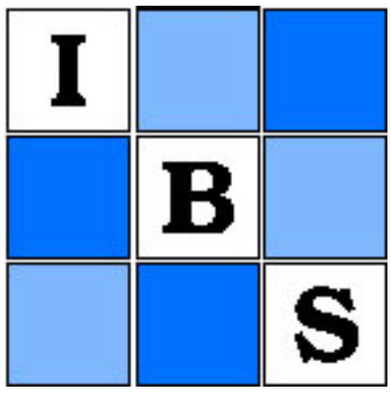

BIOMETRICS 52, 536–544
June 1996

Meta-Analysis of Published Data Using a Linear Mixed-Effects Model

Daniel O. Stram

Department of Preventive Medicine, University of Southern California, Los Angeles, California 90033, U.S.A.

SUMMARY

This paper describes the use of a linear mixed-effects regression model as a framework for the meta-analysis of published data. It generalizes the random-effects models used by DerSimonian and Laird (1986, *Controlled Clinical Trials* **7,** 177–188) and Begg and Pilote (1991, *Biometrics* **47,** 899–906), and describes the use of the model using examples from these papers and the data given by Tori et al. (1992, *Journal of the National Cancer Institute* **84,** 407–414).

1. Introduction

The summarization of the results of many studies in a meta-analysis has become increasingly important. Typically a meta-analysis in the clinical-trials context will identify a particular treatment which has been applied as the treatment of interest in a variety of separate trials and, by combining the published results of each trial, estimates an overall effect of this treatment by applying statistical methods such as that of Rao, Kaplan, and Cochran (1981), which allows for pooling of statistical estimates while taking account of the sampling variability inherent in each study's estimate. Testing for homogeneity of treatment effect across studies is recognized as an important part of meta-analysis. The ability of the meta-analysis to draw conclusions when the homogeneity assumption is found to be violated is somewhat controversial, however the estimation of overall or average treatment effects have been considered in this case. For example, DerSimonian and Laird (1986) discuss random-effects models for the analysis of comparisons of two specific treatments. In the meta-analysis that they consider, each study yields an independent estimate of a treatment comparison of interest and an average treatment effect is estimated, where the averaging is over a hypothetical population of potential studies of the treatments. Begg and Pilote (1991) consider a situation where both comparative studies of two treatments and single-armed studies exist and utilize a random-effects model for the purpose of combining these two types of studies in a single meta-analysis. In this model the random effect allows for baseline outcome to differ from study to study, but it is assumed that the difference in patient outcome between treatments is fixed. Begg and Pilote discuss several extensions of their basic model, to include both study-level covariates and an additional random effect, which, as in the DerSimonian and Laird model, allows for the true treatment difference to vary from study to study.

Invariably, as meta-analyses of published data have become more and more an accepted part of the review of medical research, the models used for such analyses have become increasingly complex. For example, a recent paper (Colditz, 1994) utilized the DerSimonian and Laird model for a meta-analysis of the world literature on Calmette-Guérin bacillus (BCG) vaccine efficacy, but then extended this model to attempt to explain the observed heterogeneity in BCG efficacy using covariates measured for each study. They found that distance from the equator was positively related to BCG effect and explained approximately two-thirds of the variance in true BCG efficacy.

This paper describes the use of a mixed-effects model for meta-analysis of published results which is flexible enough to include a wide variety of the current meta-analysis models in use and, in particular, to incorporate both fixed- and random-effect terms allowing for heterogeneity

Key words: Effect heterogeneity; Random effects; Structural variances; Variance components.

between studies in the effect of the treatment of interest. The power of this approach is illustrated by reexamining the DerSimonian and Laird, Begg and Pilote, and Tori et al. (1992) models, as special, or at least closely related, cases of the general model.

2. The Mixed-Effects Meta-Analysis Model

The general model described here takes the form

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\alpha} + \mathbf{Z}_i\boldsymbol{\beta}_i + \boldsymbol{\zeta}_i + \mathbf{e}_i, \tag{1}$$

where $i = 1, 2, \ldots, N$ indexes independent studies; for each study $\mathbf{Y}_i$ is an $(n_i \times 1)$ vector of one or more related estimates of the treatments or treatment comparisons of interest; $\mathbf{X}_i$ is an $(n_i \times p)$ matrix of known covariates related to the p vector of unknown fixed effect parameters, $\boldsymbol{\alpha}$; and $\mathbf{Z}_i$ is an $(n_i \times q)$ vector of known covariates related to a $(q \times 1)$ vector of unobserved random effects, $\boldsymbol{\beta}_i$, for each study. The two remaining $n_i \times 1$ unobserved random vectors, $\boldsymbol{\zeta}_i$ and $\mathbf{e}_i$, specify two types of error in $\mathbf{Y}_i$. The $\boldsymbol{\zeta}_i$ specify the "sampling errors" in $\mathbf{Y}_i$. That is, this vector represents the error in estimating patient outcome, using a finite sized study. This random vector is assumed to go to zero as the number of subjects participating in study i becomes increasingly large, whereas $\mathbf{e}_i$ specifies other sources of error or heterogeneity between studies or between arms of the same study. The assumption made in the following is that $\boldsymbol{\beta}_i$, $\boldsymbol{\zeta}_i$, and $\mathbf{e}_i$ are each independent multivariate normal random vectors with variance–covariance matrices $\mathbf{D}$, $\mathbf{V}_i$, and, $\sigma^2\mathbf{I}$, respectively. In general the set of parameters to be estimated are the fixed-effect parameters, $\boldsymbol{\alpha}$, the variance–covariance matrix, $\mathbf{D}$, for $\boldsymbol{\beta}_i$ and σ^2. It will be assumed that $\mathbf{V}_i$ takes the form

$$\begin{bmatrix} v_{i1} & 0 & \cdots & 0 \\ 0 & v_{i2} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & v_{in_i} \end{bmatrix},$$

where each v_{ij} is assumed known and represents the sampling variability in estimating, by a finite sized study, the true value of each element, Y_{ij}, $j = 1, \ldots, n_i$, of $\mathbf{Y}_i$, from the population sampled in this study. In specific examples, each Y_{ij} will represent such quantities as the estimate for each treatment arm in a multi-armed study, of the cure rate of patients treated with each drug treatment considered, or the estimates of the differences in cure rates seen in specific treatment comparisons for each study. The Appendix describes a Fisher's scoring algorithm which can be used for fitting all the parameters in model (1) by maximum likelihood or by restricted maximum likelihood (Harville, 1977). Other algorithms such as the EM algorithm (Laird and Ware, 1982) or the Newton–Raphson method of Lindstrom and Bates (1988) may also be adapted to fit this model.

For many simple meta-analyses, not all of σ^2 and the parameters in the $q \times q$ matrix $\mathbf{D}$ are necessary or identifiable, thus the fitting algorithm applied to estimating model (1) must be able to fix unneeded variance components to zero. Testing for heterogeneity of treatment-effect across study using likelihood-ratio tests will usually involve fitting model (1) while holding either σ^2 and/or components of $\mathbf{D}$ fixed at zero for comparison with a model in which this variance parameter is left free. The following examples describe the application of this general model in three cases.

3. Examples

3.1 *The DerSimonian and Laird Model*

DerSimonian and Laird (1986) describe the use of the random effects model

$$Y_i = \alpha + \beta_i + \zeta_i \tag{2}$$

for the analysis of univariate treatment differences, Y_i. For example, in their analysis of Winship's (1978) data (Table 1) for eight clinical trials of the use of cimetidine in duodenal ulcer, Y_i is either the difference in the proportion of healed ulcers, $p_{i,t} - p_{i,c}$, between the treatment and control arms of study i or the difference between the log odds, $\ln(p_{i,t}/(1-p_{i,t})) - \ln(p_{i,c}/(1-p_{i,c}))$. In this model α is the average treatment difference (on either the additive or log odds scale) of cimetidine over a hypothetical population of possible studies, and $\alpha + \beta_i$ gives the true treatment difference between cimetidine and the control in the population sampled by study i. Here β_i is assumed to be Gaussian with mean zero and unknown variance. The sampling error, ζ_i, in estimating $\alpha + \beta_i$ using a finite study (with $n_{i,t}$ patients receiving treatment and $n_{i,c}$ patients receiving the control) has variance $p_{i,t}(1-p_{i,t})/n_{i,t} + p_{i,c}(1-p_{i,c})/n_{i,c}$ for the difference-scale model and is approximately $1/(p_{i,t}(1-p_{i,t})n_{i,t}) + 1/(p_{i,c}(1-p_{i,c})n_{i,c})$ for the log odds analysis. DerSimonian and Laird discuss

Table 1
Data from Winship (1978), *use of cimetidine in duodenal ulcer*

Study	Number treated	Number healed	Number controls	Number healed
1	19	16	19	8
2	30	26	14	5
3	20	17	20	12
4	20	17	18	5
5	65	47	24	7
6	21	13	21	4
7	43	36	42	16
8	130	74	142	55

the use of the EM algorithm for maximum likelihood or restricted maximum likelihood estimation of α and $\text{var}(\beta_i)$. In their procedure, $\text{var}(\zeta_i)$ is first estimated for each study using the above formulae with each $p_{i,t}$ and $p_{i,c}$ replaced by its estimated value, and these variances are then assumed to be known when estimating α and $\text{var}(\beta_i)$.

Model (2) may be regarded as a special case of model (1) with $\mathbf{X}_i = \mathbf{Z}_i = 1$, $\mathbf{D} = \text{var}(\beta_i)$ and $\sigma^2 = 0$. Likelihood ratio testing for heterogeneity of the β_i is accomplished by first fitting models constraining $\mathbf{D} = 0$ and then allowing $\mathbf{D} > 0$. Hypothesis testing is accomplished by comparing the resulting change in log likelihood (denoted $-2\Delta l$) to a 50:50 mixture of a χ_1^2 and χ_0^2 statistic, since the alternative hypothesis, $\mathbf{D} > 0$, is inherently one sided (Stram and Lee, 1994). (Here χ_0^2 denotes a random variable which always takes the value zero.) This kind of treatment difference analysis is not restricted to the analysis of differences in proportions. It can be considered whenever each study in the meta-analysis tests both treatment arms of interest and a single statistic, such as a regression parameter estimate or estimated odds ratio, summarizes the result of each study's comparison, and the published data includes estimates of the sampling variability of this estimate.

3.2 *The Begg and Pilote Model*

Begg and Pilote (1991) use a random effects model for the purpose of including single treatment historical controls as well as comparative trials in a treatment effect assessment, and provide, as an example, an analysis of allogeneic bone marrow transplantation (BMT) versus chemotherapy in acute nonlymphocytic leukemia. The data that they analyze (Table 2) consist of 20 estimates of the probability of 2-year disease-free survival of patients treated either with BMT or with chemotherapy from 16 different studies, as well as standard errors for each estimate. Four of these studies are comparative in the sense that both BMT and chemotherapy arms are present for the same study, whereas the remaining 12 are single armed studies. The model that they describe for these data can again be written as

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\alpha} + \beta_i + \zeta_i. \tag{3}$$

The mean-effect parameter $\boldsymbol{\alpha}$ is a 2×1 vector, which might be parameterized as $\boldsymbol{\alpha} = (\alpha_0, \delta)$ where α_0 is the average 2-year disease-free survival (DFS) of patients treated by chemotherapy and δ is the mean "improvement" that BMT represents over the survival of patients treated on chemotherapy, i.e., δ is the treatment difference. The random baseline effect β_i varies independently from study to study but is shared by both arms of the comparative studies.

Here $\mathbf{X}_i$ and $\mathbf{Y}_i$ take three possibilities:

1. For the historical controls of chemotherapy, $\mathbf{Y}_i$ contains the estimated 2-year DFS of patients on study i. The design matrix $\mathbf{X}_i$ is a 1×2 vector $= (1, 0)$ indicating that $\mathbf{Y}_i$ contributes to the estimate of mean survival α_0 on the chemotherapy treatment.
2. For the historical controls of BMT, $\mathbf{Y}_i$ contains the estimated 2-year survival for BMT patients, and the design matrix is $\mathbf{X}_i = (1, 1)$ indicating that $\mathbf{Y}_i$ contributes to an estimate of the expected 2-year survival, $\alpha_0 + \delta$, of the BMT arm.
3. For the comparative trials of chemotherapy and BMT, $\mathbf{Y}_i$ is a 2×1 vector with one element corresponding to the disease-free survival on each of the two arms. Here both $(1, 0)$ and $(1, 1)$ appear as separate rows of $\mathbf{X}_i$.

In Begg and Pilote's model, the random baseline β_i is a mean-zero univariate normal random variable taking a distinct value for each study. Unlike the DerSimonian and Laird model, the effect

Table 2
Data from Begg and Pilote (1991), *estimated 2-year disease free survival*

	BMT		Chemotherapy	
Study	Estimate	Std. err.	Estimate	Std. err.
1	.46	.118	.25	.074
2	.50	.100	.23	.067
3	.47	.129	.42	.086
4	.70	.230	.48	.167
5	.46	.081		
6	.43	.034		
7	.49	.088		
8	.53	.079		
9			.21	.051
10			.32	.039
11			.48	.094
12			.26	.046
13			.33	.029
14			.38	.033
15			.24	.084
16			.53	.084

of treatment is fixed across studies, and only the baseline effect of chemotherapy differs for each of the study populations. In this model the expected 2-year DFS of patients treated with chemotherapy on study i is equal to $\alpha_0 + \beta_i$ and is $\alpha_0 + \beta_i + \delta$ for the patients treated with BMT. Thus, each comparative study shares the same random baseline, β_i, in its two arms. The variance, v_{ij}, of each component of $\boldsymbol{\zeta}_i$ represents the sampling variability in estimating Y_{ij} for each study or study arm, and is estimated by the (square of) the standard error of the 2-year survival of the ith study population. This variance is treated as known in the analysis.

Again, model (3) is a special case of model (1) with $\sigma^2 = 0$ and $\mathbf{Z}_i$ either a scalar 1 (for case 1 and 2) or a 2×1 vector of 1's (in case 3).

Additional covariates as fixed effects. Begg and Pilote (1991) extend their analysis by considering several study-level covariates as fixed effects in their model. Such fixed-effects are incorporated in model (1) by adding columns to the covariate matrix, $\mathbf{X}_i$, thereby increasing the number of parameters to be estimated in $\boldsymbol{\alpha}$. For example, the addition of an indicator variable, taking value 0 for the first four studies in Table 2 and value 1 for the remaining rows, permits the baseline chemotherapy survival to differ for the uncontrolled studies compared to the controlled studies. Begg and Pilote consider such covariates in order to test the equivalence of the chemotherapy in the two types of studies. Similarly, adding an indicator variable to $\mathbf{X}_i$ which is zero for all except the single armed BMT studies (studies 5–8 in Table 2) allows for testing whether the effect of BMT is different in the uncontrolled vs. the comparative studies. The analysis presented by Begg and Pilote conclude that neither of these fixed-effect covariates make a significant addition to the model based upon a likelihood ratio test.

Additional random effects. The final extension that Begg and Pilote considered was to allow the difference, δ, between the efficacy of BMT and chemotherapy to vary randomly from study to study just as in the DerSimonian and Laird model. Denoting the new random effect by γ_i, the expected 2-year DFS of patients treated with BMT on study i becomes $\alpha_0 + \delta + \beta_i + \gamma_i$ and remains $\alpha_0 + \beta_i$ for the chemotherapy patients. In their analysis of the data in Table 2, Begg and Pilote found no evidence that $\text{var}(\gamma_i) > 0$, and concluded that the effect of BMT was the same in all the studies that they examined.

The inclusion of random γ_i is accomplished in model 1 by adding an additional column to each random-effects covariate matrix $\mathbf{Z}_i$ so that it includes an indicator variable which equals 1 for the BMT arms of the comparative or the single armed BMT studies and is 0 for the chemotherapy arms. This adds a second component to $\boldsymbol{\beta}_i$ and additional terms in the covariance matrix $\mathbf{D}$. Begg and Pilote assumed that β_i and γ_i are independent of each other, which corresponds to restricting the off-diagonal component, $d_{1,2}$, of $\mathbf{D}$ to be equal to zero. In the general model, $\mathbf{D}$ is not constrained to be 0 but it is important that algorithms for fitting model (1) allow for the imposition of such restrictions on the components of the covariance matrix.

3.3 *Tori et al. Analysis of Response as a Predictor of Patient Survival*

Tori et al. (1992) analyze data from 25 randomized clinical trials (Table 3) for the purpose of investigating the relationship between response and survival in patients with advanced ovarian cancer. The emphasis is on whether response might be used as a surrogate endpoint for the purpose of designing or monitoring clinical trials. The trials that they considered each involved randomized comparisons of two or more (maximum 4) cytotoxic treatment regimens. Let $\mathbf{Y}_i$ be a $n_i \times 1$ vector containing the observed log median survival for each of the n_i arms of study i.

The model that they consider may be written as

$$\mathbf{Y}_i = \alpha_0 + \beta_i + \alpha_1\{\log(\boldsymbol{\rho}_i) - \log(1 - \boldsymbol{\rho}_i)\} + \boldsymbol{\zeta}_i + \mathbf{e}_i, \tag{4}$$

where the fixed-effect parameters α_0 and α_1 specify a linear relationship between $\mathbf{Y}_i$ and $\boldsymbol{\rho}_i$, the logit of the vector of true but unknown probabilities of response to therapy for each of the arms of study i. The univariate β_i provide for systematic differences in survival among studies, but this term is shared by each arm of study i. The difference between model (4) and the linear random-effects model (1) is that $\boldsymbol{\rho}_i$ in model (4) is taken as unobserved true response to therapy.

In the Tori et al. analysis, the errors in estimating $\boldsymbol{\rho}_i$ using the observed response rates, $\mathbf{r}_i$, among patients treated on each study arm are regarded as another source of variability in $\mathbf{Y}_i$. Because these errors occur in a predictor variable rather than in the outcome variable, this imposes an errors-in-variables structure to their analysis. The linear random-effects model does not deal with this complication, and therefore the results of using model (1) are necessarily somewhat different from those obtained by Tori et al. Tori et al.'s use of true $\boldsymbol{\rho}_i$ rather than $\mathbf{r}_i$ in the model implies that all patients who are treated on a therapy with a high $\boldsymbol{\rho}_i$ have a good expected survival, even if they themselves fail to respond to the therapy. Similarly, a patient who responds to a therapy which does not produce a response in other patients is evidently predicted to have the same poor survival as in the nonresponders.

An alternative analysis, based on a model at the individual level in which the expected log survival time for a patient differs by a constant factor depending upon whether or not she responds to therapy, would lead to an approximately linear (rather than logit) parameterization of the relationship between median log survival and r_i in the studies. Thus, an alternative to model (1) is

$$\mathbf{Y}_i = \alpha_0 + \beta_i + \alpha_1\mathbf{r}_i + \boldsymbol{\zeta}_i + \mathbf{e}_i. \tag{5}$$

Table 4 reports the results of two different reanalyses of the data in Table 3 using the linear mixed-effects model (1). In both analyses $\boldsymbol{\alpha}^{\mathrm{T}} = (\alpha_0, \alpha_1)$, $\boldsymbol{\beta}_i = \beta_i$, and $\mathbf{Z}_i = \mathbf{1}_i$. For the first reanalysis (using the logit model)

$$\mathbf{X}_i = \begin{bmatrix} 1 & \log(r_{i,1}) - \log(1 - r_{i,1}) \\ 1 & \log(r_{i,2}) - \log(1 - r_{i,2}) \\ \vdots & \vdots \\ 1 & \log(r_{i,n_i}) - \log(1 - r_{i,n_i}) \end{bmatrix}$$

and for the second (the linear model)

$$\mathbf{X}_i = \begin{bmatrix} 1 & r_{i,1} \\ 1 & r_{i,2} \\ \vdots & \vdots \\ 1 & r_{i,n_i} \end{bmatrix}.$$

The primary interests in either of these models are the estimate of α_1, which relates the observed response rate to the differences in log median survival, and in testing whether $\sigma^2 = 0$. The fit of the logit model essentially reproduces the findings of Tori et al.: there is a very strongly significant, positive relationship between response and survival. Tori et al. report an estimate of $\alpha_1 = 0.23$ in model (4) (95% confidence interval [CI] = 0.13–0.35). The estimate of α_1 given here for the logit model is 0.19 (95% CI 0.11–0.26). The comparison of these results is as expected: the errors-in-variables correction employed by Tori et al. somewhat increases both the point estimate of α_1 and its standard error relative to the results given here.

The other parameter of interest, σ^2, allows for additional variability between treatment arms beyond what is explained by response alone. This variance parameter is fixed at zero in the results shown in Table 4. Attempting to free this parameter does not alter the results; the Fisher's scoring

Table 3

Data from Tori et al. (1992), *median survival times for* 25 *studies of ovarian cancer*

Study no.	Response	No. of deaths	Median survival, mo
1	53.1	46	12.0
1	34.7	47	12.0
2	75.0	18	20.0
2	54.0	26	17.0
3	36.1	34	12.0
3	31.4	34	12.0
4	66.7	17	19.0
4	31.8	20	19.0
5	37.1	82	12.0
5	21.7	93	11.0
6	30.8	15	20.0
6	80.0	16	19.0
6	35.7	14	11.0
7	62.1	27	11.0
7	35.7	27	10.0
8	69.2	10	18.0
8	23.1	10	17.0
9	37.5	75	14.0
9	51.5	120	13.0
9	48.6	92	12.0
10	37.8	72	30.0
10	31.0	59	26.0
11	78.6	54	31.0
11	50.0	71	20.0
12	75.5	30	19.0
12	67.6	30	12.0
13	48.7	72	15.3
13	55.8	74	14.4
13	44.4	66	12.4
13	35.0	71	11.8
14	70.0	40	24.0
14	66.1	43	23.0
15	46.6	70	14.0
15	27.1	58	12.0
16	65.0	34	13.0
16	23.3	38	11.0
17	75.7	154	19.3
17	47.5	168	16.4
18	56.3	29	26.0
18	54.3	37	22.0
19	74.6	34	25.6
19	67.6	33	21.0
20	79.5	57	31.0
20	74.5	63	24.0
21	71.0	103	23.8
21	61.5	119	21.4
21	51.4	119	19.4
22	54.8	70	18.5
22	26.7	65	10.3
23	62.5	11	26.5
23	87.5	12	14.0
23	87.5	14	12.5
24	47.4	13	19.0
24	47.1	15	18.0
25	72.8	24	31.0
25	61.0	34	21.0

Table 4
Results of fitting random effects models to ovarian cancer study-level data

Model	Parameter	Estimate	Std. Err.	Est/Std. Err.	log lik
Logit	σ^2	0.	Fixed	—	3.29
	D	0.053728	0.018799	2.858072	
	α_0	2.808998	0.052161	53.852844	
	α_1	0.186744	0.037601	4.966475	
Linear	σ^2	0.	Fixed	—	3.64
	D	0.052849	0.018862	2.801929	
	α_0	2.393767	0.101639	23.551773	
	α_1	0.832747	0.165203	5.040759	

algorithm finds that the fit of the model cannot be improved by increasing σ^2 (the maximum value of the likelihood is at $\sigma^2 = 0$). A profile likelihood-based confidence procedure for σ^2 gives an approximate 95% CI from zero to 0.0058 since at $\sigma^2 = 0.0058$ the likelihood is reduced below the critical value of 2.28 based on an asymptotic distribution of $-2\Delta l$ as a 50:50 mixture of a χ^2_1 and a χ^2_0 statistic. At $\sigma^2 = 0.0058$, the estimate of α_1 is essentially unchanged (estimated = 0.19, 95% CI 0.10–0.27).

The linear model gives a very slightly better fit to the data in Table 3, as judged by the log likelihood, but the results are quite similar to the logit model overall. The relationship between survival and response is very strong and the maximum likelihood estimate for σ^2 is again equal to zero, indicating that all the between-treatment-arm variability of log median survival is apparently explained by the differences in response rates.

4. Discussion

The linear mixed-effects model (1) is an extension of the model discussed by Laird and Ware (1982), and is a special case of the linear structured variance components model (see Appendix). Besides the meta-analysis framework described here, this model may be used in other types of "two-stage" analyses of summary statistics where each summary is calculated from a series of observations on one of many independent subjects. The use of these methods is dependent upon the sample sizes for the individual studies or subjects being large enough that normality may be assumed for the summary statistics and that the variances of these estimates may be treated as known. At present there is little general guidance available concerning how large individual studies should be before these assumptions can be willingly accepted. In specific cases it may be worthwhile to conduct simulation studies of these issues. There has been increasing interest in meta-analysis in which raw, rather than summary, data from all the existent studies of a particular issue are assembled in a joint analysis. While some issues, such as whether the variance of the summary estimates may be treated as known, are no longer important in this setting, these meta-analyses still may involve complex model specification tasks, including the evaluation of random effects models for differences between studies in treatment responses. In the case of continuous data, the approach of Laird and Ware (1982) or the multilevel models of Goldstein (1986) may be taken as a starting point. For categorical data with a random effects structure, methods such as that of Breslow and Clayton (1993) may be considered.

In any particular meta-analysis, not all the variance components, σ^2, and the elements of $\mathbf{D}$ in model (1) may actually be needed (and can be set to zero in the Fisher's scoring procedure). In the analyses described here, only one example (the reanalysis of the ovarian cancer data) utilized σ^2 and $\mathbf{D}$ simultaneously, and in only one case (the elaboration of the Begg and Pilot model) was $\mathbf{D}$ allowed to be a 2×2 matrix rather than a scalar. As meta-analysis of published literature becomes increasingly complex, however, the flexibility of this model will become increasingly attractive.

There are several widely available software packages or programs which can fit either the linear mixed-effects model (SAS, 1995) or the general linear structured variance components model (BMPD 5V; see Dixon, 1988). Neither of these programs appears to have the ability to include the known variances, $\mathbf{V}_i$, in the model; thus, most use of random effects in meta-analysis has employed specially written software. Implementation of the Fisher's scoring methods described in the Appendix in programs written in GAUSS (Aptech Systems, 1991) are available from the author. These procedures have recently been incorporated into the MS DOS-based clinical trials and epidemiology package, Epilog (see Epicenter, 1994).

Acknowledgment

This research was supported by National Cancer Institute Grant CA13539.

Résumé

Cet article décrit l'utilisation d'un modèle mixte de régression linéaire comme un cadre pour les méta-analyses de données publiées. Il généralise les modèles à effets aléatoires utilisés par DerSimonian et Laird (1986, *Controlled Clinical Trials* **7,** 177–188), Begg et Pilote (1991, *Biometrics* **47,** 899–906), et décrit l'utilisation du modèle à partir d'exemples provenant de ces articles et avec une nouvelle analyse des données fournies par Tori et al (1992, *Journal of the National Cancer Institute* **84,** 407–414).

References

Aptech Systems. (1991) *GAUSS Reference Manual.* Kent, Washington: Aptech Systems Inc.

Anderson, T. W. (1973). Asympotic efficient estimation of covariance matrices with linear structure. *Annals of Statistics* **1,** 135–141.

Begg, C. B. and Pilote, L. (1991). A model for incorporating historical controls into a meta-analysis. *Biometrics* **47,** 899–906.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88,** 9–25.

Colditz, G. A., Brewer, T. F., Berkey, C. S., Wilson, M. E., Burdick, E., Fineberg, H. V., and Mosteller F. (1994). Efficacy of BCG vaccine in the prevention of tuberculosis, meta-analysis of the published literature. *Journal of the American Medical Association* **271,** 698–702.

DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* **7,** 177–188.

Dixon, W. J. (ed). (1988). *BMDP Statistical Software Manual,* Volume 2. Berkeley: University of California Press.

Epicenter. (1994). *Epilog Plus, Statistics Package for Epidemiology and Clinical Trials.* Pasadena, California: Epicenter Software.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika* **73,** 43–56.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* **72,** 320–338.

Jennrich, R. I. and Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* **42,** 805–820.

Laird, N. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38,** 963–974.

Lindstrom, M. J. and Bates, D. M. (1988). Newton–Raphson algorithms for linear mixed-effects models for repeated measures data. *Journal of the American Statistical Association* **83,** 1014–1022.

Rao, P. S. R. S., Kaplan, J., and Cochran, W. G. (1981). Estimators for the one-way random effects model with unequal error variances. *Journal of the American Statistical Association* **76,** 89–97.

SAS. (1995). *Technical Report P-229, SAS/STAT Software Changes and Enhancements*, Release 6.07. Cary, North Carolina: SAS Institute Inc.

Stram, D. O. and Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics* **50,** 1171–1177.

Szatrowski, T. H. (1978). Explict solutions, one iteration convergence, and averaging in the multivariate normal estimation problem for patterned means and covariances. *Annals of the Institute of Statistical Mathematics, Part A* **30,** 81–88.

Tori, V., Simon, R., Russek-Cohen, E., Midthune, D., and Friedman, M. (1992). Statistical model to determine the relationship of response and survival in patients with advanced ovarian cancer treated with chemotherapy. *Journal of the National Cancer Institute* **84,** 407–414.

Winship, D. A. (1978). Cimetidine in the treatment of duodenal ulcer. *Gastroenterology* **74,** 402–406.

Received July 1994; *revised May* 1995; *accepted August* 1995.

Appendix
Fisher's Scoring Algorithm

The linear mixed-effects model used in this paper is a special case of the linear structured variance components model which has been extensively studied by a number of statistical authors (Anderson, 1973; Szatrowski, 1978; Jennrich and Schluchter, 1986). Such a model specifies a linear relationship, $E(Y_i) = X_i\alpha$, between known covariate matrices X_i and the mean of Y_i, and between known "variance covariate" matrices, G_{ij}, and the covariance, Σ_i, of Y_i as in

$$\Sigma_i = \sum_{g=0}^{\omega} \theta_j G_{ig}. \tag{6}$$

In model (1)

$$\Sigma_i = \sigma^2 I + Z_i D Z_i^{\mathrm{T}} + V_i. \tag{7}$$

Now $Z_i D Z_i^{\mathrm{T}}$ can be written as

$$d_{11} Z_i K_{11} Z_i^{\mathrm{T}} + d_{12} Z_i K_{12} Z_i^{\mathrm{T}} + d_{22} Z_i K_{22} Z_i^{\mathrm{T}} + \cdots + d_{qq} Z_i K_{qq} Z_i^{\mathrm{T}},$$

where K_{ij} has a one in the i, j position and is zero elsewhere. Therefore each Σ_i is of linear form with $\omega = q(q+1)/2 + 1$, $\theta_0 = \sigma^2$, $G_{i0} = I$, $\theta_1 = d_{11}$, $G_{i1} = Z_i K_{11} Z_i^{\mathrm{T}}$, $\theta_2 = d_{12},$, $G_{i2} = Z_i K_{12} Z_i^{\mathrm{T}}, \ldots,$ $\theta_{q(q+1)/2} = d_{q,q}$, $G_{i,\omega} = V_i$ for $i = 1, \ldots, n$ and $\theta_w \equiv 1$. The score equations for estimating α and $\theta = (\theta_0, \theta_1, \ldots, \theta_\omega)^t$ are

$$\sum_{i=1}^{n} X_i^{\mathrm{T}} \widehat{\Sigma}_i^{-1} X_i \widehat{\alpha} - X_i^{\mathrm{T}} \widehat{\Sigma}_i^{-1} Y_i = 0$$

and

$$\sum_{i=1}^{n} -\frac{1}{2} tr G_{ig} \widehat{\Sigma}_i^{-1} + \frac{1}{2} tr G_{ig} \widehat{\Sigma}_i^{-1} G_{ig} \widehat{C}_i = 0 \tag{8}$$

for $g = 1, \ldots, \omega$, where $\widehat{C}_i = (Y_i - X_i\widehat{\alpha})(Y_i - X_i\widehat{\alpha})^{\mathrm{T}}$. Anderson (1973) notes that solving Equation (8) may be accomplished by iteratively solving the linear equations

$$[\sum_{i=1}^{n} tr \widehat{\Sigma}_i^{-1} G_{ig} \widehat{\Sigma}_i^{-1} G_{ih}] \widehat{\theta} - (\sum_{i=1}^{n} tr \widehat{\Sigma}_i^{-1} G_{ig} \widehat{\Sigma}_i^{-1} \widehat{C}_i) = 0, \tag{9}$$

where $[\Sigma_{i=1}^{n} tr \widehat{\Sigma}_i^{-1} G_{ig} \widehat{\Sigma}_i^{-1} G_{ih}]$ is a $\omega \times \omega$ matrix and $(\Sigma_{i=1}^{n} tr \widehat{\Sigma}_i^{-1} G_{ig} \widehat{\Sigma}_i^{-1} \widehat{C}_i)$ is a $\omega \times 1$ column vector. Any variance components which are to be fixed at known values (such as $\widehat{\theta}_\omega$, which is identically equal to 1) are deleted from $\widehat{\theta}$ and the dimensions of $[\Sigma_{i=1}^{n} tr \widehat{\Sigma}_i^{-1} G_{ig} \widehat{\Sigma}_i^{-1} G_{ih}]$ and $(\Sigma_{i=1}^{n} tr \widehat{\Sigma}_i^{-1} G_{ig} \widehat{\Sigma}_i^{-1} \widehat{C}_i)$ are correspondingly reduced. Resticted maximum likelihood estimation is accomplished in one of two ways: either by replacing Y_i, X_i, and Σ_i by $R_i Y_i$, $R_i X_i = 0$, and $R_i \Sigma_i R_i^T$ in Equation (9) and redefining the G_{ig} as $R_i G_{ig} R_i^{\mathrm{T}}$, where R_i is a full rank $(n_i - p) \times n_i$ matrix orthogonal to X_i, or by directly adding the derivative of $-1/2 \ln \det(\Sigma_{i=1}^{n} X_i^{\mathrm{T}} \widehat{\Sigma}_i^{-1} X_i)$ with respect to θ to the scores in Equation (8).