Research Article

Research Synthesis Methods

Received 29 April 2010, Revised 18 April 2011, Accepted 26 April 2011 Published online 10 June 2011 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/jrsm.35

Fixed effects and variance components estimation in three-level meta-analysis

Spyros Konstantopoulos*†

Meta-analytic methods have been widely applied to education, medicine, and the social sciences. Much of meta-analytic data are hierarchically structured because effect size estimates are nested within studies, and in turn, studies can be nested within level-3 units such as laboratories or investigators, and so forth. Thus, multilevel models are a natural framework for analyzing meta-analytic data. This paper discusses the application of a Fisher scoring method in two-level and three-level meta-analysis that takes into account random variation at the second and third levels. The usefulness of the model is demonstrated using data that provide information about school calendar types. SAS proc mixed and HLM can be used to compute the estimates of fixed effects and variance components. Copyright © 2011 John Wiley & Sons, Ltd.

Keywords: multilevel models; meta-analysis; variance components; effect sizes

The sheer volume of research related to various topics of scientific interest poses the question of how to organize and summarize findings to identify and use what is known as well as focus research on promising areas (Garvey & Griffith, 1971). This need for accumulating research evidence has led to the development of systematic methods for quantitative synthesis of research (Cooper *et al*., 2009). Currently, the use of quantitative methods to summarize results from various empirical studies that test the same hypothesis is widespread in education, psychology, medicine, and social science research. Meta-analysis is a statistical method used to combine evidence from different primary research studies that test comparable hypotheses for the purposes of summarizing evidence and drawing general conclusions (Cooper *et al*., 2009; Glass, 1976; Hedges & Olkin, 1985; Lipsey & Wilson, 2001). Meta-analytic methods involve first describing the results of individual studies via numerical indexes that are commonly called effect size estimates (e.g., correlation coefficient, standardized mean difference, odds ratio) and second combining these estimates across studies to obtain a summary statistic such as a mean (e.g., a standardized mean difference or an average association).

Meta-analytic data are naturally hierarchically structured. For instance, effect sizes are nested within studies, which can be nested within investigators, and so forth. Hence, multilevel models can provide a useful framework for analyzing meta-analytic data and take into account variation in all levels of the hierarchy. Multilevel models have been used extensively over the last 20 years (Goldstein, 1987; Longford, 1993; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999), and their applications to meta-analytic data with a two-level structure have been demonstrated in the literature (DerSimonian & Laird, 1986; Goldstein *et al*., 2000; Hedges & Olkin, 1985; Hox & de Leeuw, 2003; Raudenbush & Bryk, 2002). In this paper, the application of a Fisher scoring algorithm to univariate two-level and three-level meta-analysis is discussed. This algorithm computes the fixed effects and the variance components of the random effects within a maximum likelihood framework (Konstantopoulos, 2003; Longford, 1987, 1993). The meta-analytic data with a two-level and three-level nested structure that includes information about the effects of modified school calendars on student achievement to show the usefulness of the models are used.

Statistical models for meta-analysis

Two statistical models have been developed for inference about effect size data from a collection of studies: the random or mixed and the fixed effects models for meta-analysis (Hedges & Vevea, 1998; Konstantopoulos, 2007).

Michigan State University, East Lansing, MI, United States

**Correspondence to: Spyros Konstantopoulos, Michigan State University, East Lansing, MI, USA*

†E-mail: spyros@msu.edu

Both models are appropriate for computing estimates in meta-analysis, and the choice of model depends on the data structure and the assumptions about the statistical model. Fixed effects models treat an effect size parameter as unknown but fixed and assume that the between-study heterogeneity of the study-specific estimates is virtually zero (Hedges, 1982). Random effects models, however, treat the effect size parameters as if they were a random sample from a population of effect size parameters (DerSimonian & Laird, 1986; Hedges, 1983; Raudenbush & Bryk, 2002). The random effects model introduces heterogeneity among the effect size parameters that is estimated by the between-study variance (Hedges & Vevea, 1998).

The simplest random effects model follows typically a two-level structure and introduces a source of variation at the second level, by taking into account the between-study variance of the study-specific estimates (Hedges & Olkin, 1985; Raudenbush & Bryk, 2002). Random effects models with a two-level structure have already been developed using the method of moments or maximum likelihood methods (Goldstein *et al.*, 2000; Hedges & Vevea, 1998; Raudenbush & Bryk, 2002). Two-level models for multivariate meta-analysis that take into account that the dependency in the data have also been discussed in the literature (Berkey *et al.*, 1996; Gleser & Olkin, 2009). In addition, a recent study discussed the dependency of effect sizes within a study using both a random effects and a hierarchical Bayes approach (Stevens & Taylor, 2009). The key underlying principle in two-level models is that the effect size parameter is not fixed; instead, it has its own distribution and is treated as a random variable at the second level. The between-studies model introduces the inconsistency or heterogeneity in study effects across the sample of studies.

Although a two-level model captures the random variation between studies, it does not account for higher levels of nesting in meta-analytic data. Sometimes, however, higher level nesting takes place in meta-analytic data as with other kinds of data. For example, some researchers have shown empirically the importance of modeling achievement data using three-level models that take into account nesting effects at the second level (e.g., classroom) and at the third level (e.g., school) (Bryk & Raudenbush, 1988; Nye *et al.*, 2004). Similarly, workers can be nested within departments and firms, and patients can be nested within clinics and hospitals. Meta-analytic data could also have a three-level structure. For example, effect sizes are nested within studies, and studies are nested within level-3 units such as laboratories or investigators. Third-level units could also be firms, hospitals, neighborhoods, cities, etc., and the choice of the third-level unit depends on the nature and structure of the data. The idea is that studies conducted by the same investigator, for example, will likely produce estimates that are correlated, and this dependency needs to be taken into account in the analysis. In other words, studies are clustered into investigators. In this case, random variation is evident both at the second and third levels, because study-specific and investigator-specific effects can be modeled as random effects at the second and third levels, respectively.

In a three-level model, the random variation is divided into two parts: the between-study within-level-3 unit variation and the between-level-3 unit variation. This decomposition of the variance is important and informative because it indicates where most of the random variation lies, within or between level-3 units. If the third-level variance is considerable, then it should be included in the estimation process instead of being omitted or collapsed at the second level. In the three-level model, level-3 unit estimates form a distribution of effects with a variance that shows differences in effect sizes between level-3 units.

It is difficult to know exactly the optimal number of units that are needed to compute variance components in two-level or three-level models. As in the two-level case, three-level models involve computations of variances of random effects, and in principle, larger sample sizes at the third level are preferred because more information is used in the estimation. For example, when there are 10 or more level-3 units, a three-level model may be warranted. In contrast, when the number of level-3 units is very small (e.g., 2–4), perhaps a two-level model should be used and the effects of the third-level units could be modeled as fixed effects (via dummy indicators) at the second level. In addition, ideally, each level-3 unit should include multiple studies. Note that the question about the optimal number of units needed to compute variances of random effects is not inherent to meta-analysis and applies to any two-level or three-level model.

In this study, multilevel models for meta-analysis and focus on three-level models are discussed. An iterative computational algorithm called Fisher scoring was used to obtain maximum likelihood estimates for two-level and three-level models. This method updates the estimates of fixed effects and variance components in each iteration using the expected information matrix (see Longford, 1993).

Example

To illustrate the usefulness of the second-level and third-level models, let us consider an example about modified school calendars. The data include studies on schools that modified their calendars without extending the length of the school year (see Cooper *et al.*, 2003). The sample of studies used here is somewhat different than that used in the study by Cooper *et al.*, but it suffices for the purposes of the exercise. Overall, 56 studies were included in the sample. First, consider the two-level case, where the first level involves a within-study model and the second level involves a between-study model. Each study provided information that allowed Cooper *et al.* to construct effect sizes and their standard errors. In addition, there was information about the year of the study that can be modeled as a study-specific predictor. Now, consider the three-level case. The 56 studies were conducted by school districts, and thus, studies were nested within districts. Overall, there were 11 school districts (nearly five studies per district). In this case, the first level involves a within-study model, the second level involves a between-study within-district model, and the third level involves a between-district model. The year of study can be included in the model as a predictor in the third level because it is a district-specific variable. The data are summarized in Table 1. Of course, one could imagine a similar structure for health data. For instance, suppose that neighborhoods conduct studies about treatment effects for patients in hospitals in these neighborhoods.

Two-level meta-analysis

First, the method in the simplest case that involves two levels is illustrated. For simplicity, suppose that there is only one outcome in each study and that one effect size estimate is computed in each study (i.e., univariate case). Suppose now that there are k effect size population parameters $\vartheta_1, \ldots, \vartheta_k$ and therefore k corresponding independent effect size estimates $T_1, \ldots, T_k$ with known sampling variances $v_1, \ldots, v_k$. In our example, the T_is are given in column three of Table 1, the v_is are given in column four, and the total number of studies k is 56. We assume that these effect size estimates T_i are independently and normally distributed about ϑ_i with a mean of ϑ_i and variance v_i. The variances (the v_is) are unknown, but they are estimated using a consistent estimator, and therefore, they are assumed to be known (see Table 1). The first level of the hierarchy the within-study model is

$$T_i = \vartheta_i + \varepsilon_i, \tag{1}$$

where the error term is normally distributed with a mean of zero and a variance v_i. At the second level of the hierarchy, the between-study model, the population parameter varies around an overall mean, namely

$$\vartheta_i = \beta_0 + \eta_i, \tag{2}$$

where η_i is a study-specific random effect that is normally distributed with a mean of zero and variance τ ($\tau > 0$). In a single-level notation, the model is written as

$$T_i = \beta_0 + \eta_i + \varepsilon_i. \tag{3}$$

The second level can also include p predictors (e.g., study characteristics such as the year of the study) namely

$$\vartheta_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_p X_{pi} + \eta_i, \tag{4}$$

and in this regression model, the residual variance of the random effect η is τ_R. In our example, one predictor is year of study and is reported in the last column of Table 2. In the study by Cooper *et al.*, other study level predictors are also reported such as whether the study was conducted by an internal of external evaluator.

For the within-study model, we assume that the variances of the stochastic errors are different for each study (i.e., heterogeneity of the sampling error), whereas for the between-study model, we assume that the random effects are distributed identically (i.e., homogeneity of random effects). The sampling error variances in meta-analytic data cannot be expected to be identical across studies because they typically depend on the sample size of each study, and hence, the heterogeneity assumption seems reasonable. The units at each level are independently distributed, and thus, the error terms ε_i and η_i at the first and second levels, respectively, are uncorrelated, that is, $\mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0$ and $\mathrm{cov}(\eta_i, \eta_j) = 0$. In a single-level equation, the two-level model with second-level predictors is

$$T_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_p X_{pi} + \eta_i + \varepsilon_i. \tag{5}$$

The effect sizes T_is are normally distributed with a mean $\beta_0 + \sum_{j=1}^{p} \beta_j X_{ij}$ and a variance $v_i + \tau_R$, and when there are no predictors in the model, the between-study variance is τ.

Estimation

Estimates of the fixed effects, the regression coefficients, and the variance components of the random effects at the second level using maximum likelihood estimation were computed. A Fisher scoring algorithm was used to compute the maximum likelihood estimates following Longford (1987, 1993). For simplicity, the simplest two-level case with no predictors at the second level was discussed. The advantage of the simplest case is that the estimates have simple algebraic expressions. Even the inclusion of one predictor in the model complicates the expressions considerably. The more general case that can include predictors is expressed in matrix notation and is illustrated in Appendix A. The idea is to maximize a log-likelihood function to estimate the fixed effects and the variance components of the random effects (see Appendix A). The Fisher scoring algorithm involves the computation of the first-order and second-order derivatives of the parameter estimates, the fixed effects, and variance components.

Table 1. Data used in the analysis

District ID	Study ID	Effect size	Variance	Year
11	1	−0.18	0.118	1976
11	2	−0.22	0.118	1976
11	3	0.23	0.144	1976
11	4	−0.30	0.144	1976
12	5	0.13	0.014	1989
12	6	−0.26	0.014	1989
12	7	0.19	0.015	1989
12	8	0.32	0.024	1989
18	9	0.45	0.023	1994
18	10	0.38	0.043	1994
18	11	0.29	0.012	1994
27	12	0.16	0.020	1976
27	13	0.65	0.004	1976
27	14	0.36	0.004	1976
27	15	0.60	0.007	1976
56	16	0.08	0.019	1997
56	17	0.04	0.007	1997
56	18	0.19	0.005	1997
56	19	−0.06	0.004	1997
58	20	−0.18	0.020	1976
58	21	0.00	0.018	1976
58	22	0.00	0.019	1976
58	23	−0.28	0.022	1976
58	24	−0.04	0.020	1976
58	25	−0.30	0.021	1976
58	26	0.07	0.006	1976
58	27	0.00	0.007	1976
58	28	0.05	0.007	1976
58	29	−0.08	0.007	1976
58	30	−0.09	0.007	1976
71	31	0.30	0.015	1997
71	32	0.98	0.011	1997
71	33	1.19	0.010	1997
86	34	−0.07	0.001	1997
86	35	−0.05	0.001	1997
86	36	−0.01	0.001	1997
86	37	0.02	0.001	1997
86	38	−0.03	0.001	1997
86	39	0.00	0.001	1997
86	40	0.01	0.001	1997
86	41	−0.10	0.001	1997
91	42	0.50	0.010	2000
91	43	0.66	0.011	2000
91	44	0.20	0.010	2000
91	45	0.00	0.009	2000
91	46	0.05	0.013	2000
91	47	0.07	0.013	2000
108	48	−0.52	0.031	2000
108	49	0.70	0.031	2000
108	50	−0.03	0.030	2000
108	51	0.27	0.030	2000
108	52	−0.34	0.030	2000
644	53	0.12	0.087	1995
644	54	0.61	0.082	1995
644	55	0.04	0.067	1994
644	56	−0.05	0.067	1994

In the simplest case, where there are no study-level predictors included in the model, the objective is to compute one overall mean, the intercept, and the second-level random effects variance component τ. For example, the fixed effects estimate or overall mean in this case is a product of sums as shown in the Fisher scoring equation

$$\left\{\sum_{i=1}^{k}(v_i+\tau)^{-1}\right\}^{-1}\sum_{i=1}^{k}T_i(v_i+\tau)^{-1}, \tag{6}$$

where $\sum_{i=1}^{k}T_i(v_i+\tau)^{-1}$ is the scoring function and $\sum_{i=1}^{k}(v_i+\tau)^{-1}$ is the expected information function. The Fisher scoring equation when predictors are included in the model is given in Appendix A (A-3). Similarly, the between-study variance component at the second level is updated as

$$\tau=\tau_0-2\left\{\sum_{i=1}^{k}(v_i+\tau)^{-2}\right\}^{-1}\frac{1}{2}\left\{\sum_{i=1}^{k}(v_i+\tau)^{-1}-\sum_{i=1}^{k}e_i^2(v_i+\tau)^{-2}\right\}, \tag{7}$$

where e is a residual defined in Appendix A, τ_0 is the initial estimate of the second-level variance, $1/2\sum_{i=1}^{k}(v_i+\tau)^{-2}$ is the expected information function, and $-1/2\left\{\sum_{i=1}^{k}(v_i+\tau)^{-1}-\sum_{i=1}^{k}e_i^2(v_i+\tau)^{-2}\right\}$ is the scoring function. The variance of the fixed effect estimate is given by $\left\{\sum_{i=1}^{k}(v_i+\tau)^{-1}\right\}^{-1}$, and the variance of the variance component is given by $2\left\{\sum_{i=1}^{k}(v_i+\tau)^{-2}\right\}^{-1}$. Convergence is achieved when the log-likelihood remains unchanged for several decimal places.

Three-level meta-analytic model

The computations in the three-level model are more complicated mainly because of the estimation of the variance components at the third level. Again, suppose that there is only one outcome and one effect size estimate per study. For simplicity, let us consider first the simplest case where no predictors are included at levels 2 and 3. The computation involves an overall mean estimate and two variance component estimates at levels 2 and 3. Simple algebraic expressions are not always possible in the three-level model. The model for the first level of the hierarchy (the within-study model) for effect size estimate T_i is identical to Equation (1).

In the second level of the hierarchy (the between-study within-level-3-unit model), the unknown effect-size parameter ϑ varies around a level-3 unit g mean, namely

$$\vartheta_{ig}=\beta_{0g}+\eta_{ig}, \tag{8}$$

where $g=1,\ldots,m$ represents the level-3 units (e.g., school district). Finally, at the third level, the level-3 unit means vary around an overall mean γ_{00}, namely

$$\beta_{0g}=\gamma_{00}+\upsilon_{0g}, \tag{9}$$

where υ_{0g} is a level-3 unit-specific random effect that is normally distributed with a mean of zero and variance ω ($\omega>0$). In our example, ω is the between-district variance. In a single-level notation, the model is written as

$$T_{ig}=\gamma_{00}+\upsilon_{0g}+\eta_{ig}+\varepsilon_{ig}. \tag{10}$$

Now when p predictors are included at the second level, the model is

$$\vartheta_{ig}=\beta_{0g}+\beta_{1g}X_{1ig}+\ldots+\beta_{pg}X_{pig}+\eta_{ig}, \tag{11}$$

where $X_{1ig},\ldots,X_{pig}$ are study-specific predictors (e.g., year of study), $\beta_{0g},\beta_{1g},\ldots,\beta_{pg}$ are unknown regression coefficients that need to be estimated, and η_{ig} is a level-2 random effect or residual. The residuals at the second level are independently, identically, and normally distributed with a mean of zero and a residual variance τ_R. The third-level model for the level-3 unit mean (or intercept) β_{0g} when q level-3 predictors are included in the model is

$$\beta_{0g}=\gamma_{00}+\gamma_{01}W_{1g}+\ldots+\gamma_{0q}W_{qg}+\upsilon_{0g}, \tag{12}$$

where subscript R indicates residual variance, $W_{1g},\ldots,W_{qg}$ are level-3 unit-specific predictors (e.g., school district characteristics), $\gamma_{00},\gamma_{01},\ldots,\gamma_{0q}$ are unknown regression coefficients that need to be estimated, and υ_{0g} is a level-3 random effect or residual that is normally distributed with a mean of zero and residual variance ω_R. In this model, the year of study can be included in the third level as a school district predictor, and ω_R is the residual between-district variance.

The model illustrated in Equation (12) is also used to model the level-2 slopes in Equation (11). The study-specific characteristics are modeled either as fixed or random effects at the third level. Level-3 predictors can be used to model level-2 slopes as shown in Equation (12). In our example, a study characteristic is whether the study was conducted by an internal or an external evaluator (see Cooper *et al.*, 2003). The effect of this variable could vary by school district (the third-level unit) perhaps because the internal evaluator effects are more pronounced in some districts but weaker in others. As a result, it is possible that the evaluator effect varies by district. The evaluator effect is in this case a random effect and is known as a cross-level interaction between the evaluator effect and school districts. When the variance of this random effect is significant, then there is evidence of interaction.

In another example, suppose that a good number of researchers study differences in achievement between small and regular classes. Also, suppose that each researcher conducted multiple studies on this topic, that is, studies are nested within investigators. One study characteristic of interest is the type of research design the study used, whether for example, the study was a randomized experiment or not. Let us assume that each researcher conducted both experiments and non-experiments. The research design effect may interact with researchers, and as a result, it may vary across researchers. That is, the research design effect could be smaller for some researchers and larger for others. Study characteristics can also interact with specific level-3 unit characteristics, and in this case, the cross-level interaction is modeled as a fixed effect. In this example, the research design effect may interact with the experience or the training of the researcher in the field.

When p level-3 unit-specific slopes are regressed on predictors at the third level and are treated as random at the third level, these random effects are normally distributed with a mean of zero and a variance-covariance matrix $\mathbf{\Omega}_{(3)R}$ with $p+1$ residual variance components in the diagonal and covariances among these variances in the off diagonal, and υ_g is the vector of the $p+1$ level-3 random effects within level-3 unit g. The omega matrix is a $(p+1)\times(p+1)$ symmetric matrix.

The variance of an effect size estimate T_i in the simplest three-level model assuming no predictors at any level is

$$Var(T_i) = v_i + \tau + \omega, \tag{13}$$

where τ is the level-2 variance component and ω is the level-3 variance component. However, when predictors are included at the second level and are treated as random effects at the third level, the variance of an effect size estimate T_i is

$$Var(T_i) = v_i + \tau_R + \mathbf{z}_{ig}^T\mathbf{\Omega}_{(3)R}\mathbf{z}_{ig}. \tag{14}$$

where $\mathbf{z}_{ig}$ is a $(p+1)\times 1$ vector assuming $p+1$ level-2 slopes modeled as random effects at the third level. For example, if the intercept and evaluator effect are modeled as random effects at the third level the vector $\mathbf{z}$ is a 2×1 vector and the omega matrix is a 2×2 variance matrix. The random effects at the third level are the level-3 intercepts and slopes. The data are nested in the third level because each level-3 unit (e.g., school district) will include multiple level-2 units (e.g., studies).

Estimation

In the three-level model, we also need to maximize the log-likelihood function to estimate the fixed effects and the variance components of the random effects (see Appendix A). The log-likelihood equation is identical to that in the two-level model. However, the design matrix $\mathbf{X}$ of the fixed effects or predictors and the design matrix $\mathbf{Z}$ of the random effects are more complicated in the three-level case because more predictors and random effects are introduced in the regression equation (see Appendix A). In the following discussion, the estimation for the simplest case that involves an overall fixed effect, the weighted mean, and two variance components, one at the second and one at the third level, is illustrated.

Consider the simplest error structure at the third level, where only level-3 unit intercepts are random at the third level and there are no predictors at any level. Then, the design matrix of the random effects for unit g at the third level is $\mathbf{Z}_{(3,g)} = \mathbf{1}_{(n_g)}$, where $\mathbf{1}_{(n_g)}$ is a vector of ones that has as many elements as the number of studies within unit g and $\mathbf{Z}_{(3,g)}\mathbf{\Omega}_{(3)}\mathbf{Z}_{(3,g)}^T = \omega\mathbf{J}_{(n_g)}$, where $\mathbf{J}_{(n_g)} = \mathbf{1}_{(n_g)}\mathbf{1}_{(n_g)}^T$ is a $n_g\times n_g$ matrix of ones. To illustrate the variance structure, suppose that the level-3 unit g has three studies, then the structure of the variance-covariance matrix $\mathbf{V}_{(3,g)}$ within a level-3 unit g when predictors are not included at any level is

$$\mathbf{V}_{(3,g)} = \begin{bmatrix} v_1+\tau & 0 & 0 \\ 0 & v_2+\tau & 0 \\ 0 & 0 & v_3+\tau \end{bmatrix} + \omega\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} v_1+\tau+\omega & \omega & \omega \\ \omega & v_2+\tau+\omega & \omega \\ \omega & \omega & v_3+\tau+\omega \end{bmatrix}.$$

The variance structure illustrated above is similar when level-2 and level-3 predictors are included in the model as fixed effects. The only difference is that the level-2 and level-3 variances are now residual variances τ_R, ω_R. When level-2 slopes are treated as random effects at the third level, however, the variance structure is more complicated and includes the additional variance components and the design matrixes of these random effects (see Appendix A).

The computation of the log-likelihood involves the inverse of the variance-covariance matrix $\mathbf{V}$ as well as its determinant. When predictors are not included at any level, the between-study variance is τ and the between-level-3 unit variance is ω. Now consider a $n_g \times n_g$ diagonal matrix $\mathbf{D}$, defined as $\mathbf{D}_{(n_g)} = diag(v_i + \tau, i = 1, \ldots, n_g)$, and n_g represents the number of studies within a third-level unit g. Then, the inverse of the variance-covariance matrix for a third-level unit g is

$$\mathbf{V}^{-1}_{(3,g)} = \mathbf{D}^{-1}_{(n_g)} - \left(\omega^{-1} + \sum_{i=1}^{n_g}(v_i+\tau)^{-1}\right)^{-1} \mathbf{D}^{-1}_{(n_g)}\mathbf{J}_{(n_g)}\mathbf{D}^{-1}_{(n_g)}. \tag{15}$$

The diagonal elements of matrix $\mathbf{V}^{-1}_{(3,g)}$ are

$$(v_i+\tau)^{-1}\left\{1-(v_i+\tau)^{-1}\left(\omega^{-1}+\sum_{i=1}^{n_g}(v_i+\tau)^{-1}\right)^{-1}\right\},$$

and the non-diagonal elements of matrix $\mathbf{V}^{-1}_{(3,g)}$ are

$$-(v_i+\tau)^{-1}(v_j+\tau)^{-1}\left(\omega^{-1}+\sum_{i=1}^{n_g}(v_i+\tau)^{-1}\right)^{-1}.$$

The determinants of matrix $\mathbf{V}$ are $|\mathbf{V}_{(2,g)}| = \prod_{i=1}^{n_g}(v_i+\tau)$, $|\mathbf{\Omega}_{(3)R}| = \omega$, $|\mathbf{\Omega}^{-1}_{(3)R} + \mathbf{Z}^T_{(3,g)}\mathbf{V}^{-1}_{(2,g)}\mathbf{Z}_{(3,g)}| = \omega^{-1} + \sum_{i=1}^{n_g}(v_i+\tau)^{-1}$ and as a result

$$|\mathbf{V}_{(3,g)}| = \prod_{i=1}^{n_g}(v_i+\tau)\left(1+\omega\sum_{i=1}^{n_g}(v_i+\tau)^{-1}\right), \tag{16}$$

where $\mathbf{V}^{-1}_{(2,g)}$ is defined in the appendix. In the simplest case where we need to compute the overall mean and the second-level and third-level variance components, the matrix $\mathbf{X}$ is a column vector of ones, and the overall fixed effect or mean is computed using the following algebraic forms of the expected information and scoring functions, namely

$$\begin{aligned}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X} = &\sum_{g=1}^{m}\left\{\sum_{i=1}^{n_g}(v_i+\tau)^{-1}\right\}_g - \\ &\sum_{g=1}^{m}\left\{\left(\omega^{-1}+\sum_{i=1}^{n_g}(v_i+\tau)^{-1}\right)^{-1}_g\left(\sum_{i=1}^{n_g}(v_i+\tau)^{-2}+\sum_{i=1}^{n_g(n_g-1)}(v_i+\tau)^{-1}(v_j+\tau)^{-1}\right)_g\right\},\end{aligned} \tag{17}$$

and

$$\begin{aligned}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{T} = &\sum_{g=1}^{m}\left\{\sum_{i=1}^{n_g}T_i(v_i+\tau)^{-1}\right\}_g - \\ &\sum_{g=1}^{m}\left\{\left(\omega^{-1}+\sum_{i=1}^{n_g}(v_i+\tau)^{-1}\right)^{-1}_g\left(\sum_{i=1}^{n_g}T_i(v_i+\tau)^{-2}+\sum_{i=1}^{n_g(n_g-1)}T_i(v_i+\tau)^{-1}(v_j+\tau)^{-1}\right)_g\right\},\end{aligned} \tag{18}$$

where $i \neq j$, $j = 1, \ldots, n_g$. Equations (17) and (18) produce scalars and using Equation (A-3) one can compute the fixed effects estimate that is essentially the product of Equations (17) and (18).

The scoring and expected information functions for the second-level variance component τ are

$$-\frac{1}{2}\left\{\sum_{g=1}^{m} tr\left\{\mathbf{V}^{-1}_{(3,g)}\right\}_g - \sum_{g=1}^{m}\left\{\left(\mathbf{e}^T_{(3,g)}\mathbf{V}^{-1}_{(3,g)}\right)\left(\mathbf{V}^{-1}_{(3,g)}\mathbf{e}_{(3,g)}\right)\right\}_g\right\}, \tag{19}$$

and

$$\frac{1}{2}\left\{\sum_{g=1}^{m}\left\{tr\left\{\mathbf{V}^{-1}_{(3,g)}\mathbf{V}^{-1}_{(3,g)}\right\}\right\}_g\right\}. \tag{20}$$

respectively, where $\mathbf{e}$ is a row vector defined in Appendix A, and $tr(\mathbf{A})$ is the trace of matrix $\mathbf{A}$, the sum of diagonal elements of $\mathbf{A}$. The derivative $\frac{\partial \mathbf{V}}{\partial \tau}$ produces an identity matrix that is not used in the computation. Equations (19) and (20) produce scalars and the variance of τ is computed by taking the inverse of Equation (20). In the simplest case,

$$tr\left\{\mathbf{V}^{-1}_{(3,g)}\right\} = \sum_{i=1}^{n_g}\left\{(v_i+\tau)^{-1}\left(1-(v_i+\tau)^{-1}\left(\omega^{-1}+\sum_{i=1}^{n_g}(v_i+\tau)^{-1}\right)^{-1}\right)\right\}. \tag{21}$$

The scoring function for the third-level variance component matrix $\mathbf{\Omega}_{(3)R}$ for each variance or unique covariance is

$$-\frac{1}{2}\left\{\sum_{g=1}^{m}\left\{\left(\mathbf{Z}_{(3,g)}^{T}\mathbf{V}_{(3,g)}^{-1}\mathbf{Z}_{(3,g)}\right)_{jl}\right\}_{g}-\sum_{g=1}^{m}\left\{\left(\mathbf{Z}_{(3,g)}^{T}\mathbf{V}_{(3,g)}^{-1}\mathbf{e}_{(3,g)}\right)_{j}\left(\mathbf{Z}_{(3,g)}^{T}\mathbf{V}_{(3,g)}^{-1}\mathbf{e}_{(3,g)}\right)_{l}\right\}_{g}\right\} \tag{22}$$

where j,l indicate the jth row and lth column element of the matrix. The derivative $\frac{\partial \boldsymbol{\Omega}_{(3)R}}{\partial \omega_{Rjl}}$ produces an incidence matrix with zeros everywhere and one for element j,l (Longford, 1993).

The diagonal elements (variances) of the expected information matrix for the third-level variance components (variances and lower triangular matrix covariances) are computed as

$$\frac{1}{2}\left\{\sum_{g=1}^{m}\left\{\left(\mathbf{Z}_{(3,g)}^{T}\mathbf{V}_{(3,g)}^{-1}\mathbf{Z}_{(3,g)}\right)_{jl}\left(\mathbf{Z}_{(3,g)}^{T}\mathbf{V}_{(3,g)}^{-1}\mathbf{Z}_{(3,g)}\right)_{jl}\right\}_{g}\right\}. \tag{23}$$

A similar equation is used to compute the off-diagonal elements, the covariances between the variance and the covariance components, namely

$$\sum_{g=1}^{m}\left\{\left(\mathbf{Z}_{(3,g)}^{T}\mathbf{V}_{(3,g)}^{-1}\mathbf{Z}_{(3,g)}\right)_{jm}\left(\mathbf{Z}_{(3,g)}^{T}\mathbf{V}_{(3,g)}^{-1}\mathbf{Z}_{(3,g)}\right)_{kl}\right\}_{g}, \tag{24}$$

where j,m and k,l are elements of the matrices.

We assume that the random effects at different levels have a zero covariance, that is, they are not correlated. In the simplest case, the derivative $\frac{\partial \mathbf{V}_{(3,g)}}{\partial \omega}=\mathbf{J}_{n_g}$ is a $n_g \times n_g$ matrix of ones, and hence, $tr\left(\frac{\mathbf{V}_{(3,g)}^{-1}\partial \mathbf{V}_{(3,g)}}{\partial \omega}\right)=tr\left(\mathbf{V}_{(3,g)}^{-1}\mathbf{J}_{n_g}\right)=\sum_{i=1}^{n_g^2} a_i$ is the sum of all elements of $\mathbf{V}_{(3,g)}^{-1}$, and a_i represents an element of $\mathbf{V}_{(3,g)}^{-1}$.

Reasonable starting values (indicated by subscript 0) for the fixed effects estimates are estimates produced from ordinary least squares. Following Longford (1993), reasonable starting solutions for τ_R and the diagonal elements of the variance components matrix $\boldsymbol{\Omega}_{(3)R}$ at the third level are functions of the residuals, namely

$$\tau_{(0)R}=\frac{1}{k}\sum_{i=1}^{k} e_i^2$$

and

$$\boldsymbol{\Omega}_{(0)Rj}=\frac{1}{m}\frac{\sum_{g=1}^{m}\left(\mathbf{e}_{(3,g)}^{T}\mathbf{Z}_{(3,g)}^{(j)}\right)^2}{\sum_{g=1}^{m}\left(\mathbf{Z}_{(3,g)}^{(j)}\right)^{T}\mathbf{Z}_{(3,g)}^{(j)}},$$

where the superscript j represents the jth diagonal element of the third-level variance components matrix $\boldsymbol{\Omega}$.

Data

In the school calendar example, the studies were nested within school districts and thus a third level, the school district, was introduced in the model (Cooper *et al.*, 2003). In the data that were analyzed, each school district included at least three studies. The first level involves a within-study model, the second level involves a between-study within-district model, and the third level involves a between-district model. Criteria for selection were complete data that provided information on effect size estimates and type of calendar. All studies assessed students from grade 1 to grade 9 and reported achievement differences between students attending schools that follow a year-round calendar and schools that follow the traditional nine-month calendar. The achievement differences were expressed in standard deviation units to ensure that all estimates were on the same scale. The data included information on reading achievement and included 56 studies nested within 11 school districts. Positive effect sizes indicated that students in schools that followed a year-round calendar performed higher on average than students in schools that followed the traditional nine-month calendar. I first ran a two-level and then a three-level model. I also ran two different specifications. The first specification was an unconditional model with no predictors. This model estimates an overall mean as a fixed effect, and the variances at the second and third levels. The second specification added a predictor, the year the study was conducted. The year of study was used at the top level in the two-level or the three-level model.

Using software to compute estimates

To compute the fixed effects and the variance components, one can use either the SAS (Institute Inc. Cary NC USA) or the HLM (Scientific Software International Inc. Lincolnwood IL USA) software. The procedure proc mixed in SAS is well suited for two-level and three-level univariate meta-analysis. Similarly, the windows version of HLM produces estimates for two-level and three-level univariate meta-analysis. SAS proc mixed is a general purpose routine that can be used for fitting random effects models (Singer, 1998). The codes used to analyze the data with SAS proc mixed are presented in Appendix A for unconditional models. Predictors can of course be included in the regression equation. However, to obtain an intercept that represents an adjusted, by the predictors, average effect size estimate the predictors need to be centered around their grand mean before they are included in the equation. Detailed information about the proc mixed procedure is provided by Littell *et al.* (1996), Konstantopoulos and Hedges (2004), and Singer (1998). Alternatively, one could use the HLM software that is designed especially for fitting multilevel models. The windows version of HLM can be used to fit two-level and three-level meta-analysis models (Raudenbush *et al.*, 2004). The code created by HLM for two-level and three-level meta-analysis is also reported in Appendix A for unconditional models. In HLM, the user can choose to grand-mean center predictors to obtain a meaningful intercept, the overall effect size adjusted by predictors. The user needs to specify in the estimation settings that the first-level variances are known and that the weighting variable is the variance of the effect size. Detailed information about the HLM software is provided by Raudenbush *et al.* (2004).

Results

The studies included in the sample were conducted between 1976 and 2000 (see Table 1). Approximately 34% of the samples were from studies conducted in 1976, whereas 27% were from studies conducted in 1997. The remaining 39% were from studies conducted in 1989 (7%), 1994 (9%), 1995 (3.5%), and 2000 (19.5%). Approximately 61% of the samples were obtained from dissertations or theses, whereas the remaining samples were obtained from journal articles (18%), school reports (7%), or studies by the research departments of school districts (14%). Nearly 44% of the samples were obtained from studies conducted in large urban areas, another 35% from studies in small urban or suburban areas, and the remaining 21% from studies in rural areas. The samples included studies conducted in grades 1 through 9. All standardized mean differences or effect sizes used in the data do not reflect adjustments for covariates and thus were unadjusted differences between school calendar and traditional calendar schools.

The effect size estimates ranged from −0.52 to 1.19 with a mean of 0.12 and a standard deviation of 0.33 (see Table 2). Negative effect sizes indicate that students attending traditional (nine-month) calendar schools outperformed their counterparts in year-round schools. In contrast, positive effect sizes point to higher student achievement in year-round schools. The sample sizes ranged from 28 to 4403 students with a mean of 913 students. About 52% of the samples were from year-round schools on a nine-week instruction followed by a three-week break schedule, 12.5% from schools on a 12th-week instruction followed by a four-week break schedule, and nearly 35.5% from schools on other types of schedules.

The data set has an unbalanced structure. There were 11 level-3 units (districts), and within each district, the number of studies ranged from 3 to 11 with an average of 5.1. Table 3 reports means and standard deviations for effect size estimates by district. It appears that there is considerable variability within as well as between districts. District 71 had the highest mean and second highest standard deviation, whereas district 11 had the lowest mean. District 108 had the largest standard deviation and district 86 had the lowest standard deviation.

The results from the two-level meta-analysis are summarized in Table 4. Specifically, Table 4 reports the estimates of the fixed effects and the variance components, their standard errors, and 95% confidence intervals (CIs)[1] around the estimates. The overall effect size estimate was 0.128 and significant indicating that on average, students in schools that follow year-round calendars outperformed their peers in schools that follow traditional calendars. The between-study variance component was 0.088 and significantly different from zero, which indicates that the effect sizes varied across studies. The range of the 95% CI was 0.059–0.146. When the year of study was included in the model at the study level, the effect size estimate was 0.126 and still significant. The variance component estimate did not change, and the year of study was not a significant predictor of the effect sizes and did not explain any of the level-2 variance. The estimate of year of study was very close to zero. Now, using the same data, the estimate of the overall mean using fixed effects models was nearly one-half as large as the estimate of the mean using a two-level random effects model because of the different structure of the weight matrix. The estimate of year of study was still very close to zero and insignificant. In addition, the standard error of the weighted mean in the two-level random effects model was 50% larger than that in the fixed effects model.

Table 2. Summary statistics

	Reading			
	Mean	Standard deviation	Minimum	Maximum
Effect size estimate	0.120	0.326	−0.52	1.19
Type of calendar	51.79%	0.504	0.00	1.00
Sample size across studies	913.018	1459.680	28.00	4403.00

Table 3. Descriptive statistics by district

	Reading			
District ID	Mean	Standard deviation	Minimum	Maximum
11	−0.118	0.237	−0.30	0.23
12	0.095	0.250	−0.26	0.32
18	0.373	0.080	0.29	0.45
27	0.443	0.227	0.16	0.60
56	0.061	0.103	−0.06	0.19
58	−0.077	0.126	−0.28	0.07
71	0.823	0.465	0.30	1.19
86	−0.029	0.042	−0.10	0.02
91	0.247	0.271	0.00	0.66
108	0.016	0.487	−0.52	0.70
644	0.180	0.295	−0.05	0.61

Table 4. Two-level estimates of fixed effects and variance components

	Unconditional model			Including year of study		
	Estimates	Standard error	95% confidence interval	Estimates	Standard error	95% confidence interval
Fixed effects						
Intercept	0.128*	0.044	(0.040–0.216)	0.126*	0.043	(0.040–0.212)
Year of study	–	–	–	0.005	0.004	(−0.003–0.013)
Variance components						
Second level	0.088*	0.020	(0.059–0.146)	0.088*	0.020	(0.059–0.146)

*$p < 0.05$.

The three-level analysis estimates are summarized in Table 5. The structure of Table 5 is the same as the structure in Table 4. The overall effect size estimate is now 0.184 and significant. The overall weighted mean estimate is different than that in the two-level model or the fixed effects model, because a different weight matrix $\mathbf{V}^{-1}$ is used in the three-level model computation. The variance–covariance matrix is block diagonal in the three-level model, whereas it is a diagonal matrix in the two-level model (and includes the between-study variance τ). In the fixed effects model case, the weight matrix is also diagonal but includes only the v_is. A change in the magnitude of the overall mean should be expected whenever random effects are introduced in the model because the variance components of these random effects are included in the weight matrix. For example, using the same data, the estimate of the overall weighted mean using fixed effects models was one-half to one-third as large as the estimate of the mean using a three-level random effects model. In addition, the standard errors of the weighted mean and the year of study using fixed effects models were at least one-half as large as those in the three-level model. Similarly, the standard errors of the weighted mean and the year of study in the three-level model were twice as large as those in the two-level case.

The second-level variance was 0.033 and was significantly different from zero. The 95% CI was 0.020–0.070. The third-level variance was almost twice as large and also significantly different from zero. The 95% CI was 0.027–0.256. Hence, most of the random variation was between districts not between studies within districts. The advantage of the third-level model is that it provided a more accurate picture of the dependencies in the data through the variance decomposition. When the level-3 variance is significant and non-trivial in magnitude, it should be included in the computation of the regression estimates and their standard errors. In the two-level model, all random variation was assumed to be between-study variation, where in fact nearly 60% of this variation is due to district differences. When the year of study was included in the model at the third level, the effect size estimate was 0.183 and was still significant. The variance component estimates did not change much, and the year of study was not a significant predictor of the effect sizes and did not explain any of the level-2 or level-3 variance. The estimate of the year effect was similar to that in two-level model, small and insignificant. The standard errors of the fixed effects estimates were larger in the three-level model, as expected.

When the third level is not included in the analysis, the overwhelming majority of the third-level variance is part of the second-level variance. Results from previous work on multilevel models have indicated that when the top level is omitted, almost all of its variance is captured by the immediate lower level that is present (Moerbeek, 2004). In our example, when the district level was omitted, the between-district variance was captured by the between-study variance.

Table 5. Three-level estimates of fixed effects and variance components

	Unconditional model			Including year of study		
	Estimates	Standard error	95% confidence interval	Estimates	Standard error	95% confidence interval
Fixed effects						
Intercept	0.184*	0.080	(0.006–0.362)	0.183*	0.080	(0.002–0.364)
Year of study	–	–	–	0.005	0.009	(−0.015–0.025)
Variance components						
Second level	0.033*	0.010	(0.020–0.070)	0.033*	0.010	(0.020–0.070)
Third level	0.058*	0.030	(0.027–0.256)	0.056*	0.030	(0.027–0.315)

*$p < 0.05$.

Conclusion

Multilevel models have been used widely over the last two decades, and one of their main advantages is that they take into account the clustering or dependencies in the data. Conceptually, one way to think about clustering is via sampling. If one assumes for example that schools are sampled first, and then classrooms within schools are sampled, a three-level model seems appropriate for analyzing such data. However, if sampling does not take place at different levels, fixed effects models seem reasonable. Multilevel methods also seem natural methods for analyzing meta-analytic data, and three-level models have advantages over two-level models when there is variability, or clustering, at the third level. In this case, first level-3 units (e.g., school districts) would be sampled first, and then studies within districts would be sampled. Other times however, we are interested empirically in whether clustering takes place at different levels because that clustering indicates level-specific effects (e.g., study or district effects). In addition, sometimes, conceptually, we are interested in modeling predictors at the appropriate level. In the calendar data example, one could include for instance the evaluator variable as a study-specific predictor or district year or size as a third-level predictor. So long as there are enough studies and districts, the model should run with no problems.

One advantage of using three-level models is that such models allow the estimation of variance components of random effects at the third level. For example, a three-level model will estimate the between-researcher variance of study-specific effect size estimates. This model is useful and appropriate when the between-level-3 unit variance is different from zero and there are enough level-3 units in the sample to be able to estimate that variance. When the above conditions hold, the third-level variance should be included in the estimation of regression estimates and their standard errors. If the third-level variance is not different from zero, however, or when there are only a few level-3 units, one could argue that a two-level model is more appropriate because the level-3 units can be treated as fixed effects (e.g., binary indicators). In the same vein, if the between-study variance in two-level models is not different from zero, or is assumed to be zero by design, then a fixed effects model is appropriate. Nonetheless, when variance components are included in the computation of the fixed effects estimates and their standard errors, these estimates will be different than in the fixed effects models case. Specifically, the standard errors of the estimates in the random effects models that take into account variance components should be larger than those in the fixed effects models.

Another advantage of the three-level model is that study-specific variables, or slopes, at the second level can be modeled as random effects at the third level (Raudenbush & Bryk, 2002). Again the assumption is that there are enough level-3 units to compute such variance components. The example illustrated earlier was about the evaluator effect that could vary across districts. Now, consider another example that is conceptually a meta-analytic problem. Suppose that the researcher is interested in whether small classes affect the classroom variance in achievement differently than regular classes and whether this effect varies across schools. The classrooms are nested within schools. Each classroom variance (i.e., the effect size) has a known asymptotic variance (i.e., v_i) (see Raudenbush & Bryk, 1987). The main independent variable has two categories (small or regular class), is included at the study level, and captures class size effects on the classroom achievement variance. One of the researcher's objectives is whether the class size effects on the classroom achievement variance differ across schools. That is, the class size effects may not be consistent across schools and may interact with school context. In a two-level model, the estimate of the class size effect cannot be treated as random effect. In a three-level model, however, the class size estimate can be modeled as a random effect at the third level, the school, and thus, the variability of the class size effects across schools can be estimated. In this example, the research question can be addressed by using three-level models that allow the treatment effect to vary across schools.

To conclude, multilevel models are appropriate for modeling meta-analytic data with nested structure and dependencies in the data. Two-level or three-level models could be used to model meta-analytic data, and the choice of the model should be supported by the data structure and the sample sizes at each level of the hierarchy as well as the assumptions about the sampling that takes place at each level.

Appendix A: The log-likelihood function is

$$L = -\frac{1}{2}n\ \log\ (\pi) - \frac{1}{2}\ \log\ (|\mathbf{V}|) - \frac{1}{2}\mathbf{e}^T\mathbf{V}^{-1}\mathbf{e}, \tag{A-1}$$

where log is the natural logarithm, π is a mathematical constant, $\mathbf{V}$ is a $k \times k$ variance–covariance matrix, $\mathbf{e}$ is a $k \times 1$ column vector of residuals, $\mathbf{e} = \mathbf{T} - \mathbf{X}\boldsymbol{\beta}$, $\mathbf{T}$ is a $k \times 1$ vector of effect sizes, $\mathbf{X}$ is a $k \times k$ matrix of predictors, $\boldsymbol{\beta}$ is a $k \times 1$ vector of fixed effects estimates, and $|\mathbf{V}|$ is the determinant of $\mathbf{V}$.

Two-level estimation

In the two-level case, $\mathbf{V}_2$ is $k \times k$ diagonal variance–covariance matrix with diagonal elements $v_i + \tau_R$. When no predictors are included at the second level, the diagonal elements are $v_i + \tau$. The computation of the log-likelihood involves essentially the computation of the inverse and the determinant of matrix $\mathbf{V}_2$. The inverse of matrix $\mathbf{V}_2$ is also a diagonal matrix with diagonal elements $(v_i + \tau_R)^{-1}$.

Similarly, following standard results for determinants of matrices (Harville, 1997), the determinant of the diagonal matrix $\mathbf{V}_2$ is the product of the diagonal elements of $\mathbf{V}_2$

$$|\mathbf{V}_2| = \prod_{i=1}^{k}\ (v_i + \tau_R) \tag{A-2}$$

and when there are no covariates in the model $|\mathbf{V}_2| = \prod_{i=1}^{k}\ (v_i + \tau)$. In each iteration, the Fisher scoring algorithm updates the fixed effects vector

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{V}_2^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}_2^{-1}\mathbf{T}. \tag{A-3}$$

The asymptotic variance matrix of the fixed effects is computed from the inverse of the expected information matrix, namely

$$\text{var}\left(\hat{\boldsymbol{\beta}}\right) = \left(\mathbf{X}^T\mathbf{V}_2^{-1}\mathbf{X}\right)^{-1}. \tag{A-4}$$

In the simplest case, the computation of the overall mean is updated by

$$\mathbf{X}^T\mathbf{V}_2^{-1}\mathbf{X} = \sum_{i=1}^{k}\ (v_i + \tau)^{-1}, \mathbf{X}^T\mathbf{V}_2^{-1}\mathbf{T} = \sum_{i=1}^{k}\ T_i(v_i + \tau)^{-1}. \tag{A-5}$$

The Fisher scoring equation for the variance component when predictors are included in the second level is

$$\tau_R = \tau_{R0} - 2\left\{\sum_{i=1}^{k}\ \left(v_i + \tau_R\right)^{-2}\right\}^{-1}\frac{1}{2}\left\{\sum_{i=1}^{k}\ \left(v_i + \tau_R\right)^{-1} - \mathbf{e}^T\mathbf{V}_2^{-2}\mathbf{e}\right\}.$$

Three-level estimation

When we follow the general formulation for the linear mixed model, the three-level meta-analytic model in a single-level equation is

$$\mathbf{T} = \mathbf{XB} + \mathbf{Z}\xi + \boldsymbol{\varepsilon}, \tag{A-6}$$

where $\mathbf{T}$ is the vector of effect-size estimates, $\mathbf{X}$ is the design matrix of the fixed effects that includes the level-2 and level-3 predictors, $\mathbf{B}$ is the vector of fixed effects at the second and third levels that need to be estimated, $\mathbf{Z}$ is a design matrix of the random effects at the second and third levels, ξ represents the vectors of random effects at the second and third levels, and ε is the vector of the level-1 sampling errors. The matrix $\mathbf{Z}$ at the second level is $\mathbf{Z}_{(2)} = \mathbf{I}_k$, where $\mathbf{I}$ is an identity matrix, and at the third level (assuming no predictors at any level), the matrix $\mathbf{Z}_{(3)} = \mathbf{1}_k$ is a vector of ones because the third-level random effect involves the level-3 unit intercepts. When level-2 slopes are treated as random effects at the third level, the $\mathbf{Z}$ matrix is complicated and includes columns that represent these random slopes.

The variance–covariance matrix $\mathbf{V}_3$ is now a $k \times k$ block diagonal matrix

$$\mathbf{V}_3 = \mathbf{I}_m \otimes \left\{\mathbf{V}_{(3,g)}\right\} = \text{diag}\left\{\mathbf{V}_{(3,g_1)}, \ldots, \mathbf{V}_{(3,g_m)}\right\},$$

assuming m level-3 units. Each block matrix is

$$\mathbf{V}_{(3,g)} = \mathbf{V}_{(2,g)} + \mathbf{Z}_{(3,g)}\boldsymbol{\Omega}_{(3)R}\mathbf{Z}_{(3,g)}^T, \tag{A-7}$$

where $\mathbf{V}_{(2,g)} = \mathbf{I}_{n_g} \otimes \left\{ V_{(2,n_g)} \right\} = \text{diag}\left(v_1 + \tau_R, \ldots, v_{n_g} + \tau_R\right)$ is a diagonal matrix with elements $v_i + \tau_R$, and n_g is the sample size or the number of studies within a level-3 unit g, and $\otimes$ is the Kronecker product. When no predictors are included at any level, the diagonal elements are $v_i + \tau$. From standard matrix algebra, we know that

$$\mathbf{V}_3 = \mathbf{I}_m \otimes \left\{ \mathbf{V}_{(2,g)} \right\} + \mathbf{I}_m \otimes \left\{ \mathbf{Z}_{(3,g)} \mathbf{\Omega}_{(3)R} \mathbf{Z}^T_{(3,g)} \right\}, \tag{A-8}$$

where $\mathbf{Z}_{(3,g)}$ is the design matrix of the random effects at the third level within level-3 unit g (the subscript $3, g$ indicates unit g at the third level) and $\mathbf{\Omega}_{(3)R}$ is the matrix of the third-level residual variance components and covariances.

The computation of the log-likelihood involves the computation of the inverse and the determinant of matrix $\mathbf{V}_3$. To facilitate these computations, I use standard results from matrix algebra (see Harville, 1997; Longford, 1987, 1993). The inverse of matrix $\mathbf{V}_3$ is also a block diagonal matrix expressed as

$$\mathbf{V}_3^{-1} = \mathbf{I}_m \otimes \left\{ \mathbf{V}^{-1}_{(3,g)} \right\} \tag{A-9}$$

assuming a total number of m level-3 units where each block is

$$\mathbf{V}^{-1}_{(3,g)} = \mathbf{V}^{-1}_{(2,g)} - \mathbf{V}^{-1}_{(2,g)} \mathbf{Z}_{(3,g)} \left(\mathbf{\Omega}^{-1}_{(3)R} + \mathbf{Z}^T_{(3,g)} \mathbf{V}^{-1}_{(2,g)} \mathbf{Z}_{(3,g)} \right)^{-1} \mathbf{Z}^T_{(3,g)} \mathbf{V}^{-1}_{(2,g)}, \tag{A-10}$$

and

$$\mathbf{V}^{-1}_{(2,g)} = \mathbf{I}_{n_g} \otimes \left\{ V^{-1}_{(2,n_g)} \right\}, V^{-1}_{(2,n_g)} = (v_i + \tau_R)^{-1},$$

that is, $\mathbf{V}^{-1}_{(2,g)}$ is a diagonal matrix with elements $(v_i + \tau_R)^{-1}$.

Similarly, following standard results for determinants of matrices (Harville, 1997), the determinant of the block-diagonal matrix $\mathbf{V}_3$ is

$$|\mathbf{V}|_3 = \prod_{g=1}^{m} |\mathbf{V}_{(3,g)}|,$$

and the determinant of $\mathbf{V}_{(3,g)}$ is

$$|\mathbf{V}_{(3,g)}| = |\mathbf{V}_{(2,g)}||\mathbf{\Omega}_{(3)R}||\mathbf{\Omega}^{-1}_{(3)R} + \mathbf{Z}^T_{(3,g)} \mathbf{V}^{-1}_{(2,g)} \mathbf{Z}_{(3,g)}|, \tag{A-11}$$

where

$$|\mathbf{V}_{(2,g)}| = \prod_{i=1}^{n_g} V_{(2,n_g)} = \prod_{i=1}^{n_g} (v_i + \tau_R).$$

The first-order and second-order derivatives of the log-likelihood equation with respect to the second-level variance component τ are

$$-\frac{1}{2} \left\{ \sum_{g=1}^{m} tr \left\{ \mathbf{V}^{-1}_{(3,g)} \right\}_g - \sum_{g=1}^{m} \left\{ \left(\mathbf{e}^T_{(3,g)} \mathbf{V}^{-1}_{(3,g)} \right) \left(\mathbf{V}^{-1}_{(3,g)} \mathbf{e}_{(3,g)} \right) \right\}_g \right\} \tag{A-12}$$

and

$$\frac{1}{2} \sum_{g=1}^{m} tr \left\{ \mathbf{V}^{-1}_{(3,g)} \mathbf{V}^{-1}_{(3,g)} \right\}_g, \tag{A-13}$$

respectively, where $tr(\mathbf{A})$ is the trace of matrix $\mathbf{A}$, the sum of diagonal elements of $\mathbf{A}$. The derivative $\frac{\partial \mathbf{V}}{\partial \tau}$ produces an identity matrix that is not used in the computation.

The first-order derivatives of the log-likelihood equation of the third-level variance component matrix $\mathbf{\Omega}_{(3)R}$ for each variance or unique covariance is

$$-\frac{1}{2} \left\{ \sum_{g=1}^{m} \left\{ \left(\mathbf{Z}^T_{(3,g)} \mathbf{V}^{-1}_{(3,g)} \mathbf{Z}_{(3,g)} \right)_{jl} \right\}_g - \sum_{g=1}^{m} \left\{ \left(\mathbf{Z}^T_{(3,g)} \mathbf{V}^{-1}_{(3,g)} \mathbf{e}_{(3,g)} \right)_j \left(\mathbf{Z}^T_{(3,g)} \mathbf{V}^{-1}_{(3,\mathbf{g})} \mathbf{e}_{(3,\mathbf{g})} \right)_l \right\}_g \right\}, \tag{A-14}$$

where j,l indicates, respectively, the jth row and lth column element of the matrix. The derivative $\frac{\partial \mathbf{\Omega}_{(3)R}}{\partial \omega_{Rjl}}$ produces an incidence matrix with zeros everywhere and one for element j,l (Longford, 1993).

The diagonal elements (variances) of the expected information matrix for the third-level variance components (i.e., the variances and lower triangular matrix covariances) are computed as

$$\frac{1}{2}\sum_{g=1}^{m}\left\{\left(\mathbf{Z}_{(3,g)}^{T}\mathbf{V}_{(3,g)}^{-1}\mathbf{Z}_{(3,g)}\right)_{jl}\left(\mathbf{Z}_{(3,g)}^{T}\mathbf{V}_{(3,g)}^{-1}\mathbf{Z}_{(3,g)}\right)_{jl}\right\}_{g}. \tag{A-15}$$

To compute the off-diagonal elements (covariances) between the variance and the covariance components we use

$$\sum_{g=1}^{m}\left\{\left(\mathbf{Z}_{(3,g)}^{T}\mathbf{V}_{(3,g)}^{-1}\mathbf{Z}_{3,g}\right)_{jm}\left(\mathbf{Z}_{3,g}^{T}\mathbf{V}_{(3,g)}^{-1}\mathbf{Z}_{(3,g)}\right)_{kl}\right\}_{g}. \tag{A-16}$$

Two-level unconditional meta-analysis using proc mixed in SAS

```
proc mixed data=temp covtest;
        class studyid;
        model effectsize = / solution ddfm = bw notest;
        random int / sub = studyid;
        repeated / group = studyid;
        parms (0.1)
( 0.118 ) ( 0.118 ) ( 0.144 ) ( 0.144 ) ( 0.014 ) ( 0.014 ) ( 0.015 ) ( 0.024 ) ( 0.023 ) ( 0.043 )
( 0.012 ) ( 0.020 ) ( 0.004 ) ( 0.004 ) ( 0.007 ) ( 0.019 ) ( 0.007 ) ( 0.005 ) ( 0.004 ) ( 0.020 )
( 0.018 ) ( 0.019 ) ( 0.022 ) ( 0.020 ) ( 0.021 ) ( 0.006 ) ( 0.007 ) ( 0.007 ) ( 0.007 ) ( 0.007 )
( 0.015 ) ( 0.011 ) ( 0.010 ) ( 0.001 ) ( 0.001 ) ( 0.001 ) ( 0.001 ) ( 0.001 ) ( 0.001 ) ( 0.001 )
( 0.001 ) ( 0.010 ) ( 0.011 ) ( 0.010 ) ( 0.009 ) ( 0.013 ) ( 0.013 ) ( 0.031 ) ( 0.031 ) ( 0.030 )
( 0.030 ) ( 0.030 ) ( 0.087 ) ( 0.082 ) ( 0.067 ) ( 0.067 )
        / eqcons=2 to 57;
run;
```

Three-level unconditional meta-analysis using proc mixed in SAS

```
proc mixed data=temp covtest;
        class districtid studyid;
        model effectsize = / solution ddfm = bw notest;
        random int / sub = districtid;
        random int / sub = studyid(districtid);
        repeated / group = studyid(districtid);
        parms (0.1) (0.1)
( 0.118 ) ( 0.118 ) ( 0.144 ) ( 0.144 ) ( 0.014 ) ( 0.014 ) ( 0.015 ) ( 0.024 ) ( 0.023 ) ( 0.043 )
( 0.012 ) ( 0.020 ) ( 0.004 ) ( 0.004 ) ( 0.007 ) ( 0.019 ) ( 0.007 ) ( 0.005 ) ( 0.004 ) ( 0.020 )
( 0.018 ) ( 0.019 ) ( 0.022 ) ( 0.020 ) ( 0.021 ) ( 0.006 ) ( 0.007 ) ( 0.007 ) ( 0.007 ) ( 0.007 )
( 0.015 ) ( 0.011 ) ( 0.010 ) ( 0.001 ) ( 0.001 ) ( 0.001 ) ( 0.001 ) ( 0.001 ) ( 0.001 ) ( 0.001 )
( 0.001 ) ( 0.010 ) ( 0.011 ) ( 0.010 ) ( 0.009 ) ( 0.013 ) ( 0.013 ) ( 0.031 ) ( 0.031 ) ( 0.030 )
( 0.030 ) ( 0.030 ) ( 0.087 ) ( 0.082 ) ( 0.067 ) ( 0.067 )
        / eqcons=3 to 58;
run;
```

Two-level unconditional meta-analysis using HLM

```
NUMIT:100
STOPVAL:0.0000010000
NONLIN:n
LEVEL1:EFFECT_S=INTRCPT1+RANDOM
LEVEL2:INTRCPT1=INTRCPT2+RANDOM/
LEVEL1WEIGHT:NONE
LEVEL2WEIGHT:NONE
VARIANCEKNOWN:VARIANCE
RESFIL1:N
RESFIL2:N
HETEROL1VAR:n
```

```
ACCEL:5
LVR:N
LEV1OLS:10
MLF:n
HYPOTH:n
FIXSIGMA2:1.000000
FIXTAU:3
CONSTRAIN:N
OUTPUT:C:\hlm2.txt
FULLOUTPUT:N
TITLE:no title
```

Three-level unconditional meta-analysis using HLM

```
NUMIT:100
STOPVAL:0.0000010000
NONLIN:n
LEVEL1:EFFSIZE=INTRCPT1+RANDOM
LEVEL2:INTRCPT1=INTRCPT2+RANDOM/
LEVEL3:INTRCPT2=INTRCPT3+RANDOM/
LEVEL1WEIGHT:NONE
LEVEL2WEIGHT:NONE
LEVEL3WEIGHT:NONE
VARIANCEKNOWN:VARIANCE
RESFIL1:N
RESFIL2:N
RESFIL3:N
FISHERTYPE:2
HYPOTH:n
FIXSIGMA2:1.000000
FIXTAU2:3
FIXTAU3:3
CONSTRAIN:N
OUTPUT:C:\hlm3.txt
FULLOUTPUT:N
ACCEL:5
LVR-BETA:N
```

Acknowledgement

The author thanks Larry Hedges for his valuable comments. Thanks are also due to Harris Cooper and Jeff Valentine for providing the data.

References

Berkey CS, Anderson JJ, Hoaglin DC. Multiple-outcome meta-analysis of clinical trials. *Stat Med* 1996; **15**: 537–557.
Bryk AS, Raudenbush SW. Toward a more appropriate conceptualization of research on school effects. A three-level hierarchical linear model. *Am J Educ* 1988; **97**: 65–108.
Burdick RK, Graybill FA. *Confidence intervals on variance components*. Marcel Dekker, New York, 1992
Cooper H, Valentine JC, Charlton K, Melson A. The effects of modified school calendars on student achievement and on school and community attitudes: a research synthesis. *Rev Educ Res* 2003; **73**: 1–52.
Cooper H, Hedges LV, Valentine JC. 2009. *The handbook of research synthesis and meta-analysis* (2nd edn). Russell Sage, New York.
DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; **7**: 177–188.
Garvey W, Griffith B. Scientific communication: its role in the conduct of research and creation of knowledge. *Am Psychol* 1971; **26**: 349–361.
Glass GV. Primary, secondary, and meta-analysis. *Educ Res* 1976; **5**: 3–8.
Gleser LJ, Olkin I. Stochastically dependent effect sizes. In *The handbook of research synthesis and meta-analysis*, Cooper H, Hedges LV, Valentine JC (eds), Russell Sage, New York, 2009; 357–376.

Goldstein H. *Multilevel models in educational and social research*. Oxford University Press, London, 1987.
Goldstein H, Yang M, Omar R, Turner R, Thompson S. Meta-analysis using multilevel models with an application to the study of class size effects. *J R Stat Soc C Appl Stat* 2000; **49**: 399–412.
Harville DA. *Matrix algebra from a statistician's perspective*. Springer, New York, 1997.
Hedges LV. Estimation of effect size from a series of independent experiments. *Psychol Bull* 1982; **92**: 490–499.
Hedges LV. A random effects model for effect sizes. *Psychol Bull* 1983; **93**: 388–395.
Hedges LV, Olkin I. *Statistical methods for meta-analysis*. Academic Press, Orlando, FL, 1985.
Hedges LV, Vevea JL. Fixed and random effects models in meta analysis. *Psychol Meth* 1998; **3**: 486–504.
Hox JJ, de Leeuw ED. Multilevel models for meta-analysis. In *Multilevel Modeling: Methodological Advances, Issues, and Applications*, Reise SP, Duan N (eds), Lawrence Erlbaum Associates, Mahwah, NJ, 2003; 90–111.
Konstantopoulos S. *Variance components estimation in three-level meta-analysis*. Unpublished doctoral dissertation, 2003.
Konstantopoulos S. Introduction to meta-analysis. In *Best Practices in Quantitative Methods*, Osbourne J (ed.), Thousand Oaks, CA, Sage, 2007; 177–194.
Konstantopoulos S, Hedges LV. Meta-analysis. In *Handbook of Quantitative Methodology for the Social Sciences* Kaplan D (ed.), Sage, New York, 2004; 281–297.
Lipsey MW, Wilson DB. *Practical meta-analysis*. Sage, Thousand Oaks, CA, 2001.
Littell RC, Milliken GA, Stroup WW, Wolfinger RD. *SAS System for Mixed Models*. SAS Institute Inc., Cary, NC, 1996.
Longford NT. A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, 1987; **74**: 817–827.
Longford NT. *Random Coefficient Models*. Oxford University Press, New York, 1993.
Moerbeek M. The consequences of ignoring a level of nesting in multilevel analysis. *Multivariate Behav Res*, 2004; **39**: 129–149.
Nye B, Konstantopoulos S, Hedges LV. How large are teacher effects? *Educ Eval Policy Anal*, 2004; **26**: 237–257.
Raudenbush SW, Bryk AS. Examining correlates of diversity. *J Educ Stat*, 1987; **12**: 241–269.
Raudenbush SW, Bryk AS. *Hierarchical linear models*. Sage, Newbury Park, CA, 2002.
Raudenbush SW, Bryk A, Cheong YF, Congdon R. *HLM 6: hierarchical linear and onlinear modeling*. Scientific Software International, Lincolnwood, IL, 2004.
Singer JD. Using SAS PROC MIXED to fit multilevel growth models, hierarchical models, and individual growth models. *J Educ Behav Stat*, 1998; **24**: 323–355.
Snijders TAB, Bosker RJ. *Multilevel analysis*. Sage, London, 1999.
Stevens JR, Taylor AM. Hierarchical dependence in meta-analysis. *J Educ Behav Stat*, 2009; **34**: 46–73.

ENDNOTES

1. The 95% CI for the variance components was constructed using methods by Burdick and Graybill (1992). For example, the 95% CI for the second-level variance component was constructed as $v\hat{\tau}/\chi^2_{v,1-\alpha/2} \leq \tau \leq v\hat{\tau}/\chi^2_{v,\alpha/2}$, where $v = 2(\tau/SE(\tau))^2$. This formula is used in SAS proc mixed.