STATISTICS IN MEDICINE, VOL. 14, 395–411 (1995)

# A RANDOM-EFFECTS REGRESSION MODEL FOR META-ANALYSIS

C. S. BERKEY, D. C. HOAGLIN,* F. MOSTELLER† AND G. A. COLDITZ‡

*Technology Assessment Group, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, U.S.A.*

## SUMMARY

Many meta-analyses use a random-effects model to account for heterogeneity among study results, beyond the variation associated with fixed effects. A random-effects regression approach for the synthesis of $2 \times 2$ tables allows the inclusion of covariates that may explain heterogeneity. A simulation study found that the random-effects regression method performs well in the context of a meta-analysis of the efficacy of a vaccine for the prevention of tuberculosis, where certain factors are thought to modify vaccine efficacy. A smoothed estimator of the within-study variances produced less bias in the estimated regression coefficients. The method provided very good power for detecting a non-zero intercept term (representing overall treatment efficacy) but low power for detecting a weak covariate in a meta-analysis of 10 studies. We illustrate the model by exploring the relationship between vaccine efficacy and one factor thought to modify efficacy. The model also applies to the meta-analysis of continuous outcomes when covariates are present.

## INTRODUCTION

The literature of medical research has become too massive for individual researchers and clinicians to digest, particularly when studies provide seemingly contradictory conclusions.[1] Thus the quantitative synthesis of previous research has become an important part of the scientific method. 'Overview', 'research synthesis' and 'meta-analysis' are general terms applied to these techniques for the aggregation and synthesis of prior research.[2] Two recent reports[3,4] discuss current statistical methods for data synthesis.

Rubin[5] and others have criticized conventional meta-analysis techniques that average the outcomes of available studies. Instead, Rubin[5] suggests a need to move beyond current approaches to understand the underlying science. He proposes that meta-analysis models estimate the effect of treatment versus control as a function of a set of scientific factors that influence efficacy (for example, age, race, gender) and a set of design factors (sample size, randomization, blindness, etc.). As a step toward this goal, we present (in a specific context) a random-effects regression model for meta-analysis based on ideas presented by Morris.[6] Although the basic method applies to a continuous outcome variable, the present paper also describes a regression approach to supplement the random-effects model of DerSimonian and Laird[7] for synthesis of $2 \times 2$ tables. The National Research Council report[3] consistently

* Also at the Department of Statistics, Harvard University, Cambridge, MA 02138, U.S.A. and Abt Associates Inc., 55 Wheeler Street, Cambridge, MA 02138, U.S.A.

† Also at the Department of Statistics, Harvard University, Cambridge, MA 02138, U.S.A.

‡ Also at the Channing Laboratory, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, U.S.A.

CCC 0277–6715/95/040395–17

*Received October 1993*
*Revised February 1994*

recommends the use of random-effects approaches for meta-analysis and the exploration of sources of variation in study results.

Our motivation was a need to synthesize a series of $2 \times 2$ tables, each extracted from a randomized controlled trial of a vaccine, Bacillus Calmette-Guérin (BCG), for the prevention of tuberculosis (TB). This vaccine has been in use outside the U.S.A. since 1921, for routine vaccination at birth in many countries worldwide, yet debate over its efficacy continues. The published trials that we extracted from the literature varied widely in ways that were expected *a priori* to provide different *true* BCG efficacies, and also to provide different *estimates* of BCG efficacy. In addition to our overall quantitative estimate of BCG efficacy, our analysis needed a way to explore the contribution of certain covariates to the observed variation in vaccine efficacy and, when appropriate, to obtain estimates of vaccine efficacy that adjusted for them. The methods developed in this paper were applied to studies of the efficacy of the BCG TB vaccine, and the results are reported elsewhere.[8]

Simulation studies empirically assessed the performance of the model in this particular application and suggested modifications that were successful in this context. To illustrate the method, we present a random-effects regression meta-analysis that adjusts for distance of a trial from the equator (absolute latitude), a variable with a suspected association with TB vaccine efficacy.[9] Latitude accounts for variation in rainfall, humidity, environmental mycobacteria that may produce natural immunity, temperature, and other factors that may have biological implications for vaccine efficacy. Preparation of the live vaccine requires refrigeration; unrefrigerated, it would spoil more quickly in warmer temperatures. Furthermore, direct exposure of the vaccine to sunlight may reduce counts of live bacteria.

Berlin *et al.*[10] consider the (fixed-effects and random-effects) regression meta-analysis of epidemiologic dose–response data (slopes), where each study provides separate results for a (possibly different) series of doses. Our model cannot directly accommodate this type of data from dose–response studies, but our model has the capacity to include a dose variable, in the same manner as any other study-level covariate, if each study provides results on a single dose.

If the results from studies are homogeneous, then fixed-effects approaches may be used for deriving overall estimates of treatment efficacy, and regression models are not needed. Fortunately, the random-effects method of DerSimonian and Laird[7] and the random-effects regression method developed here both reduce to a fixed-effects analysis when the data are homogeneous.

## METHODS

### Model

Using the notation of DerSimonian and Laird[7], let $y_i = \log_e(RR_i)$, where $RR_i$ is the relative risk,[11] denote the observed measure of treatment effect in study $i$, $i = 1, \ldots, k$. In the no-covariate situation, we assume that $y_i | \theta_i \sim \mathrm{N}(\theta_i, \sigma_i^2)$ and $\theta_i \sim \mathrm{N}(\mu, D)$. ($\theta_i$ is the true log-relative-risk of the $i$th study, that is, the mean of the distribution of $y_i$, $\sigma_i^2$ is the variance of $y_i$, and $\mu$ is the mean and $D$ the variance of the distribution of $\theta_i$ across studies.) In introducing $\theta_i$, $\mu$ and $D$ in this way, we aim primarily to take explicit overall account of some sources of variation among studies. We have no special attachment to the formal structure of random sampling. In practice (perhaps through selection processes), the variation that one can observe among the studies in a meta-analysis may often be less than one might encounter if other studies (actual or contemplated) were also available or if the treatment were applied outside the well-controlled structures of randomized trials.

To incorporate study-level covariates and thus account for heterogeneity among studies, we may further specify $\mu$ by $\mathbf{X}_i\boldsymbol{\alpha}$, where $\mathbf{X}_i$ is a row vector that contains the values of the covariates for study $i$ and $\boldsymbol{\alpha}$ is a column vector of regression coefficients, so that $\theta_i \sim \mathrm{N}(\mathbf{X}_i\boldsymbol{\alpha}, D)$. (It is convenient to specify the Gaussian distribution, from the class of two-parameter distributions, but most of the analysis may proceed without assuming normality. We use normality only in computing probabilities.) We assume the design vector $\mathbf{X}_i$ and within-study variance $\sigma_i^2$ are known (in practice, one uses data from individual studies to estimate $\sigma_i^2$), and that we need to estimate $\boldsymbol{\alpha}$ and $D$ from the data. Thus we may write $y_i|\theta_i = \theta_i + e_i$ with $e_i \sim \mathrm{N}(0, \sigma_i^2)$ and $\theta_i = \mathbf{X}_i\boldsymbol{\alpha} + \delta_i$ with $\delta_i \sim \mathrm{N}(0, D)$. Combining these components yields the meta-analysis regression model

$$y_i = \mathbf{X}_i\boldsymbol{\alpha} + \delta_i + e_i.$$

We take $\delta_i$ and $e_i$ independent, and thus $\mathrm{var}(y_i) = D + \sigma_i^2$. The $\delta_i$ represents the $i$th trial's true deviation from the true mean of all trials having the same covariate values (specified in $\mathbf{X}_i$). The predicted value in the random-effects regression model corresponds to the mean value of $y$ (or $\log(RR)$) for all studies with a specified combination of covariates. The predicted value from a fixed-effects model, in contrast, corresponds to the single true value for that combination of covariates[10] (there is no residual among-trial heterogeneity, $D = 0$).

Notice that if $D$ is large relative to the within-study variances $\sigma_i^2$ ($\mathrm{var}(y_i) \approx D$) or if the studies have $\sigma_i^2$ that do not vary much ($\mathrm{var}(y_i) \approx D + \sigma^2$), then $\mathrm{var}(y_i)$ is approximately the same for all studies, so that an ordinary least-squares regression model is approximately correct. At the other extreme, if $D$ is small relative to the within-study variances, then $\mathrm{var}(y_i) \approx \sigma_i^2$ and the fixed-effects weighted-least-squares (WLS) regression model is approximately correct. For situations between these extremes, a random-effects model that takes into account both sources of variation is appropriate. A good estimation procedure for the random-effects model will approximate either the ordinary least-squares model or the fixed-effects weighted-least-squares model when the circumstances warrant.

If $D$ and the $\sigma_i^2$ were known, we could apply weighted-least-squares regression using weights proportional to $(D + \sigma_i^2)^{-1}$ to provide a minimum-variance unbiased estimator of $\boldsymbol{\alpha}$. $D$, however, is unknown, as is $\sigma_i^2$, although we use the estimated value $s_i^2$ as a known value of $\sigma_i^2$. Note that one will usually estimate $D$ with more uncertainty than $\sigma_i^2$. The estimation of $D$ is guided by the number of studies, $k$, which is generally smaller than the numbers of subjects within the individual studies that provide the sample sizes for estimation of the $\sigma_i^2$.

Morris[6] (page 53) presents an iterative scheme for estimating the between-study variance $D$ ($A$ in his notation, often $\tau^2$ in the meta-analysis literature) when the individual studies have known unequal variances $\sigma_i^2$. His approach (and ours) iterates between estimating the regression coefficients $\boldsymbol{\alpha}$ via weighed least squares, where the weights incorporate the current estimate of $D$, and estimating the between-study variance $D$.

Assume that $y_i|\boldsymbol{\alpha}, D \sim \mathrm{N}(\mathbf{X}_i\boldsymbol{\alpha}, D + s_i^2)$. If $\hat{D}$ is an approximately unbiased estimate of $D$, then a weighted-least-squares estimate of $\boldsymbol{\alpha}$ is

$$\hat{\boldsymbol{\alpha}} = (X^{\mathrm{T}}VX)^{-1}X^{\mathrm{T}}V\mathbf{Y}$$

with $V = \mathrm{diag}(W_1, \ldots, W_k)$, the diagonal matrix of weights

$$W_i = 1/(\hat{D} + s_i^2).$$

Table I. SAS PROC IML code for fitting random-effects regression model for meta-analysis with a continuous dependent variable $y_i$ and one covariate $x_{1i}$ (to fit a model with two covariates, insert 'read all var {x_2i} into xcov2;' after the seventh line below and replace '$x$ = one || xcov1' with '$x$ = one || xcov1 || xcov2')

```
data rereg;
    ...
    run;
proc iml;     use rereg;
read all var {y_i} into y;
read all var {s2_i} into vsubi;
read all var {x_1i} into xcov1;
*y_i is the continuous dependent variable in the regression model;
*x_1i is a continuous or binary (0/1) covariate;
*s2_i is the variance of y_i;
* Code by Berkey, Hoaglin, Mosteller, Colditz in 'A random-effects
regression model for meta-analysis', Statistics in Medicine (1994);
k = nrow (y); one = j (k, 1, 1); x = one || xcov1;
      r = ncol(x); print k r;
  * r is the number of regression coefficients;
  * k is the number of trials in the meta-analysis;
A = 0;     * A is the between-trial variance (called D in text);
start b; niters = 20;
do iter = 1 to niters;
  wsubi = 1/(vsubi + A);
  d = diag (wsubi);
  covbeta = inv(x` * d * x);
  beta = covbeta * (x` *d*y);
  sebeta = sqrt (vecdiag (covbeta));
  z = beta/sebeta;
    if iter = 1 then print 'FIXED-EFFECTS MODEL' A beta sebeta z;
newa = (sum((wsubi) # ((k/(k - r)) # (y - x * beta) # (y - x * beta) - vsubi)))/
sum(wsubi);
diff = a - newa; a = newa; if (a < 0.0) then a = 0;
  if (abs(diff) < .00001) then goto exit;
end;
  exit: print ' **** converged at iteration' iter '***';
print '*between var A not change by more than .00001 between iters*';
print iter a k vsubi wsubi yxcovbeta;
print '**********RANDOM-EFFECTS REGRESSION MODEL**********';
  print beta sebeta z;
finish; run b;
print '***IF Y = LOG (RR) THEN USE t WITH DF = k - r - 3***************';
  quit; run;
```

Morris[6] approximates the variance $D$ by

$$\hat{D} = \frac{\sum_i W_i\{(k/(k-r))(y_i - \mathbf{X}_i\hat{\boldsymbol{a}})^2 - s_i^2\}}{\sum_i W_i}$$

where $k$ is the number of studies and $r$ is the dimension of the vector $\boldsymbol{a}$. (Morris also obtains the restricted maximum-likelihood (REML) estimator of $D$ by replacing $W_i$ in the numerator and denominator of $\hat{D}$, but not in $V$, by $W_i^2$. In our analyses for this paper, on $2 \times 2$ tables, however,

the REML estimator consistently tended to underestimate $D$ and provided 95 per cent confidence intervals that were too narrow.) We force $\hat{D} \geqslant 0$ when the estimate is negative. We begin by setting $\hat{D} = 0$, and we estimate the $W_i$, $\boldsymbol{a}$ and $D$ iteratively until the estimate of $D$ converges, whereupon $\hat{\boldsymbol{a}} \sim \mathrm{N}(\boldsymbol{a}, (X^{\mathrm{T}} V X)^{-1})$. The iterative software for fitting this model is written in SAS PROC IML[12] (Table I).

The presentation up to this point pertains primarily to the analysis of a continuous variable $y_i$. In our application to vaccine efficacy, $y_i$ is the natural logarithm of the relative risk (for prospective studies). The log-relative-risk is symmetric, and its variance is well understood. We estimate the log-relative-risk from the $2 \times 2$ table from study $i$

| | disease | no disease | |
|---|---|---|---|
| vaccine | $a_i$ | $b_i$ | $n_{i+}$ |
| no vaccine | $c_i$ | $d_i$ | $n_{i-}$ |

by $y_i = \log_e(RR_i) = \log_e[(a_i/n_{i+})/(c_i/n_{i-})]$. (The plus or minus subscripts on the row totals indicate whether the vaccine was present ( + ) or absent ( − ).) The variance of $\log_e(RR_i)$ is customarily estimated[11] by $[b_i/a_i n_{i+}] + [d_i/c_i n_{i-}]$. Later, we present our tactic for dealing with any $a_1$ or $c_i$ that is zero.

Emerson *et al.*[13] have demonstrated in a similar (though no-covariate, $r = 1$) situation that, because each study's estimate of risk difference (we use relative risk) and the corresponding estimated variance $s_i^2$ are not independent (the variances of the individual proportions depend upon the proportions), the DerSimonian–Laird approach (among others) may produce a biased estimate of overall treatment efficacy. For odds ratios, Raghunathan and Ii[14] also found substantial bias in treatment effects estimated by the DerSimonian and Laird[7] method. For the same reason, bias may be an issue in the random-effects regression model when the continuous dependent variable is log-relative-risk.

Our simulations (described in the 'Results' section) indicate that the correlation between $\log_e(RR_i)$ and $\widehat{\mathrm{var}}\,(\log_e(RR_i))$, estimated in the manner described above, does indeed slightly bias the estimates of $\boldsymbol{a}$ toward the null and $D$ toward zero. We therefore also considered an alternative estimator for the variance of $\log_e(RR_i)$. Because only the sample sizes $n_{i+}$ and $n_{i-}$ of the $i$th study appear explicitly (the other quantities are means over the $k$ studies), the smoothed estimator

$$\widehat{\mathrm{var}}\,[\log_e(RR_i)] = \left[\sum_{i=1}^{k} (b_i/a_i)\right] \Big/ kn_{i+} + \left[\sum_{i=1}^{k} (d_i/c_i)\right] \Big/ kn_{i-}$$

reduces the correlation between $\log_e(RR_i)$ and the estimated $\mathrm{var}(\log_e(RR_i))$. (Whenever $a_i$ or $c_i$ is 0, we use the adjustment described later.) The smoothed estimator of variance for the $i$th trial is based upon the mean of the $b_i/a_i$ across the trials, and similarly for the $d_i/c_i$. This form implicitly assumes that the true error variances differ between studies solely because of differences in sample size.

We may also apply this method to case-control studies by using estimated odds ratios (OR) instead of relative risks. In the $2 \times 2$ table of vaccine status by disease status, consider the 'disease' column to represent the cases or deaths from the disease and the 'no disease' column to represent the matched controls in a case-control study. Then $y_i = \log_e(OR_i) = \log_e[(a_i d_i)/(c_i b_i)]$. One customarily estimates its variance[11] by $(1/a_i) + (1/b_i) + (1/c_i) + (1/d_i)$. (Again, empty cells require some form of adjustment.) To reduce the correlation between $\log_e(OR_i)$ and its estimated

variance, however, we estimate the variance by

$$\widehat{\text{var}}\,[\log_e(OR_i)] = \left[(a_i + c_i)\left(\sum_{i=1}^{k}(a_i/(a_i + c_i))\right)\Big/k\right]^{-1} + \left[(a_i + c_i)\left(1 - \left(\sum_{i=1}^{k}(a_i/(a_i + c_i))\right)\Big/k\right)\right]^{-1} + \left[(b_i + d_i)\left(\sum_{i=1}^{k}(b_i/(b_i + d_i))\right)\Big/k\right]^{-1} + \left[(b_i + d_i)\left(1 - \left(\sum_{i=1}^{k}(b_i/(b_i + d_i))\right)\Big/k\right)\right]^{-1}.$$

For study $i$, the total numbers of cases $(a_i + c_i)$ and controls $(b_i + d_i)$ then directly influence its variance, but its proportions of cases and controls that are vaccinated do not affect its variance except through the study's contribution to a $k$-study mean. This estimator replaces the $a_i$ in the usual estimator by the number of cases who would have been observed in that cell (diseased and vaccinated) if the $i$th study had the mean proportion (observed across studies) of diseased (cases) who were also vaccinated. The $b_i$, $c_i$ and $d_i$ are similarly replaced.

Our objective is that, on average, this procedure for regression in random-effects meta-analysis is compatible with the DerSimonian and Laird[7] random-effects model. This objective is exemplified as follows. If we perform a meta-analysis by the DerSimonian and Laird[7] method on studies of boys, and a separate meta-analysis on girls, and if a standard $t$-test that compares the results of the two meta-analyses shows that their estimates differ significantly, then the random-effects regression model would also declare a significant sex effect when we analyse the boys' and girls' $\log(RR)$ together with a binary covariate for gender, and vice versa. (If the between-study variance $D$ for boys differed greatly from that for girls, then we might expect discrepancies between the DerSimonian and Laird[7] and regression model conclusions.) If we consider a continuous covariate, rather than a binary one such as sex, or if we wish to consider two or more covariates simultaneously, then the DerSimonian and Laird[7] meta-analysis becomes more difficult to use, because it requires categorization of continuous variables and cross-stratification in analyses. The random-effects regression method is then an efficacious alternative.

Table I gives code in SAS PROC IML[12] for fitting the random-effects regression model with one covariate $x_{1i}$, when the outcome (dependent) variable $y_i$ ($\log(RR_i)$ in our application) is continuous and one has derived the variances $s_i^2$.

## Simulation studies

We focus on features (number of studies to be combined, sample size of each study, number of predictor variables in the regression model, etc.) relevant to the vaccine meta-analysis that motivated this methodological work. In particular, we deal only with relative risks. We drew the values assumed for $a$, $D$, $k$, the $n_i$, and the probability of disease among the control subjects from preliminary data collected early in the vaccine project. The random numbers for the simulations came from the SAS functions RANUNI (uniform distribution), RANNOR (normal distribution) and RANBIN (binomial distribution).[15]

### *No-covariate model*

We simulated 1000 meta-analyses. Each included a set of $k = 10$ generated studies, so that we created 10,000 studies (or $2 \times 2$ tables). At each replication, we selected the sample size $n_i$ for the

$i$th study randomly from a pool of preliminary sample sizes observed among the BCG vaccine trials. By requiring that each $n_i$ be even (adding 1 to odd sample sizes), we then assumed that each treatment group (vaccine/no vaccine) had sample size $n_i/2$, the usual goal in randomized controlled trials.

The model that generated the data (the $y_i$) for the $i$th study, for inclusion in a meta-analysis of 10 studies, was

$$y_i = \log_e(RR_i) = -0{\cdot}8387 + \delta_i$$

for $i = 1$ to 10, with $\delta_i \sim N(0, 0{\cdot}211)$. Thus, with only a constant term, $\alpha_0$, we have $\mathbf{X}_i\boldsymbol{\alpha} = 1\alpha_0 = \mu = -0{\cdot}8387$ and $D = 0{\cdot}211$. The choice $\alpha_0 = -0{\cdot}8387$ represents a ballpark estimate for the weighted mean of the $\log(RR_i)$. The variance 0·211 is a value that we might reasonably expect to obtain from the BCG studies, according to the preliminary data.

After randomly selecting the sample size $n_i$ for the $i$th $2 \times 2$ table, and setting $n_{i+} = n_i/2$ and $n_{i-} = n_i/2$ (see the $2 \times 2$ table for study $i$ in the 'Model' section), we randomly selected the value $c_i$ from a binomial ($n_i/2$, $p = 0{\cdot}05$) distribution (that is, the true probability of disease in those not vaccinated was 0·05) and set $d_i = n_i/2 - c_i$. In the next step of simulation, we randomly drew a $\delta_i$ from the normal distribution with mean 0 and variance 0·211. To generate a table with true $\log_e(RR_i) = -0{\cdot}8387 + \delta_i$, we randomly selected $a_i$ from binomial ($n_i/2$, $p = 0{\cdot}05 \exp(-0{\cdot}8387 + \delta_i)$). Finally, $b_i = n_i/2 - a_i$. (Note that if $\delta_i = 0$, the $a_i$ cell counts would be generated assuming the true probability 0·05 exp( − 0·8387), and the $c_i$ would be generated independent of the $a_i$ assuming the true probability 0·05. Then the true $\log(RR_i) = \log(0{\cdot}05 \exp(-0{\cdot}8387)/0{\cdot}05) = \log(\exp(-0{\cdot}8387)) = -0{\cdot}8387$.)

Whenever the generated $a_i$ was zero, we added 0·5 to $a_i$, 0·5 to $b_i$, and 1·0 to the row total $n_{i+}$; we treated a zero $c_i$ in the same manner. We later, instead, added 1/6 to the $a_i$ and $b_i$ in such cases (see Mosteller and Tukey,[16] Chapter 5), rather than 1/2, to determine whether our findings were sensitive to the method used for adjusting empty cells.

From these four simulated counts in the $2 \times 2$ table, we computed the observed $\log(RR_i) = \log[(a_i/n_{i+})/(c_i/n_{i-})]$ for study $i$ and estimated $\sigma_i^2$ by the two methods presented earlier:

*Study-specific*

$$\widehat{\text{var}}\,[\log(RR_i)] = [b_i/a_i n_{i+}] + [d_i/c_i n_{i-}]$$

*Smoothed*

$$\widehat{\text{var}}\,[\log(RR_i)] = \left[\sum_{i=1}^{k} (b_i/a_i)\right] \Big/ kn_{i+} + \left[\sum_{i=1}^{k} (d_i/c_i)\right] \Big/ kn_{i-}.$$

The smoothed estimator uses the unweighted mean of the $b_i/a_i$ and the mean of the $d_i/c_i$ over the $k = 10$ studies.

Using this strategy, we generated a sample of ten estimated $\log(RR_i)$, each with an $s_i^2$, that become the data used for estimating $\alpha_0$ and $D$ by the random-effects regression model. The regression model provided an estimate of $\alpha_0$ (a pooled $\log(RR)$ from 10 studies), the estimated $\text{se}(\hat{\alpha}_0)$ derived from $(X'VX)^{-1/2}$, an estimate of the between-study variance ($D$), and finally a 95 per cent confidence interval for $\alpha_0$. We replicated this process 1000 times, producing 1000 $\hat{\alpha}_0$, 1000 standard errors, 1000 $\hat{D}$ and 1000 confidence intervals. Hence, we can compare these against the true values that generated the data. The estimated 'true' value for the standard error of $\alpha_0$ is the empirical standard deviation of the 1000 estimates of $\alpha_0$.

### One-covariate model

The simulation strategy followed the process described above with the following differences. The model that generated the true log(*RR*) for the *i*th study was

$$\log_e(RR_i) = \alpha_0 + \alpha_1 x_i + \delta_i$$

or, more specifically,

$$\log_e(RR_i) = -0{\cdot}8387 + 0{\cdot}3\,x_i + \delta_i$$

for $i = 1$ to 10, with $\delta_i \sim N(0, 0{\cdot}211)$. The $x_i$ were randomly generated as N(0, 1). (We chose a normally distributed covariate for our simulation, but in an application $x_i$ may be binary or may come from a non-normal continuous distribution.) Therefore, an increase of one standard deviation in the predictor variable ($x_i = 1{\cdot}0$) raises the predicted *RR* from 0·432 (= exp(− 0·8387) when $x = 0$) to 0·584 (= exp(− 0·8387 + 0·3) when $x = 1$), and a decrease of one standard deviation ($x_i = -1{\cdot}0$) decreases the predicted *RR* to 0·320 (= exp(− 0·8387 − 0·3) when $x = -1$). (As stated earlier, under the random-effects model the predicted *RR* represents the mean of a distribution.) This represents a weak modifier of vaccine efficacy, but one of clinical interst if it were true.

We generated the $2 \times 2$ tables similarly to the no-covariate situation, except that to generate a table with true $\log(RR_i) = -0{\cdot}8387 + 0{\cdot}3\,x_i + \delta_i$, we selected the $a_i$ randomly as binomial $(n_i/2, p = 0{\cdot}05 \exp(-0{\cdot}8387 + 0{\cdot}3\,x_i + \delta_i))$.

We fit a random-effects regression model to the generated data from the 10 studies in each of the 1000 meta-analyses. We then compared the 1000 estimated $\alpha$, their 95 per cent confidence intervals, and the 1000 estimated $D$ with the true values to assess bias and efficiency of the methodology. We estimated power from the simulation studies for both the no-covariate model and the single-covariate model.

### Two misspecified models

We sought to examine the consequences of fitting a model with no covariates when a variable is present that modifies the outcome. The 10,000 $2 \times 2$ tables (for 1000 meta-analyses of 10 tables each) simulated for the one-covariate model (true $\alpha_1 = 0{\cdot}3$) provided such a basis.

A misspecified model also arises when one fits a one-covariate model by including a variable that has no true association with the outcome. Here, we used the same $2 \times 2$ tables generated for the no-covariate situation (true $\alpha_1 = 0$), and we generated an independent randomly distributed N(0, 1) covariate that served as $x_1$ in the model fitting. (We did not use this covariate in generating the $2 \times 2$ tables.)

## RESULTS

### No-covariate model

*Adequacy of simulation*

The 10,000 generated $\delta_i$ values (10 $\delta_i$'s for each of 1000 meta-analyses) had mean 0·004 and variance 0·210. The generating theoretical values were mean 0·0 and variance 0·211.

*Results*

Using the study-specific estimator of the $\sigma_i^2$, the mean of 1000 estimates of $\alpha_0$ was − 0·7928 (true value − 0·8387) (Table II). Because both values represent a similar relative risk (0·452 versus

Table II. Random-effects regression analyses of simulated meta-analyses, each consisting of ten 2 × 2 tables. The random-effects regression model estimated the $\alpha_0$, $\alpha_1$, and $D$ for each meta-analysis. The 0·5 adjustment for empty cells and the $t_{k-r-3}$ distribution were used. One thousand meta-analyses were simulated for the no-covariate model and 1000 for the single-covariate model using the true model parameters shown

| True model | No covariates | | | One covariate | | |
|---|---|---|---|---|---|---|
| | $\alpha_0 = -0{\cdot}8387$ | $\alpha_1 = 0$ | $D = 0{\cdot}211$ | $\alpha_0 = -0{\cdot}8387$ | $\alpha_1 = 0{\cdot}3$ | $D = 0{\cdot}211$ |
| *Study-specific estimator of* $\sigma_i^2$ | | | | | | |
| Mean estimate | −0·7928 | 0·206 | | −0·7882 | 0·2881 | 0·204 |
| (sd) | (0·1779) | (0·150) | | (0·1954) | (0·2281) | (0·159) |
| Mean of se's | (0·1706) | | | (0·1817) | (0·2012) | |
| 95% coverage $t_{k-r-3}$ | 94·9% | | | 94·6% | 94·8% | |
| 90% coverage $t_{k-r-3}$ | 88·8% | | | 88·2% | 90·0% | |
| 80% coverage $t_{k-r-3}$ | 78·3% | | | 77·3% | 79·7% | |
| 50% coverage $t_{k-r-3}$ | 48·9% | | | 48·0% | 52·2% | |
| *Smoothed estimator of* $\sigma_i^2$ | | | | | | |
| Mean estimate | −0·8603 | 0·249 | | −0·8543 | 0·3030 | 0·244 |
| (sd) | (0·1916) | (0·194) | | (0·2139) | (0·2486) | (0·202) |
| Mean of se's | (0·1851) | | | (0·1968) | (0·2163) | |
| 95% coverage $t_{k-r-3}$ | 95·4% | | | 95·5% | 94·4% | |
| 90% coverage $t_{k-r-3}$ | 90·8% | | | 90·4% | 89·5% | |
| 80% coverage $t_{k-r-3}$ | 78·9% | | | 78·3% | 80·7% | |
| 50% coverage $t_{k-r-3}$ | 50·6% | | | 48·4% | 50·5% | |

0·432), the resulting estimator of $\alpha_0$ appears mildly biased toward the null ($RR = 1$) for this particular situation. The empirical standard deviation of the 1000 $\hat{\alpha}_0$'s was 0·1779. This is reasonably well approximated by the estimated standard error, whose average over the 1000 replications was 0·1706. The mean estimated variance $D$ from the 1000 meta-analyses was 0·2061 (true value 0·211).

We computed 95 per cent confidence intervals around the estimate of $\alpha_0$ from each of the 1000 meta-analyses, using the critical value (2·262) from $t_{k-1}$. Larholt[17] found that the $t$ distribution performed better than the normal distribution in a similar situation. Using the $t$ distribution ($k - 1 = 9$ d.f.), 93·6 per cent of the confidence intervals covered the true value $\alpha_0 = -0{\cdot}8387$. (The normal distribution consistently provided even more under coverage.) The $t_{k-2}$ and $t_{k-3}$ distributions provided 93·7 per cent and 94·3 per cent coverage. The $t$ distribution with $k - 4$ degrees of freedom, however, provided more appropriate coverage, with 94·9 per cent of the confidence intervals enclosing the true $\alpha_0$. Table II also presents results for 90 per cent confidence intervals. (This empirical result pertains to our TB vaccine application, but we have no theoretical justification for it. It may not hold different $n_i$, $D$, etc.)

Using the smoothed estimator of $\sigma_i^2$, intended to reduce the correlation between $\sigma_i^2$ and $\log(RR_i)$, and the same simulated 2 × 2 tables, the mean $\alpha_0$ was −0·8603 (Table II), which represents a relative risk ($RR$) of 0·423 versus the true value 0·432. The empirical standard

deviation of the $\hat{\alpha}_0$'s was 0·1916; again this is reasonably well approximated by the estimated standard errors, whose average was 0·1851. The mean of the estimates of $D$ was 0·249 (versus 0·211). The 95 per cent confidence intervals based on $t_{k-1}$ provided the same coverage as above, 93·6 per cent, and $t_{k-2}$ and $t_{k-3}$ provided coverages of 94·1 per cent and 94·5 per cent, respectively. Again, the coverage for $t_{k-4}$, 95·4 per cent, was closest to the theoretical value 95 per cent.

We conclude that, for the simplest case with no covariates in the model, the estimates of $\alpha_0$ and $D$ appear nearly unbiased, and coverage at 95 per cent is well approximated using the $t$ distribution with $k$-4 degrees of freedom. We have a mild preference for use of the smoothed estimator of variance because it provided slightly less bias in $\alpha_0$ and also because it performed better in the next model.

### Model with one covariate

*Adequacy of simulation*

The 10,000 generated $\delta_i$ values had mean 0·0038 and variance 0·2117. The generating theoretical values were mean 0·0 and variance 0·211.

*Results*

Using the study-specific estimator of the $\sigma_i^2$, the mean of the 1000 estimates of $\alpha_0$ was $-0{\cdot}7882$ (true value $-0{\cdot}8387$) (sd = 0·20, Table II). Both values represent a similar relative risk (0·455 versus 0·432), but again this estimator of $\alpha_0$ appears slightly biased toward the null. The mean of 1000 estimates of $\alpha_1$ was 0·2881 (true value 0·3). The mean estimated variance $D$ from the 1000 meta-analyses was 0·204 (true value of $D$ = 0·211) (Table II).

We computed 95 per cent confidence intervals around the estimates of $\alpha_0$ and $\alpha_1$ from each of the 1000 meta-analyses, using $t_{k-2} = 2{\cdot}306$ in the computations. The true value $\alpha_0 = -0{\cdot}8387$ was covered by 92·1 per cent of the confidence intervals, and the true $\alpha_1 = 0{\cdot}3$ was covered by 92·4 per cent of the confidence intervals. As we saw in the model with no covariates, reducing the degrees of freedom by 3, thus using $t_{k-5}$, gave coverage closer to the nominal 95 per cent. Thus using $t_{k-5}$ for $\alpha_0$, 94·6 per cent of these confidence intervals covered the true value, and for $\alpha_1$, 94·8 per cent covered the true value (Table II).

Using the smoothed estimator of $\sigma_i^2$ and the same simulated $2 \times 2$ tables, the mean of 1000 estimates of $\alpha_0$ was $-0{\cdot}8543$, representing a relative risk of 0·426 versus the true value 0·432. The mean of the 1000 estimated $D$ was 0·244 (versus 0·211). Using the $t_{k-2}$ distribution for 95 per cent confidence intervals provided 93·3 per cent coverage of the true $\alpha_0$, and the $t_{k-5}$ distribution provided coverage of 95·5 per cent.

The mean estimated $\alpha_1$ was 0·303 (true value 0·3). The 1000 95 per cent confidence intervals computed using $t_{k-2}$ provided 92 per cent coverage of the true $\alpha_1$, and the intervals using $t_{k-5}$ provided 94·4 per cent coverage (Table II).

The need for this adjustment in the degrees of freedom may arise from our use of the $s_i^2$, estimated from data of the individual studies, in place of the true known $\sigma_i^2$. The substantial variance in the estimated $D$'s may also play some role. The actual degrees-of-freedom penalty may vary by application. It might differ according to the numbers of studies in a meta-analysis, the magnitude of the within-study variances $\sigma_i^2$ relative to the between-study variance $D$ (if the $\sigma_i^2$ are very small relative to $D$, then we are approximately in the OLS regression situation, where the degrees of freedom are the usual $k - r$), the ratio of the largest $s_i^2$ to the smallest $s_i^2$ among the $k$ studies in a meta-analysis, or the true parameter values ($D$ and $\boldsymbol{\alpha}$). In our simulation study, the

Table III. Summary of simulated meta-analyses of ten 2 × 2 tables each, including two misspecified models (on off-diagonal). The random-effects regression model provided the estimates of $\alpha_0$, $\alpha_1$, and $D$ for each meta-analysis. The smoothed estimator of $\sigma_i^2$, the $t_{k-r-3}$ distribution, and the 0·5 replacement for empty cells were used. One thousand meta-analyses were simulated for the no-covariate model, and 1000 for the single-covariate model. using the true model parameters shown

| Estimated model | True model | | | | | |
|---|---|---|---|---|---|---|
| | *No covariates* | | | *One covariate* | | |
| | $\alpha_0 = -0{\cdot}8387$ | $\alpha_1 = 0$ | $D = 0{\cdot}211$ | $\alpha_0 = -0{\cdot}8387$ | $\alpha_1 = 0{\cdot}3$ | $D = 0{\cdot}211$ |
| *No covariates* | | | | | | |
| Mean estimate | −0·8603 | | 0·249 | −0·8630 | | 0·332 |
| (sd) | (0·1916) | | (0·194) | (0·2174) | | (0·242) |
| 95% coverage $t_{k-4}$ | 95·4% | | | 95·1% | | |
| Power | 97% | | | 93·3% | | |
| *One covariate* | | | | | | |
| Mean estimate | −0·8597 | 0·0076 | 0·249 | −0·8543 | 0·3030 | 0·244 |
| (sd) | (0·2077) | (0·2444) | (0·205) | (0·2139) | (0·2486) | (0·202) |
| 95% coverage $t_{k-5}$ | 95·4% | 96·2% | | 95·5% | 94·4% | |
| Power | 92·5% | na | | 92% | 22% | |

within-study variances $s_i^2$ are a little smaller on average than the between-study variance $D$; there is generally great disparity (an average 90-fold difference) among the within-study variances $s_i^2$ in a single simulated meta-analysis, and the estimated $D$'s have a large standard deviation (around 0·20).

### Rarer disease

To see whether smaller incidence rates of TB in the simulation study would have altered our conclusions, we replicated the one-covariate simulation study, with the assumption that the true probability of TB in the control group was 0·025 rather than 0·05. The major difference from the earlier analysis was that power decreased (from 92 per cent to 87·7 per cent for $\alpha_0$ and from 22 per cent to 18·2 per cent for $\alpha_1$). We discuss power after the next subsection. The biases we noted earlier from use of the study-specific estimator of $\sigma_i^2$ persisted for a rarer disease.

### Empty cells

We repeated our analysis of the 2 × 2 tables generated from a one-covariate model (assuming the true probability of TB in the control group was 0·05) with use of an alternative adjustment for empty cells. Use of the 1/6 correction[16] for empty cells (adding 1/6 instead of 1/2 to each cell of the relevant row of the 2 × 2 table) provided essentially the same conclusions outlined earlier. This held for both the study-specific estimator of $\sigma_i^2$ and the smoothed estimator. For comparison with the lower right corner of Table III, which uses the 1/2 correction and the smoothed estimator of $\sigma_i^2$, the 1/6 correction provided a mean estimated $D$ of 0·2777, mean estimated $\alpha_0$ of −0·8646, mean estimated $\alpha_1$ of 0·3099, coverages of the two $\alpha$'s of 95·6 per cent and 94·6 per cent, and powers of 90 per cent and 20·8 per cent. We had feared more sensitivity to choice of adjustment. We found much greater differences attributable to choice of the smoothed estimator of $\sigma_i^2$ or the study-specific estimator. If our simulated probability of TB had been closer to zero, we might have observed more sensitivity. The value we simulated (0·05), however, is near the mean value 0·045 observed in 13 real trials extracted from the literature (see 'Illustration' section).

## Power

We use the simulations to estimate the power of our random-effects regression model for meta-analysis. We estimate power here for the version of the model that uses the smoothed estimator of $\sigma_i^2$, and with the 3-degrees-of-freedom penalty (d.f. $= k - r - 3$) in the $t$ distribution. To test the null hypothesis, we divided each estimate of $\alpha_0$ and $\alpha_1$ by its standard error and referred the result to the $t$ distribution.

### *No covariates* ($r = 1$)

The random-effects regression model with no covariates ($r = 1$ coefficient in the model) provides an estimator (similar to that of DerSimonian and Laird) of the overall or combined $RR$. We found that if the true $RR = 0{\cdot}43$, then the power of the test of the null hypothesis $H_0: RR = 1$ was 97 per cent (Table III, upper left). In other words, 97 per cent of the 1000 simulated meta-analyses provided a statistically significant ($p < 0{\cdot}05$ for $H_0: \alpha_0 = 0$) estimate of $\alpha_0$.

### *One or more covariates* ($r \geqslant 2$)

We expect, however, to lose power as we add covariates to a model estimated from a small number of studies (10 in the simulations). This number of studies is in the range of meta-analyses often used in medicine and public health. According to our simulations, we maintained good power (92 per cent) for evaluating overall vaccine efficacy (the $\alpha_0$ term) when estimated by a model that includes a weak covariate (Table III, lower right). Our power (only 22 per cent) for detecting the weak covariate itself ($\alpha_1 = 0{\cdot}3$), however, was low. The conservative approach that we used for computing the 95 per cent confidence intervals (the 3 d.f. penalty to maintain nominal coverage) naturally diminishes power.

Because of cost-effectiveness considerations and side-effects, society may not consider the implementation of a vaccine worthwhile unless the vaccine provides considerably more than zero protection. Investigators may therefore have more interest in the power of the test of the null $RR = 0{\cdot}75$ (a 25 per cent protective effect) rather than $RR = 1$ (zero protective effect). In the one-covariate model, when the true $RR = 0{\cdot}43$, we estimated this power as 58·5 per cent, much less than 92 per cent power observed above.

## Misspecified models

When we fitted the model with no covariate term to the simulated $2 \times 2$ tables generated with a weak covariate, we obtained estimates of $D$ (the mean $D = 0{\cdot}332$) that were larger than the true value because the real variability due to the covariate was not being modelled. Yet we still retained good power (93·3 per cent) for evaluating overall vaccine efficacy ($\alpha_0$ in Table III, upper right panel). The coverage was good (95·1 per cent), but the estimated $\alpha_0$ was slightly biased ($-0{\cdot}8630$ versus the true $-0{\cdot}8387$) away from the null. The penalty for ignoring the weak covariate (the cost of underfitting) appears small.

For the other situation, where no covariate was present in the simulating model but we estimated the effect of an independent $N(0, 1)$ variable in the model, the penalty (cost of overfitting) again seemed quite minor (Table III, lower left panel). The mean estimate of $D$ (0·249) was the same as when we fit the appropriate no-covariate model to the simulated meta-analyses, and the 95 per cent coverage ($t_{k-5}$) for $\alpha_0$ was unchanged at 95·4 per cent. We lost a little power, 97 per cent power from correct model versus 92·5 per cent from the one-covariate model. The coverage for $\alpha_1$ (true $\alpha_1 = 0$) was 96·2 per cent.

Table IV. Data from clinical trials of BCG vaccine efficacy (the full SAS code for fitting the random-effects regression model to these data is available from CSB)

| Trial | Latitude* | Vaccinated | | Not vaccinated | | *RR* | 95% CI |
|---|---|---|---|---|---|---|---|
| | | Disease | No Dis | Disease | No Dis | | |
| 1 | 44° | 4 | 119 | 11 | 128 | 0·41 | 0·13–1·26 |
| 2 | 55 | 6 | 300 | 29 | 274 | 0·20 | 0·09–0·49 |
| 3 | 42 | 3 | 228 | 11 | 209 | 0·26 | 0·07–0·92 |
| 4 | 52 | 62 | 13,536 | 248 | 12,619 | 0·24 | 0·18–0·31 |
| 5 | 13 | 33 | 5,036 | 47 | 5,761 | 0·80 | 0·52–1·25 |
| 6 | 44 | 180 | 1,361 | 372 | 1,079 | 0·46 | 0·39–0·54 |
| 7 | 19 | 8 | 2,537 | 10 | 619 | 0·20 | 0·08–0·50 |
| 8 | 13 | 505 | 87,886 | 499 | 87,892 | 1·01 | 0·89–1·14 |
| 9 | −27 | 29 | 7,470 | 45 | 7,232 | 0·63 | 0·39–0·99 |
| 10 | 42 | 17 | 1,699 | 65 | 1,600 | 0·25 | 0·15–0·43 |
| 11 | 18 | 186 | 50,448 | 141 | 27,197 | 0·71 | 0·57–0·89 |
| 12 | 33 | 5 | 2,493 | 3 | 2,338 | 1·56 | 0·37–6·53 |
| 13 | 33 | 27 | 16,886 | 29 | 17,825 | 0·98 | 0·58–1·66 |

* In the regression model, the absolute latitude is centred at the mean 33·4615

## ILLUSTRATION

A meta-analysis of trials that evaluated the efficacy of the BCG vaccine for preventing TB motivated our work on the random-effects regression model for meta-analysis. The efficacy of this vaccine has been controversial; estimates of vaccine efficacy have ranged from detrimental to no benefit to an 80 per cent protective effect (protective effect = $1 - RR$). The emergence of multiple drug-resistant strains of TB, the resurgence of TB in parallel with the spread of AIDS, and institutional outbreaks of TB have prompted the reconsideration of broadened use of BCG in the U.S.A.[18]

Details of the search for trials, published and unpublished, and the inclusion criteria appear elsewhere,[8,19] so we summarize them briefly here. We identified 13 randomized controlled trials that evaluated the efficacy of BCG vaccination against tuberculosis. From each trial we extracted details of vaccine efficacy in the form of a $2 \times 2$ table of vaccine status (BCG vaccinated or not) by TB status (TB diagnosis or not during the period of follow-up). These 13 $2 \times 2$ tables contained no empty cells. We also extracted data on covariates that might explain heterogeneity among study results.

Some covariates (age of subjects, dose of vaccine) may influence the true efficacy, and other factors (study design, quality of follow-up) may influence *estimates* of true efficacy. For illustration, consider latitude, one of several factors historically suspected as associated with true vaccine efficacy. Our analysis examined the relationship between vaccine efficacy and absolute latitude, or distance of each trial from the equator, which may serve as a surrogate for the presence of environmental mycobacteria that provide a certain level of natural immunity against TB. Latitude is one of a small group of variables analysed, and several appeared important.[8] Table IV summarizes data from these trials. The complete SAS code for fitting the random-effects regression model to these data, and its executed output, is available from the authors (CSB).

The mean rate of TB among unvaccinated control groups in these trials was 4·5 per cent (the simulation study assumed 5 per cent). The log ($RR$) for the BCG vaccine tends to decline with

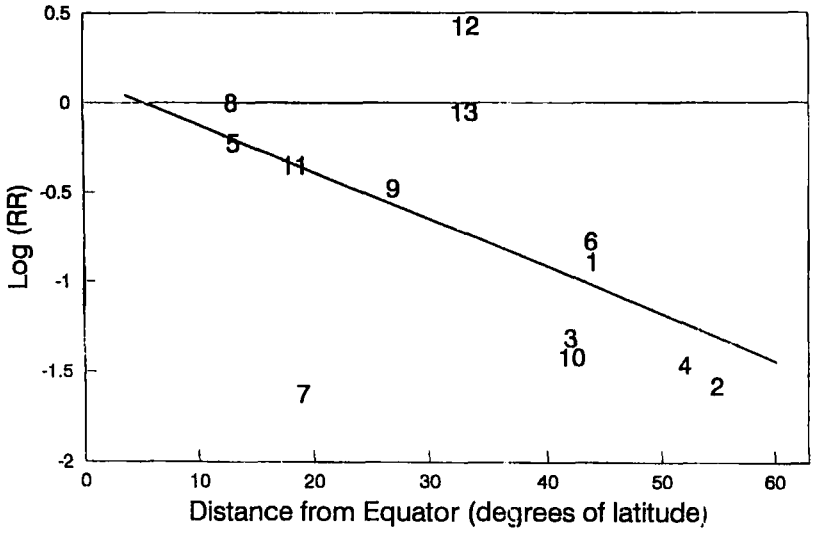


Figure 1. Log (*RR*) and distance from equator from 13 RCT's. A lower log(*RR*) corresponds to greater vaccine efficacy. The number of each trial (from Table IV) and the fitted random-effects regression line are shown

increasing distance form the equator in the 13 trials (Figure 1). The apparent outlier (lower left corner) had the fourth largest variance and thus was down-weighted in the analysis. Studies 12 and 13 (Figure 1) had a variety of methodologic shortcomings, some of which could account for the apparent lack of vaccine efficacy. When we fitted the random-effects regression model with no covariates, we estimated the between-trial variance as 0·268 and the pooled estimate of log(*RR*) as $-0{\cdot}5429$ with standard error 0·1842. Inclusion of the covariate distance from the equator, centred at 33·46°, provided a smaller between-trial variance of $D = 0{\cdot}157$ for the model $\log(RR) = -0{\cdot}6304 - 0{\cdot}0268\,(x - 33{\cdot}46)$, where $x$ was the distance from the equator (in degrees latitude). This covariate accounted for considerable (41 per cent) heterogeneity. The standard errors of the estimated coefficients were 0·1591 and 0·01102. If we had not simply begun with a single covariate (and instead had chosen it after exploration among several potential covariates), we would need to take into account the multiplicity of questions asked, for example, by using Bonferroni's inequality (see Ingelfinger *et al.*,[2] page 166).

We interpret this model for illustration only. It suggests that at 33·46° above (for example, southern U.S.A.) or below the equator, the mean relative risk of TB for vaccinated individuals is 0·53, for a mean protective effect of 47 per cent. At 42° north (for example, Chicago), the estimated mean *RR* is 0·42, for a mean protective effect of 58 per cent. Recall the many possible factors that latitude may be a proxy for, so that we regard the application to these particular geographic areas as merely illustrative.

We briefly consider the alternative conclusions that we would make under the simpler fixed-effects model, where the among-trial variance was assumed to be zero (rather than the random-effects model estimate of 0·157). The fitted fixed-effects regression model was $\log(RR) = -0{\cdot}5950 - 0{\cdot}0282\,(x - 33{\cdot}46)$. The fixed-effects line is similar to the random-effects line shown in Figure 1, and it provides latitude-specific estimates comparable to those of the random-effects model. The estimated *RR* for the southern U.S.A. equals 0·55 instead of 0·53, and the estimated *RR* for Chicago equals 0·43 rather than 0·42. The standard errors provided by the fixed-effects model (0·0696 for $\alpha_0$ and 0·0039 for $\alpha_1$), however, are only 44 per cent and 36 per cent

of the corresponding standard errors from the random-effects model. This considerable underestimation of the standard errors would foster overconfidence with major consequences in significance tests and reported confidence intervals for $\alpha_0$ and for $\alpha_1$. Furthermore, the fixed-effects model implies, for example, that future studies performed in the southern U.S.A. will arise from a single true *RR*, estimated 0·55, whereas the random-effects model implies that these future studies will not arise from a single true *RR* but from a distribution of true *RR*'s ($\log(RR) \sim N(-0{\cdot}63, 0{\cdot}157)$). Thus, we expect future studies to observe relative risks over a much broader range of *RR*'s than the range enclosed by the usual 95 per cent confidence interval of the estimated *RR*.

## DISCUSSION

The random-effects regression model provided a good empirical approximation in the vicinity of the parameters found in the TB vaccine data. The estimation procedure typically converged in fewer than 10 iterations, with use of a tolerance of 0·00001 for differences between the estimated $D$ in subsequent iterations. The computing program itself is uncomplicated, so that this method is relatively simple and computationally inexpensive.

In the future, one might potentially eliminate the small biases in the estimates of $a$ and $D$ from the random-effects regression model (analysis of $2 \times 2$ tables) by use of some alternative estimator of the $\sigma_i^2$, because the two estimators that we considered provided biases in opposite directions, for both the no-covariate and the single-covariate model. The magnitude and direction of the biases, however, do not present a problem for the meta-analysis for BCG efficacy. Because the biases for $a$ were smaller when we used the smoothed estimator of $\sigma_i^2$, we apply this version in the analysis of the efficacy of BCG for the prevention of tuberculosis.

The random-effects regression estimator of $\alpha_0$ (representing overall vaccine efficacy with $x_1$ centred at the sample mean) is quite powerful, but the small number of studies that we have for analysis limits our power to detect a contribution $\alpha_1$ by a weak covariate $x_1$. Because the simulation study focused on meta-analyses with 10 studies, whereas our application had 13, our meta-analysis of real vaccine data may have somewhat more power than that estimated by the simulations.

We found that confidence intervals and significance tests based on a $t$ distribution with $k - r - 3$ degrees of freedom maintained nominal coverages close to the 90 per cent and 95 per cent levels most often reported in statistical analyses. This contrasts with the use by Berlin *et al.*[10] and Grizzle *et al.*[20] of $z$ statistics for regression parameters with group-level data. The $t$ distribution often arises in linear regression settings, but the estimated $D$ and $s_i^2$ both offer sources of variation beyond the usual $k - r$ degrees of freedom; the choice of $k - r - 3$ degrees of freedom happened to work well in our context, even as we modified various aspects of the simulation studies. In further work we hope to learn how to adapt this adjustment to other situations. In the absence of a general rule, we may need a new simulation study for each application. A simulation study geared to the particular application would give a basis for choice.

Our findings on coverage are similar to those obtained by Larholt,[17] who used simulated meta-analyses of log-odds-ratios and estimation by the DerSimonian and Laird[7] method. For true $D = (0{\cdot}5)^2 = 0{\cdot}25$ (we used 0·211) and with 7 studies in each simulated meta-analysis (we used 10), her confidence intervals (based on the normal distribution: see her Table 5, page 67) provided 89·35 per cent coverage, and her confidence intervals based on $t_{k-1}$ provided 94·1 per cent coverage. Raghunathan and Ii,[14] in comparing methods for combining log-odds-ratios in multicentre clinical trials, also performed a simulation study of the DerSimonian and Laird method. Using a $t$ distribution with $k - 2$ degrees of freedom, they obtained for nominal 95 per cent

coverages results ranging from 88 per cent to 99 per cent for the DerSimonian and Laird method, and two alternative methods provided results as low as 34 per cent. Those authors[14] propose methods for including covariates, but they give no comparative results for them.

We have proposed a random-effects regression method that allows for covariates in the exploration of sources of variation in the meta-analysis of $2 \times 2$ tables. In the model, one may include interaction terms between covariates to investigate whether one covariate modifies the effect of the other covariate on reported treatment efficacy. Our method is not, however, directly applicable to all regression situations, such as when each study, in a meta-analysis of dose–response studies, reports separate results for a series of doses.[10] Our simulation studies provide evidence that our method works well for model parameters in the vicinity of a meta-analysis of the efficacy of a TB vaccine. Use of a smoothed estimator of the $\sigma_i^2$, rather than a study-specific estimator, produced less bias in the estimated regression coefficients. This happened because the smoothed estimator reduced the correlation between the $\log(RR_i)$ from the $2 \times 2$ table and its estimated variance. The method maintained its performance when we simulated a rarer disease. It did not depend on the adjustment used for empty cells in the $2 \times 2$ table. The method provided very good power for detecting a non-zero intercept term (representing overall $\log(RR)$), but low power for detecting a weak covariate in meta-analyses of 10 studies. Our examination of two misspecified models indicated minor penalties for under-fitting or over-fitting. The illustration demonstrates a situation in which one may relate variation in the results from different studies to a study-level covariate.

ACKNOWLEDGEMENTS

This work was supported by a collaborative agreement with the Centers for Disease Control and Prevention through the Association of Schools of Public Health and by grant HS05936 from the Agency for Health Care Policy and Research. We are especially grateful to Carl Morris for his encouragement and helpful advice, to Jesse Berlin for many valuable suggestions, to the reviewers, and to Timothy F. Brewer, Mary E. Wilson, Elisabeth Burdick, and Harvey V. Fineberg, our colleagues on the tuberculosis vaccine project.

## REFERENCES

1. Antman, E. M., Lau, J., Kupelnick, B., Mosteller, F. and Chalmers, T. C. 'A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts', *Journal of the American Medical Association*, **268**(2), 240–248 (1992).
2. Ingelfinger, J. A., Mosteller, F., Thibodeau, L. A. and Ware, J. H. *Biostatistics in Clinical Medicine*, 3rd edn, McGraw-Hill, New York, 1994.
3. National Research Council. *Combining Information: Statistical Issues and Opportunities for Research*, National Academy Press, Washington, D.C., 1992.
4. Cooper, H. and Hedges, L. V. (eds.) *Handbook of Research Synthesis*, Russell Sage Foundation, New York, 1994.
5. Rubin, D. 'A new perspective', *in* Wachter, K. W. and Straf, M. L. (eds.), *The Future of Meta-Analysis*, Russell Sage Foundation, New York, 1990, Chapter 14, pp. 155–165.
6. Morris, C. N. 'Parametric empirical Bayes inference: theory and applications', *Journal of the American Statistical Association*, **78**, 47–55 (1983).
7. DerSimonian, R. and Laird, N. 'Meta-analysis in clinical trials', *Controlled Clinical Trials*, **7**, 177–188 (1986).
8. Colditz, G. A., Brewer, T. F., Berkey, C. S., Wilson, M. E., Burdick, E., Fineberg, H. V. and Mosteller, F. 'Efficacy of BCG vaccine in the prevention of tuberculosis: Meta-analysis of the published literature', *Journal of the American Medical Association*, **271**, 698–702 (1994).
9. Bates, J. H. 'Tuberculosis: susceptibility and resistance', *American Review of Respiratory Disease*, **135** (part 2), 20–24 (1982).

10. Berlin, J. A., Longnecker M. P. and Greenland, S. 'Meta-analysis of epidemiologic dose-response data', *Epidemiology*, **4**, 218–228 (1993).
11. Rothman, K. J. *Modern Epidemiology*, Little Brown, Boston, MA, 1986.
12. SAS Institute, Inc. *SAS/IML User's Guide, Version 5 Edition*, SAS Institute, Inc., Cary, NC, 1985.
13. Emerson, J. D., Hoaglin, D. C. and Mosteller, F. 'A comparison of procedures for combining risk differences in sets of $2 \times 2$ tables from clinical trials', *Journal of the Italian Statistical Society* (in press).
14. Raghunathan, T. E. and Ii, Y. 'Analysis of binary data from a multicentre clinical trial', *Biometrika*, **80**, 127–139 (1993).
15. SAS Institute Inc. *SAS User's Guide: Basics, Version 5 Edition*, SAS Institute Inc., Cary, NC, 1985.
16. Mosteller, F. and Tukey, J. W. *Data Analysis and Regression*, Addison-Wesley, Reading, MA, 1977, Chapter 5.
17. Larholt, K. M. *Statistical Methods and Heterogeneity in Meta-Analysis*, unpublished Sc. D. thesis, Department of Biostatistics, Harvard School of Public Health, Boston, MA, 1989.
18. Centers for Disease Control. 'Nosocomial transmission of multidrug-resistant tuberculosis among HIV-infected persons-Florida and New York, 1988–1991', *Morbidity and Mortality Weekly Reports*, **40**, 858–891 (1991).
19. Colditz, G. A., Brewer, T., Berkey, C., Wilson, M., Burdick, E., Fineberg, H. V. and Mosteller, F. 'The Efficacy of BCG in the Prevention of Tuberculosis: Meta-analyses of the Published Literature', Final report to the Centers for Disease Control and Prevention, Atlanta, Harvard School of Public Health, Boston, MA, 1993.
20. Grizzle, J. E., Starmer, C. F. and Koch, G. G. 'Analysis of categorical data by linear models', *Biometrics*, **25**, 489–504 (1969).