

METHODOLOGICAL DEVELOPMENTS

Random-Effects Regression Models for Clustered Data With an Example from Smoking Prevention Research

Donald Hedeker, Robert D. Gibbons, and Brian R. Flay

A random-effects regression model is proposed for analysis of clustered data. Unlike ordinary regression analysis of clustered data, random-effects regression models do not assume that each observation is independent but do assume that data within clusters are dependent to some degree. The degree of this dependency is estimated along with estimates of the usual model parameters, thus adjusting these effects for the dependency resulting from the clustering of the data. A maximum marginal likelihood solution is described, and available statistical software for the model is discussed. An analysis of a dataset in which students are clustered within classrooms and schools is used to illustrate features of random-effects regression analysis, relative to both individual-level analysis that ignores the clustering of the data, and classroom-level analysis that aggregates the individual data.

Clustered designs occupy an important role in clinical and public health research. In these designs, the program target is typically the individual, whereas randomization and implementation of the intervention often occurs at the cluster level, which may be a hospital, clinic, firm, or school. For example, a smoking prevention intervention may be applied and randomized at the clinical level, and the primary outcome is assessing the smoking behavior of subjects within the clinics. Usually the number of individuals within the clusters is not constant but varies with each cluster. Much has been written concerning the appropriate level of statistical analysis for this type of clustered data (Burstein, 1980; Haney, 1980; Hopkins, 1982; Jacobs, Jeffery, & Hannan, 1989; Koepke & Flay, 1989; McKinlay, Stone, & Zucker, 1989; Murray, Hannan, & Zucker, 1989), with varying recommendations. When the purpose of the intervention is to produce an effect on individuals, analysis of data at the individual level is more readily interpretable than analysis at the cluster level. Furthermore, there is usually interest in assessing effects of both cluster-level variables (e.g., intervention and classroom size) and individual-level variables (e.g., sex, age, and IQ), and additionally, interaction of the two (e.g., sex \times inter-

vention), on the individual-level outcome. Because individuals are observed within clusters, however, individual-level analysis should take this clustering of the data into account.

If the sample size within clusters is balanced, one may analyze individual data using analysis of variance models that account for nesting of subjects within clusters, using a randomized block or split-plots model, for example. These models are referred to as mixed-models because they include both fixed (i.e., intervention group) and random (i.e., cluster) effects. Nesting of subjects within clusters is accounted for by assigning the appropriate error term for tests of main effects and interactions within the model. As Hopkins (1982) indicated, when sample size within clusters is the same, proper use of mixed-model analysis of variance yields identical results for the test regarding the effect of the cluster-level intervention, regardless of whether cluster means or individual observations are analyzed. Thus, for balanced designs, the mixed-model resolves the issue concerning the level of analysis. Unfortunately, the number of subjects within clusters is usually unbalanced, and use of the mixed-model analysis of variance approach becomes problematic (Searle, 1987).

In analysis of unbalanced clustered data, one approach has been to analyze data at the level of the individual, ignoring the clustering of the data using common statistical techniques such as linear regression or fixed-effects analysis of variance models. Because these models assume that all observations are independent, this type of analysis at the individual level does not take into account dependency in the data resulting from subjects having been clustered. The amount of dependency in the data that is observable because of the clustering is measured by the intraclass correlation. For certain variables, intraclass correlation levels have been observed at from 5% to 12% for data from spouse pairs and 0.05% to 0.85% for data clustered by counties (Donner, 1982). As the intraclass correlation increases, the amount of independent information from the data decreases, inflating the Type I error rate of an analysis that ignores this

Donald Hedeker and Brian R. Flay, School of Public Health and Prevention Research Center, University of Illinois at Chicago; Robert D. Gibbons, Department of Psychiatry and School of Public Health, University of Illinois at Chicago.

Preparation of this article was supported by National Institute of Mental Health Grant MH44826-01A2 and National Institute of Drug Abuse Grant DA06307.

Portions of this article were presented in November 1991 at the American Public Health Association Continuing Education Program, Atlanta, Georgia.

Correspondence concerning this article should be addressed to Donald Hedeker, Division of Epidemiology and Biostatistics (M/C 922), Room 510, School of Public Health, University of Illinois at Chicago, 2121 West Taylor Street, Chicago, Illinois 60612-7260.

correlation (Blair, Higgins, Topping, & Mortimer, 1983). Thus, individual-level analysis yields tests of significance that are generally too liberal.

A second approach has been to analyze data at the cluster level. Here, the outcome variable is an aggregate of individual outcomes; for example, mean reading level of students within a class, or the percentage of employees practicing a given behavior within a firm. Several problems are inherent with this type of aggregate analysis. First, individual differences are lost with this approach, and the effect of other individual variables (e.g., sex or motivation) on the outcome cannot be examined in a straightforward way. The power of statistical tests at this level can be low because the number of clusters within each intervention condition is typically small. Thus, analysis at this level can increase the Type II error rate because with a limited sample size, there is statistical power to detect only large differences.

Recently, random-effects regression models (RRM) have been developed and proposed for analysis of unbalanced clustered data in educational (Aitkin & Longford, 1986; Bock, 1989a; DeLeeuw & Kreft, 1986; Goldstein, 1987; Raudenbush & Bryk, 1986), psychological (Bock, 1983; Bryk & Raudenbush, 1987), biological (Laird & Ware, 1982; Longford, 1987), epidemiological (Donner, 1985), and psychiatric (Gibbons, Hedeker, Waternaux, & Davis, 1988; Hedeker, Gibbons, & Davis, 1991; Hedeker, Gibbons, Waternaux, & Davis, 1989) literatures. RRMs are useful in the analysis of clustered data because outcomes at the individual level are modeled in terms of both individual- and cluster-level variables while concurrently estimating and adjusting for the amount of intraclass correlation present in the data. Furthermore, these models make no assumption regarding cluster sample size, allowing for a varying number of subjects within each cluster.

Because data can be clustered at more than a single level (e.g., students within classrooms within schools), some researchers have generalized the RRM for multiple levels of clustering (Goldstein, 1987; Longford, 1987). Because of a focus on the levels of clustering in the data, as opposed to an emphasis on the statistical nature of the model, these generalized models are sometimes referred to simply as multilevel models and the data as multilevel data. With this terminology, data clustered within a single unit (e.g., classes, firms, or clinics) are referred to as two-level data and the statistical model as a two-level model, because the individuals (Level 1) are measured within clusters (Level 2). With emphasis on the clustering, or nesting, of observations within various data levels, Longford's development (1987) indicates how the multilevel model extends the balanced random and mixed analysis of variance models, like those described by Hopkins (1982), for unbalanced data. As such, use of multilevel models or RRMs can be seen as resolving the unit of analysis issue for unbalanced clustered data.

Although the RRM provides a powerful statistical tool in analysis of clustered data, its use has been limited in prevention and clinical research. Part of the problem is that this type of model is not readily known or understood by researchers in these fields. In particular, researchers are often unclear as to how the random-effects model relates to statistical models with which they are already familiar (e.g., regression and analysis of variance). Another problem is that computer software to perform random-effects regression analysis has been only recently

developed, so researchers are generally unaware of the potential for its use. In this article, we attempt to address these problems by concentrating on describing the main features of the statistical model, showing the connection between the RRM and the usual linear regression model, illustrating use of the random-effects model in comparison with more conventional analyses, and providing a means to perform the analysis using readily available statistical software. As we discuss application of both two- and three-level RRMs to aid in the understanding of the model, we limit the statistical development to the two-level RRM for clustered data. As a result, this article is intended more as a descriptive introduction to the RRM; for reviews and details regarding more general random-effects models, see the edited collections of Bock (1989b) or Raudenbush and Willms (1991) or the texts of Goldstein (1987) or Bryk and Raudenbush (1992).

For the two-level model, we outline a solution of the model parameters using empirical Bayesian and maximum marginal likelihood estimation procedures. Solutions using the EM algorithm (Dempster, Rubin, & Tsutakawa, 1981) and the Fisher scoring method (Longford, 1987) are presented. We illustrate the application of both the two- and three-level models, using an example from a school smoking-prevention study (Flay et al., 1989), and we compare and discuss the results from the RRM. Finally, computer programs that are available for random-effects analysis are described, including a two-level program using SPSS program code.¹

The RRM

In the present context, we consider patients clustered within clinics, with the intervention being performed in the clinics. For now, we ignore the possibility that clinics can be clustered within hospitals, hospitals within counties, and so forth. Our interest is in modeling the outcome of patients (y) in terms of a clinic-level variable (x_1) and a patient-level variable (x_2). Consider the following model for patient j (where $j = 1, 2, \dots, n_i$ for clinic i) within clinic i ($i = 1, \dots, N$ clinics in the sample),

$$y_{ij} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2ij} + \alpha_i + \epsilon_{ij}, \quad (1)$$

where y_{ij} equals the observed outcome variable for patient j within clinic i , β_0 is the intercept of the regression model, β_1 is the coefficient for the clinic-varying covariate x_{1i} , β_2 is the coefficient for the patient-varying covariate x_{2ij} , α_i represents the effect caused by clinic i , and ϵ_{ij} is an independent residual distributed normally, $N(0, \sigma^2)$, in the population of patients.

As sampled clinics are thought representative of a larger population of clinics, α_i is considered a random parameter, and the assumed distribution for α is a normal distribution with mean 0 and variance σ_α^2 ; that is, $N(0, \sigma_\alpha^2)$. Notice that it is the inclusion of the random term α_i that separates this model from the usual (fixed-effects) multiple regression model and that it is this term that represents the effect of the patient clustering within clinics on the outcome variable. To the degree that patient clustering within clinics has little effect on the outcome data, estimates of

¹ The two-level program using SPSS program code can be obtained from the authors.

α_i will all be near 0 and the estimate of clinic variance (σ_α^2) will approach 0. If, on the other hand, patient clustering within clinics has a strong effect on the outcome data, estimates of α_i will deviate from 0 and differ for each clinic i ; thus, clinic variance (σ_α^2) will increase in value.

The model can be written in a slightly more general form using matrix notation. For this, consider the following model with p covariates for the $n_i \times 1$ response vector \mathbf{y} for clinic i , $i = 1, 2, \dots, N$,

$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{1}_i\alpha_i + \epsilon_i, \quad (2)$$

where \mathbf{y}_i = the $n_i \times 1$ vector of responses for clinic i , \mathbf{X}_i = a known $n_i \times (p + 1)$ covariate matrix, β = a $(p + 1) \times 1$ vector of unknown population parameters, $\mathbf{1}_i$ = a $n_i \times 1$ vector of ones, α_i = the unknown individual clinic effect distributed $N(0, \sigma_\alpha^2)$, and ϵ_i = a $n_i \times 1$ vector of independent residuals distributed as $N(0, \sigma^2 I_{n_i})$. Because the clinic subscript i is present for the \mathbf{y} vector and the \mathbf{X} matrix, each clinic can have a varying number of patients nested within. Furthermore, the covariate matrix \mathbf{X} can include covariates measured at either the clinic or patient level, with the total number of covariates equal to p . The number of columns in \mathbf{X} is $p + 1$ because the intercept is the first parameter of the vector β , so the first column of \mathbf{X} consists only of ones, and remaining p columns are for covariates.

The response vectors from the clinics, denoted \mathbf{y}_i , are distributed as independent normals with mean $\mathbf{X}_i\beta$ and variance-covariance matrix $\sigma_\alpha^2 \mathbf{1}_i \mathbf{1}_i' + \sigma^2 I_{n_i}$ (by *independent*, it is meant that \mathbf{y}_i is independent of $\mathbf{y}_{i'}$, where $i \neq i'$). The elements of the mean vector given by $\mathbf{X}_i\beta$ agree in form with the usual means (for the patients within clinic i) that would be obtained if an ordinary multiple regression model at the individual level was performed. However, because the parameter estimates of β will not, in general, be the same from both approaches, it is the form (i.e., $\mathbf{X}_i\beta$) and not the actual values of the mean vector that is in agreement from both approaches. The variance-covariance structure, however, explicitly takes into account the clustering of patients within clinics and is the same form as that corresponding to a randomized block analysis of variance model (see Kirk, 1982, pp. 253–257). As such, the interdependency of the patient responses within a clinic is expressed by the variance-covariance structure. Specifically, for a given clinic i , the variance associated with each of the $j = 1, \dots, n_i$ patient observations is assumed to be the same with form $\sigma^2 + \sigma_\alpha^2$, whereas the covariance of any two patient observations within a clinic is assumed to be σ_α^2 . This structure is sometimes referred to as the *compound symmetry* form, because within a clinic, variances are assumed to be homogeneous ($\sigma^2 + \sigma_\alpha^2$), as are all covariances (σ_α^2). The ratio of the clinic (or cluster) variance σ_α^2 to the total variance $\sigma^2 + \sigma_\alpha^2$ is the *intraclass correlation*, which indicates proportion of variance in the data attributable to the clinic or cluster.

Estimation

In terms of parameter estimation, a combination of two complementary methods has been proposed (Bock, 1989a; Laird & Ware, 1982). For estimation of clinic effects α_i , empirical Bayes (EB) methods have been recommended, whereas maximum marginal likelihood (MML) methods are recommended for

estimation of variance parameters, σ^2 and σ_α^2 , and covariate effects β .

EB estimates of clinic effects are sometimes termed *EAP* (“expected a posteriori”) estimates, because they are derived as the mean of the posterior distribution of α , given \mathbf{y}_i . Denoting the EAP estimate of α_i as $\tilde{\alpha}_i$ to distinguish it from subsequent MML estimates, and given the aforementioned assumptions, we get the following EAP estimator of clinic parameters,

$$\tilde{\alpha}_i = \rho_{n_i} \frac{1}{n_i} \mathbf{1}_i' (\mathbf{y}_i - \mathbf{X}_i\beta) = \rho_{n_i} \frac{1}{n_i} \sum_{j=1}^{n_i} (y_{ij} - \mathbf{x}_{ij}\beta), \quad (3)$$

where \mathbf{x}_{ij} is the covariate vector for a patient j within clinic i , and ρ_{n_i} is equivalent to the Spearman-Brown reliability formula (Guilford, 1954), given as $\rho_{nn} = nr/[1 + (n - 1)r]$ with r as the intraclass correlation.

A property of EB estimation is that $\tilde{\alpha}_i$ is a function of both the actual data and the empirical prior distribution specified for α_i . As information about a clinic increases (i.e., the reliability ρ_{n_i} increases toward 1), by increasing data interdependency within the clinic (increasing r), increasing clinic sample size (n_i), or both, the clinic estimate approaches the average patient deviation for that clinic, $(\sum_{j=1}^{n_i} y_{ij} - \mathbf{x}_{ij}\beta)/n_i$. Alternatively, as information about a clinic decreases (i.e., ρ_{n_i} decreases toward 0), by decreasing data interdependency within the clinic, decreasing clinic sample size, or both, the clinic estimate approaches the posited mean of the empirical prior distribution of α_i , namely 0.

In addition to the EB estimate of the posterior mean, the variance of the posterior distribution of α is given as

$$\sigma_{\alpha|y_i}^2 = \sigma_\alpha^2 (1 - \rho_{n_i}). \quad (4)$$

Again, the form reveals the nature of this EB estimator of the posterior variance: As information about the clinic increases, the posterior variance becomes a fraction of the empirical prior variance (σ_α^2), whereas as information about the clinic decreases, this variance approaches the empirical prior variance.

For estimating covariate effects β and variance parameters σ_α^2 and σ^2 , we recommend MML estimation. The MML estimation procedure is more fully presented in the appendix. We describe it by using two numerical algorithms: the EM algorithm solution (Dempster et al., 1981) and the Fisher scoring solution (Longford, 1987). From the EM algorithm solution, one can see how the RRM can be viewed as a generalization of the ordinary multiple linear-regression model. Namely, the following equations are used in the iterative EM algorithm solution:

$$\hat{\beta} = \left[\sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i \right]^{-1} \left[\sum_{i=1}^N \mathbf{X}_i' (\mathbf{y}_i - \mathbf{1}_i \tilde{\alpha}_i) \right] \quad (5)$$

$$\hat{\sigma}_\alpha^2 = \frac{1}{N} \sum_i \tilde{\alpha}_i^2 + \sigma_{\alpha|y_i}^2 \quad (6)$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_i (\mathbf{y}_i - \mathbf{X}_i \hat{\beta} - \mathbf{1}_i \tilde{\alpha}_i)' (\mathbf{y}_i - \mathbf{X}_i \hat{\beta} - \mathbf{1}_i \tilde{\alpha}_i) + n_i \sigma_{\alpha|y_i}^2 \quad (7)$$

with the solution proceeding by iterating between EB Equations 3 and 4 and MML Equations 5–7 until convergence. Notice, as estimates of clinic effects α_i and variances $\sigma_{\alpha|y_i}^2$ approach zero, the cluster variance estimate (σ_α^2) approaches zero, and the equations for regression coefficients β and residual variance σ^2

approach the maximum likelihood solution of these parameters in the usual fixed-effects regression model, namely, $\hat{\beta} = [\sum_{i=1}^N X_i'X_i]^{-1} \sum_{i=1}^N X_i'y_i$ and $\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (y_i - X_i\hat{\beta})(y_i - X_i\hat{\beta})$.

Thus, as dependency caused by clustering of the data decreases, the RRM solution approaches the individual-level analysis solution using an ordinary multiple linear-regression model.

For determining the significance of specific model parameters, the Fisher scoring solution provides the large-sample variances and covariances of the MML estimates, which can be used to construct confidence intervals and tests of hypotheses for the model parameters (Wald, 1943). Standard errors for the MML estimates are used to determine asymptotically normal test statistics (estimate divided by its standard error) for each parameter. These test statistics can then be compared with a standard normal frequency table to test the null hypothesis that a given parameter equals 0.

To test for statistical difference between alternative models, one can use the likelihood ratio chi-square test (Silvey, 1975) in certain cases. This test is appropriate when a model (model B, for example) includes all the parameters of another model (model A, for example) plus some additional terms. The likelihood ratio test compares the relative fit of the data provided by models B and A, and thus determines the significance of including these additional terms into the statistical model of the data. For this, it can be shown that the log likelihood is given as

$$\log L = -\frac{1}{2} \sum_{i=1}^N [n_i \log(2\pi\sigma^2) + \log \sigma_\alpha^2 - \log \sigma_{\alpha|y_i}^2 + \tilde{\alpha}_i^2/\sigma_\alpha^2 + \mathbf{u}_i'\mathbf{u}_i/\sigma^2],$$

where $\mathbf{u}_i = \mathbf{y}_i - X_i\hat{\beta} - \mathbf{1}_i\tilde{\alpha}_i$. Evaluating this likelihood using the estimated parameters of the two models yields $\log L_B$ and $\log L_A$. The significance of the additional terms in model B is determined by comparing $-2(\log L_B - \log L_A)$ to a table of the chi-square distribution with degrees of freedom equal to the number of additional parameters in model B. If this likelihood ratio statistic exceeds the critical value of the chi-square distribution, the additional terms significantly improve model fit.

Example

The Data Set

The Television School and Family Smoking Prevention and Cessation Project (TVSFP) study was designed to test independent and combined effects of a school-based social-resistance curriculum and a television-based program in terms of tobacco use prevention and cessation.² The initial study sample consisted of seventh-grade students who were pretested in January 1986. Students who took the pretest completed an immediate postintervention questionnaire in April 1986, a 1-year follow-up questionnaire (in April 1987), and a 2-year follow-up (in April 1988). The study involved students of schools from Los Angeles and San Diego, California. Randomization to various design conditions was at the school level, whereas much of the intervention was delivered to students within classrooms.

For this illustration of the random-effects model, a subset of the TVSFP data was used. We concentrated on students from

Table 1
Tobacco and Health Knowledge Scale: Subgroup Descriptive Statistics at Pretest and Postintervention

Time point, <i>M</i> , and <i>SD</i>	No CC condition		CC condition	
	Without TV (<i>n</i> = 421)	With TV (<i>n</i> = 416)	Without TV (<i>n</i> = 380)	With TV (<i>n</i> = 383)
Pretest				
<i>M</i>	2.152	2.087	2.050	1.979
<i>SD</i>	1.182	1.288	1.285	1.286
Postintervention				
<i>M</i>	2.361	2.539	2.968	2.823
<i>SD</i>	1.296	1.437	1.405	1.312
Difference	0.209	0.452	0.918	0.844

Note. CC = social-resistance classroom curriculum; TV = a mass media (television) intervention.

28 Los Angeles schools, where the schools were randomized to one of four study conditions: (a) a social-resistance classroom curriculum (CC); (b) a media (television) intervention (TV); (c) a combination of CC and TV conditions; and (d) a no-treatment control group. These conditions form a 2 × 2 (CC [used or not used] × TV [used or not used]) design. A tobacco and health knowledge scale (THKS) score was one of the primary study outcome variables, and the one chosen for this analysis. The scale consisted of seven questionnaire items used to assess student tobacco and health knowledge. A student's score on this scale was defined as the sum of the items that the student answered correctly. Only data from the pretest and postintervention time points were analyzed; therefore, subjects were included if they had complete data on the THKS at these two time points. In all, there were 1,600 students from 135 classrooms and 28 schools who met these criteria. The resulting data set was unbalanced with a range of 1 to 13 classrooms per school and 2 to 28 students per classroom. Student means and standard deviations for the THKS, broken down by condition and time, are given in Table 1.

A Comparison of Regression Models

Several regression models were fit to these data. Results from these analyses are given in Table 2. In all cases, the postintervention THKS score was modeled in terms of baseline THKS score and effects of CC, TV, and CC × TV interaction. The first two columns of Table 2 give results for the usual regression analysis at class level and student level. Class-level analysis aggregated

² The design and intervention of this study is fully described by Flay et al. (1989). The data used for this article constitute a subset of the study's dataset; namely, in this article we examined the data from only one of the many study outcome variables, one of the two study sites, four of the five study conditions, and two of the four study time points. This subset was chosen to provide an illustration of the use of RRM on a fairly representative, although not overly complex, data set. The more complete analysis of the data from this study is included in a forthcoming article (Flay et al., in press).

Table 2
Tobacco and Health Knowledge Scale (THKS) Postintervention Scores: Linear Regression Estimates (and Standard Errors)

Variable	Ordinary regression model				Random-effects regression model					
	Class level (<i>n</i> = 135)		Student level (<i>n</i> = 1,600)		Students in classes (<i>n</i> = 1,600)		Students in schools (<i>n</i> = 1,600)		Three levels (<i>n</i> = 1,600)	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Constant	1.3087**	0.229	1.6613**	0.084	1.6776**	0.099	1.6952**	0.115	1.6970**	0.117
Baseline THKS	0.4962**	0.097	0.3252**	0.026	0.3116**	0.026	0.3103**	0.026	0.3072**	0.026
Condition										
Classroom curriculum (CC)	0.5749**	0.153	0.6406**	0.092	0.6330**	0.119	0.6601**	0.144	0.6392**	0.147
Television (TV)	-0.0150	0.150	0.1987*	0.090	0.1597	0.117	0.2023	0.140	0.1781	0.144
Interaction (CC × TV)	-0.0485	0.216	-0.3216*	0.130	-0.2747	0.168	-0.3696	0.201	-0.3204	0.206
Residual variance	0.3924	—	1.6929	—	1.6030**	0.059	1.6522**	0.059	1.6020**	0.059
Class variance	—	—	—	—	0.0870**	0.028	—	—	0.0636*	0.028
School variance	—	—	—	—	—	—	0.0372*	0.018	0.0258	0.020

* $p < .05$. ** $p < .01$.

student scores and thus modeled the average classroom postintervention THKS score in terms of CC and TV, controlling for average classroom THKS baseline score. The student-level analysis ignored nesting of students within classrooms within schools; consequently, students were treated as independent observations. Interestingly, although both analyses indicate the positive effect of the CC condition, different conclusions would be drawn regarding the effects of TV and the CC × TV interaction: Both TV terms are statistically significant by student-level analysis and clearly nonsignificant by class-level analysis. Thus, student-level analysis suggests that, although TV intervention is effective in increasing THKS scores for those not receiving the CC component, it has a negative effect on those exposed to both components. As a result of vastly different sample sizes used by these two analyses, standard error estimates are also quite different.

The next two columns of Table 2 list results obtained from analyses using the two-level RRM. Two 2-level models were examined: a students-within-classrooms model and a students-within-schools model. Results from these two 2-level RRMs are very similar and are not vastly different from those obtained from ordinary regression analysis at the student level. However, a difference does emerge that could influence the interpretation drawn from the data. Unlike ordinary student-level analysis, random-effects regression analysis indicates that the interaction of CC × TV is not statistically significant at the $p < .05$ level.

The final column of Table 2 lists results from the three-level RRM, which considers students nested within classrooms, which are nested within schools. Results from this model are similar to those obtained from the students-within-classrooms two-level model; again, the CC × TV interaction was not statistically significant. The additional variance parameter for the nesting of classrooms within schools was also not statistically significant (likelihood ratio $\chi^2_1 = 2.58$); therefore, it is reasonable that these models yield similar results. Thus, the additional variance term that the three-level model provides does not significantly improve the fit of the data over the two-level students-within-classroom model.

Discussion

This example has illustrated that choice of statistical models can influence conclusions drawn from a given data set. In the current context, all of the statistical models assessed significance of the main effects of a CC intervention, a TV intervention, and the interaction of these two components. Although all models examined in this article indicated a significant positive main effect of the CC intervention, these models yielded different results concerning significance of the CC × TV interaction. Using a conservative aggregate approach at the classroom level, we observed a nonsignificant CC × TV interaction. On the other hand, ordinary regression analysis at the student level indicated that this interaction was statistically significant. On the basis of random-effects regression analysis, this interaction was not statistically significant. Conclusions based on the traditional approaches (conservative aggregate and ordinary student-level analysis) showed disagreement with random-effects regression analysis.

Interestingly, regarding parameter estimates, the estimate of two-way interaction from the student-level analysis was reasonably similar to the estimate obtained from RRMs (the estimate of -0.3216 was from 87% to 117% of the RRM estimates). However, the estimate of the standard error of this interaction was much smaller for ordinary regression analysis, compared with RRM analyses (from 63% to 77% smaller). Underestimation of standard errors by ordinary regression analysis was observed for nearly all the model parameters and is expected because ordinary regression assumes independence of observations. Because the observations were not independent the amount of independent information available in parameter estimation is inflated.

That the amount of dependency present in the data as a result of the clustering of students was statistically significant, as evidenced by significant cluster variance terms for both students-within-classrooms and students-within-schools models. For the students-within-classrooms model, intraclass correlation ($\hat{r} = \hat{\sigma}^2_{\alpha} / [\hat{\sigma}^2_{\alpha} + \hat{\sigma}^2_{\epsilon}]$) was equal to .052, indicating that clustering of students within classrooms accounted for roughly 5% of the vari-

ability in the data not explained by the model covariates. The amount of intraclass correlation for the students-within-schools model equaled .022, representing about 2% of total unexplained variance in the data. Taken together, this indicates that students were more homogeneous in their postintervention scores when clustered by classrooms than by schools. The three-level RRM further supported the observation that the classroom variance is larger than school variance, and it indicated that when classroom variance was included in the model, school variance was no longer statistically significant. On the basis of the three-level model, the random classroom effect accounted for 3.8% of total unexplained variance, whereas the random school effect accounted for 1.5% of total unexplained variance.

Levels of intraclass correlation observed in these data were of moderate size and are consistent with reported levels from the literature (Donner, 1982; Jacobs et al., 1989). It is important to realize that the level of intraclass correlation can vary depending on several factors. First, the intraclass correlation can depend on which response variable is being modeled. Also, when cluster and residual variance terms are concurrently estimated along with other model parameters, the terms included in the model can affect the estimated value of intraclass correlation. Finally, the sample size at various levels can influence precision of the variance terms, and thus affect precision of intraclass correlation as well.

For instance, given these same THKS data, intraclass correlation equaled .101 for a students-within-classrooms model that included only the random classroom intercept term and .090 when the effect of baseline THKS scores was added to this model. The intraclass correlation levels from these two simple models are considerably higher than the .052 level for the students-within-classrooms model given in Table 2. Likewise, the students-within-schools model yielded intraclass correlations of .068 and .058, respectively for these two simple models, as opposed to .022 from the students-within-schools model given in Table 2. Thus, the addition of the school-level variables CC, TV, and the $CC \times TV$ interaction accounted for approximately 4% of what was explained as intraclass correlation in the models that did not include these terms. In general, adding variables into the model that vary at the school or class level will reduce the amount of unexplained variability caused by the cluster (the cluster variance term) and thus reduce the intraclass correlation. Likewise, adding variables into the model that vary at the student level will reduce the amount of unexplained variability at that level (the residual variance term) but also, in general, reduce the cluster variance, thus affecting the amount of intraclass correlation that is estimated on the basis of the model.

With the same data, using the postintervention-minus-pretest change score as the dependent variable also influenced the observed value of the intraclass correlation. When the dependent variable was the THKS change score, intraclass correlation equaled .0183 and .0017, respectively, for the students-within-classrooms and students-within-schools models, for the RRM with fixed terms for CC, TV, and the $CC \times TV$ interaction. The large decrease in intraclass correlation when analyzing change scores was mainly attributable to the residual variance which, relative to the covariate-adjusted models, was much larger. As a result, the denominator in the intraclass correlation equation was much larger as well.

Another context for clustered data is in the area of repeated measures data. Here, the repeated measures (Level 1) are observed clustered within individuals (Level 2). Notice, that in our example, there were two repeated observations clustered within students (baseline and postintervention THKS scores). With only two repeated observations, the longitudinal nature of the data can be modeled by analyzing change scores or by analyzing the postintervention score covarying out the baseline score; however, with more repeated observations per student, a longitudinal approach is more appropriate and informative. If there is no clustering of individuals (i.e., the individuals are independent), the model given in Equation 2 can be used to analyze repeated measures data by considering the y_i vector to represent the n_i repeated observations from individual i , and the X_i matrix to represent the effects corresponding to the repeated measurements (e.g., orthogonal polynomial contrasts) as well as time-invariant and time-varying covariates. Unfortunately, this approach cannot be recommended in general because, as noted, this model assumes the "compound symmetry" form for the variance-covariance structure, and although this form is reasonable for clustered data, it is generally not reasonable for longitudinal data. In the longitudinal setting, this structure specifies that variances across time points are homogeneous, as are all covariances across time points; however, it is more likely that observations from closely aligned time points are more correlated than those more separated in time. Also, variances often increase over time, for example, when subjects are more alike at baseline than at the end of the study, violating the homogeneity-of-variance assumption. In general, random-effects models for longitudinal data extend the model given in Equation 2 by including multiple random effects to account for the clustering of observations within individuals. For instance, one may allow individuals to vary in terms of both their intercept, or starting point, and their trend across time, yielding a model with two random effects at the individual level (Level 2). As others have noted, these more general random-effects models provide a powerful approach to the analysis of longitudinal data (Bock, 1983; Bryk & Raudenbush, 1987; Gibbons et al., 1988; Laird & Ware, 1982).

Finally, when dealing with multilevel data, the number of levels of data must be considered. Often, pooling higher-order levels is determined before the analysis for pragmatic or conceptual reasons; at other times, the decision can be empirically tested. Notice, in the example, three potential levels of the data were considered—students (Level 1), classrooms (Level 2), and schools (Level 3)—whereas other potential higher-order levels (e.g., school districts) were implicitly ignored. Even with regard to the three levels in the example, one could empirically argue that three-level analysis is unnecessary because the variance attributable to nesting of classrooms within schools (Level 3) is not significant and thus justify pooling this additional level of the data and proceeding with a two-level analysis (students within classrooms). From a design perspective, on the other hand, because schools were the treatment assignment level, one could argue that the random school term (Level 3) must remain in the model regardless of significance and, therefore, a three-level analysis is necessary. When the variance attributable to a higher-order level is observed to be small and nonsignificant and the sample is of moderate size, the parameter estimates and

standard errors will not differ greatly whether the higher-order level is included in the model or not. In this example, because difference between the three-level analysis and the students-within-classrooms two-level analysis was fairly small, both models gave similar results. Unfortunately, software for higher-order (greater than two levels) RRM is not as readily available, and so comparisons between two-level and higher-order models is not always feasible. In this case, a reasonable approach is to choose the two-level model, which provides the best fit of the data based on the log-likelihood value for the model. This ensures that the level accounting for the largest share of total variance (the largest amount of intradependency in the data) will be included in the model. Further development of higher-order models will eliminate this dilemma in the future.

Computer Programs

The development of computer software for the RRM has been relatively recent, although at an increasing rate. At present, several commercially available software programs exist for computing the RRM, although, except where noted, all perform only two-level analysis. The BMDP3V (Jennrich & Sampson, 1988) and BMDP5V (Schluchter, 1988) procedures are available as part of the general BMDP statistical package for both mainframe and disk operating system (DOS)-based personal computers. Taken together, these procedures are fairly complete and take advantage of the extensive data-handling capabilities of the BMDP environment. The 3V procedure is suited primarily for clustered data, whereas 5V is better suited for repeated-measures data. Recently, Statistical Analysis Systems (SAS) introduced the MIXED procedure, which can handle a variety of both clustered and longitudinal two-level models in an accessible way. At present, MIXED is included in Version 6.07 of SAS, which runs only on mainframe computers; however this procedure will soon be available for PC versions of SAS as well. Other software programs—ML3 (Prosser, Rasbash, & Goldstein, 1991), HLM (Bryk, Raudenbush, Seltzer, & Congdon, 1989), and VARCL (Longford, 1986)—are stand-alone programs that are primarily suited for use on DOS-based PCs, although available for some types of mainframe computers, and run mainly in interactive sessions with the user. As the name implies, ML3 can perform three-level analysis as well as fit some models for dichotomous outcome variables. Additionally, ML3 has excellent data-handling capabilities for a stand-alone statistical program, and its manual is complete and clearly written. A strength of HLM is that it has the ability to read data from SYSTAT files, allowing one to take advantage of SYSTAT data-handling capabilities before the analysis. VARCL is unique in that it allows for up to nine levels of nesting (with one random effect per level). All of these programs were written by researchers with many published articles on random-effects models, and the manuals refer to these articles as examples and to clarify the use of RRM. Not surprisingly, use of the programs is made simpler with reference to the published articles. For a detailed comparison of the stand-alone programs, see Kreft, de Leeuw, & Kim (1990).

Alternatively, programming facilities of general statistical packages (e.g., Statistical Packages for the Social Sciences [SPSS] SAS, or both) can be used to develop tailor-made sub-

programs to perform random-effects regression analysis and interface directly with data sets from these packages. To this end, we have programmed an SPSS matrix subprogram that can perform a two-level random-effects regression analysis.³ Because this program is written in the SPSS matrix language it can be directly used within SPSS, which additionally supports data transfer from SAS and BMDP system files as well as several spreadsheet and database program files. The matrix language is currently available on the mainframe, Macintosh, and Windows versions of SPSS; however, it is not available on the DOS version of SPSS.

Conclusion

Our example has highlighted features of analysis using RRM and how analysis by more traditional regression models can lead to biased estimates of uncertainty and different conclusions. As demonstrated, RRM provides a powerful tool for analysis of clustered data where the number of individuals within clusters varies. Alternatively, as noted by many researchers (Burstein, 1980; Jacobs, Jeffery, & Hannan, 1989; Koepke & Flay, 1989; McKinlay, Stone, & Zucker, 1989; Murray, Hannan, & Zucker, 1989; Raudenbush & Bryk, 1988), traditional analysis at the individual or cluster level is problematic. Traditional analysis at the individual level ignores dependency in the data that results from clustering, whereas analysis at the cluster level does not permit a straightforward analysis of individual-level characteristics. By controlling for and estimating the degree of dependency resulting from the clustering of data, RRM allows testing of relationships at the individual level in a manner consistent with data collection; namely, from individuals within clusters. Furthermore, RRM can include independent variables at either the cluster or individual level and estimate their effects as well as interactions. It is our hope that, by using one of the statistical programs that performs RRM or the available SPSS matrix language code, researchers will find these models more accessible.

³ The SPSS matrix subprogram is available from the authors.

References

- Aitkin, M., & Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society, 149*(Series A), 1-43.
- Blair, R. C., Higgins, J. J., Topping, M. E., & Mortimer, A. L. (1983). An investigation of the robustness of the *t* test to unit of analysis violations. *Educational and Psychological Measurement, 43*, 69-80.
- Bock, R. D. (1983). The discrete Bayesian. In H. Wainer & S. Messick (Eds.), *Modern advances in psychometric research* (pp. 103-115). Hillsdale, NJ: Erlbaum.
- Bock, R. D. (1989a). Measurement of human variation: A two-stage model. In R. D. Bock (Ed.), *Multilevel analysis of educational data* (pp. 319-342). San Diego, CA: Academic Press.
- Bock, R. D. (Ed.). (1989b). *Multilevel analysis of educational data*. San Diego, CA: Academic Press.
- Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101*, 147-158.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

- Bryk, A. S., Raudenbush, S. W., Seltzer, M., & Congdon, R. (1989). *An introduction to HLM: Users' guide*. Chicago: Scientific Software, Inc.
- Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. In D. Berliner (Ed.), *Review of research in education* (Vol. 8, pp. 158–233). Washington, DC: American Educational Research Association.
- DeLeeuw, J., & Kreft, I. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics*, 11, 57–85.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society*, 39(Series B), 1–38.
- Dempster, A. P., Rubin, D. B., & Tsutakawa, R. K. (1981). Estimation in covariance component models. *Journal of the American Statistical Society*, 76, 341–353.
- Donner, A. (1982). An empirical study of cluster randomization. *International Journal of Epidemiology*, 11, 283–286.
- Donner, A. (1985). A regression approach to the analysis of data arising from cluster randomization. *International Journal of Epidemiology*, 14, 322–326.
- Flay, B. R., Brannon, B. R., Johnson, C. A., Hansen, W. B., Ulene, A. L., Whitney-Saltiel, D. A., Gleason, L. R., Sussman, S., Gavin, M., Glowacz, K. M., Sobol, D. F., & Spiegel, D. C. (1989). The Television, School and Family Smoking Cessation and Prevention Project: I. Theoretical basis and program development. *Preventive Medicine*, 17, 585–607.
- Flay, B. R., Miller, T. Q., Hedeker, D., Siddiqui, O., Britton, C. F., Brannon, B. R., Johnson, C. A., Hansen, W. B., Sussman, S., & Dent, C. (in press). *The Television, School and Family Smoking Cessation and Prevention Project: VIII. Student outcomes and mediating variables*.
- Gibbons, R. D., Hedeker, D., Waternaux, C., & Davis, J. M. (1988). Random regression models: A comprehensive approach to the analysis of longitudinal psychiatric data. *Psychopharmacology Bulletin*, 24, 438–443.
- Goldstein, H. (1987). *Multilevel models in educational and social research*. New York: Oxford University Press.
- Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York: McGraw-Hill.
- Haney, W. (1980). Units and levels of analysis in large-scale evaluation. In K. H. Roberts & L. Burstein (Eds.), *Issues in aggregation. New directions for methodology of social and behavioral science* (No. 6, pp. 1–15). San Francisco: Jossey-Bass.
- Hedeker, D., Gibbons, R. D., Waternaux, C., & Davis, J. M. (1989). Investigating drug plasma levels and clinical response using random regression models. *Psychopharmacology Bulletin*, 25, 227–231.
- Hedeker, D., Gibbons, R. D., & Davis, J. M. (1991). Random regression models for multicenter clinical trials data. *Psychopharmacology Bulletin*, 27, 73–77.
- Hopkins, K. D. (1982). The unit of analysis: Group means versus individual observations. *American Educational Research Journal*, 19, 5–18.
- Jacobs, D. R., Jeffery, R. W., & Hannan, P. J. (1989). Methodological issues in worksite health intervention research: II. Computation of variance in worksite data. In K. Johnson, J. H. LaRosa, C. J. Scheirer, & J. M. Wolle (Eds.), *Proceedings of the 1988 methodological issues in worksite research conference* (pp. 77–88). Airlie, VA: U.S. Department of Health and Human Services.
- Jennrich, R. I., & Sampson, P. F. (1988). 3V: General mixed model analysis of variance. In W. J. Dixon (Ed.), *BMDP statistical software manual* (Vol. 2, pp. 1025–1043). Berkeley, CA: University of California Press.
- Kirk, R. E. (1982). *Experimental design* (2nd ed.). Monterey, CA: Brooks/Cole.
- Koepke, D., & Flay, B. R. (1989). Levels of analysis. In M. T. Braverman (Ed.), *Evaluating health promotion programs. New directions for program evaluation* (No. 43, pp. 75–87). San Francisco: Jossey-Bass.
- Kreft, I. G., de Leeuw, J., & Kim, K. (1990). *Comparing four different statistical packages for hierarchical linear regression: Genmod, HLM, ML2, and VARCL* (Statistical Series No. 50). Los Angeles: University of California Press.
- Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38, 963–974.
- Longford, N. T. (1986). VARCL—Interactive software for variance component analysis. *The Professional Statistician*, 74, 817–827.
- Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, 74, 817–827.
- Longford, N. T. (1989). Fisher scoring algorithm for variance component analysis of data with multilevel structure. In R. D. Bock (Ed.), *Multilevel analysis of educational data* (pp. 297–310). San Diego, CA: Academic Press.
- McKinlay, S. M., Stone, E. J., & Zucker, D. M. (1989). Research design and analysis issues. *Health Education Quarterly*, 16, 307–313.
- Murray, D. M., Hannan, P. J., & Zucker, D. M. (1989). Analysis issues in school-based health promotion studies. *Health Education Quarterly*, 16, 315–320.
- Prosser, R., Rasbash, J., & Goldstein, H. (1991). *ML3 users' guide for v.2*. London: Institute of Education.
- Raudenbush, S. W., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59, 1–17.
- Raudenbush, S. W., & Bryk, A. S. (1988). Methodological advances in analyzing the effects of schools and classrooms on student learning. In E. Z. Rothkopf (Ed.), *Review of research in education* (Vol. 15, pp. 423–475). Washington, DC: American Educational Research Association.
- Raudenbush, S. W., & Willms, J. D. (Eds.). (1991). *Schools, classrooms, and pupils: International studies of schooling from a multilevel perspective*. San Diego, CA: Academic Press.
- Schluchter, M. D. (1988). 5V: Unbalanced repeated measures models with structured covariance matrices. In W. J. Dixon (Ed.), *BMDP statistical software manual* (Vol. 2, pp. 1081–1114). Berkeley, CA: University of California Press.
- Searle, S. R. (1987). *Linear models for unbalanced data*. New York: Wiley.
- Silvey, S. D. (1975). *Statistical inference*. London: Chapman and Hall.
- Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54, 426–482.

Appendix

Maximum Marginal Likelihood (MML) Estimation

To estimate covariate effects β and variance parameters σ_α^2 and σ^2 , we outline the MML solution. For this, the marginal density of the data y_i in the population, abbreviated as $h(y_i)$, is expressed as the integral,

$$h(y_i) = \int_{\alpha} f(y_i | \alpha; \beta, \sigma^2) g(\alpha; 0, \sigma_\alpha^2) d\alpha$$

which is a function of two normal densities. The first, the likelihood or measurement probability function [abbreviated as $f(y_i | \alpha; \beta, \sigma^2)$] is given as follows:

$$f(y_i | \alpha; \beta, \sigma^2) = \frac{1}{\sqrt{n_i(2\pi)^{n_i} \sigma}} \exp[-(y_i - X_i\beta)(y_i - X_i\beta)/2\sigma^2]$$

and the second, the empirical prior density, denoted by $g(\alpha; 0, \sigma_\alpha^2)$, is the normal density with mean 0 and variance σ_α^2 , that is,

$$g(\alpha; 0, \sigma_\alpha^2) = \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} \exp[-\alpha^2/2\sigma_\alpha^2].$$

The marginal log-likelihood of a sample of N independently responding clusters is given as: $\log L = \sum_i^N \log h(y_i)$, and differentiating the marginal log-likelihood with respect to each parameter yields the following first derivatives:

$$\frac{\partial \log L}{\partial \beta} = \sigma^{-2} \sum_{i=1}^N X_i' u_i \tag{8}$$

$$\frac{\partial \log L}{\partial \sigma_\alpha^2} = \frac{1}{2} \sigma_\alpha^{-4} \sum_i^N \tilde{\alpha}_i^2 + (\sigma_{\alpha|y_i}^2 - \sigma_\alpha^2) \tag{9}$$

$$\frac{\partial \log L}{\partial \sigma^2} = \frac{1}{2} \sigma^{-4} \sum_i^N u_i' u_i + n_i (\sigma_{\alpha|y_i}^2 - \sigma^2) \tag{10}$$

where $u_i = y_i - X_i\beta - \mathbf{1}_i \tilde{\alpha}_i$.

Setting expressions for the first derivatives to zero and solving yields Equations 5–7. As mentioned, the EM solution proceeds by iterating between EB Equations 3 and 4 and MML Equations 5–7 until convergence.

As the EM iterative process uses only information from first derivatives, it is termed a first-order solution, and as a result, it can be slow to converge under certain circumstances (see discussion of Dempster, Laird, & Rubin, 1977; Laird & Ware, 1982). Thus, at some point in the iterative procedure, it is often desirable to switch to a second-order solution (e.g., using second derivatives in addition to first derivatives), for example, the Fisher scoring solution. The Fisher scoring solution is an iterative process that uses first derivatives and expectations of the second derivatives (the negative of the information matrix) of the likelihood of the data with respect to the estimated parameters. Specifically, multiplying the vector of first derivatives by the inverse of the information matrix provides the vector of corrections that, added to parameter values, yield improved estimates. From the improved estimates, values for the first derivatives and information matrix are reobtained, yielding further improved estimates. This process is repeated until convergence.

For the Fisher scoring solution, partitions of the information matrix need to be specified. It can be shown that both the information for β with σ_α^2 and the information for β with σ^2 equal zero (Bock, 1989; Long-

ford, 1987). Thus, the information matrix is given by the following form:

$$I = \begin{bmatrix} I(\beta) & 0 & 0 \\ 0 & I(\sigma_\alpha^2) & I(\sigma^2, \sigma_\alpha^2) \\ 0 & I(\sigma^2, \sigma_\alpha^2) & I(\sigma^2) \end{bmatrix}$$

where,

$$I(\beta) = \sigma^{-4} \sum_i^N X_i' (\sigma^2 I_{n_i} - \sigma_{\alpha|y_i}^2 \mathbf{1}_i \mathbf{1}_i') X_i \tag{11}$$

$$I(\sigma_\alpha^2) = \frac{1}{2} \sigma_\alpha^{-8} \sum_i^N (\sigma_\alpha^2 - \sigma_{\alpha|y_i}^2)^2 \tag{12}$$

$$I(\sigma^2) = \frac{1}{2} \sigma^{-8} \sum_i^N n_i (\sigma^2 - \sigma_{\alpha|y_i}^2)^2 \tag{13}$$

$$I(\sigma^2, \sigma_\alpha^2) = \frac{1}{2} \sigma^{-4} \sigma_\alpha^{-4} \sum_i^N n_i \sigma_{\alpha|y_i}^4 \tag{14}$$

At convergence, the inverse of the information matrix, denoted I^{-1} , provides the large-sample variances and covariances of the MML estimates which can be used to construct confidence intervals and tests of hypotheses for the structural and population parameters (Wald, 1943). The square root of each diagonal element of I^{-1} provides standard errors for the MML estimates, which are used to determine asymptotically normal test statistics (estimate divided by its standard error) for each parameter. There are some concerns in using the standard errors in constructing a hypothesis test for the random-effect variance term (i.e., the cluster variance), particularly when the variance is near zero and the Level 2 sample size is small (Bryk & Raudenbush, 1992); in this case, the use of the likelihood ratio chi-square test may be used by comparing a model ignoring cluster variance with one that includes it.

Finally, in terms of numerical computation, estimation of variance terms likely to be small or near zero can be problematic in an iterative second-order solution (i.e., Fisher scoring). As a result, it is better to reparameterize the cluster variance term using the exponential transform, namely, $e^\tau = \sigma_\alpha^2$, and to estimate τ during the iterative Fisher scoring solution. This reparameterization helps keep the provisional values of the cluster variance positive during the iterations. An alternative reparameterization that is often considered for this purpose is to estimate σ_α rather than σ_α^2 (e.g., see Longford, 1989). To modify the earlier equations for the estimation of τ (using the chain rule of calculus and the property that $\partial \sigma_\alpha^2 / \partial \tau = \sigma_\alpha^2$): The derivative in Equation 9 and the information term in Equation 14 are multiplied by σ_α^2 , and the information term in Equation 12 is multiplied by σ_α^4 . Once the solution has converged, it is beneficial to revert to the original equations to estimate the standard error for the parameter σ_α^2 and not the reparameterized term τ . Even with the exponential transformation, if the cluster effect has no variation in the population, the Fisher scoring steps will eventually fail as τ goes to $-\infty$. At this point, the term will have to be considered a fixed effect and the solution reestimated.

Received September 11, 1991
 Revision received September 17, 1993
 Accepted October 15, 1993 ■