Multivariate Behavioral Research

The Consequence of Ignoring a Level of Nesting in Multilevel Analysis

Mirjam Moerbeek

Available online: 10 Jun 2010

PLEASE SCROLL DOWN FOR ARTICLE

The Consequence of Ignoring a Level of Nesting in Multilevel Analysis

Mirjam Moerbeek
Department of Methodology and Statistics
Utrecht University

Multilevel analysis is an appropriate tool for the analysis of hierarchically structured data. There may, however, be reasons to ignore one of the levels of nesting in the data analysis. In this article a three level model with one predictor variable is used as a reference model and the top or intermediate level is ignored in the data analysis. Analytical results show that this has an effect on the estimated variance components and that standard errors of regression coefficients estimators may be overestimated, leading to a lower power of the test of the effect of the predictor variable. The magnitude of these results depends on the ignored level and the level at which the predictor variable varies, and on the values of the variance components and the sample sizes.

Introduction

In many studies data have a nested structure, which means that individuals are nested within clusters, which may themselves be nested within higher order clusters, and so on. Examples of such studies are school-based smoking prevention or HIV/STD and pregnancy prevention interventions with pupils nested within schools (Ausems, Mesters, Van Breukelen, & De Vries, 2002; Carvajal, Baumler, Harrist, & Parcel, 2001), school effectiveness studies with pupils nested within schools (Aitkin & Longford, 1986), and community-based drug abuse prevention programs with individuals nested within communities within states (Livert, Rindskopf, Saxe, & Stirratt, 2001).

When data have a nested structure, the common assumption of independency of outcomes does not necessarily hold. In school-based smoking prevention interventions, for instance, smoking behavior of pupils within the same school is likely to be correlated due to mutual influence,

smoking behavior of teachers, and school policy towards smoking. Ignoring the nested data structure and using traditional regression analysis may lead to incorrect estimators of model parameters and their standard errors, and hence to incorrect test statistics and thereby incorrect conclusions. Moerbeek, van Breukelen, and Berger (2003) show that ignoring the nested data structure may lead too an under- or overestimation of the standard error of the treatment effect estimator in an intervention study and consequently inflated type I or type II error rates, respectively. Wampold and Serlin (2000) show that this inflation may be substantial, especially when the dependency of outcome variables of individuals within the same cluster is large. Such errors may be avoided by using the multilevel model (Goldstein, 1995; Hox, 2002; Kreft & De Leeuw, 1998; Snijders & Bosker, 1999), sometimes also referred to as hierarchical linear model (Raudenbush & Bryk, 2002) or random coefficient model (Longford, 1993). With multilevel modeling a random error term at each level of the hierarchy is included into the regression model that relates the outcome variable of interest to fixed predictor variables. The total variance of the outcome is partitioned into variance components, which are the variances of these error terms. Since the multilevel model contains both fixed and random terms it is called a mixed model.

Users of multilevel models should be very cautious and lay out clear rationale for why they decide to model a particular higher-level of variation. This rationale should be based on theory, logic or study design (e.g. cluster randomized trials, Raudenbush, 1997). A possible rationale is provided by the underlying assumption of multilevel analysis that (higher order) clusters represent a random sample from a larger population of (higher order) clusters, so that the results of the study are generalizable to this larger population if their effects are treated as random in the model. Another possible rationale for regarding a cluster as a source of higher-level of variation is that once clusters are established they will become differentiated, implying that both the cluster and its members are influenced by the cluster membership (Goldstein, 1995).

Even when the multilevel model is used, incorrect conclusions may be drawn when higher-order levels of variation are ignored, which may occur when the data set does not include identifiers on all possible levels of nesting. So far, little is known on the consequence of ignoring a level of nesting in multilevel models with more than two levels. Hutchison and Healy (2001) and Tranmer and Steel (2001) showed the effect of ignoring a level of nesting on the estimated variance components. Opdenakker and Van Damme (2000) analyzed an existing data set with four levels of nesting to show the effect of ignoring a top or intermediate level on the parameters estimates and their standard errors. They conclude that ignoring a level of nesting can lead

to different research conclusions, and they stress the need for further research on this topic. The purpose of the present article is to give a systematic overview of the consequence of ignoring a level of nesting in multilevel analysis. It is assumed the underlying data structure has three levels of nesting; the corresponding multilevel model is given in the next section. Data sets are analyzed to show the consequence of ignoring a level of nesting. The results are explained and it is shown how variation at the ignored level is redistributed to the other levels in the model and how test statistics and power levels are influenced if a level of nesting is ignored. Attention is also paid to unbalanced designs, and conclusions and a discussion are given at the end of the article.

Multilevel Model and Test Statistic

In this article a three-level data structure is assumed. For the sake of concreteness, units at the bottom, intermediate and top level are called pupils, classes, and schools, respectively. Of course, words relevant to any other field of science may be substituted. For pupil i within class j within school k the three level model that relates a quantitative (continuous) pupil level outcome variable y_{ijk} to a (continuous or binary) fixed predictor variable x_{ijk} is given by

$$(1) \qquad y_{ijk} = \gamma_0 + \gamma_1 x_{ijk} + v_k + u_{jk} + e_{ijk},$$

were γ_0 and γ_1 are the intercept and slope. v_k, u_{jk}, and e_{ijk} are the random effects at the school, class, and pupil level, respectively. These random terms are independently and normally distributed with zero mean and variances σ_v^2, σ_u^2, σ_e^2, respectively. These variances are called variance components since they contribute to the total variance of the outcome variable, given the fixed part $\gamma_0 + \gamma_1 x_{ijk}$:

$$(2) \qquad Var(y_{ijk}) = \sigma_v^2 + \sigma_u^2 + \sigma_e^2.$$

Outcomes of pupils i and i' within the same class are dependent; their covariance is

$$(3) \qquad Cov(y_{ijk}, y_{i'jk}) = \sigma_v^2 + \sigma_u^2, \text{ for } i \neq i'.$$

Likewise, the dependency of outcomes of pupils within the same school but within different classes is given by

M. Moerbeek

$$(4) \qquad Cov(y_{ijk}, y_{i'j'k}) = \sigma_v^2, \text{ for } j \neq j'.$$

The intra-school, intra-class, and intra-pupil correlation coefficients measure the amount of variance that is at the school, class, and pupil level and are calculated as

$$(5) \quad \rho_{school} = \frac{\sigma_v^2}{\sigma_e^2 + \sigma_u^2 + \sigma_v^2}, \quad \rho_{class} = \frac{\sigma_u^2}{\sigma_e^2 + \sigma_u^2 + \sigma_v^2}, \quad \rho_{pupil} = \frac{\sigma_e^2}{\sigma_e^2 + \sigma_u^2 + \sigma_v^2},$$

respectively (Davis & Scott, 1995).

In order to derive relatively simple formulae that are of practical use the following assumptions are made. The design is balanced. The number of schools in the sample is denoted n_3. Within each school n_2 classes are sampled, and within each class n_1 pupils are sampled. The predictor variable may be measured at the pupil, class or school level. A predictor variable that is measured at the pupil level is assumed to vary at the pupil level only and is centered to have zero mean within each class. Likewise, a variable measured at the class level is assumed to vary at the class level only and is centered to have zero mean within each school. A school level predictor variable is centered to have grand mean zero.

The slope γ_1 is of main interest since it determines the effect of the predictor variable x_{ijk} on outcome y_{ijk}. Since the outcomes are correlated the regression coefficients and variance components should be estimated with a suitable estimation method for multilevel analysis, such as Iterative Generalized Least squares (IGLS) (Goldstein, 1986). For a three level balanced design with one predictor variable this gives

$$(6) \qquad \hat{\gamma}_1 = \frac{\sum_{ijk} x_{ijk} y_{ijk}}{\sum_{ijk} x_{ijk}^2}.$$

For a binary predictor variable (with $x_{ijk} = -0.5$ for group A and $+0.5$ for group B) this simply reduces to $\hat{\gamma}_1 = \bar{y}_B - \bar{y}_A$, that is the difference in mean outcomes in the two groups. The variance of the slope estimator, $\hat{\gamma}_1$, depends on the variance components at and below the level at which the predictor variable varies (Moerbeek, van Breukelen, & Berger, 2000, 2001; Snijders, 2001). For three levels of nesting it can be found in Table 1.

In the next sections it will be shown that ignoring a level of nesting has an effect on the value of the test statistic for the slope γ_1, and hence on type I or type II error rates of the corresponding statistical test. The null

Table 1

Estimated Variance of the Slope Estimator without Ignoring a Level of Nesting

Level of variation	$\hat{var}(\hat{\gamma}_1)$
Pupil	$\dfrac{\hat{\sigma}_e^2}{n_1 n_2 n_3 s_x^2}$
Class	$\dfrac{\hat{\sigma}_e^2 + n_1 \hat{\sigma}_u^2}{n_1 n_2 n_3 s_x^2}$
School	$\dfrac{\hat{\sigma}_e^2 + n_1 \hat{\sigma}_u^2 + n_1 n_2 \hat{\sigma}_v^2}{n_1 n_2 n_3 s_x^2}$

Note. s_x^2 is the variance of the predictor variable.

hypothesis of zero slope (H_0: $\gamma_1 = 0$) is tested with the test statistic $z = \hat{\gamma}_1 / \sqrt{\hat{var}(\hat{\gamma}_1)}$. The computer programs MLwiN (Rasbash et al., 2001) and MIXREG (Hedeker & Gibbons, 1996) assume that this test statistic has the standard normal distribution under the null hypothesis. Given this assumption, the power $1 - \beta$ of two-sided tests with alternative H_a: $\gamma_1 \neq 0$ at a significance level α may be calculated from

$$(7) \qquad \sqrt{var(\hat{\gamma}_1)} = \frac{\Delta}{z_{1-\alpha/2} + z_{1-\beta}}$$

where Δ is the minimal relevant deviation of γ_1 from zero, and $z_{1-\alpha/2}$ and $z_{1-\beta}$ are the $100(1 - \alpha/2)\%$ and $100(1 - \beta)\%$ standard normal deviates.

The value of Δ may be difficult to specify. For a binary predictor variable the effect size (ES) for two independent means as defined by (Cohen, 1992) may be used:

$$(8) \qquad ES = \frac{\bar{y}_B - \bar{y}_A}{\text{standard deviation of outcome}}.$$

The effect size is the degree to which the null hypothesis of no difference between the two group means is believed to be false. It is scale free and ranges upwards from zero. An effect size equal to zero corresponds with H_0. Cohen uses 0.2, 0.5, and 0.8 for small, medium, and large effect sizes, respectively. A medium effect size represents an effect likely to be visible to the naked eye of a careful researcher. The numerator of Equation 8 is simply equal to γ_1, see Equation 6 and the succeeding sentence. The standard deviation of the outcome is equal to $\sqrt{\sigma_v^2 + \sigma_u^2 + \sigma_e^2}$, see Equation 2. Once the effect size is chosen and the values of the variances components are known or estimated from previous studies or a pilot, Δ follows from Equation 8 as $\Delta = \text{ES} \times \sqrt{\sigma_v^2 + \sigma_u^2 + \sigma_e^2}$ and may be substituted into Equation 7 to calculate the power.

Numerical Example

An example is presented in Table 2 to illustrate the problems associated with ignoring the top or intermediate level of nesting. A binary predictor variable is used; it may vary at the pupil, class or school level. For each of these levels a data set was generated using the three level model shown by Equation 1 with $n_1 = n_2 = n_3 = 10$, and parameter values $\gamma_0 = 0$, $\gamma_1 = 0.2$, $\sigma_v^2 = 0.1$, $\sigma_u^2 = 0.1$, and $\sigma_e^2 = 0.8$. The data were analyzed with a three level model and models ignoring one level of nesting, using IGLS as implemented in the computer program MLwiN for multilevel analysis (Rasbash et al., 2001). Note that only the cases in which the ignored level is not the level at which the predictor variable varies are considered, since it is realistic to assume that all identifiers at the latter level are well registered and available in the data set.

As follows from Table 2 ignoring a level of nesting has an effect on the estimated variance components, and this effect does not depend on the level at which the predictor variable varies. Ignoring the school level results in an unchanged estimated σ_e^2, but a higher estimated σ_u^2. In fact, the estimated σ_v^2 is added to the estimated σ_u^2. Ignoring the class level results in higher estimated σ_v^2 and σ_e^2. Ignoring a level of nesting does not have an effect on the intercept and slope estimators in case of a balanced design, but it does have an effect on their standard errors. Ignoring the class level results in a too high standard error of the slope estimator for a pupil level predictor variable, and hence to a too high *p*-value, but does not have an effect on the standard error of the slope estimator for a school level predictor variable. Ignoring the school level results in a too high standard error of the slope estimator for a class level predictor variable, and again a too high *p*-value, but does not have an effect on the standard error of the slope estimator for

Table 2
<u>Results for the Three Level Model and Models Ignoring a Level of Nesting</u>

	pupils in classes in schools			pupils in classes			pupils in schools		
	estimate	*SE*	*p*	estimate	*SE*	*p*	estimate	*SE*	*p*
Pupil level predictor variable									
γ_0	0.127	0.076		0.127	0.056		0.127	0.076	
γ_1	0.128	0.059	0.029	0.128	0.059	0.029	0.128	0.062	0.039
σ_v^2	0.029			—			0.039		
σ_u^2	0.113			0.142			—		
σ_e^2	0.863			0.863			0.966		
Class level predictor variable									
γ_0	0.175	0.103		0.175	0.070				
γ_1	0.165	0.084	0.049	0.165	0.099	0.096			
σ_v^2	0.070			—					
σ_u^2	0.096			0.166					
σ_e^2	0.799			0.799					
School level predictor variable									
γ_0	–0.016	0.118					–0.016	0.118	
γ_1	0.302	0.167	0.071				0.302	0.167	0.071
σ_v^2	0.052						0.059		
σ_u^2	0.078						—		
σ_e^2	0.958						1.030		

Note. The design is balanced with $n_1 = n_2 = n_3 = 10$.

M. Moerbeek

a pupil level predictor variable. In the next two sections it will be shown that these results hold for balanced designs in general.

Effect of Ignoring a Level of Nesting on Estimated Variance Components

The estimation method IGLS will be used to explore the effect of ignoring a level of nesting on the estimators of the variance components. IGLS produces maximum likelihood estimates in the normal case. It iterates between two steps. In the first step the fixed effects are estimated from the current estimates of the (co-) variance components using Generalized Least Squares (GLS):

$$(9) \qquad \hat{\boldsymbol{\gamma}} = (\mathbf{X'V^{-1}X})^{-1}\mathbf{X'V^{-1}Y},$$

where $\boldsymbol{\gamma}$ is the vector of regression coefficients, $\mathbf{X}$ is the design matrix for the fixed part, $\mathbf{Y}$ is the response vector, and $\mathbf{V}$ is the covariance matrix of the responses. The raw residual matrix is calculated as $\tilde{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\boldsymbol{\gamma}$. The crossproduct matrix $\mathbf{Y}^* = \tilde{\mathbf{Y}}\tilde{\mathbf{Y}}'$ has expectation $\mathbf{V}$. Define $\mathbf{Y}^{**} = \text{vec}(\mathbf{Y}^*)$, where vec is the vector operator stacking the columns of $\mathbf{Y}^*$ underneath each other. Then, $E(\mathbf{Y}^{**}) = \mathbf{Z}^*\boldsymbol{\sigma}$, where $\mathbf{Z}^*$ is the design matrix for the random part and $\boldsymbol{\sigma}$ is the vector of (co-) variance components. $\text{Var}(\mathbf{Y}^{**}) = \mathbf{V}^* = \mathbf{V} \otimes \mathbf{V}$, where $\otimes$ is the Kronecker product. In the second step the (co-) variance components are estimated from the current estimates of the fixed effects, again using GLS:

$$(10) \qquad \hat{\boldsymbol{\sigma}} = (\mathbf{X}^{*'}\mathbf{V}^{*-1}\mathbf{X}^*)^{-1}\mathbf{X}^{*'}\mathbf{V}^{*-1}\mathbf{Y}^{**}.$$

The estimation procedure iterates between these two steps until convergence is achieved. For balanced designs as described below Equation 5 and models with fixed slopes convergence is reached after one iteration and closed form estimators for the fixed effects and variance components are available.

For Model 1 with three levels of nesting application of Equation 10 gives

$$(11) \quad \begin{pmatrix} \hat{\sigma}_v^2 \\ \hat{\sigma}_u^2 \\ \hat{\sigma}_e^2 \end{pmatrix} = \frac{1}{n_1^2 n_2 n_3 (n_1-1)(n_2-1)} \begin{bmatrix} (n_1-1)SSC_s - (n_1-1)SSC_c \\ -(n_1-1)SSC_s + (n_1 n_2-1)SSC_c - n_1(n_2-1)SS_p \\ -n_1(n_2-1)SSC_c + n_1^2(n_2-1)SS_p \end{bmatrix},$$

136   MULTIVARIATE BEHAVIORAL RESEARCH

where SSC_s, and SSC_c are the sums of squares and crossproducts of residuals of pupils within the same school and class, respectively, and SS_p is the sums of squares of the residuals. Ignoring the class level gives

$$(12) \qquad \begin{pmatrix} \hat{\tilde{\sigma}}_v^2 \\ \hat{\tilde{\sigma}}_e^2 \end{pmatrix} = \frac{1}{\tilde{n}_1 \tilde{n}_3 (\tilde{n}_1 - 1)} \begin{pmatrix} SSC_s - SS_p \\ -SSC_s + \tilde{n}_1 SS_p \end{pmatrix}.$$

Note that the variance components and sample sizes are indicated with a tilde in order to distinguish them from those for three levels of nesting. In this case $\tilde{n}_3$ is the number of schools, and $\tilde{n}_1$ is the number of pupils per school, that is $\tilde{n}_1 = n_1 n_2$. Ignoring the school level results in

$$(13) \qquad \begin{pmatrix} \hat{\tilde{\sigma}}_u^2 \\ \hat{\tilde{\sigma}}_e^2 \end{pmatrix} = \frac{1}{\tilde{n}_1 \tilde{n}_2 (\tilde{n}_1 - 1)} \begin{pmatrix} SSC_c - SS_p \\ -SSC_c + \tilde{n}_1 SS_p \end{pmatrix}$$

where $\tilde{n}_2 = n_2 n_3$ is the total number of classes, and $\tilde{n}_1 = n_1$ is the number of pupils per class.

Now it can be easily shown that ignoring the class level results in

$$(14) \qquad \begin{aligned} \hat{\tilde{\sigma}}_v^2 &= \hat{\sigma}_v^2 + \frac{n_1 - 1}{n_1 n_2 - 1} \hat{\sigma}_u^2 \\ \hat{\tilde{\sigma}}_e^2 &= \hat{\sigma}_e^2 + \frac{n_1 n_2 - n_1}{n_1 n_2 - 1} \hat{\sigma}_u^2 \end{aligned}.$$

So, the estimated variance component at the class level is distributed over the two other estimated variance components. For any fixed n_1, the fraction of $\hat{\sigma}_u^2$ that is added to $\hat{\sigma}_e^2$ is equal to 0 if $n_2 = 1$, and increases if n_2 increases. For any fixed n_2, the fraction of $\hat{\sigma}_u^2$ that is added to $\hat{\sigma}_e^2$ is equal to 1 if $n_1 = 1$ and decreases if n_1 increases. Of course, the change in the estimated variance components at the pupil and school level is low if $\hat{\sigma}_u^2$ is low.

The situation is less complicated when ignoring the school level:

$$(15) \qquad \begin{aligned} \hat{\tilde{\sigma}}_u^2 &= \hat{\sigma}_v^2 + \hat{\sigma}_u^2 \\ \hat{\tilde{\sigma}}_e^2 &= \hat{\sigma}_e^2 \end{aligned}.$$

In this case, the variance component at the school level is added to that at the class level, and the variance component at the pupil level remains

unchanged. These results are in agreement with those of Tranmer and Steel (2001) for the moment estimator, and those of Hutchison and Healy (2001) in a random effects ANOVA.

It should be noted that relations of Equations 14 and 15 hold for any number and type of predictor variables (i.e. continuous or binary) as long as IGLS is used. A drawback of this estimation method is that it produces biased estimates for small samples. This problem may be overcome by using Restricted Iterative Generalized Least Squares (Goldstein, 1989). For this estimation method the crossproduct matrix $\mathbf{Y}^* = \tilde{\mathbf{Y}}\tilde{\mathbf{Y}}'$ has expectation $\mathbf{V} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'$. Consequently, closed form relations between estimated variance components in models without or with ignoring a level of nesting depend on the values of the predictor variables, and are not that easily derived. For large sample sizes, however, these relations may be expected to be approximately equal to relations of Equations 14 and 15 for IGLS.

Effect of Ignoring a Level of Nesting on Test Statistics

Ignoring a level of nesting also has an effect on the variance of the slope estimator. Table 3 shows this variance, $\hat{\text{var}}(\hat{\gamma}_1)$, for a predictor variable that varies at the pupil, class, or school level. As in the numerical example only the cases where the ignored level is not the level at which the predictor variable varies are considered. The third column of Table 3 shows the $\hat{\text{var}}(\hat{\gamma}_1)$ expressed in terms of $\hat{\tilde{\sigma}}_e^2, \hat{\tilde{\sigma}}_u^2, \hat{\tilde{\sigma}}_v^2, \tilde{n}_1, \tilde{n}_2, \tilde{n}_3$, which are the estimated variance components and the sample sizes when ignoring a level of nesting. The relationships between these and their counterparts in a three level model depend on the ignored level and can be found in the previous section. Substituting these relationships into the $\hat{\text{var}}(\hat{\gamma}_1)$ expressed in terms of $\hat{\tilde{\sigma}}_e^2, \hat{\tilde{\sigma}}_u^2, \hat{\tilde{\sigma}}_v^2, \tilde{n}_1, \tilde{n}_2, \tilde{n}_3$ gives the $\hat{\text{var}}(\hat{\gamma}_1)$ expressed in terms of $\hat{\sigma}_e^2, \hat{\sigma}_u^2, \hat{\sigma}_v^2, n_1, n_2, n_3$ as found in the last column of Table 3.

Pupil Level Predictor Variable

As follows from Table 3 the $\hat{\text{var}}(\hat{\gamma}_1)$ for a pupil level predictor variable is equal to $\hat{\sigma}_e^2 / (n_1 n_2 s_x^2)$ if the school level is ignored, which is exactly equal to the $\hat{\text{var}}(\hat{\gamma}_1)$ in a three level model, see Table 1. Thus, ignoring the school level does not have an effect on the $\hat{\text{var}}(\hat{\gamma}_1)$ of a pupil level predictor variable in case of a balanced design.

Table 3 also shows that the $\hat{\text{var}}(\hat{\gamma}_1)$ for a pupil level predictor variable is equal to

Table 3
Estimated Variance of the Slope Estimator with Ignoring a Level of Nesting

Level of variation	Ignored level	$\text{vâr}(\hat{\gamma}_1)$ in terms of $\hat{\tilde{\sigma}}_e^2, \hat{\tilde{\sigma}}_u^2, \hat{\tilde{\sigma}}_v^2, \tilde{n}_1, \tilde{n}_2, \tilde{n}_3$	$\text{vâr}(\hat{\gamma}_1)$ in terms of $\hat{\sigma}_e^2, \hat{\sigma}_u^2, \hat{\sigma}_v^2, n_1, n_2, n_3$
pupil	school	$\dfrac{\hat{\tilde{\sigma}}_e^2}{\tilde{n}_1 \tilde{n}_2 s_x^2}$	$\dfrac{\sigma_e^2}{n_1 n_2 n_3 s_x^2}$
pupil	class	$\dfrac{\hat{\tilde{\sigma}}_e^2}{\tilde{n}_1 \tilde{n}_3 s_x^2}$	$\dfrac{\hat{\sigma}_e^2 + \dfrac{n_1 n_2 - n_1}{n_1 n_2 - 1}\hat{\sigma}_u^2}{n_1 n_2 n_3 s_x^2}$
class	school	$\dfrac{\hat{\tilde{\sigma}}_e^2 + \tilde{n}_1 \hat{\tilde{\sigma}}_u^2}{\tilde{n}_1 \tilde{n}_2 s_x^2}$	$\dfrac{\hat{\sigma}_e^2 + n_1(\hat{\sigma}_u^2 + \hat{\sigma}_v^2)}{n_1 n_2 n_3 s_x^2}$
school	class	$\dfrac{\hat{\tilde{\sigma}}_e^2 + \tilde{n}_1 \hat{\tilde{\sigma}}_v^2}{\tilde{n}_1 \tilde{n}_3 s_x^2}$	$\dfrac{\hat{\sigma}_e^2 + n_1 \hat{\sigma}_u^2 + n_1 n_2 \hat{\sigma}_v^2}{n_1 n_2 n_3 s_x^2}$

Note. s_x^2 is the variance of the predictor variable.

$$\left(\hat{\sigma}_e^2 + \frac{n_1 n_2 - n_1}{n_1 n_2 - 1}\hat{\sigma}_u^2 \right) \Big/ \left(n_1 n_2 n_3 s_x^2 \right)$$

if the class level is ignored. The term

$$\frac{n_1 n_2 - n_1}{n_1 n_2 - 1}\hat{\sigma}_u^2$$

does not appear in the numerator of the $\text{vâr}(\hat{\gamma}_1)$ for three levels of nesting, see Table 1. As written below Equation 14, this term increases if n_2 increases and if n_1 decreases. It is also larger if $\hat{\sigma}_u^2$ is larger. So, ignoring the class level leads to a larger $\text{vâr}(\hat{\gamma}_1)$, and hence to a smaller test statistic

for γ_1 and therefore a lower power for the statistical test of the effect of the pupil level predictor variable on the outcome.

To gain insight in the magnitude of loss in power some calculations were done. Assume a three level data structure with $\rho_{pupil} = 0.7$, $\rho_{class} = 0.15$, and $\rho_{school} = 0.15$, and sample sizes $n_1 = 4$ pupils per class and $n_2 = 4$ classes per school. The predictor variable varies at the pupil level and is binary with coding scheme –0.5 and +0.5 (for instance, gender). Aim is to detect a small effect size with power $1 - \beta = 0.9$ at a significance level of $\alpha = 0.05$ in a two sided test. From Equation 8 it follows that $\Delta = \text{ES} \times \sqrt{\sigma_v^2 + \sigma_u^2 + \sigma_e^2} = 0.2 \times \sqrt{\sigma_v^2 + \sigma_u^2 + \sigma_e^2}$, and from Equation 7 it follows that

$$(16) \quad \text{var}(\hat{\gamma}_1) = \Delta^2 / (z_{1-\alpha/2} + z_{1-\beta})^2 = 0.2^2 \times (\sigma_v^2 + \sigma_u^2 + \sigma_e^2)/(1.96 + 1.28)^2.$$

From Table 1 it follows that

$$(17) \quad \hat{\text{var}}(\hat{\gamma}_1) = \sigma_e^2 / (n_1 n_2 n_3 s_x^2) = 4 \times \rho_{pupil} \times (\sigma_e^2 + \sigma_u^2 + \sigma_v^2)/(n_1 n_2 n_3)$$

since $s_x^2 = 0.25$ for a binary predictor coded –0.5 and +0.5, and $\sigma_e^2 = \rho_{pupil}(\sigma_e^2 + \sigma_u^2 + \sigma_v^2)$, see Equation 5. The $\hat{\text{var}}(\hat{\gamma}_1)$ as given by Equation 17 should be equal to $\hat{\text{var}}(\hat{\gamma}_1)$ as calculated from Equation 16 in order to reach the desired power. The number of schools, n_3, is the only unknown since the term $(\sigma_e^2 + \sigma_u^2 + \sigma_v^2)$ in Equations 16 and 17 cancels out. For the given n_1, n_2, and ρ_{pupil} the number of schools is almost equal to 46 ($n_3 = 45.97$). Assume now that the class level is ignored. The sample sizes and values of variance components are plugged into the $\hat{\text{var}}(\hat{\gamma}_1)$ ignoring the class level as given in Table 3 and $1 - \beta$ is subsequently calculated from Equation 7. In this case $1 - \beta = 0.850$, which is somewhat lower than the nominal power obtained without ignoring the class level.

These calculations were performed for different values of the sample sizes and the intra-pupil, -class, and -school correlation coefficients. Table 4 shows results for three different values of the number of pupils per class ($n_1 = 4, 8$, or 16), three different values for the number of classes per school ($n_2 = 4$, 8, or 16), and three different values of the intra-pupil correlation coefficient ($\rho_{pupil} = 0.7, 0.8$, or 0.9). For each value of ρ_{pupil}, the three values of the intra-class correlation coefficient were chosen such that $\rho_{class} = \lambda (1 - \rho_{pupil})$, where $\lambda = 0.5, 0.7$, or 0.9. The intra-school correlation coefficient follows from $\rho_{school} = 1 - \rho_{pupil} - \rho_{class}$. These values were chosen to represent the fact that the ρ at the bottom level is generally (much) larger than the ρ at the intermediate level, which in its turn is generally (much) larger than the ρ at the top level (see Gulliford, Ukoumunne, & Chinn, 1999; Siddiqui, Hedeker, Flay, & Hu, 1996, for example values of the ρs at different levels).

Table 4
Power Levels of the Test of the Effect of a Binary Pupil Level Predictor Variable when Ignoring the Class Level

	$\rho_{pupil} = 0.7$			$\rho_{pupil} = 0.8$			$\rho_{pupil} = 0.9$		
	$\rho_{class} =$ 0.15	$\rho_{class} =$ 0.21	$\rho_{class} =$ 0.27	$\rho_{class} =$ 0.10	$\rho_{class} =$ 0.14	$\rho_{class} =$ 0.18	$\rho_{class} =$ 0.05	$\rho_{class} =$ 0.07	$\rho_{class} =$ 0.09
$n_1 = 4$									
$n_2 = 4$	0.850	0.829	0.809	0.871	0.859	0.847	0.887	0.882	0.877
$n_2 = 8$	0.843	0.820	0.797	0.867	0.854	0.840	0.886	0.880	0.874
$n_2 = 16$	0.840	0.816	0.792	0.865	0.851	0.837	0.885	0.879	0.872
$n_1 = 8$									
$n_2 = 4$	0.851	0.832	0.812	0.872	0.860	0.849	0.888	0.883	0.878
$n_2 = 8$	0.844	0.821	0.799	0.868	0.854	0.841	0.886	0.880	0.874
$n_2 = 16$	0.840	0.816	0.792	0.866	0.851	0.837	0.885	0.879	0.873
$n_1 = 16$									
$n_2 = 4$	0.852	0.833	0.813	0.872	0.861	0.850	0.888	0.883	0.878
$n_2 = 8$	0.844	0.822	0.800	0.868	0.855	0.842	0.886	0.880	0.874
$n_2 = 16$	0.841	0.817	0.793	0.866	0.852	0.838	0.885	0.879	0.873

Note. Nominal power level = 0.9, ES = 0.2, α = 0.05.

Table 4 shows that ignoring the class level has some effect on the power, but this effect is not very tremendous for the chosen sample sizes and values of the intra-pupil, -class, and -school correlation coefficients. The lowest power level is 0.792, which is 12% lower than the nominal power level of 0.9 that would have been obtained with three levels of nesting. As follows from Table 4 the power increases with increasing ρ_{pupil} (and hence decreasing ρ_{class}) and n_1, and with decreasing ρ_{class} and n_2. This is in concordance with what is to be expected from the change in variance components given by Equation 14.

Class Level Predictor Variable

As follows from Table 3 the $\mathrm{v\hat{a}r}(\hat{\gamma}_1)$ for a class level predictor variable is equal to $\left[\hat{\sigma}_e^2 + n_1(\hat{\sigma}_u^2 + \hat{\sigma}_v^2)\right]/\left(n_1 n_2 n_3 s_x^2\right)$ if the school level is ignored. The term $n_1\hat{\sigma}_v^2$ does not appear in the numerator of the $\mathrm{v\hat{a}r}(\hat{\gamma}_1)$ for three levels of nesting, see Table 1. So, ignoring the school level results in a higher $\mathrm{v\hat{a}r}(\hat{\gamma}_1)$ and thus in a lower power of the test of the effect of a class level predictor on the outcome. The power decreases with increasing n_1 and $\hat{\sigma}_v^2$, but does not depend on n_2.

To gain insight in the magnitude of loss in power, the same calculations as in the previous subsection were performed, again using a small effect size, a significance level $\alpha = 0.05$, and a nominal power level of 0.9. The results are given in Table 5. As follows from Table 5 the consequences of ignoring the school level on the power of the test of a class level predictor variable are not trivial. Power levels may be as low as 0.682, which is 24% smaller than the nominal power level of 0.9. Tables 5 shows that the power increases with increasing ρ_{pupil} and ρ_{class} (and hence decreasing ρ_{school}), and with decreasing n_1, but does not depend on n_2. This is in concordance with what is to be expected from the change in variance components given in Equation 15.

School Level Predictor Variable

The $\mathrm{v\hat{a}r}(\hat{\gamma}_1)$ of a school level predictor variable is equal to $(\hat{\sigma}_e^2 + n_1\hat{\sigma}_u^2 + n_1 n_2\hat{\sigma}_v^2)/(n_1 n_2 n_3 s_x^2)$ if the class level is ignored, see Table 3. Comparison with Table 1 shows that it is equal to the $\mathrm{v\hat{a}r}(\hat{\gamma}_1)$ without ignoring the class level. Thus, ignoring the class level does not have an effect on $\mathrm{v\hat{a}r}(\hat{\gamma}_1)$ of a school level predictor variable in the case of a balanced design.

Table 5
Power Levels of the Test of the Effect of a Binary Class Level Predictor Variable when Ignoring the School Level

	$\rho_{pupil} = 0.7$			$\rho_{pupil} = 0.8$			$\rho_{pupil} = 0.9$		
	$\rho_{class} =$ 0.15	$\rho_{class} =$ 0.21	$\rho_{class} =$ 0.27	$\rho_{class} =$ 0.10	$\rho_{class} =$ 0.14	$\rho_{class} =$ 0.18	$\rho_{class} =$ 0.05	$\rho_{class} =$ 0.07	$\rho_{class} =$ 0.09
$n_1 = 4$	0.765	0.831	0.881	0.802	0.848	0.885	0.847	0.870	0.891
$n_1 = 8$	0.718	0.811	0.876	0.754	0.826	0.879	0.809	0.852	0.886
$n_1 = 16$	0.682	0.795	0.872	0.709	0.807	0.875	0.762	0.830	0.880

Note. Nominal power level = 0.9, ES = 0.2, α = 0.05. The results do not depend on n_2.

M. Moerbeek

A Note on Repeated Measures Data

The values of ρ_{pupil}, ρ_{class}, and ρ_{school} that were used in Tables 4 and 5 were chosen to represent the fact that for multilevel data with individuals nested within clusters within higher order clusters the ρ at the bottom level is generally (much) larger than the ρ at the intermediate level, which in its turn is generally (much) larger than the ρ at the top level. This fact does not hold for three-level models for repeated measures data where the bottom, intermediate, and top levels represent repeated measures, persons, and clusters (e.g. therapy groups, or schools). With such data individual differences account for substantial variation among all the levels modeled, and values of ρ at the intermediate (person) level as large as 0.5 of even larger are not unrealistic.

The effects of ignoring a level of nesting in a repeated measures study are similar to those presented in Equations 14-15 and Tables 3-5 with the terms pupil, class, and school replaced by the terms repeated measure, person, and cluster, respectively. It should, however, be noted that these results only hold if the errors e_{ijk} are uncorrelated, an assumption that does not always hold for repeated measures data. It is uncommon to ignore the person level in a repeated measures study. The power when ignoring the cluster level may be calculated as explained in the first subsection of this section with the $v\hat{a}r(\hat{\gamma}_1)$ for a top level predictor used in Equation 17. For instance, assume that $\rho = 0.4$ at the repeated measures level, $\rho = 0.5$ at the person level, $n_1 = 4$ measurements per person are taken, and $n_2 = 40$ persons per cluster are available. Then, $n_3 = 15.76$ clusters would be needed to detect a small effect size of a binary person level predictor with power $1 - \beta = 0.9$ and a significance level $\alpha = 0.05$ in a two sided test. Ignoring the cluster level would result in a power of 0.851, a loss in power of 5.4%.

Unbalanced Designs

It should be noted that a balanced design was assumed in the previous sections since that leads to relatively simple formulae for the change in variance components (Equations 14 and 15) when ignoring a level of nesting and to relatively simple formulae for the $v\hat{a}r(\hat{\gamma}_1)$ in a three or two level design (Tables 1 and 3). Moreover, a balanced design leads to a smaller $v\hat{a}r(\hat{\gamma}_1)$ than a design with the same total number of pupils but with varying class and school sizes (Manatunga, Hudges, & Chen, 2001) and thus to a higher power for the test of the effect of a predictor variable on the outcome.

The assumption of a balanced design is not always met in practice. In repeated measures designs the number of repeated measures per person may be under experimental control, but in educational studies, the number of

pupils per class and the number of classes per school varies by nature. If sample sizes vary, their mean values may be substituted into Equations 14 and 15 and the formulae in Tables 1 and 3, which then only hold approximately. The approximation is, of course, less accurate if the variability in sample sizes is large.

In order to show the effects of ignoring a level of nesting on parameter estimates and their standard errors in case of an unbalanced design, data sets where generated and analyzed as described for the numerical example for the balanced design, using the same parameter values. The numbers of pupils per class and the numbers of classes per school were drawn from the normal distribution with mean 10 and standard deviation 5. The number of schools was equal to ten. The results of the data analyses with three and two levels of nesting are given in Table 6. The effects of ignoring a level of nesting on estimated variance components are similar to those in a balanced design: the estimated variance components at both the school and pupil level increase if the class level is ignored, and only the estimated variance component at the class level increases if the school level is ignored. Ignoring a level of nesting also has an effect on the estimated slopes, especially of those of class and school level predictors. Furthermore, the standard error of the estimated slope of a school level predictor is slightly overestimated if the class level is ignored. As was shown previously the latter two results did not hold for balanced designs. From the examples in Table 6 it should be clear that with an unbalanced design one should take care not to ignore a level of nesting in a multilevel model since otherwise the point estimates and associated standard errors of many, if not all, model parameters may be incorrect and hence conclusions drawn from p-values are not to be trusted.

Conclusions and Discussion

This article explored the consequences of ignoring a top or intermediate level in a three level data structure. Analytical results showed that, in case the top level is ignored, the variance component at the top level is added to that at the intermediate level while the variance component at the bottom level remains unchanged. The variance component at the intermediate level is distributed over the variance components at the top and bottom level if the intermediate level is ignored. This distribution depends on the sample sizes at the intermediate and bottom level.

Ignoring a level of nesting also has an effect on the power level of a statistical test of the effect of a predictor variable on the outcome. Only the cases in which the predictor variable of interest does not vary at the ignored level were considered. Analytical results showed that ignoring the top level

Table 6
Results for the Three Level Model and Models Ignoring a Level of Nesting

	pupils in classes in schools			pupils in classes			pupils in schools		
	estimate	SE	p	estimate	SE	p	estimate	SE	p
Pupil level predictor variable									
γ_0	0.115	0.080		0.136	0.045		0.113	0.078	
γ_1	0.086	0.052	0.110	0.084	0.052	0.110	0.084	0.055	0.123
σ_v^2	0.046			—			0.052		
σ_u^2	0.088			0.139			—		
σ_e^2	0.768			0.767			0.849		
Class Level Predictor Variable									
γ_0	–0.051	0.128		–0.058	0.058				
γ_1	0.123	0.091	0.178	0.085	0.115	0.463			
σ_v^2	0.139			—					
σ_u^2	0.152			0.296					
σ_e^2	0.757			0.757					
School Level Predictor Variable									
γ_0	–0.126	0.091					–0.133	0.093	
γ_1	0.253	0.182	0.202				0.195	0.186	0.325
σ_v^2	0.053						0.069		
σ_u^2	0.087						—		
σ_e^2	0.784						0.861		

Note. The design is unbalanced with $n_1 \sim N(10,5)$, $n_2 \sim N(10,5)$, and $n_3 = 10$.

does not have an effect on testing the effect of a bottom level predictor variable, while power would be to low if the effect of an intermediate level predictor variable is to be tested. The loss in power becomes small if n_1 is small and/or if the variance component at the top level is small in relation to the variance components at the intermediate and bottom level. Ignoring the intermediate level variable does not have an effect on the results of the statistical test of a top level predictor variable, while power would be too low if the effect of a bottom level predictor variable is to be tested. In this case, the loss in power becomes small if n_1 is large and/or if n_2 is small and/or if the variance component at the intermediate level is small in relation to the variance component at the bottom level. In conclusion: ignoring a level of nesting in a three level model does not create substantial problems with respect to the test of a predictor variable if the variability of the outcome variable at the ignored level is relatively small, and does not create problems at all if the predictor variable of interest does not vary at the level immediately below the ignored level.

These findings, however, only hold for balanced designs. Analyses of data sets showed that for unbalanced designs not only the standard error of the estimated effect of a predictor variable on the outcome may be overestimated, also the estimated effect itself may be incorrect. So with unbalanced designs one should expect the estimates and associated standard errors of many, if not all, model parameters to be incorrect if a level of nesting is ignored in the data analysis, and hence the conclusions drawn are not to be trusted.

For simplicity a model with just one predictor variable was used in this article, but it can be shown that the results also hold for models with more predictor variables as long as each predictor variable varies at just one level of the multilevel data structure. A predictor variable x_{ijk} that varies at more that one level may be split up into orthogonal components that each vary at just one level (Neuhaus & Kalbfleisch, 1998). In a three level model the orthogonal components are calculated as $x_{ijk} - \bar{x}_{.jk}$, $\bar{x}_{.jk} - \bar{x}_{..k}$, and $\bar{x}_{..k}$. The first varies within classes and has zero class mean, the second varies between classes within school and has zero school mean, and the third varies between schools. As Neuhaus and Kalbfleisch (1998) point out the effects of these three components of the predictor variable on the outcome are not necessarily the same. They show that a model that incorrectly assumes these effects are the same can lead to very misleading results of the effect of the predictor variable on the outcome.

Of course the results only hold if the same predictors are used in the model with three levels of nesting and in the model with ignoring a level of nesting. The situation becomes more complex if predictor variables at the ignored level are also removed from the multilevel model since then the

variance components, and hence standard errors of regression coefficients estimators, are not only affected by ignoring the level of nesting but also by the removal of these predictor variables from the multilevel model (Snijders & Bosker, 1994).

References

Aitkin, M. & Longford, N. T. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society Series B Statistical Methodology, 149*, 1-43.

Ausems, M., Mesters, I., Van Breukelen, G., & De Vries, H. (2002). Short-term effects of a randomized computer-based out-of-school smoking prevention trial aimed at elementary school children. *Preventive Medicine, 34*, 581-589.

Carvajal, S. C., Baumler, E., Harrist, R. B., & Parcel, G. S. (2001). Multilevel models and unbiased tests for group based interventions: Examples from the safer choices study. *Multivariate Behavioral Research, 36*, 185-205.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Davis, P. & Scott, A. (1995). The effect of interviewer variance on domain comparisons. *Survey Methodology, 21*, 99-106.

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

Goldstein, H. (1989). Restricted unbiased iterative generalized least squares estimation. *Biometrika, 76*, 622-623.

Goldstein, H. (1995). *Multilevel statistical models* (2nd Ed.). London: Edward Arnold.

Gulliford, M. C., Ukoumunne, O. C., & Chinn, S. (1999). Components of variance and intraclass correlations for the design of community-based surveys and intervention studies: data from the Health Survey for England 1994. *American Journal of Epidemiology, 149*, 876-883.

Hedeker, D. & Gibbons, R. D. (1996). MIXREG: a computer program for mixed-effects regression analysis with autocorrelated errors. *Computer Methods and Programs in Biomedicine, 49*, 229-252.

Hox, J. (2002). *Multilevel analysis. Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hutchison, D. & Healy, M. (2001). The effect of variance component estimates of ignoring a level in a multilevel model. *Multilevel Modelling Newsletter, 13*, 4-5.

Kreft, I. & De Leeuw, J. (1998). *Introducing multilevel modeling*. London: Sage Publications.

Livert, D., Rindskopf, D., Saxe, L., & Stirratt, M. (2001). Using multilevel modeling in the evaluation of community-based treatment programs. *Multivariate Behavioral Research, 36*, 155-183.

Longford, N. T. (1993). *Random coefficient models*. Oxford: Clarendon Press.

Manatunga, A. K., Hudges, M. G., & Chen, S. (2001). Sample size estimation in cluster randomized studies with varying cluster size. *Biometrical Journal, 43*, 75-86.

Moerbeek, M., van Breukelen, G. J. P., & Berger, M. P. F. (2000). Design issues for experiments in multilevel populations. *Journal of Educational and Behavioral Statistics, 25*, 271-284.

Moerbeek, M., van Breukelen, G. J. P., & Berger, M. P. F. (2001). Optimal experimental designs for multilevel models with covariates. *Communications in Statistics, Theory and Methods, 30*, 2683-2697.

Moerbeek, M., van Breukelen, G. J. P., & Berger, M. P. F. (2003). A comparison between traditional methods and multilevel regression for the analysis of multi-center intervention studies. *Journal of Clinical Epidemiology, 56*, 341-350.

Neuhaus, J. M. & Kalbfleisch, J. D. (1998). Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics, 54*, 638-645.

Opdenakker, M.-C. & Van Damme, J. (2000). The importance of identifying levels in multilevel analysis: An illustration of the effects of ignoring the top or intermediate levels in school effectiveness research. *School Effectiveness and School Improvement, 11*, 103-130.

Rasbash, J., Goldstein, H., Browne, W., Yang, M., Plewis, I., Healy, M., et al. (2001). *A user's guide to MLwiN (Version 2.1c)*. London: Institute of Education.

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods, 2*, 173-185.

Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods*. Thousand Oaks: Sage Publications.

Siddiqui, O., Hedeker, D., Flay, B. R., & Hu, F. B. (1996). Intraclass correlation estimates in a school-based smoking prevention study. *American Journal of Epidemiology, 144*, 425-433.

Snijders, T. A. B. (2001). Sampling. In A. H. Leyland & H. Goldstein (Eds.), *Multilevel modelling of health statistics* (pp. 159-174). New York: John Wiley & Sons.

Snijders, T. A. B. & Bosker, R. J. (1994). Modeled variance in two-level models. *Sociological Methods and Research, 22*, 342-363.

Snijders, T. A. B. & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage Publications.

Tranmer, M. & Steel, D. G. (2001). Ignoring a level in a multilevel model: evidence from UK census data. *Environment and Planning A, 33*, 941-948.

Wampold, B. E. & Serlin, R. C. (2000). The consequence of ignoring a nested factor on measures of effect size in analysis of variance. *Psychological Methods, 5*, 425-433.