Proc Mixed are based on the t-distribution rather than on the standard normal distribution, as is done in the standard likelihood approach. The number of degrees of freedom of the t-distribution is determined by Proc Mixed according to some algorithm. One can choose between several algorithms, but one can also specify in the model statement the number of degrees of freedom to be used for each covariable, except for the intercept. To get the standard Wald confidence interval and p-value for the intercept, the number of degrees of freedom used for the intercept should be specified to be ∞, which can be accomplished by making a new intercept covariate equal to 1 and subsequently specifying 'no intercept' (`'noint'`). The SAS statement to be used is then:

```
model ln_or = int / s cl noint ddf = 1000;
```

(the variable `'int'` is a self-made intercept variable equal to 1).

3.1.2. Simple random effects model, maximum likelihood. The analysis under heterogeneity can be carried out by executing the following SAS statements. Unlike the previous model where we read in the within-trial variances from the datafile, we now specify the within-trial variances explicitly in the 'parms' statement. This has to be done because we want to define a grid of values for the first covariance parameter, that is, the between-trial variance, to get the profile likelihood function for the between-trial variance to get its likelihood ratio based 95 per cent confidence interval. Of course, one could also give only one starting value and read the data from an SAS datafile like we did before.

```
# THE RANDOM EFFECTS MODEL (MAXIMUM LIKELIHOOD);
   Proc mixed cl method=ml data=BCG_data;   #call of procedure; 'cl' asks for confidence
                                            intervals of covariance parameters;
   class trial;                             #trial is classification variable;
   model ln_or= / s cl;                     #an intercept only model. print fixed effect
                                            solution 's' and its confidence limits 'cl';
   random int/ subject=trial s;             #trial is specified as random effect; 's'
                                            asks for the empirical Bayes estimates;
   repeated /group=trial;                   #each trial has its own within trial
                                            variance;
   parms (0.01 to 2.00 by 0.01)(0.35712)    #defines grid of values for between trial
   (0.20813)(0.43341)(0.02031)(0.05195)     variance (from 0.01 to 1.00), followed by the
   (0.00991)(0.22701)(0.00401)(0.05698)     13 within trial variances which are assumed
   (0.07542)(0.01253)(0.53416)(0.07164)     to be known and must be kept fixed;
   /eqcons=2 to 14;
   make 'Parms' out=Parmsml;                #in the dataset 'Parms' the maximum log
                                            likelihood for each value of the grid
                                            specified for the between trial variance is
                                            stored, in order to read off the profile
                                            likelihood based 95% CI for the between trial
                                            variance;
   run;
```


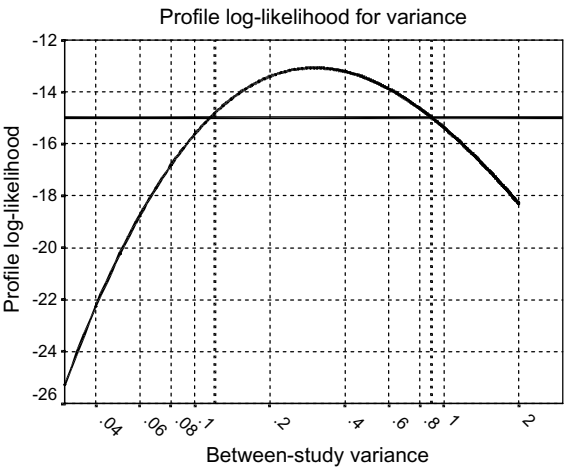

Figure 1. The 95 per cent confidence interval of the between-trial variance σ^2 based on the profile likelihood funtion: $(0.12, 0.89)$.

Running this program gives the following output:

```
                          The MIXED Procedure

(...)
                  Covariance Parameter Estimates (MLE)

    Cov Parm    Subject  Group         Estimate  Alpha     Lower     Upper
    INTERCEPT   TRIAL                0.30245716   0.05    0.1350    1.1810

(...)

                         Solution for Fixed Effects

Effect         Estimate     Std Error    DF       t  Pr > |t|  Alpha     Lower   Upper
INTERCEPT   -0.74197023    0.17795376    12   -4.17    0.0013   0.05   -1.1297 -0.3542
```

The ML-estimate of the mean log-odds ratio is -0.742 with standard error 0.178. The standard Wald based 95 per cent confidence interval is -1.091 to -0.393. (SAS Proc Mixed gives a slightly wider confidence interval based on a t-distribution with d.f. $= 12$). This corresponds to an estimated odds ratio of 0.476 with a 95 per cent confidence interval from 0.336 to 0.675.

The ML-estimate of the between-trial variance σ^2 is equal to 0.302. For each value of the grid specified in the 'Parms' statement for the between-trial variance (in the example the grid runs from 0.01 to 2.00 with steps of 0.01), the maximum log-likelihood value is stored as variable 'LL' in the SAS file '`Parmsml.sd2`'. Plotting the maximum log-likelihood values against the grid of between-trial variances gives the profile likelihood plot for the between-trial variance presented in Figure 1. From this plot or a listing of the data set '`Parmsml.sd2`' one can read off the profile likelihood based 95 per cent confidence interval for the between-trial

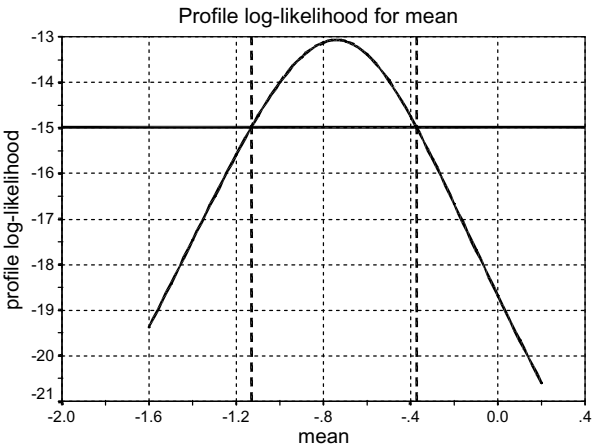

Figure 2. The 95 per cent confidence interval of the treatment effect (log-odds ratio) θ based on the profile likelihood funtion: $(-1.13, -0.37)$.

variance σ^2. This is done by looking for the two values of the between-trial variance with a corresponding log-likelihood of 1.92 lower than the maximum log-likelihood. The 95 per cent profile likelihood based confidence interval for σ^2 is $(0.12, 0.89)$. (SAS Proc Mixed gives a Satterthwaite approximation based 95 per cent confidence interval running from 0.135 to 1.180.)

Notice that by comparing the maximum log-likelihood of this model with the previous fixed effects model, one gets the likelihood ratio test for homogeneity (the p-value has to be halved, because $\sigma^2 = 0$ is on the boundary of the parameter space).

A profile likelihood based confidence interval for the mean treatment effect ϑ can be made by trial and error by defining the variable `y=ln_or-c` as dependent variable for various values of c and specifying a model without intercept (add '`noint`' after the slash in the model statement). Then look for the two values of c that decrease the maximum log-likelihood by 1.92. The profile log-likelihood plot for ϑ is given in Figure 2. The 95 per cent confidence interval for the log-odds ratio ϑ is $(-1.13, -0.37)$, slightly wider than the simple Wald approximation given above. This corresponds with a 95 per cent confidence interval for the odds ratio of 0.323 to 0.691.

Remark

In Proc Mixed one can also choose the restricted maximum likelihood (REML) estimate (specify `method=reml` instead of `method=ml`). Then the resulting estimate for the between-trial variance σ^2 is identical to the iterated DerSimonian–Laird estimator [5]. However, in this case the profile likelihood function should not be used to make a confidence interval for the log-odds ratio ϑ. The reason is that differences between maximized REML likelihoods cannot be used to test hypotheses concerning fixed parameters in a general linear mixed model [20].

The observed and corresponding empirical Bayes estimated log-odds ratios with their 95 per cent standard Wald, respectively, the 95 per cent posterior confidence intervals per trial, are presented in Figure 3. This figure shows the shrinkage of the empirical Bayes estimates

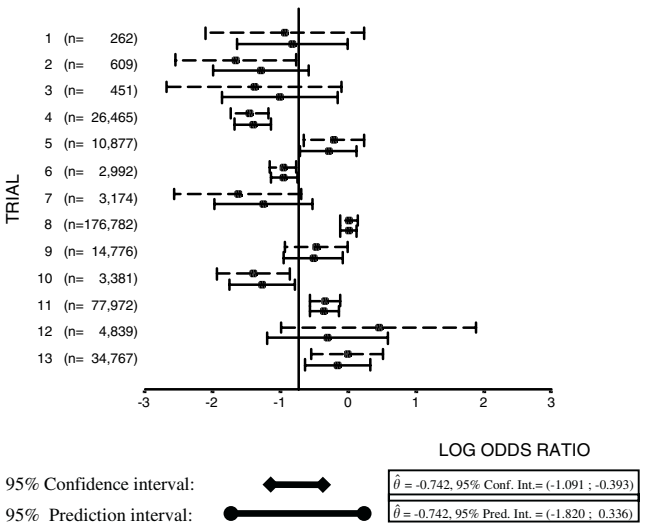

Figure 3. Forrest plot with the estimated log-odds ratios of tuberculosis with their 95 per cent confidence intervals in the trials included in the meta-analysis. The dashed horizontal lines indicate the standard Wald confidence intervals. The solid horizontal lines indicate the posterior or so-called empirical Bayes confidence intervals. The vertical line indicates the ML-estimate of the common (true) log-odds ratio. Below the figure the 95 per cent confidence interval for the mean log-odds ratio and the 95 per cent prediction interval for the true log-odds ratio are presented.

towards the estimated mean log-odds ratio and their corresponding smaller posterior confidence intervals. The overall confidence interval of the mean true treatment effect and the overall prediction interval of the true treatment effect are given at the bottom of the figure. The 95 per cent prediction interval indicates the interval in which 95 per cent of the true treatment effects of new trials are expected to fall. It is calculated as the ML-estimate plus and minus 1.96 times the estimated between-trial standard deviation s and is here equal to $(-1.820$ to $0.336)$. The estimated probability for a new trial having a positive true treatment effect is $\Phi(0.742/0.302) = 0.993$.

4. BIVARIATE APPROACH

In the previous section the parameter of interest was one-dimensional. In many situations it can be bivariate or even multivariate, for instance when there are more treatment groups or more outcome variables. In this section we discuss the case of a two-dimensional parameter of interest. We introduce the bivariate approach with special reference to the situation where one is interested in 'control rate regression', that is, relating the treatment effect size to the risk of events in the control group. However, the approach applies generally.

Many studies show considerable variation in what is called the baseline risk. The baseline risk indicates the risk for patients under the control condition, which is the average risk of the patients in that trial when the patients were treated with the control treatment. One might wonder if there is a relation between treatment effect and baseline risk. Considering only the

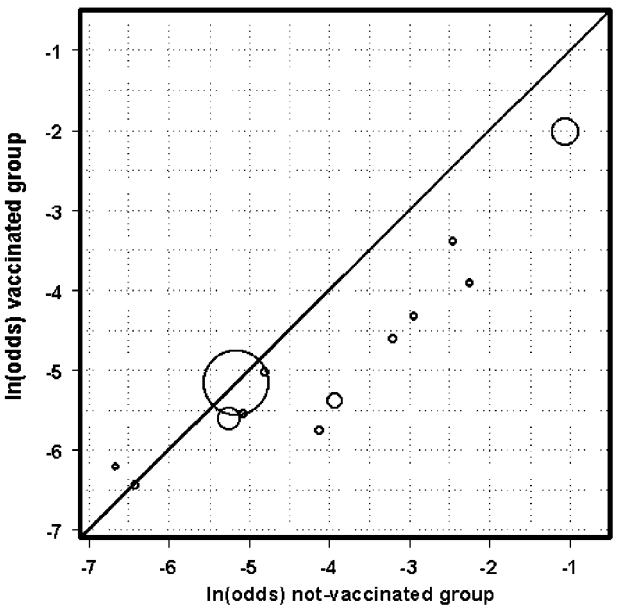

Figure 4. L'Abbé plot of observed log(odds) of the not-vaccinated trial arm versus the vaccinated trial arm. The size of the circle is an indication for the inverse of the variance of the log-odds ratio in that trial. Below the $x = y$ line, the log-odds in the vaccinated are lower than the log-odds in the not-vaccinated arm, indicating that the vaccination works. On or above the $x = y$ line, vaccination does not work beneficially.

differences between the study arms may hide a lot information. Therefore, we think it is wise to consider the pair of outcomes of the two treatments. This is nicely done in the l'Abbé plot [21], that gives a bivariate representation of the data by plotting the log-odds in arm A versus the log-odds in arm B. We show the plot in Figure 4 for the data of our example with A the vaccinated arm and B the not-vaccinated arm. The size of each circle represents the inverse of the variance of the log-odds ratio in that trial. Points below the line of identity correspond to trials with an observed positive effect of vaccination.

The graph shows some effect of vaccination especially at the higher incidence rates. A simple (approximate) bivariate model for any observed pair of arm specific outcome measures $\omega_i = (\hat{\omega}_{A,i}, \hat{\omega}_{B,i})$ with standard errors $(s_{A,i}, s_{B,i})$ in trial i is

$$\begin{pmatrix} \hat{\omega}_{A,i} \\ \hat{\omega}_{B,i} \end{pmatrix} \sim N\left(\begin{pmatrix} \omega_{A,i} \\ \omega_{B,i} \end{pmatrix}, \begin{pmatrix} s^2_{A,i} & 0 \\ 0 & s^2_{B,i} \end{pmatrix} \right) \quad (i = 1, \ldots, n)$$

where $\omega_i = (\omega_{A,i}, \omega_{B,i})$ is the pair of true arm specific outcome measures for trial i. The conditional independence of $\hat{\omega}_A$ and $\hat{\omega}_B$ given the true ω_A and ω_B is a consequence of the randomized parallel study design and the fact that ω_A and ω_B are arm specific. In general, for instance in a cross-over study, or when ω_A and ω_B are treatment effects on two different outcome variables, the estimates might be correlated.

The mixed model approach assumes the pair $(\omega_{A,i}, \omega_{B,i})$ to follow a bivariate normal distribution, where, analogous to the univariate random effects model of Section 3, the true outcome measures for both arms in the trials are normally distributed around some common mean treatment-arm outcome measure with a between-trial covariance matrix Σ:

$$\begin{pmatrix} \omega_{A,i} \\ \omega_{B,i} \end{pmatrix} \sim N\left(\begin{pmatrix} \omega_A \\ \omega_B \end{pmatrix}, \Sigma \right) \quad \text{with } \Sigma = \begin{pmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{AB} & \Sigma_{BB} \end{pmatrix}$$

Σ_{AA} and Σ_{BB} describe the variability among trials in true risk under the vaccination and control condition, respectively. Σ_{AB} is the covariance between the true risk in the vaccination and control group.

The resulting marginal model is

$$\begin{pmatrix} \hat{\omega}_{A,i} \\ \hat{\omega}_{B,i} \end{pmatrix} \sim N\left(\begin{pmatrix} \omega_A \\ \omega_B \end{pmatrix}, \Sigma + C_i \right)$$

with C_i the diagonal matrix with the s_i^2's.

Maximum likelihood estimation for this model can be quite easily carried out by a self-made program based on the EM algorithm as described in reference [1], but more practically convenient is to use appropriate mixed model software from statistical packages, such as the SAS procedure Proc Mixed.

Once the model is fitted, the following derived quantities are of interest:

(i) The mean difference $(\omega_A - \omega_B)$ and its standard error

$$\sqrt{\{(\text{var}(\omega_A) + \text{var}(\omega_B) - 2\,\text{cov}(\omega_A, \omega_B))\}}$$

(ii) The population variance of the difference $\text{var}(\omega_A - \omega_B) = \Sigma_{AA} + \Sigma_{BB} - 2\Sigma_{AB}$.

(iii) The shape of the bivariate relation between the (true) ω_A and ω_B. That can be described by ellipses of equal density or by the regression lines of ω_A on ω_B and of the ω_B on ω_A. These lines can be obtained from classical bivariate normal theory. For example, the regression line of ω_A on ω_B has slope $\beta = \Sigma_{AB}/\Sigma_{BB}$ and residual variance $\Sigma_{AA} - \Sigma_{AB}^2/\Sigma_{BB}$. The regression of the difference $(\omega_A - \omega_B)$ on either ω_A or ω_B can be derived similarly. At the end of this section we come back to the usefulness of these regression lines.

The standard errors of the regression slopes can be calculated from the covariance matrix of the estimated covariance parameters by the delta method or by Fieller's method [22].

4.1. Example (continued): bivariate random effects model

As an example we carry out a bivariate meta-analysis with ω_A and ω_B the log-odds of tuberculosis in the vaccinated and the not-vaccinated control arm, respectively. To execute a bivariate analysis in the SAS procedure Proc Mixed, we have to change the structure of the data set. Each treatment arm of a trial becomes a row in the data set, resulting in twice as many rows as in the original data set. The dependent variable is now the estimated log-odds

in a treatment arm instead of the log-odds ratio. The new data set is called `BCGdata2.sd2` and the observed log-odds is called `lno`. The standard error of the observed log-odds, estimated by taking the square root of minus the inverse of the second derivative of the log-likelihood, is equal to $\sqrt{\left(\frac{1}{x}+\frac{1}{n-x}\right)}$, where n is the sample size of a treatment arm and x is the number of tuberculosis cases in a treatment arm. These standard errors are also stored in the SAS data set `covvars2.sd2`.

The bivariate random effects analysis can be carried out by running the SAS commands given below. In the data step, the new data set `BCGdata2.sd2` is made out of the data set `BCGdata2.sd2`, and the covariates are defined on trial arm level. The variable `exp` is 1 for the vaccinated (experimental) arms and 0 for the not-vaccinated (control) arms. The variable `con` is defined analogously with experimental and control reversed. The variable `arm` identifies the 26 unique treatment arms from the 13 studies (here from 1 to 26); `latcon`, `latexp`, `yearcon` and `yearexp` are covariates to be used later. For numerical reasons we centralized the four variables `latcon`, `latexp`, `yearcon` and `yearexp` by substracting the mean.

```
# THE DATA STEP (BIVARIATE ANALYSIS)
data bcgdata2;set bcg_data;
treat=1; lno=log(vd/vwd); var=1/vd+1/vwd; n=vd+vwd; output;
treat=0; lno=log(nvd/nvwd); var=1/nvd+1/nvwd; n=nvd+nvwd; output;
keep trial lno var n treat latitude--alloc;
run;
data bcgdata2;set bcgdata2;
arm=_n_; exp=(treat=1); con=(treat=0);
latcon=(treat=0)*(latitude-33); latexp=(treat=1)*(latitude-33);
yearcon=(treat=0)*(year-66); yearexp=(treat=1)*(year-66);
proc print noobs;run;
```

Running these SAS commands gives the following output:

TRIAL	LATITUDE	YEAR	ALLOC	TREAT	LNO	VAR	ARM	EXP	CON	LATCON	LATEXP	YEARCON	YEAREXP
1	44	48	1	1	-3.39283	0.25840	1	1	0	0	11	0	-18
1	44	48	1	0	-2.45413	0.09872	2	0	1	11	0	-18	0
2	55	49	1	1	-3.91202	0.17000	3	1	0	0	22	0	-17
2	55	49	1	0	-2.24583	0.03813	4	0	1	22	0	-17	0
3	42	60	1	1	-4.33073	0.33772	5	1	0	0	9	0	-6
3	42	60	1	0	-2.94444	0.09569	6	0	1	9	0	-6	0
4	52	77	1	1	-5.38597	0.01620	7	1	0	0	19	0	11
4	52	77	1	0	-3.92953	0.00411	8	0	1	19	0	11	0


```
 5   13   73   2   1  -5.02786   0.03050    9   1   0    0   -20    0    7
 5   13   73   2   0  -4.80872   0.02145   10   0   1  -20     0    7    0
 6   44   53   2   1  -2.02302   0.00629   11   1   0    0    11    0  -13
 6   44   53   2   0  -1.06490   0.00361   12   0   1   11     0  -13    0
 7   19   73   1   1  -5.75930   0.12539   13   1   0    0   -14    0    7
 7   19   73   1   0  -4.12552   0.10162   14   0   1  -14     0    7    0
 8   13   80   1   1  -5.15924   0.00199   15   1   0    0   -20    0   14
 8   13   80   1   0  -5.17126   0.00202   16   0   1  -20     0   14    0
 9   27   68   1   1  -5.55135   0.03462   17   1   0    0    -6    0    2
 9   27   68   1   0  -5.07961   0.02236   18   0   1   -6     0    2    0
10   42   61   3   1  -4.60458   0.05941   19   1   0    0     9    0   -5
10   42   61   3   0  -3.20337   0.01601   20   0   1    9     0   -5    0
11   18   74   3   1  -5.60295   0.00540   21   1   0    0   -15    0    8
11   18   74   3   0  -5.26210   0.00713   22   0   1  -15     0    8    0
12   33   69   3   1  -6.21180   0.20040   23   1   0    0     0    0    3
12   33   69   3   0  -6.65844   0.33376   24   0   1    0     0    3    0
13   33   76   3   1  -6.43840   0.03710   25   1   0    0     0    0   10
13   33   76   3   0  -6.42106   0.03454   26   0   1    0     0   10    0
```

#THE PROCEDURE STEP (BIVARIATE RANDOM EFFECTS ANALYSIS)

`Proc mixed cl method=ml data=BCGdata2 asycov;`	#call procedure; 'asycov' asks for asymptotic covariance matrix of covariance parameters
`class trial arm;`	#trial and arm are classification variables;
`model lno= exp con / noint s cl covb ddf=1000, 1000;`	#model with indicator variables 'exp' and 'con' as explanatory variables for log-odds; confidence intervals and p-values for coefficients of 'exp' and 'con' should be based on standard normal distribution (i.e. t-distribution with df $=\infty$). 'covb' asks for covariance matrix of fixed effects parameters.
`random exp con/ subject=trial type=un s;`	#experimental and control treatment are random effects, possibly correlated within a trial, and independent between trials; covariance matrix (Σ) is unstructured; print empirical Bayes estimates 's';
`repeated /group=arm;`	#each study-arm in each trial has its own within study-arm variance (matrix C_i); within study estimation errors are independent (default);
`estimate 'difference' exp 1 con -1/cl df=1000;`	#the 'estimate' command produces estimates of linear combinations of the fixed parameters with standard error computed from the covariance matrix of the estimates. Here we ask for the estimate of mean log-odds ratio;
`parms /parmsdata=covvars2 eqcons=4 to 29;`	#data file covvars2.sd2 contains the variable 'est' with starting values for the three covariance parameters of the random effects together with the 26 within study-arm variances. The latter are assumed to be known and should be kept fixed;
`run;`	

Running this program gives the following output:

```
                          The MIXED Procedure

(...)
                  Covariance Parameter Estimates (MLE)

     Cov Parm  Subject  Group       Estimate  Alpha     Lower     Upper
     UN(1,1)   TRIAL              1.43137384   0.05    0.7369    3.8894
     UN(2,1)   TRIAL              1.75732532   0.05    0.3378    3.1768
     UN(2,2)   TRIAL              2.40732608   0.05    1.2486    6.4330
(...)

                       Solution for Fixed Effects

  Effect      Estimate     Std Error    DF       t  Pr > |t|  Alpha     Lower    Upper
 EXP       -4.83374538    0.33961722  1000  -14.23    0.0001   0.05   -5.5002   -4.1673
 CON       -4.09597366    0.43469692  1000   -9.42    0.0001   0.05   -4.9490   -3.2430



            Covariance Matrix for Fixed Effects

                 Effect   Row          COL1          COL2
                 EXP        1    0.11533985    0.13599767
                 CON        2    0.13599767    0.18896142

(...)
                       ESTIMATE Statement Results

Parameter    Estimate    Std Error    DF    t     Pr > |t|  Alpha   Lower    Upper
difference  -0.73777172  0.17973848  1000  -4.10  0.0001    0.05  -1.0905   -0.3851
```

The fixed parameter estimates $\hat{\omega} = (\hat{\omega}_A, \hat{\omega}_B) = (-4.834, -4.096)$ represent the estimated mean log-odds in the vaccinated and non-vaccinated group, respectively. The between-trial estimated variance of the log-odds is $\hat{\Sigma}_{AA} = 1.431$ in the vaccinated groups and $\hat{\Sigma}_{BB} = 2.407$ in the not-vaccinated groups. The between-trial covariance is estimated to be $\hat{\Sigma}_{AB} = 1.757$. Thus, the estimated correlation between the true vaccinated and true control log-odds is $\hat{\Sigma}_{AB}/(\sqrt{\hat{\Sigma}_{AA}} \cdot \sqrt{\hat{\Sigma}_{BB}}) = 0.947$. The estimated covariance matrix for the ML-estimates $\hat{\omega}_B$ and $\hat{\omega}_A$ is

$$\begin{pmatrix} \mathrm{var}(\hat{\omega}_A) & \mathrm{cov}(\hat{\omega}_A, \hat{\omega}_B) \\ \mathrm{cov}(\hat{\omega}_B, \hat{\omega}_A) & \mathrm{var}(\hat{\omega}_B) \end{pmatrix} = \begin{pmatrix} 0.115 & 0.136 \\ 0.136 & 0.189 \end{pmatrix}$$

The estimated mean vaccination effect, measured as the log-odds ratio, is equal to $(\hat{\omega}_A - \hat{\omega}_B) = (-4.834 - (-4.096)) = -0.738$. The standard error of the mean vaccination effect is equal to $\sqrt{\{\mathrm{var}(\hat{\omega}_A) + \mathrm{var}(\hat{\omega}_B) - 2\,\mathrm{cov}(\hat{\omega}_A, \hat{\omega}_B)\}} = \sqrt{(0.115 + 0.189 - 2 \cdot 0.136)} = 0.180$, almost identical to the result of the univariate mixed model. This corresponds to an estimated odds ratio of $\exp(-0.738) = 0.478$ with a 95 per cent confidence interval equal to $(0.336; 0.680)$, again strongly suggesting an average beneficial vaccination effect. The slope of the regression line to predict the log-odds in the vaccinated group from the log-odds in the not-vaccinated

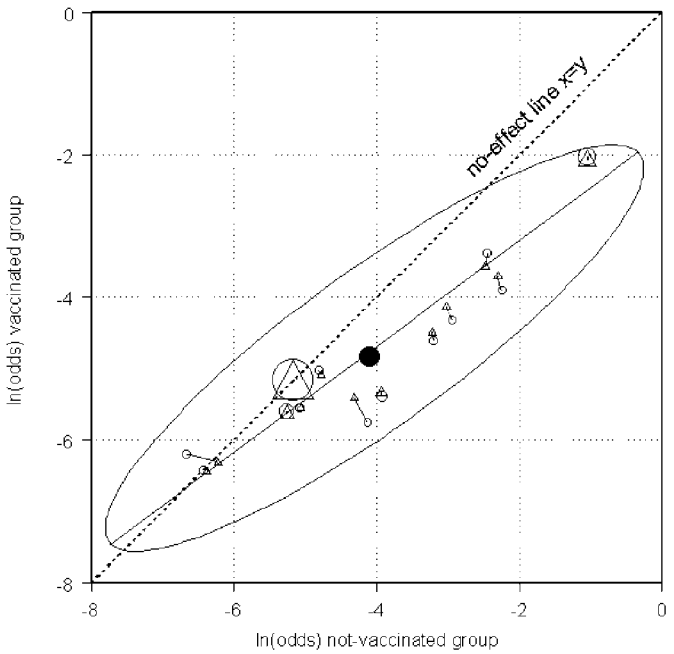

Figure 5. The 95 per cent coverage region for the pairs of true log-odds under vaccination and non-vaccination. The diagonal line is the line of equality between the two log-odds. Observed data from the trials are indicated with ○, the empirical Bayes estimates are with △. The common mean is indicated with the • central in the plot. The ellipse is obtained from a line plot based on the equation $(x-\hat{\omega})\hat{\Sigma}^{-1}(x-\hat{\omega})' = 5.99$.

group is equal to $\beta_{AB} = \hat{\Sigma}_{AB}/\hat{\Sigma}_{BB} = (1.757/2.407) = 0.730$. The slope of the reverse relationship is equal to $\beta_{BA} = \hat{\Sigma}_{AB}/\hat{\Sigma}_{AA} = (1.757/1.431) = 1.228$. The variance of the treatment effect, measured as the log-odds ratio, calculated from $\hat{\Sigma}$ is $(1.431 + 2.407 - 2 \cdot 1.757) = 0.324$, which is only slightly different from what we found earlier in the univariate random effects analysis. The conditional variance of the true log-odds, and therefore also of the log-odds ratio, in the vaccinated group given the true log-odds in the not-vaccinated group is $(\Sigma_{AA} - \Sigma_{AB}^2/\Sigma_{BB}) = (1.431 - 1.757^2/2.407) = 0.149$, which is interpreted as the variance between treatment effects among trials with the same baseline risk. So baseline risk, measured as the true log-odds in the not-vaccinated group, explains $(0.324 - 0.149)/0.324 = 54$ per cent of the heterogeneity in vaccination effect between the trials. The 95 per cent coverage region of the estimated bivariate distribution can be plotted in the so-called l'Abbé plot [21] in Figure 5.

Figure 5 nicely shows that the vaccination effect depends on the baseline risk (log-odds in not-vaccinated group) and that the heterogeneity in the difference between the log-odds in the vaccinated versus the not-vaccinated treatment arms is for a large part explained by the regression coefficient being substantially smaller than 1. It also shows the shrinkage of the empirical Bayes estimates towards the main axis of the ellipse. In this example we specified the model in Proc Mixed as a model with two random intercepts in which the fixed parameters

correspond to ω_A and ω_B. An alternative would be to specify the model as a random-intercept random-slope model, in which the fixed parameters correspond to ω_B and the mean treatment effect $\omega_A - \omega_B$. Then the SAS commands should be modified as follows:

```
model lno=treat/s cl covb ddf=1000;
random int treat/subject=trial type=un s;
```

Here `int` refers to a random trial specific intercept.

4.2. *Relation between effect and baseline risk*

The relation between treatment effect and baseline risk has been very much discussed in the literature [6–9, 23–30]. There are two issues that complicate the matter:

1. The relation between 'observed difference A – B' and 'observed baseline risk B' is prone to spurious correlation, since the measurement error in the latter is negatively correlated with measurement error in the first. It would be better to study B versus A or B – A versus (A+B)/2.
2. Even in the regression of 'observed risk in group A' on 'observed baseline risk in group B', which is not hampered by correlated measurement errors, the estimated slope is attenuated due to measurement error in the observed baseline risk [31].

For an extensive discussion of these problems see the article of Sharp *et al.* [32].
In dealing with measurement error there are two approaches [31, 33]:

(i) The *functional equation* approach: true regressors as nuisance parameters.
(ii) The *structural equation* approach: true regressors as random quantities with an unknown distribution.

The usual likelihood theory is not guaranteed to work for the functional equation approach because of the large number of nuisance parameters. The estimators may be inconsistent or have the wrong standard errors. The bivariate mixed model approach to meta-analysis used in this paper is in the spirit of the structural approach. The likelihood method does work for the structural equation approach, so in this respect our approach is safe. Of course, the question of robustness of the results against misspecification of the mixing distribution is raised. However, Verbeke and Lesaffre [34] have shown that, in the general linear mixed model, the fixed effect parameters as well as the covariance parameters are still consistently estimated when the distribution of the random effects is misspecified, so long as the covariance structure is correct. Thus our approach yields (asymptotically) unbiased estimates of slope and intercept of the regression line even if the normal distribution assumption is not fulfilled, although the standard errors might be wrong. Verbeke and Lesaffre [34] give a general method for robust estimation of the standard errors.

The mix of many fixed and a few random effects as proposed by Thompson *et al.* [8] and the models of Walter [9] and Cook and Walter [29] are more in the spirit of the functional approach. These methods are meant to impose no conditions on the distribution of the true baseline risks. The method of Walter [9] was criticized by Bernsen *et al.* [35]. Sharp and Thompson [30] use other arguments to show that Walter's method is seriously flawed. In a letter to the editor by Van Houwelingen and Senn [36] following the article of Thompson

et al. [8], Van Houwelingen and Senn [36] argue that putting Bayesian priors on all nuisance parameters, as done by Thompson *et al*., does not help solving the inconsistency problem. This view is also supported in the chapter on Bayesian methods in the book of Carroll *et al.* [31]. It would be interesting to apply the ideas of Caroll *et al.* [31] in the setting of meta-analysis, but that is beyond the scope of this paper. Arends *et al.* [10] compare, in a number of examples, the approach of Thompson *et al.* [8] with the method presented here and the results were in line with the remarks of Van Houwelingen and Senn [36]. Sharp and Thompson [30], comparing the different approaches in a number of examples, remark that whether or not to assume a distribution for the true baseline risks remains a debatable issue.

Arends *et al.* [10] also compared the approximate likelihood method as presented here with an exact likelihood approach where the parameters are estimated in a Bayesian manner with vague priors and found no relevant differences.

5. META-REGRESSION

In case of substantial heterogeneity between the studies, it is the statistician's duty to explore possible causes of the heterogeneity [15, 37–39]. In the context of meta-analysis that can be done by covariates on the study level that could 'explain' the differences between the studies. The term meta-regression to describe such analysis goes back to papers by Bashore *et al.* [40], Jones [41], Greenland [42] and Berlin and Antman [37]. We consider only analyses at the aggregated meta-analytic level. Aggregated information (mean age, percentage males) can describe the differences between studies. We will not go into covariates on the individual level. If such information exists, the data should be analysed on the individual patient level by hierarchical models. That is possible and a sensible thing to do, but beyond the scope of this paper. We will also not consider covariates on the study arm level. That can be relevant in non-balanced observational studies. Such covariates could both correct the treatment effect itself in case of confounding as well as explain existing heterogeneity between studies. Although the methods presented in this paper might be applied straightforwardly, we will restrict attention to balanced studies in which no systematic difference between the study arms is expected.

Since the number of studies in a meta-analysis is usually quite small, there is a great danger of overfitting. The rule of thumb of one explanatory variable for each 5 (10) 'cases' leaves only room for a few explanatory variables in a meta-regression. In the example we have three covariates available: latitude; year of study, and method of treatment allocation. Details are given in Table I.

In the previous section we have seen that heterogeneity between studies can be partly explained by differences in baseline risk. Thus, it is also important to investigate whether covariates on the study level are associated with the baseline risk. That asks for a truly multivariate regression with a two-dimensional outcome, but we will start with the simpler regression for the one-dimensional treatment effect difference measure.

5.1. Regression for difference measure

Let X_i stand for the (row)vector of covariates of study i including the constant term. Meta-regression relates the true difference ϑ_i to the 'predictor' $X_i\beta$. This relation cannot be expected to be perfect; there might be some residual heterogeneity that could be modelled by a normal

distribution once again, that is $\vartheta_i \sim \mathrm{N}(X_i\beta, \sigma^2)$. Taking into account the imprecision of the observed difference measure $\hat{\vartheta}_i$ we get the marginal approximate model

$$\hat{\vartheta}_i \sim \mathrm{N}(X_i\beta, \sigma^2 + s_i^2)$$

This model could be fitted by iteratively reweighted least squares, where a new estimate of σ^2 is used in each iteration step or by full maximum likelihood with appropriate software. In the following we will describe how the model can be fitted by SAS.

5.2. *Example (continued)*

A graphical presentation of the data is given in Figure 6. Latitude and year of publication both seem to be associated with the log-odds ratio, while latitude and year are also correlated. Furthermore, at first sight, the three forms of allocation seem to have little different average treatment effects.

5.2.1. Regression on latitude. The regression analysis for the log-odds ratio on latitude can be carried out by running the following mixed model in SAS:

```
Proc mixed cl method=ml                  #call procedure;
data=BCG_data;
class trial;                             #trial is classification variable;
model ln_or=latitude / s cl covb;        #latitude is only predictor variable;
random int/ subject=trial s;             #random trial effect;
repeated /group=trial;                   #each trial has its own within study
                                         variances;
parms /parmsdata=covvars3 eqcons=2 to    #data set covvars3 contains a starting value
14;                                      for between study variance and 13 within
                                         study variances which should be kept fixed;
run;
```

Running this program gives the following output:

```
                    The MIXED Procedure

(...)
                Covariance Parameter Estimates (MLE)

    Cov Parm    Subject  Group         Estimate  Alpha     Lower     Upper
    INTERCEPT   TRIAL                0.00399452   0.05    0.0004  1.616E29

(...)

                    Solution for Fixed Effects

Effect          Estimate     Std Error    DF       t  Pr > |t|  Alpha     Lower  Upper
INTERCEPT     0.37108745    0.10596655    11    3.50    0.0050   0.05    0.1379 0.6043
LATITUDE     -0.03272329    0.00337134     0   -9.71         .   0.05         .      .

                Covariance Matrix for Fixed Effects

                Effect      Row          COL1          COL2
                INTERCEPT     1    0.01122891   -0.00031190
                LATITUDE      2   -0.00031190    0.00001137
```


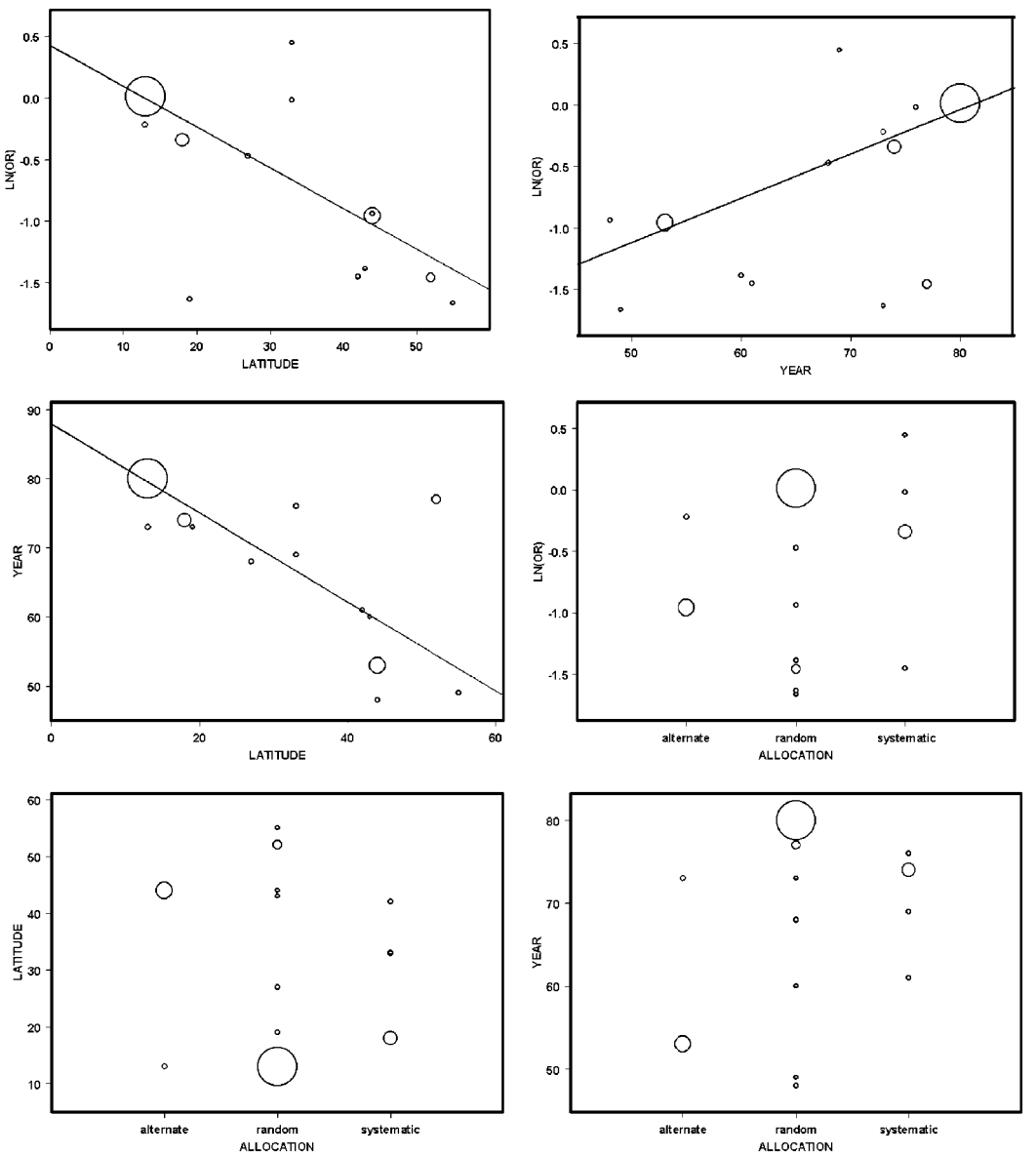

Figure 6. Graphical relationships between the variables with a weighted least squares regression line. The size of the circle corresponds to the inverse variance of the log-odds ratio in that trial.

The residual between-study variance in this analysis turns out to be 0.004, which is dramatically smaller than the between-study variance of 0.302 in the random effect model above without the covariate latitude in the model. Thus latitude explains 98.7 per cent of the between-trial variance in treatment effects differences. The regression coefficients for the intercept and for latitude are 0.371 (standard error = 0.106) and −0.033 (standard error = 0.003), respectively.

The estimated correlation between these estimated regression coefficients is -0.873.

Just for comparison we give the results of an ordinary weighted linear regression. The weights are equal to the inverse squared standard error of the log-odds ratio, instead of the correct weights equal to the inverse squared standard error of the log-odds ratio plus $\hat{\sigma}^2$. The intercept was 0.395 (SE $= 0.124$) and the slope -0.033 (SE $= 0.004$). The results are only slightly different, which is explained by the very small residual between-study variance.

5.2.2. Regression on year. Running the same model as above, only changing latitude into year, the residual between-study variance becomes 0.209. Thus year of publication explains 30.8 per cent of the between-trial variance in treatment effects differences, much less than the variance explained by the covariate latitude. The regression coefficients for the intercept and for year are -2.800 (standard error $= 1.031$) and 0.030 (standard error $= 0.015$), respectively. The estimated correlation between these estimated regression coefficients is -0.989.

Again, just for comparison, we also give the results of the ordinary weighted linear regression. The intercept was -2.842 (SE $= 0.876$) and the slope 0.033 (SE $= 0.012$). Like in the previous example, the differences are relatively small.

5.2.3. Regression on allocation. Running the model with allocation as only (categorical) covariate (in the SAS commands, specify: `class trial alloc;`), gives a residual between-study variance equal to 0.281. This means that only 7 per cent of the between-trial variance in the treatment effect differences is explained by the different forms of allocation. The treatment effects (log-odds ratio) do not differ significantly between the trials with random, alternate and systematic allocation ($p = 0.396$).

5.2.4. Regression on latitude and year. When both covariates latitude and year are put into the model, the residual between-study variance becomes only 0.002, corresponding with an explained variance of 99.3 per cent, only slightly more than by latitude alone. The regression coefficients for the intercept, latitude and year are, respectively, 0.494 (standard error $= 0.529$), -0.034 (standard error $= 0.004$) and -0.001 (standard error $= 0.006$).

We conclude that latitude gives the best explanation of the differences in vaccination effect between the trials, since it already explains 98 per cent of the variation. Since the residual variance is so small, the regression equation in this example could have been obtained by ordinary weighted linear regression under the assumption of homogeneity. In the original medical report [4] on this meta-analysis the authors mentioned the strong relationship between treatment effect and latitude as well. They speculated that the biological explanation might be the presence of non-tuberculous myobacteria in the population, which is associated with geographical latitude.

Goodness-of-fit of the model obtained above can be checked as in the weighted least squares approach by individual standardization of the residuals $(\hat{\vartheta}_i - X_i\hat{\beta})/\sqrt{(\sigma^2 + s_i^2)}$ and using standard goodness-of-fit checks.

In interpreting the results of meta-regression analysis, it should be kept in mind that this is all completely observational. Clinical judgement is essential for correct understanding of what is going on. Baseline risk may be an important confounder and we will study its effect below.

5.3. Bivariate regression

The basis of the model is the relation between the pair $(\omega_{A,i}, \omega_{B,i})$, for example, (true log-odds in vaccinated group, true log-odds in control group) and the covariate vector X_i. Since the covariate has influence on both components we have a truly multivariate regression problem in the classical sense that can be modelled as

$$\begin{pmatrix} \omega_{A,i} \\ \omega_{B,i} \end{pmatrix} \sim \mathrm{N}(BX_i, \Sigma)$$

Here, the matrix B is a matrix of regression coefficients: the first row for the A-component and the second row for the B-component. Taking into account the errors in the estimates we get the (approximate) model

$$\begin{pmatrix} \hat{\omega}_{A,i} \\ \hat{\omega}_{B,i} \end{pmatrix} \sim \mathrm{N}(BX_i, \Sigma + C_i)$$

Fitting this model to the data can again be done by a self-made program using the EM algorithm or by programs such as SAS Proc Mixed. The hardest part is the interpretation of the model. We will discuss the interpretation for the example.

So far we have shown for our leading example the univariate fixed effects model, the univariate random effect without covariates, the bivariate random effects model without covariates and eventually the univariate random effects model with covariates. We end this section with a bivariate random effects model with covariates.

5.4. Example (continued): bivariate meta-analysis with covariates

To carry out the bivariate regression analyses in SAS Proc Mixed we again need the data set `BCGdata2.sd2` which was organized on treatment arm level. In this example we take latitude as the covariate. The model can be fitted using the SAS code given below, where the variables `exp, con` and `arm` have the same meaning as in the bivariate analysis above without covariates. The variable `latcon` is for the not-vaccinated (control) groups equal to the latitude value of the trial and zero for the vaccinated (experimental) groups. The variable `latexp,` is defined analogously with vaccinated and non-vaccinated reversed.

```
Proc mixed cl method=ml data=BCGdata2;          # call procedure;
class trial arm;                                # trial and treatment arm are
                                                  defined as classification variables;
model lno= con exp latcon latexp/noint s cl     # model with indicator variables
ddf=1000,1000,1000,1000;                          'exp' and 'con' together with
                                                  latitude as explanatory variable
                                                  for log-odds in both treatment groups;
random con exp / subject=trial type=fa0(2) s ;  # control arm and experimental
                                                  trial arm are specified as random
                                                  effects; covariance matrix is
                                                  unstructured, parameterized as
                                                  factor analytic;
repeated /group=arm;                            #each study-arm in each trial has
                                                  its own within study-arm error variance;
```



```
parms /parmsdata=covvars4 eqcons=4 to 29;        #in the data file covvars4 three
                                                  starting values are given for the
                                                  between study covariance matrix,
                                                  together with the 26 within
                                                  study-arm variances. The latter
                                                  are assumed to be known and kept fixed;
estimate 'difference slopes' latexp 1 latcon -1  #estimate of the difference in
/cl df=1000;                                      slope between the vaccinated
                                                  and not-vaccinated groups;
run;
```

Remark

In the program above we specified `type=fa0(2)` instead of `type=un` for Σ. If one chooses the latter, the covariance matrix is parameterized as

$$\begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_3 \end{bmatrix}$$

and unfortunately the program does not converge if the estimated correlation is (very near to) 1, as is the case here. If one chooses the former, the covariance matrix is parameterized as

$$\begin{bmatrix} \alpha_{11}^2 & \alpha_{11}\alpha_{12} \\ \alpha_{11}\alpha_{12} & \alpha_{12}^2 + \alpha_{22}^2 \end{bmatrix}$$

and the program converges even if the estimated correlation is 1, that is, if $\alpha_{22} = 0$.

Running the program gives the following output:

```
                          The MIXED Procedure
(...)
                  Covariance Parameter Estimates (MLE)

     Cov Parm  Subject  Group       Estimate  Alpha     Lower     Upper
     FA(1,1)   TRIAL              1.08715174   0.05    0.7582    1.6896
     FA(2,1)   TRIAL              1.10733154   0.05    0.6681    1.5466
     FA(2,2)   TRIAL             -0.00000000      .         .         .

(...)
                       Solution for Fixed Effects

  Effect      Estimate     Std Error    DF       t  Pr > |t|  Alpha     Lower   Upper
  CON      -4.11736845    0.30605608  1000  -13.45    0.0001   0.05   -4.7180  -3.5168
  EXP      -4.82570990    0.31287126  1000  -15.42    0.0001   0.05   -5.4397  -4.2118
  LATCON    0.07246261    0.02192060  1000    3.31    0.0010   0.05    0.0294   0.1155
  LATEXP    0.03913388    0.02239960  1000    1.75    0.0809   0.05   -0.0048   0.0831

                       ESTIMATE Statement Results

Parameter          Estimate     Std Error    DF     t  Pr > |t|  Alpha   Lower   Upper

difference slopes -0.03332874  0.00284902  1000 -11.70  0.0001   0.05  -0.0389 -0.0277
```

In Figure 7 the relationship between latitude and the log-odds of tuberculosis is presented for the vaccinated treatment arms A as well as for the non-vaccinated treatment arms B. For

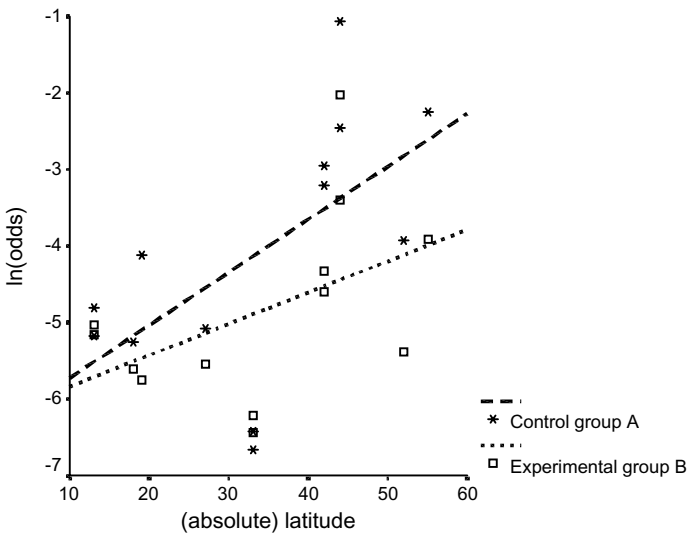

Figure 7. Log-odds versus latitude for control group A and experimental group B.

the not-vaccinated trial arms the regression line is log(odds) = $-0.4117 + 0.072$ (latitude $- 33) = -6.509 + 0.072$ latitude (standard errors of intercept and slope are 0.794 and 0.022, respectively). Notice that latitude was centralized at latitude = 33 (see Section 4.1). For the vaccinated trial arms the regression line is log(odds) = -0.483 (latitude $- 33) = -6.117 +$ 0.039 latitude (standard errors of intercept and slope are 0.809 and 0.022, respectively). We see that latitude has a strong effect, especially on the log-odds of the non-vaccinated study group.

The between-study covariance matrix $\hat{\Sigma}$ is equal to the nearly singular matrix

$$\begin{bmatrix} 1.1819 & 1.2038 \\ 1.2038 & 1.2262 \end{bmatrix}$$

The estimated regression line of the treatment difference measure on latitude is log-odds $\text{ratio}_{\text{A vs B}} = 0.392 - 0.033$ latitude, with standard errors 0.093 and 0.003 for intercept and slope, respectively. This regression line is almost identical to the one resulting from the univariate analysis in the previous example. The estimated residual between-study variance is only 0.0003, meaning that latitude explains almost all heterogeneity in the treatment effects.

The regression line of the difference measure on both latitude and baseline risk is: log-odds $\text{ratio}_{\text{A vs B}} = 0.512 - 0.039$ latitude $+ 0.019$ log-odds_{B}. The standard errors can be calculated by the delta method. We see that the regression coefficient of the baseline log-odds is quite small compared to the analysis without any covariates.

The results of this bivariate regression and the results of the simple bivariate model without covariates of Section 4 are summarized in Table II. By explaining variation in treatment effects by latitude, hardly any residual variation is left. Although this is all observational, we

Table II. Residual variance of treatment effect in different meta-regression models.

Explanatory variables in the model	Residual variance of treatment effect
No covariates	0.324
Baseline	0.149
Latitude	0.0003
Baseline + latitude	0.0001

come to the tentative conclusion that the effect of vaccination depends on latitude rather than on baseline risk.

6. EXTENSIONS: EXACT LIKELIHOODS, NON-NORMAL MIXTURES, MULTIPLE ENDPOINTS

The approximate likelihood solutions may be suspect if the sample sizes per study are relatively small. There are different approaches to repair this and to make the likelihoods less approximate. We will first discuss the bivariate analysis where things are relatively easy and then the analysis of difference measures.

6.1. More precise analysis of bivariate data

Here, the outcome measures per study arm are direct maximum likelihood estimates of the relevant parameter. The estimated standard error is derived from the second derivative of the log-likelihood evaluated at the ML-estimate. Our approach is an approximation for fitting a generalized linear mixed model (GLMM) by the maximum likelihood method. The latter is hard to carry out. A popular approximation is by means of the second-order Laplace approximation or the equivalent PQL method [43], that is, based on an iterative scheme where the second derivative is evaluated at the posterior mode. This can easily be mimicked in the SAS procedure Proc Mixed by iteratively replacing the estimated standard error computed from the empirical Bayes estimate as yielded by the software. For the analysis of log-odds as in the example, one should realize that the variance of log-odds is derived from the second derivative of the log-likelihood evaluated at the ML-estimate of p, and is given by $1/(np(1-p))$. In the first iteration, p is estimated by the fraction of events in the study arm. In the next iteration p is replaced by the value derived from the empirical Bayes estimate for log-odds. This is not very hard to do and easy to implement in a SAS macro that iteratively uses Proc Mixed (see the example below; the macro version is available from the authors).

This will help for intermediate sample sizes and moderate random effect variances. There are, however, possible situations (small samples, large random effect variances) in which the second-order approximations do not work [44] and one has to be very careful in computing and maximizing the likelihoods. Fortunately, that is much more of a problem for random effects at the individual level than at the aggregated level we have here.

6.2. Example (continued)

After running the bivariate random effects model discussed in Section 4, the empirical Bayes estimates can be saved by adding the statement:

```
make 'Predicted' out = Pred;
```

in the Proc Mixed command and adding a 'p' after the slash in the model statement. In this way the empirical Bayes estimates for log-odds are stored as variable _PRED_ in the new SAS data file Pred.sd2. The within-trial variances in the next iteration of the SAS procedure Proc Mixed are derived from these empirical Bayes estimates in the way we described above. The three starting values needed for the between-trial variance matrix are stored as variable est in the SAS file covvars5.sd2.

Thus, after running the bivariate random effects model once and saving the empirical Bayes estimates for log-odds, one can run the two data steps described below to compute the new estimates for the within-trial variances, use these within-trial variances in the next bivariate mixed model, save the new empirical Bayes estimates and repeat the whole loop. This iterative process should be continued until the parameter estimates converge.

```
# DATA STEP TO COMBINE EMPIRICAL BAYES ESTIMATES AND ORIGINAL DATAFILE FROM SECTION 4 AND
TO CALCULATE THE NEW WITHIN-TRIAL VARIANCES;
data Pred1;
merge BCGdata2 Pred;
pi=exp(_PRED_)/(1+exp(_PRED_));
est=1/(n*pi*(1-pi));
run;

# DATA STEP TO CREATE THE TOTAL DATAFILE THAT IS NEEDED IN THE PARMS-STATEMENT (BETWEEN-
AND WITHIN-TRIAL VARIANCES);
data Pred2;
set covvars5 Pred1;
run;

# PROCEDURE STEP TO RUN THE BIVARIATE RANDOM EFFECTS MODEL WITH NEW WITHIN-TRIAL
VARIANCES, BASED ON THE EMPIRICAL BAYES ESTIMATES.
proc mixed cl method=ml data=BCGdata2 asycov;
class trial arm;
model lno= exp con / p noint s cl covb ddf=1000, 1000;
random exp con/ subject=trial type=un s;
repeated /group=arm subject=arm;
estimate 'difference' exp 1 con -1 / cl df=1000;
parms / parmsdata=Pred2 eqcons=4 to 29;
run;
```

Running the data steps and the mixed model iteratively until convergence is reached gives the following output:

```
                              The MIXED Procedure
(...)
                    Covariance Parameter Estimates (MLE)

Cov Parm  Subject  Group       Estimate  Alpha     Lower     Upper
UN(1,1)   TRIAL              1.43655989   0.05    0.7392    3.9084
UN(2,1)   TRIAL              1.76956270   0.05    0.3395    3.1996
UN(2,2)   TRIAL              2.43849037   0.05    1.2663    6.4991
```


Solution for Fixed Effects

Effect	Estimate	Std Error	DF	t	Pr > |t|	Alpha	Lower	Upper
EXP	-4.84981269	0.34001654	1000	-14.26	0.0001	0.05	-5.5170	-4.1826
CON	-4.10942999	0.43736103	1000	-9.40	0.0001	0.05	-4.9677	-3.2512

Covariance Matrix for Fixed Effects

Effect	Row	COL1	COL2
EXP	1	0.11561125	0.13690215
CON	2	0.13690215	0.19128467

ESTIMATE Statement Results

Parameter	Estimate	Std Error	DF	t	Pr > |t|	Alpha	Lower	Upper
difference	-0.74038270	0.18191102	1000	-4.07	0.0001	0.05	-1.0974	-0.3834

The mean outcome measures (log-odds) for arms A and B are, respectively, -4.850 (standard error $= 0.340$) and 4.109 (standard error $= 0.437$). The between-trial variance of the log-odds in the vaccinated treatment arm A is $\hat{\Sigma}_{AA} = 1.437$ and $\hat{\Sigma}_{BB} = 2.438$ in the not-vaccinated arm B. The estimate of the between-trial covariance is equal to $\hat{\Sigma}_{AB} = 1.770$. The estimated mean vaccination effect in terms of the log-odds ratio is -0.740 (standard error $=$ 0.182). In this example, convergence was already reached after one or two iterations. The final estimates are very similar to the original bivariate random effects analysis we have discussed in Section 4, where the mean outcome measures $\hat{\omega}_A$ and $\hat{\omega}_B$ were, respectively, -4.834 (SE $= 0.340$) and -4.096 (SE $= 0.434$). Of course, when the number of patients in the trials were smaller, the benefit and necessity of this method would be more substantial.

Another possibility if the approximate likelihood solutions are suspect is to use the exact likelihood, based on the binomial distribution of the number of events per treatment arm, and to estimate the parameters following a Bayesian approach with vague priors in combination with Markov chain Monte Carlo (MCMC) methods [45]. Arends *et al.* [10] give examples of this approach. In their examples the difference with the approximate likelihood estimates turned out to be very small.

6.3. *More precise analysis of difference measures*

The analysis of difference measures, that is, one summary measure per trial characterizing the difference in efficacy between treatments, is a bit more complicated because the baseline value is considered to be a nuisance parameter. Having this nuisance parameter can be avoided and a lot of 'exactness' in the analysis can be gained by suitable conditioning on ancillary statistics. In the case of binary outcomes one can condition on the marginals of the 2×2 tables and end up with the non-central hypergeometric distribution that only depends on the log-odds ratio. Details are given in Van Houwelingen *et al.* [1].

However, the hypergeometric distribution is far from easy to handle and it does not seem very attractive to try to incorporate covariates in such an analysis as well. The bivariate analysis is much easier to carry out at the price of the assumption that the baseline parameter follows a normal distribution. However, that assumption can be relaxed as well and brings us to the next extension: the non-normal mixture.

6.4. Non-normal mixtures

The assumption of a normal distribution for the random effects might not be realistic. Technically speaking it is not very hard to replace the normal mixture by a fully non-parametric mixture. As is shown by Laird [46], the non-parametric maximum likelihood estimator of the mixing distribution is always a discrete mixture and can easily be estimated by means of the EM algorithm [47]. An alternative is to use the software C.A.MAN of Böhning *et al.* [48]. However, just fitting a completely non-parametric mixture is no good way of checking the plausibility of the normal mixture. The non-parametric estimates are always very discrete even if the true mixture is normal. A better way is to see whether a mixture of two normals (with the same variance) fits better than a single normal. This model can describe a very broad class of distributions: unimodal as well as bimodal, symmetric as well as very skewed [19]. Another way is to estimate the skewness of the mixture somehow and mistrust the normality if the skewness is too big. It should be realized, however, that estimating mixtures is a kind of ill-posed problem and reliable estimates are hard to obtain [49]. To give an impression we fitted a non-parametric mixture with the homemade program based on the EM algorithm described in Van Houwelingen *et al.* [1] to the log-odds ratio of our example using approximate likelihoods. The results were as follows:

atom	probability
−1.4577	0.3552
−0.9678	0.1505
−0.3296	0.2980
0.0023	0.1963

corresponding mean:	−0.761
corresponding variance:	0.349

The first two moments agree quite well with the normal mixture. It is very hard to tell whether this four-point mixture gives any evidence against normality of the mixture. The bivariate normal mixture of Section 4 is even harder to check. Non-parametric mixtures are hard to fit in two dimensions. An interesting question is whether the estimated regression slopes are robust against non-normality. Arends *et al.* [10] modelled the baseline distribution with a mixture of two normal distributions and found in all their examples a negligible difference with modelling the baseline parameter with one normal distribution, indicating that the method is robust indeed [10]. However, this was only based on three examples and we do not exclude the possibility that in some other data examples the regression slopes might be more different.

6.5. Multiple outcomes

In a recent paper Berkey *et al.* [50] discussed a meta-analysis with multiple outcomes. A similar model was used in the context of meta-analysis of surrogate markers by Daniels and Hughes [51] and discussed by Gail *et al.* [52]. In the simplest case of treatment difference measures for several outcomes, the situation is very similar to the bivariate analysis of Sections 4 and 5. The model

$$\begin{pmatrix} \omega_{\mathrm{A},i} \\ \omega_{\mathrm{B},i} \end{pmatrix} \sim \mathrm{N}(BX_i, \Sigma)$$

Table III.

Trial	Publication year	$\hat{\omega}_{PD,i}$	$\hat{\omega}_{AL,i}$	$\text{var}(\hat{\omega}_{PD,i})$	$\text{var}(\hat{\omega}_{AL,i})$	$\text{cov}\,ar(\hat{\omega}_{PD,i}, \hat{\omega}_{AL,i})$
1	1983	0.47	−0.32	0.0075	0.0077	0.0030
2	1982	0.20	−0.60	0.0057	0.0008	0.0009
3	1979	0.40	−0.12	0.0021	0.0014	0.0007
4	1987	0.26	−0.31	0.0029	0.0015	0.0009
5	1988	0.56	−0.39	0.0148	0.0304	0.0072

could be used, where ω_A stands for the (difference) measure on outcome A and ω_B for the measure on outcome B. It could easily be generalized to more measures C, D, etc. The main difference is that the estimated effects are now obtained in the same sample and, therefore, will be correlated. An estimate of this correlation is needed to perform the analysis. The only thing that changes in comparison with Section 5 is that the matrix C_i in

$$\begin{pmatrix} \hat{\omega}_{A,i} \\ \hat{\omega}_{B,i} \end{pmatrix} \sim N(BX_i, \Sigma + C_i)$$

is not diagonal anymore but allows within-trial covariation.

This approach can easily be adapted to the situation where there are more than two outcome variables or more treatment groups.

6.6. *Example Berkey et al. [50]*

Berkey *et al.* [50] illustrate several fixed and random (multivariate) meta-regression models using a meta-analysis from Antczak-Bouckoms *et al.* [53]. This meta-analysis concerns five randomized controlled trials, where a surgical procedure is compared with a non-surgical procedure. Per patient two outcomes are assessed: (pre- and post-treatment change in) probing depth (PD) and (pre- and post-treatment change in) attachment level (AL). Since the efficacy of the surgical procedure may improve over time, a potential factor that may influence the trial results is the year of publication [50]. The two treatment effect measures are defined as:

$$\omega_{PD} = \text{mean PD under surgical treatment} - \text{mean PD under non-surgical treatment}$$

$$\omega_{AL} = \text{mean AL under surgical treatment} - \text{mean AL under non-surgical treatment}$$

The data are given in Table III.

As an example we fit the model with year of publication as explanatory variable. Berkey *et al.* [50] fitted this model using a self-written program in SAS Proc IML. We show how it can be done with SAS Proc Mixed. The data set-up is the same as in the earlier discussed bivariate models with two data rows per trial, one for each outcome measure. Also the Proc Mixed program is completely analogous. The only difference is that in the data set containing the elements of the C_i's now the covariance between the two outcomes per trial must be

specified as well. The SAS code is:

```
proc mixed cl method=ml data=berkey;          #call procedure;
class trial type;                             #trial and outcome type (PD or
model outcome=pd al pdyear                    AL) are classification variables;
alyear/noint s cl;                            #model with indicator variables
                                              'pd' and 'al' together with
                                              publication year as
                                              explanatory variable;
random pd al / subject=trial type=un s;       #specification of among-trial
                                              covariance matrix for
                                              both outcomes;
repeated type /subject=trial                  #specification of
group=trial type=un;                          (non-diagonal) within-trial
                                              covariance matrix;
parms /parmsdata=covvars6                     #covvars6 contains: 3 starting
eqcons=4 to 18;                               values for the two between-trial
                                              variances and covariance, 10
                                              within-trial variances (5 per
                                              outcome measure) and 5
                                              covariances. The last 15
                                              parameters are assumed to be
                                              known and must be kept fixed.
run;
```

Part of the SAS Proc Mixed output is given below.

```
                          The MIXED Procedure
(...)
                  Covariance Parameter Estimates (MLE)

     Cov Parm  Subject  Group        Estimate  Alpha     Lower     Upper

     UN(1,1)   TRIAL              0.00804054   0.05    0.0018    2.0771
     UN(2,1)   TRIAL              0.00934132   0.05   -0.0113    0.0300
     UN(2,2)   TRIAL              0.02501344   0.05    0.0092    0.1857

(...)
                       Solution for Fixed Effects

 Effect       Estimate     Std Error    DF       t  Pr > |t|  Alpha     Lower    Upper
 PD         0.34867848    0.05229098     3    6.67    0.0069   0.05    0.1823   0.5151
 AL        -0.34379097    0.07912671     3   -4.34    0.0225   0.05   -0.5956  -0.0920
 PDYEAR     0.00097466    0.01543690     0    0.06         .   0.05         .        .
 ALYEAR    -0.01082781    0.02432860     0   -0.45         .   0.05         .        .
```

The estimated model is

$$\omega_{\mathrm{PD}} = 0.34887 + 0.00097^{*}(\text{year-1984})$$

$$\omega_{\mathrm{AL}} = -0.34595 - 0.01082^{*}(\text{year-1984})$$

The standard errors of the slopes are 0.0154 and 0.0243 for PD and AL, respectively. The estimated among-trial covariance matrix is

$$\hat{\Sigma} = \begin{pmatrix} 0.008 & 0.009 \\ 0.009 & 0.025 \end{pmatrix}$$

The results are identical to those of Berkey *et al.* [50] with the random-effects multiple outcomes that were estimated with the method named by Berkey as the multivariate maximum likelihood (MML) method.

6.7. Other outcome measures

Our presentation concentrates on dichotomous outcomes. Much of it carries over to other effect measures that are measured on a different scale. For instance, our methods apply if the outcome variable is continuous and an estimate of the average outcome and its standard error is available in both treatment arms. However, in some cases only a relative effect is available, such as the standardized effect measure (difference in outcome/standard deviation of the measurements in the control group) which is popular in psychological studies. In that case only the one-dimensional analysis applies. A special case is survival analysis. The log hazard ratio in the Cox model cannot be written as the difference of two effect measures. However, some measure of baseline risk, for example, one-year survival rate in the control arm, might be defined and the bivariate outcome analysis described above can be used to explore the relation between treatment effect and baseline risk. A complicating factor is that the two measures are not independent any more. However, if an estimate of the correlation between the two measures is available, the method can be applied.

6.8. Other software

Although we illustrated all our examples with the SAS procedure Proc Mixed, most if not all analyses could be carried out by other (general) statistical packages as well. A nice review of available software for meta-analysis has recently been written by Sutton *et al.* [54]. Any package like SPSS, SAS, S-plus and Stata that can perform a weighted linear regression suffices to perform a standard fixed effect meta-analysis or a fixed effects meta-regression.

For fitting random effects models with approximate likelihood, a program for the general linear mixed model (GLMM) is needed, which is available in many statistical packages. However, not all GLMM programs are appropriate. One essential requirement of the program is that one can fix the within-trial variance in the model at arbitrary values per trial.

In S-plus the function *lme* is used to fit linear mixed effects models and all the analyses carried out with Proc Mixed of SAS in our examples can also be carried out with *lme* from S-plus. The 'parms' statement used by SAS to fix the within-trial variances corresponds with 'varFixed' in S-plus [55].

Several Stata macros have been written which implement some of the discussed methods [56, 57]. The Stata program *meta* of Sharp and Sterne [56] performs a standard fixed and random effects meta-analysis without covariates. The Stata command *metareg* of Sharp [57] extends this to univariate meta-regression. We are not aware of Stata programs that are capable of fitting bivariate meta-regression models, but of course one can do an univariate meta-regression on the log-odds ratios instead of a bivariate meta-regression on the log-odds of

the two treatment arms. However, such an analysis does not give any information about the relationship between the (true) log-odds of the two arms.

Mlwin or MLn appears to be one of the most flexible methods to fit mixed-effect regression models [54]. Although we do not have experience with this package, we assume that most if not all of the discussed models can be fitted in it.

Finally, in the free available Bayesian analysis software package BUGS, one can also execute all approximate likelihood analyses that have been presented in this article. If vague prior distributions are used, the results are very similar. With BUGS it is also possible to fit the models using the exact likelihood, based on the binomial distribution of the number of events in a treatment arm. The reader is referred to Arends *et al.* [10] for examples and the required BUGS syntax.

7. BAYESIAN STATISTICS IN META-ANALYSIS

As we mentioned in Section 4, putting uninformative Bayesian priors on all individual nuisance parameters as done in Thompson *et al.* [8], Daniels and Hughes [51], Smith *et al.* [58] and Sharp and Thompson [30] can lead to inconsistent results as the number of nuisance parameters grows with the number of studies [36]. This observation does not imply that we oppose Bayesian methods. First of all, there is a lot of Bayesian flavour to random effects meta-analysis. The mixing distribution can serve as a prior distribution in the analysis of the results of a new trial. However, the prior is estimated from the data and not obtained by educated subjective guesses, that is why random effects meta-analysis can be seen as an example of the empirical Bayes approach. For each study, the posterior distribution given the observed value can be used to obtain empirical Bayes corrections.

In this paper we describe estimating the mixing distribution by maximum likelihood. The maximum likelihood method has two drawbacks. First, in complex problems, maximizing the likelihood might become far from easy and quite time-consuming. Second, the construction of confidence intervals with the correct coverage probabilities can become problematic. We proposed the profile likelihood approach in the simple setting of Section 3. For more complex problems, the profile likelihood gets very hard to implement.

When the maximum likelihood approach gets out of control (very long computing times, non-convergence of the maximization procedure), it can be very profitable to switch to a Bayesian approach with vague priors on the parameters of the model in combination with Markov chain Monte Carlo (MCMC) methods [45] that circumvent integration by replacing it by simulation. If one wants to use the MCMC technique in this context, the prior should be set on all parameters of the hierarchical model. Such a model could be described as a Bayesian hierarchical or Bayesian empirical Bayes model. For examples of this approach, see Arends *et al.* [10]. The difference with the approach of Thompson *et al.* [8, 30] is then that they assume that the true baseline log-odds are a random sample of a fully specified flat normal distribution (for example, $N(0, 10)$), while we assume that the true log-odds are sampled from a $N(\theta, \sigma)$ distribution with θ and σ parameters to be estimated, putting vague priors on them. So Thompson *et al.*'s model is a special case of our model. We prefer the parameters of the baseline risks distribution to be determined by the data. For the examples discussed in this paper, maximum likelihood was quite convenient in estimating the parameters of the model and getting a rough impression of their precision. It sufficed for the global analysis described

here. If the model is used to predict outcomes of new studies, as in the surrogate marker setting of Daniels and Hughes [51], nominal coverage of the prediction intervals becomes important and approximate methods can be misleading. MCMC can be very convenient, because the prediction problem can easily be embedded in the MCMC computations. An alternative is bootstrapping as described in Gail *et al*. [52].

8. CONCLUSIONS

We have shown that the general linear mixed model using an approximate likelihood approach is a very useful and convenient framework to model meta-analysis data. It can be used for the simple meta-analysis up to complicated meta-analyses involving multivariate treatment effect measures and explanatory variables. Extension to multiple outcome variables and multiple treatment arms is very straightforward. Suitable software is widely available in statistical packages.

ACKNOWLEDGEMENTS

This paper is based on a course on meta-analysis given by Van Houwelingen and Stijnen at the ISCB in Dundee in 1998. We thank the organizers for their stimulating invitation. We thankfully acknowledge the help of Russ Wolfinger of SAS in finding the best Proc Mixed specifications for the bivariate models. We are very grateful to three reviewers for their detailed comments and valuable suggestions, which have led to considerable improvements of this paper.

REFERENCES

1. Van Houwelingen H, Zwinderman K, Stijnen T. A bivariate approach to meta-analysis. *Statistics in Medicine* 1993; **12**:2272–2284.
2. Berkey CS, Hoaglin DC, Mosteller F, Colditz GA. Random effects regression model for meta-analysis. *Statistics in Medicine* 1995; **14**:396–411.
3. Normand S-L. Meta-analysis: formulating, evaluating, combining and reporting. *Statistics in Medicine* 1999; **18**:321–359.
4. Colditz GA, Brewer FB, Berkey CS, Wilson EM, Burdick E, Fineberg HV, Mosteller F. Efficacy of BCG vaccine in the prevention of tuberculosis. *Journal of the American Medical Association* 1994; **271**:698–702.
5. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**(3):177–188.
6. Brand R, Kragt H. Importance of trends in the interpretation of an overall odds ratio in the meta-analysis of clinical trials. *Statistics in Medicine* 1992; **11**(16):2077–2082.
7. McIntosh MW. The population risk as an explanatory variable in research syntheses of clinical trials. *Statistics in Medicine* 1996; **15**:1713–1728.
8. Thompson SG, Smith TC, Sharp SJ. Investigating underlying risk as a source of heterogeneity in meta-analysis. *Statistics in Medicine* 1997; **16**(23):2741–2758.
9. Walter SD. Variation in baseline risk as an explanation of heterogeneity in meta-analysis. *Statistics in Medicine* 1997; **16**:2883–2900.
10. Arends LR, Hoes AW, Lubsen J, Grobbee DE, Stijnen T. Baseline risk as predictor of treatment benefit: three clinical meta-re-analyses. *Statistics in Medicine* 2000; **19**(24):3497–3518.
11. Fleiss JL. Analysis of data from multiclinic trials. *Controlled Clinical Trials* 1986; **7**:267–275.
12. Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomised trials. *Progress in Cardiovascular Diseases* 1985; **27**:335–371.
13. Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of clinical trials. *Statistics in Medicine* 1991; **10**:1665–1677.
14. Hardy RJ, Thompson SG. A likelihood approach to meta-analysis with random effects. *Statistics in Medicne* 1996; **15**:619–629.
15. Greenland S. Quantitative methods in the review of epidemiologic literature. *Epidemiologic Reviews* 1987; **9**: 1–30.

16. Newcombe RG. Improved confidence intervals for the difference between binomial proportions based on paired data. *Statistics in Medicine* 1998; **17**:2635–2650.
17. Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Statistics in Medicine* 1998; **17**:873–890.
18. Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. *Statistics in Medicine* 2001; **20**:825–840.
19. Verbeke G. Linear mixed models for longitudinal data. In *Linear Mixed Models in Practice*, Verbeke G, Molenberghs G (eds). Springer-Verlag: New York, 1997; 63–153.
20. Roger JH, Kenward MG. Repeated measures using proc mixed instead of proc glm. In *Proceedings of the First Annual South-East SAS Users Group Conference.* SAS Institute: Cary NC, 1993; 199–208.
21. L'Abbé KA, Detsky AS. Meta-analysis in clinical research. *Annals of Internal Medicine* 1987; **107**(2): 224–233.
22. Armitage P, Berry G. *Statistical Methods in Medical Research.* Blackwell Scientific Publications: Oxford, 1978.
23. Davey Smith G, Song F, Sheldon TA. Cholesterol lowering and mortality: the importance of considering initial level of risk. *British Medical Journal* 1993; **306**(6889):1367–1373.
24. Senn S. Importance of trends in the interpretation of an overall odds ratio in the meta-analysis of clinical trials (letter). *Statistics in Medicine* 1994; **13**(3):293–296.
25. Brand R. Importance of trends in the interpretation of an overall odds ratio in the meta-analysis of clinical trials (letter). *Statistics in Medicine* 1994; **13**(3):293–296.
26. Hoes AW, Grobbee DE, Lubsen J. Does drug treatment improve survival? Reconciling the trials in mild-to-moderate hypertension. *Journal of Hypertension* 1995; **13**(7):805–811.
27. Egger M, Smith GD. Risks and benefits of treating mild hypertension: a misleading meta-analysis? [comment]. *Journal of Hypertension* 1995; **13**(7):813–815.
28. Senn SJ. Relation between treatment benefit and underlying risk in meta-analysis. *British Medical Journal* 1996; **313**:1550.
29. Cook RJ, Walter SD. A logistic model for trend in $2 \times 2 \times K$ tables with applications to meta-analyses. *Biometrics* 1997; **53**(1):352–357.
30. Sharp SJ, Thompson SG. Analysing the relationship between treatment effect and underlying risk in meta-analysis: comparison and development of approaches. *Statistics in Medicine* 2000; **19**:3251–3274.
31. Carroll RJ, Ruppert D, Stefanski LA. *Measurement Error in Nonlinear Models.* Chapman & Hall: London, 1995.
32. Sharp SJ, Thompson SG, Altman DG. The relation between treatment benefit and underlying risk in meta-analysis. *British Medical Journal* 1996; **313**(7059):735–738.
33. Kendall MG, Stuart A. *The Advanced Theory of Statistics. Volume II*: *Inference and Relationship.* Griffin: London, 1973.
34. Verbeke G, Lesaffre E. The effect of misspecifying the random effects distribution in linear models for longitudinal data. *Computational Statistics and Data Analysis* 1997; **23**:541–556.
35. Bernsen RMD, Tasche MJA, Nagelkerke NJD. Some notes on baseline risk and heterogeneity in meta-analysis. *Statistics in Medicine* 1999; **18**(2):233–238.
36. Van Houwelingen HC, Senn S. Investigating underlying risk as a source of heterogeneity in meta-analysis (letter). *Statistics in Medicine* 1999; **18**:107–113.
37. Berlin JA, Antman EM. Advantages and limitations of meta-analytic regressions of clinical trials data. *Online Journal of Current Clinical Trials* 1994; **134**.
38. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *British Medical Journal* 1994; **309**(6965):1351–1355.
39. Thompson SG, Sharp SJ. Explaining hetergoneity in meta-anaysis: a comparison of methods. *Statistics in Medicine* 1999; **18**:2693–2708.
40. Bashore TR, Osman A, Heffley EF. Mental slowing in elderly persons: a cognitive psychophysiological analysis. *Psychology & Aging* 1989; **4**(2):235–244.
41. Jones DR. Meta-analysis of observational epidemiological studies: a review. *Journal of the Royal Society of Medicine* 1992; **85**(3):165–168.
42. Greenland S. A critical look at some popular meta-analytic methods. *American Journal of Epidemiology* 1994; **140**(3):290–296.
43. Platt RW, Leroux BG, Breslow N. Generalized linear mixed models for meta-analysis. *Statistics in Medicine* 1999; **18**:643–654.
44. Engel B. A simple illustration of the failure of PQL, IRREML and APHL as approximate ML methods for mixed models for binary data. *Biometrical Journal* 1998; **40**:141–154.
45. Gilks WR, Richardson S, Spiegelhalter DJ. *Markov Chain Monte Carlo in Practice.* Chapman & Hall: London, 1996.
46. Laird NM. Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association* 1978; **73**:805–811.

47. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society*, *Series B* 1977; **39**(1):1–38.
48. Böhning D, Schlattman P, Linsay B. Computer-assisted analysis of mixtures. *Biometrics* 1992; **48**:283–304.
49. Eilers PHC, Marx BD. Flexible smooting using B-splines and penalized likelihood. *Statistical Science* 1996; **11**:89–121.
50. Berkey CS, Hoaglin DC, Antczak-Bouckoms A, Mosteller F, Colditz GA. Meta-analysis of multiple outcomes by regression with random effects. *Statistics in Medicine* 1998; **17**:2537–2550.
51. Daniels MJ, Hughes MD. Meta-analysis for the evaluation of potential surrogate markers. *Statistics in Medicine* 1997; **16**:1965–1982.
52. Gail MH, Pfeiffer R, van Houwelingen HC, Carroll RJ. On meta-analytic assessment of surrogate outcomes. *Biostatistics* 2000; **1**(3):231–246.
53. Antczak-Bouckoms A, Joshipura K, Burdick E, Tulloch JFC. Meta-analysis of surgical versus non-surgical method of treatment for periodontal disease. *Journal of Clinical Periodontology* 1993; **20**:259–268.
54. Sutton AJ, Lambert PC, Hellmich M, Abrams KR, Jones DR. Meta-analysis in practice: a critical review of available software. In *Meta-Analysis in Medicine and Health Policy*, Berry DA, Stangl DK (eds). Marcel Dekker: New York, 2000.
55. Pinheiro JC, Bates DM. *Mixed-effects Models in S and S-Plus.* Springer-Verlag: Berlin, 2000.
56. Sharp S, Sterne J. Meta-analysis. *Stata Technical Bulletin* 1997; **38**(sbe 16):9–14.
57. Sharp S. Meta-analysis regression. *Stata Technical Bulletin* 1998; **42**(23):16–22.
58. Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. *Statistics in Medicine* 1995; **14**:2685–2699.

