

Robustness of parameter and standard error estimates against ignoring a contextual effect of a subject-level covariate in cluster-randomized trials

Elly J. H. Korendijk · Joop J. Hox · Mirjam Moerbeek · Cora J. M. Maas

Published online: 22 April 2011

© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract In experimental research, it is not uncommon to assign clusters to conditions. When analysing the data of such cluster-randomized trials, a multilevel analysis should be applied in order to take into account the dependency of first-level units (i.e., subjects) within a second-level unit (i.e., a cluster). Moreover, the multilevel analysis can handle covariates on both levels. If a first-level covariate is involved, usually the within-cluster effect of this covariate will be estimated, implicitly assuming the contextual effect to be equal. However, this assumption may be violated. The focus of the present simulation study is the effects of ignoring the inequality of the within-cluster and contextual covariate effects on parameter and standard error estimates of the treatment effect, which is the parameter of main interest in experimental research. We found that ignoring the inequality of the within-cluster and contextual effects does not affect the estimation of the treatment effect or its standard errors. However, estimates of the variance components, as well as standard errors of the constant, were found to be biased.

Keywords Group-randomized design · Multilevel analysis · Within-cluster regression · Between-cluster regression · Hierarchical linear model · Random coefficient model

In experimental research, it is not uncommon to assign clusters rather than individuals to conditions. For example, in educational research, schools or classes may be assigned to either the treatment or control condition; in organisational psychology, it

may be companies that are assigned to conditions; in medical research, it may be hospitals or general practices; and in health psychology, it may be therapy groups. Experimental designs such as these are referred to as *cluster-randomized trials* (see, e.g., Murray, 1998). In cluster-randomized trials, the assumption of independent observations is violated. Due to shared features, shared leadership, and mutual influences, subjects within the same cluster are likely to respond more alike. When analysing the data of cluster-randomized trials, the dependency of the subjects within the same cluster is taken into account by applying a multilevel analysis. Moreover, multilevel analysis makes it possible to include covariates on the individual level as well as on the cluster level. When cluster-level covariates—that is, contextual variables—are taken into account in an ordinary regression or ANOVA model, this covariate has to be disaggregated. The consequence of disaggregation is that an artificial homogeneity is introduced, so that standard errors are biased downwards, with the consequence of an inflated Type I error. On the other hand, in contextual analyses, individual-level covariates have to be aggregated, often resulting in a loss of information and, hence, of power (see, e.g., Greenland, 2002). That aggregation may not always result in a loss of power is shown by, for instance, Hedges (2007). He presented an adjusted *t* test that accounted for clustering and obtained “reasonably accurate significance levels.” However, he recommended that this test and other tests on cluster means (i.e., Barcikowski’s test) should only be used when raw data are not available and argued that, in other cases, multilevel statistical methods are more appropriate. With multilevel analysis, the covariates are analysed without aggregation or disaggregation (e.g., Snijders & Bosker, 1999).

It is well known that by taking into account influential covariates, power can be increased (see, e.g., Moerbeek, 2006). In cluster-randomized trials, covariates on various

E. J. H. Korendijk (✉) · J. J. Hox · M. Moerbeek · C. J. M. Maas
Department of Methodology and Statistics, Faculty of Social and Behavioral Sciences, Utrecht University,
Heidelberglaan 2, De Uithof Utrecht University, P.O. Box 80140,
3508 TC Utrecht, The Netherlands
e-mail: e.j.h.korendijk@uu.nl

levels may need to be included in the study. In the present study, we focused on including a covariate on the lowest level. Since in most cluster-randomized trials the lowest-level units concern individual subjects, from here on we will refer to this level as the *subject level*.

When including a subject-level covariate, it seems obvious that one wants to assess the influence of this covariate at the level on which it is measured—that is, the subject level. For example, when taking into account pupils' IQs on math performance, the researcher expects the math performance to be influenced by the subject's IQ score and, hence, in the multilevel model will treat IQ score as a subject-level covariate. The effect of a subject-level covariate at the subject level is referred to as the *within-cluster* effect. However, in addition to the within-cluster effect, it is not uncommon that a subject-level covariate has an effect at the cluster level as well. For example, in addition to the effect of the pupil's individual IQ score on math performance, the mean IQ score in a class may also influence the individual math performance. The effect of a subject-level covariate at the cluster level is in the literature referred to as a *between-cluster* or *contextual* effect. Although in the remainder of this article we will make a distinction between these phrases, throughout this introduction we will simply use the phrase *contextual effect*.

The within-cluster effect and the contextual effect of a subject-level covariate may differ, and they may even have different signs (see, e.g., Kreft & de Leeuw, 1988; Snijders & Bosker, 1999). This phenomenon is also discussed by Begg and Parides (2003), Greenland (2002), Neuhaus and Kalbfleisch (1998), and Palta and Seplaki (2002). These authors have emphasized that when a multilevel analysis is applied, the researcher should check whether the within-cluster and contextual effects differ, and if they do, both effects should be modelled explicitly. If the separate effects are not explicitly modelled, the within-cluster and contextual effects are implicitly assumed to be equal.

However, most substantive researchers seem unaware of the possibly different within-cluster and contextual effects of a first-level covariate and do not explicitly model both effects, implicitly assuming them to be equal. That the assumption of equal within-cluster and contextual effects may be violated has been shown by, for instance, Mann, De Stravol, and Leon (2004) and Dwyer and Blizzard (2005).

The studies cited above have shown not only that the within-cluster effect may differ from the contextual effect of a subject-level covariate, but also that different analysis models—that is, assuming equal effects or modelling both effects explicitly in different ways—can affect the estimation of the parameters of the covariate itself. In an extended study, Shen, Shao, Park, and Palta (2008) discussed the effect of misspecifying an inequality of the covariate

effects, illustrating the effect on real data. However, they used data from observational studies and did not discuss the main parameter in cluster-randomized trials—that is, the treatment effect and its standard error. The present study fills this gap, by focusing on the robustness of the treatment parameter and its standard error against ignoring inequality of the within-cluster and contextual effects of the covariate within the framework of a cluster-randomized trial.

In the present study, we apply two different models to simulated data—one of them explicitly modelling within-cluster and contextual effects of the subject-level covariate, and the other implicitly assuming equal effects. Note that the latter model also can be seen as omitting a second-level covariate, the contextual effect. In a randomized design, it is assumed that all covariates—the observed as well as the omitted ones—are balanced over the conditions, and hence that the expected correlation between the omitted covariate and the treatment effect is zero. However, when the covariate is not balanced, the variability may actually increase, which will be visible in larger standard error estimates.

Simulation has the advantage of known true parameter values, as well as presenting the possibility of manipulating the total effect of the covariate and the magnitude of the difference between the within-cluster and contextual effects. Though the main interest of this study is the robustness of the estimated treatment effect and its standard errors, for completeness we will also present the other fixed parameters (i.e., the within-cluster effect, the contextual effect—if estimated—and the constant), the estimated variance components, and their standard errors.

In the next section, two models are described. The first takes into account the different within-cluster and contextual effects. From here on, we will refer to this model as the *covariate different-effects multilevel model*. This model, though not the label, is proposed by, for instance, Neuhaus and Kalbfleisch (1998) and Snijders and Bosker (1999). The second model ignores the different within-cluster and contextual effects of the covariate, hence implicitly assuming equal effects. Since the latter model is the one usually used for nested data, from here on we will refer to it as the *ordinary multilevel model*. In the third section, the simulation and an analysis of the parameter estimates are described. The results are given in the fourth section, and the article ends with a summary and discussion.

The models

The covariate different-effects multilevel model

Assume that we have data from a cluster-randomized trial with a continuous outcome, one variable at the cluster level—

that is, the treatment indicator—and one continuous subject-level covariate. As stated in the introduction, in data from a cluster-randomized trial and in other hierarchical data, a first-level covariate may have different within-cluster and contextual effects. This is taken into account by explicitly modelling two parameters associated with the subject-level covariate:

$$Y_{ij} = \gamma_{\text{const}} + \gamma_{\text{treat}}T_j + \gamma_W X_{ij} + \gamma_B \bar{X}_j + u_j + e_{ij}. \quad (1)$$

In this equation, Y_{ij} is a continuous outcome for subject i in cluster j , and γ_{const} is a constant reflecting the mean outcome when T_j and X_{ij} both equal zero. T_j is the treatment indicator, coded 1 for the treatment condition and 0 for the control condition. Coded as such, γ_{treat} represents the treatment effect, which is assumed to be the same for all clusters in the treatment condition. The subject-level covariate X is separated into a within-cluster part and a contextual part. The within-cluster part is given by X_{ij} , which is the score on covariate X_{ij} for subject i in cluster j , and by the regression of Y_{ij} on X_{ij} (i.e., the within-cluster effect, given by γ_W). The contextual part is given by the cluster mean \bar{X}_j and the regression of Y_{ij} on \bar{X}_j (i.e., the contextual effect γ_B). Finally, u_j is the cluster-level residual, and e_{ij} is the subject-level residual. These residuals are independently distributed with zero mean and the variances σ_u^2 , for the cluster-level residuals, and σ_e^2 , for the subject-level residuals. Since these variances are independent, the total residual variance is $\sigma^2 = \sigma_u^2 + \sigma_e^2$.

It should be noted that the separation of the within-cluster and contextual effects of a first-level covariate can also be accomplished by using $X_{ij} - \bar{X}_j$ instead of X_{ij} , so that Eq. 1 changes into

$$Y_{ij} = \tilde{\gamma}_{\text{const}} + \tilde{\gamma}_{\text{treat}}T_j + \tilde{\gamma}_W(X_{ij} - \bar{X}_j) + \tilde{\gamma}_B \bar{X}_j + u_j + e_{ij}. \quad (2)$$

In Eq. 2, the symbol $\tilde{\gamma}$ is used to distinguish these parameters from the corresponding parameters in Eq. 1. In Eq. 2, $\tilde{\gamma}_W$ is the regression of Y_{ij} on $(X_{ij} - \bar{X}_j)$ —that is, the within-cluster effect. This parameter is equal to the corresponding parameter in Model (1): $\tilde{\gamma}_W = \gamma_W$. However, the contextual effect $\tilde{\gamma}_B$ in Model (2) is the regression of \bar{Y}_j on \bar{X}_j and, hence, reflects the actual contextual effect. For clarity, from here on, we will refer to $\tilde{\gamma}_B$ as the *contextual effect* and γ_B [i.e., the parameter used in Model (1)] as the *between-cluster effect*. The relation between $\tilde{\gamma}_B$ and γ_B is given by $\tilde{\gamma}_B = \gamma_W + \gamma_B$ (see, e.g., Snijders & Bosker, 1999). From this, it becomes obvious that when the between-cluster effect $\gamma_B = 0$, the contextual effect $\tilde{\gamma}_B$ is equal to the within-cluster effect γ_W . Though Models (1) and (2) are equivalent, we choose to use Model (1) in the present study, because in this model both parameters—that is, the between-cluster parameter γ_B and the within-cluster parameter γ_W —reflect the regression of Y_{ij} .

The true total effect of the subject-level covariate is a weighted sum of the within-cluster and contextual effects (see, e.g., Snijders & Bosker, 1999, p. 30):

$$\gamma_{\text{total}} = \eta^2 * \tilde{\gamma}_B + (1 - \eta^2) * \tilde{\gamma}_W \quad (3)$$

In this equation, η^2 is the correlation ratio, which is defined by the intraclass coefficient and the reliability of the cluster mean \bar{X}_j . It should be noted that the reliability of the cluster mean depends on both the magnitude of the intraclass correlation coefficient (ICC) and the group size (e.g., Bliese, 2000). That is, the larger the ICC, the more likely is a single score to be a reliable estimate of the cluster mean. And by means of the law of large numbers, it is obvious that the larger the cluster size, the more likely it is that the cluster mean is a reliable estimate for the population cluster mean. Expressed in terms of Model (1), Eq. 3 changes into

$$\gamma_{\text{total}} = \eta^2 * (\gamma_B + \gamma_W) + (1 - \eta^2) * \gamma_W. \quad (4)$$

The ordinary multilevel model

In the ordinary multilevel model, the possibility of different within-cluster and contextual effects of the covariate is ignored. In other words, it is assumed that the within-cluster and contextual effects are equal, or in terms of Model (1), that the between-cluster effect is zero. If this assumption holds, it is sufficient to estimate a single parameter for the covariate X :

$$Y_{ij} = \gamma'_{\text{const}} + \gamma'_{\text{treat}}T_j + \gamma'_W X_{ij} + u_j + e_{ij}. \quad (5)$$

In this equation, the symbol γ' is used, to distinguish the parameters from the corresponding parameters used in Eqs. 1 and 2. Let Y_{ij} be a continuous outcome for subject i in cluster j and γ'_{const} be a constant, reflecting the mean outcome when T_j and X_{ij} both equal zero. T_j is the treatment indicator for cluster j (coded 0 for the control condition and 1 for the experimental condition), and γ'_{treat} is the treatment effect. γ'_W is the within-cluster effect of the first-level covariate X . It should be noted that this parameter usually is not subscripted $_W$, since it is the only estimated parameter for the subject-level covariate. However, in the context of the present study, the subscript is used to emphasize that this parameter is estimated under the assumption that γ'_B is zero. As stated in the introduction section, in the ordinary multilevel model this parameter can also be viewed as an omitted covariate.

In Model (5), as in Model (1), the residual at the cluster level is captured by the term u_j , while the residual at the subject level is captured by e_{ij} . The residuals are

assumed to be independent of each other and normally distributed, with zero mean and variances σ_u^2 and σ_e^2 , respectively. The total variance of the model is given by $\sigma^2 = \sigma_u^2 + \sigma_e^2$.

Simulation and evaluation of the estimated parameters

Simulation

Data were generated and analysed with the MLwiN software, version 2.0 (Rasbash, Steele, Browne, & Prosser, 2004). The data were generated according to Model (1)—that is, the covariate different-effects multilevel model (CDEMM). Data were then analysed with the covariate different-effects multilevel model (Eq. 1) and with the ordinary multilevel model (OMM; Eq. 5). Restricted maximum likelihood estimation (REML) was used, as is advised for small sample sizes at the cluster level in most literature, in order to get good estimates for the subject-level variance σ_e^2 (see, e.g., Hox, 2010; Raudenbush, Bryk, Cheong, Congdon, & Du Toit 2004; Snijders & Bosker, 1999).

In all conditions, the cluster size m was fixed at five. This is not an unusual number in, for example, group therapy. Furthermore, when all other requirements are met, a small sample size at the first level hardly affects the estimation of the parameters and standard errors. In order to get unbiased parameter estimates, the second-level sample size is of main importance (Maas & Hox, 2005). In the present study, we chose to have a relatively small second-level sample size of 20 clusters per condition. Such small numbers of clusters are quite common in practice, due to limited financial resources and other practical limitations. Furthermore, it was expected that a large number of clusters would not result in a more accurate estimation of the treatment effect, if it is affected by ignoring the nonequivalence of the within-cluster and contextual effects.

The intercept was fixed at 1, and the treatment effect was fixed at 0.3, which is a medium-sized effect (Cohen, 1988). The subject-level residual variance σ_e^2 was fixed at 1. The value of the cluster-level variance σ_u^2 follows from the ICC and the subject-level variance σ_e^2 . Although the ICC affects the sampling variance, it does not affect the estimation of fixed effects. Multilevel software incorporates this effect of the ICC, so the magnitude of the ICC does not affect the bias of estimators. In earlier simulation studies (e.g., Korendijk, Maas, Moerbeek, & Van der Heijden, 2008; Maas & Hox, 2005), the ICC indeed was an insignificant predictor of the biases of estimated parameters and their standard errors. For these reasons, in the present study the ICC was not manipulated but fixed at .10, which is a value that is often encountered in practice.

It was expected that the more the contextual effect differed from the within-cluster effect, the more biased the parameter estimate would be. Since substantive researchers rarely use the CDEMM, little information is available on empirical values for the magnitude of the difference between the within-cluster and contextual effects. Based on the studies by Mann, De Stravol, and Leon (2004), Neuhaus and Kalbfleisch (1998), and Palta and Seplaki (2002), we chose the within-cluster effect to be ten times as small, three times as small, three times as large, and ten times as large as the contextual effect. These different conditions in the tables are labelled in terms of the magnitude of the within-cluster effect as compared to the contextual effect; for instance, 0.1 refers to the condition in which the within-cluster effect is ten times as small as the contextual effect, and 10 refers to the condition in which the within-cluster effect is ten times as large as the contextual effect. From here on we will refer to this variable as the *inequality of covariate effects*, with the values 0.1, 0.333, 3, and 10. Finally, in order to compare the performance of the CDEMM and the OMM when the within-cluster effect is equal to the contextual effect, data with equal effects were generated, thus adding a condition labelled “1” to the inequality-of-covariate-effects variable. In this case, it was expected that both models would perform equally well, since both models fit the data when the covariate effect is the same at both levels.

Furthermore, it was expected that the stronger the total effect of the first-level covariate, the more the treatment effect and variance estimates would be affected. The total effect of the first-level covariate—that is, the weighted sum of the within-cluster and contextual effects (Eq. 3)—was chosen to be small (.1), medium (.3) and large (.5) (Cohen, 1988). From here on, we will use the term *total covariate effect* when referring to the total effect of the first-level covariate.

The above-described design consists of ($5 \times 3 =$) 15 conditions, and in each condition 3,000 data sets were generated. The large number of replications was chosen to increase the precision of the estimates (Skrondal, 2000).

Evaluation of the parameter and standard error estimates

Estimated parameters were evaluated by means of the relative bias—that is, the estimated parameter divided by the true parameter value: $RB = \frac{\hat{\theta}}{\theta}$. When estimated without bias, the parameter estimate equals the true parameter, and hence the RB will be 1. When the parameter is overestimated, the RB will be larger than 1, and underestimation will result in an RB smaller than 1. However, when the within-cluster and the contextual effects are equal, the true between-cluster effect is zero: $\gamma_B = 0$. In this situation, the bias of the estimated between-cluster effect is determined

by the deviation of the true value, which is the estimate itself ($\widehat{\gamma}_B - \gamma_B = \widehat{\gamma}_B - 0 = \widehat{\gamma}_B$). As Skrandal (2000) recommends, estimated parameters will be tested. The parameter of main interest for a researcher conducting a trial is the treatment effect. Hence, the focus of this study is on the estimated treatment effect and its standard error. Investigation of estimated covariate effects is more appropriately done by means of an observational study. For this reason, the results with respect to the estimated within- and between-cluster effects will not be elaborated on, although for completeness they will be presented in the tables, giving an overview of the simulation results in terms of relative bias. In these tables, biased parameters are indicated; however, the estimated covariate effects will not be discussed or further tested. For a discussion of the effects of misspecification on the estimated within-cluster and contextual effects, we refer the reader to the study of Shen, Shao, Park, and Palta (2008).

A number of t tests were performed to test whether the relative bias equalled one. When two or more parameter estimates within a variable (i.e., the ratio or the covariate effect) were biased, two-sample t tests or ANOVAs were performed to test whether the biases differed between conditions. Significant ANOVA results were evaluated by post hoc t tests.

Since unbalance of an omitted covariate, which in the present study is the between-cluster effect of the covariate γ_B in the ordinary multilevel model, can result in increased standard errors of the treatment effect, the distributions of the estimated standard errors of this parameter were inspected. The standard errors are expected to increase when the OMM is applied. The distribution of the estimated standard errors of the CDEMM (i.e., the appropriate model when within-cluster and contextual effects differ) was used to determine a cutoff point. This cutoff was the estimated value at the 97.5th percentile of the distribution. The estimated standard errors of the OMM exceeding this cutoff point were considered to be inflated, and percentages of inflated standard errors were determined per condition. The estimated standard errors of all parameters were evaluated by means of the coverage. The 95% confidence intervals around the estimated parameters were established, and it was determined whether or not the true parameter was in this interval, and the estimated parameter was coded 1 or 0, respectively. The mean of this parameter is the *coverage*, and ideally this proportion should equal .95 ($1 - \alpha$). Underestimation of the standard errors is reflected by coverages below .95, and overestimation by coverages over .95. Coverages were evaluated by establishing a 99.99% ($\alpha = .001$) confidence interval around .95 (see, e.g., Newcombe, 1998), corrected for the number of parameters tested simultaneously, and determining whether the coverage was within or outside this interval. The small α was

chosen because of the huge number of simulated data sets. When two or more coverages of a parameter were significant within inequality of covariate effects or the total covariate effect, χ^2 tests were performed to test whether the coverages differed between the conditions. Post hoc Fisher exact tests were performed when a χ^2 test turned out to be significant. In accordance with the treatment of the parameter estimates, no tests were conducted on biased standard error estimates associated with the covariate parameters.

When the influence of the magnitude of the covariate effect was evaluated, this was done separately for data in which the within-cluster and contextual effects were different and data in which these effects were equal. It is obvious that in the first situation the CDEMM is the appropriate model and, hence, was expected to outperform the OMM, while in the latter situation both models were expected to perform equally well.

Results

This section is divided into three subsections. In the first subsection, issues concerning convergence and inadmissible solutions of the estimation process are described. In the second subsection, the results with respect to the parameter estimates are given. As stated before, our main interest is the estimate of the treatment effect. However, we present and discuss the estimates of the constant and variance components as well. In the third subsection, the standard error estimates are presented. With respect to the presenta-

Table 1 Percentages of negative variance and standard deviations per model, by inequality of the covariate effects and by total covariate effect

	CDEMM		OMM	
	%	<i>SD</i>	%	<i>SD</i>
ICE				
0.1	4.756	0.213	1.789	0.133
0.333	4.689	0.211	2.811	0.165
1	5.178	0.222	4.489	0.207
3	5.256	0.223	3.178	0.175
10	4.744	0.213	2.811	0.165
TCE				
.1	4.889	0.216	4.400	0.205
.3	4.993	0.218	2.993	0.170
.5	4.893	0.216	1.653	0.128

CDEMM, covariate different-effects model; OMM, ordinary multilevel model; ICE, inequality of covariate effects; TCE, total covariate effect

Table 2 Pairwise comparisons of percentages of negative second-level variance for the ordinary multilevel model, by inequality of the covariate effects (top panel) and total covariate effect (bottom panel)

	<i>t</i> Value	<i>df</i>	<i>p</i> Value
ICE			
1 vs. 0.333	6.007*	17,155.773	.0000
1 vs. 3	4.583*	17,524.423	.0000
1 vs. 0.1	10.418*	15,313.840	.0000
1 vs. 10	6.007*	17,155.773	.0000
0.1 vs. 0.333	− 4.577*	17,186.871	.0000
3 vs. 10	1.443	17,934.844	.149
0.1 vs. 3	− 5.993*	16,749.115	.0000
0.333 vs. 3	− 1.443	17,934.844	.149
0.1 vs. 10	− 4.577*	17,186.871	.0000
0.333 vs. 10	Not tested, since equal percentages		
TCE			
.1 vs. .3	6.461**	29,023.762	.0000
.3 vs. .5	7.711**	27,787.050	.0000
.1 vs. .5	13.929**	25,087.387	.0000

ICE, inequality of covariate effects; TCE, total covariate effect. * Significant at the .0001 level. ** Significant at the .00033 level

tion and discussion of the standard errors, we will follow the same line as with respect to the estimated parameters.

Convergence and inadmissible solutions

Convergence was reached in all conditions. In MLwiN (Rasbash et al., 2004), it is possible that the estimation procedure will result in negative variance estimates, especially when the true value is close to zero, as is the case in the present study. In practice, such negative variance estimates are usually set to zero. However, by doing so in a

simulation study, bias would be introduced. Therefore, the negative values for the second-level variance were retained. Table 1 shows the percentages of negative second-level variance estimates found per model by condition. When the data are analysed with the covariate different-effects multilevel model, the percentage of negative second-level variance estimates is approximately 5% in all conditions [$F(4, 44995) = 1.395, p = .233$, for inequality of the covariate effects and $F(2, 44997) = 0.114, p = .892$, for the total covariate effect]. A smaller percentage is found when the data are analysed with the ordinary multilevel model, and it varies by condition [$F(4, 44995) = 29.294, p < .001$, for inequality of the covariate effects, and $F(2, 44997) = 27.193, p < .001$, for the total covariate effect; for pairwise comparisons, see Table 2]. The percentage of negative variance is largest when the within-cluster and contextual effects are equal, and it decreases when the inequality becomes more extreme. When the total effect of the covariate increases, the percentage of negative second-level variance estimates decreases (Table 2).

Parameter estimates

Tables 3 and 4 show all estimated parameters for various magnitudes of the disparity between the covariate effects and of the total covariate effect, respectively. All fixed parameters appear to be estimated without bias when the CDEMM is applied, irrespective of the magnitudes of the inequality and the total covariate effect. When the OMM is applied, the treatment effect γ'_{treat} and the constant γ'_{const} are estimated without bias in all conditions as well.

The random parameters in the CDEMM—that is, the residual cluster-level variance σ_u^2 and the residual subject-level variance σ_e^2 —are estimated without bias. The random effects in the OMM, $\sigma_u'^2$ and $\sigma_e'^2$, are estimated without bias

Table 3 Relative biases (and standard deviations) of the estimated parameters per model, by inequality of the covariate effects

		0.1		0.333		1		3		10	
		RB	SD	RB	SD	RB	SD	RB	SD	RB	SD
CDEMM	γ_{const}	0.9992	0.091	1.0001	0.091	0.9999	0.090	0.9997	0.092	1.0000	0.091
	γ_{treat}	1.0046	0.599	0.9980	0.599	1.0026	0.591	1.0020	0.604	0.9997	0.600
	γ_W	0.9781	1.943	0.9925	0.812	0.9842	0.498	0.9912	0.386	0.9999	0.347
	γ_B	1.0016	0.608	1.0049	1.159	0.0054 ^A	0.224	0.9837	1.580	1.0039	1.061
	σ_u^2	0.9962	0.641	1.0022	0.643	1.0002	0.647	0.9929	0.643	0.9922	0.640
	σ_e^2	1.0000	0.050	0.9995	0.057	1.0000	0.050	0.9999	0.050	1.0004	0.050
OMM	γ'_{const}	0.9989	0.113	1.0020	0.095	1.0003	0.089	0.9987	0.092	1.0013	0.096
	γ'_{treat}	1.0037	0.675	0.9877	0.611	0.9986	0.592	1.0091	0.599	0.9916	0.611
	γ'_W	2.0148*	1.894	1.2440*	0.771	0.9969	0.468	0.9104*	0.358	0.8910*	0.326
	$\sigma_u'^2$	1.8673*	1.131	1.2271*	0.728	0.9996	0.624	1.1220*	0.671	1.2632*	0.743
	$\sigma_e'^2$	1.0045*	0.051	1.0031*	0.051	0.9999	0.050	1.0012	0.050	1.0028*	0.050

CDEMM, covariate different-effects model; OMM, ordinary multilevel model. ^A Bias is deviation from zero. * Significant at the $p < .001$ level

Table 4 Relative biases of the estimated parameters (and standard deviations) per model, by total covariate effect

		Equal Covariate Effects Excluded						Equal covariate effects only					
		.1		.3		.5		.1		.3		.5	
		RB	SD	RB	SD	RB	SD	RB	SD	RB	SD	RB	SD
CDEMM	γ_{const}	0.9998	0.092	0.9989	0.091	1.0005	0.092	0.9972	0.090	1.0015	0.090	1.0010	0.091
	γ_{treat}	1.0002	0.602	1.0077	0.597	0.9954	0.603	1.0192	0.587	0.9935	0.590	0.9950	0.596
	γ_W	0.9773	1.750	0.9958	0.581	0.9982	0.358	0.9532	0.805	0.9998	0.265	1.0000	0.162
	γ_B	0.9933	1.861	0.9997	0.629	1.0027	0.377	0.0074 ^A	0.223	0.0033 ^A	0.222	0.0055 ^A	0.226
	σ_u^2	0.9976	0.645	0.9979	0.640	0.9921	0.640	0.9792	0.637	1.0105	0.659	1.0108	0.645
	σ_e^2	1.0004	0.050	0.9997	0.052	0.9998	0.057	1.0015	0.050	0.9992	0.051	0.9992	0.050
OMM	γ'_{const}	1.0004	0.089	0.9997	0.097	1.0006	0.110	0.9991	0.090	0.9995	0.089	1.0021	0.088
	γ'_{treat}	0.9979	0.590	1.0005	0.618	0.9958	0.664	1.0059	0.600	1.0034	0.592	0.9863	0.585
	γ'_W	1.3292	0.018	1.2363	0.728	1.2025	0.506	0.9957	0.758	0.9947	0.244	1.0003	0.150
	σ'^2_u	1.0371*	0.646	1.2702*	0.739	1.8024*	1.070	1.0096	0.626	1.0001	0.626	0.9890	0.621
	σ'^2_e	1.0004	0.050	1.0028*	0.050	1.0055*	0.052	0.9993	0.050	0.9996	0.050	1.0009	0.050

CDEMM, covariate different-effects model; OMM, ordinary multilevel model. ^A Bias is deviation from zero. * Significant at the $p < .001$ level

when the within-cluster effect equals the contextual effect. This is as expected, since in this condition the OMM is not a misspecification of the data. However, when the within-cluster and contextual effects differ, the OMM misspecifies the data, and in this situation both variances are biased. The second-level variance $\sigma_u'^2$ is overestimated in all conditions, and this overestimation is affected by the magnitudes of the inequality of the covariate effect and the total covariate effect [$F(3, 35996) = 1,419.396, p < .0001$, and $F(2, 35997) = 2,628.895, p < .0001$]. Table 5 shows that the overestimation is more severe when the magnitude of the inequality increases [$t(15123.858) = 44.321, p < .00017$, when the within-cluster is smaller than the contextual effect, and $t(11901.675) = -13.382, p < .00017$, when the within-cluster is larger than the contextual effect]. Comparisons between the conditions with, on the one hand, a small within-cluster effect to a larger contextual effect and, on the other, the

reverse situation are all significant, as well (see Table 5). Comparisons between the biases for the different conditions of the total covariate effect reveal that the overestimation of the second-level residual variance $\sigma_u'^2$ increases when the covariate effect increases; that is, the overestimation is smallest when the covariance effect is small, and it is largest when the covariate effect is large [$t(29562.393) = -23.338, p < .00033$, and $t(26691.307) = -40.695, p < .00033$, respectively; see Table 5].

As we have said, the subject-level residual variance $\sigma_e'^2$ is also biased in conditions with unequal covariate effects. It is overestimated when the within-cluster effect is smaller than the contextual effect and when the within-cluster effect is ten times as large (Table 3). The severity of the overestimation is the same in those three conditions [$F(2, 26997) = 3.021, p = .049$]. Furthermore, the subject-level variance $\sigma_e'^2$ is overestimated when the total effect of the subject-level

Table 5 Primary and post hoc t tests on the biased parameter estimates, by inequality of the covariate effects (top panel) and by total covariate effect (bottom panel)

	Parameter	Comparison	Peculiarity	t Value	df	p Value	
ICE	$\sigma_u'^2$	0.1 vs. 0.333	Post hoc	44.321**	15,123.858	.00000	
		0.1 vs. 3	Post hoc	52.731**	14,401.113	.00000	
		0.1 vs. 10	Post hoc	41.571**	15,312.580	.00000	
		0.333 vs. 3	Post hoc	10.074**	17,878.929	.00000	
		0.333 vs. 10	Post hoc	-3.293	17,998	.00093	
		3 vs. 10	Post hoc	-13.382**	11,901.675	.00000	
TCE	$\sigma_u'^2$.1 vs. .3	Post hoc	-23.338***	29,562.292	.00000	
		.3 vs. .5	Post hoc	-40.695***	26,691.307	.00000	
		.1 vs. .5	Post hoc	-60.600***	24,871.962	.00000	
		$\sigma_e'^2$.3 vs. .5	Primary test	-4.054****	29,998	.00005

ICE, inequality of covariate effects; TCE, total covariate effect. * Significant at the $p < .0005$ level. ** Significant at the $p < .00017$ level. *** Significant at the $p < .00033$ level. **** Significant at the $p < .001$ level

Table 6 Coverages evaluating the estimated standard errors per model, by inequality of the covariate effects

ICE		0.1	0.333	1	3	10
CDEMM	γ_{const}	.9902*	.9892*	.9910*	.9910*	.9918*
	γ_{treat}	.9412	.9432	.9474	.9344*	.9427
	γ_W	.9512	.9539	.9492	.9480	.9494
	γ_B	.9454	.9444	.9436	.9482	.9468
	σ_u^2	.9386	.9390	.9356*	.9364*	.9387
	σ_e^2	.9991*	.9996*	.9991*	.9989*	.9994*
OMM	γ'_{const}	.9858*	.9896*	.9930*	.9904*	.9872*
	γ'_{treat}	.9451	.9449	.9439	.9443	.9448
	γ'_W	.8290*	.8902*	.9480	.9127*	.8858*
	$\sigma_u'^2$.8453*	.9529	.9439	.9559	.9507
	$\sigma_e'^2$.9997*	.9991*	.9992*	.9996*	.9991*

ICE, inequality of covariate effects; CDEMM, covariate different-effects model; OMM, ordinary multilevel model. * Coverage outside the interval $< .9369: .9631 >$ for two estimated parameters. ** Coverage outside the interval $< .9365: .9635 >$ for three estimated parameters. *** Coverage outside the interval $< .9361: .9639 >$ for four estimated parameters

covariate is medium or large (Table 4), and the overestimation is more severe when the total effect is large [$t(11998) = -3.326, p < .0001$; Table 5].

To sum up, using the CDEMM results in an unbiasedly estimated constant and treatment effect as well as unbiasedly estimated random parameters, whether the within-cluster effect differs from the contextual effect or not. When the OMM is applied to data with unequal covariate effects, the residual variance on both levels is biased. However,

although the model is misspecified, the parameter of main interest, the treatment effect, is unbiased in all conditions.

Standard error estimates

Since the treatment effect is the parameter of main interest in trials, we start with inspection of the distribution of the standard errors of this estimated parameter. Graphs of the standard error (not presented here) showed nearly perfect normal distributions, regardless of the magnitude of the inequality, the magnitude of the total covariate effect, or the model. Furthermore, we found consistent means (0.18), standard deviations (0.019), and hence consistent values for the 97.5th percentiles (0.216), over the conditions for the CDEMM. Using the 97.5th percentile value as a cutoff score, the percentages that exceed this value in the distributions of the estimated standard errors for the OMM were determined. When the within-cluster effect was ten times as small as the contextual effect, we found 27.9% of the estimated standard errors to be larger than the cutoff score. In the other conditions of unequal covariate effects, the percentages were minor (5.7% when the within-cluster effect was three times as small, 3.1% when it was three times as large, and 6.3% when it was ten times as large as the contextual effect). In other words, only when the within-cluster effect is very small compared to the contextual effect are the standard errors associated with the treatment effect seriously inflated.

Although a comparison of the distributions of the standard errors revealed in one of the conditions for the OMM a considerable inflation of the standard errors, inspection of the coverages gave no reason to be concerned

Table 7 Coverages evaluating the estimated standard errors by total covariate effect, by unequal or equal within-cluster and contextual effects

		Unequal Within-Cluster and Contextual Covariate Effects			Equal Within-Cluster and Contextual Covariate Effects		
	Parameter	.1	.3	.5	.1	.3	.5
CDEMM	γ_{const}	.9896*	.9907*	.9914*	.9900*	.9907*	.9923*
	γ_{treat}	.9397	.9434	.9381	.9470	.9490	.9463
	γ_W	.9503	.9533	.9484	.9500	.9513	.9463
	γ_B	.9494	.9430	.9463	.9443	.9470	.9393
	σ_u^2	.9385	.9386	.9374	.9313*	.9367*	.9387
	σ_e^2	.9992*	.9994*	.9992*	.9993*	.9993*	.9987*
OMM	γ'_{const}	.9915*	.9888*	.9844*	.9930*	.9930*	.9930*
	γ'_{treat}	.9444	.9455	.9444	.9423	.9427	.9467
	γ'_W	.9368	.8832*	.8183*	.9437	.9537	.9467
	$\sigma_u'^2$.9434	.9547	.8805*	.9400	.9433	.9483
	$\sigma_e'^2$.9994*	.9993*	.9994*	.9997*	.9993*	.9987*

CDEMM, covariate different-effects model; OMM, ordinary multilevel model. * Coverage outside the interval $< .9369: .9631 >$ for two estimated parameters. ** Coverage outside the interval $< .9365: .9635 >$ for three estimated parameters. *** Coverage outside the interval $< .9361: .9639 >$ for four estimated parameters

Table 8 Results of χ^2 tests on significant coverages, by inequality of the covariate effects (top panel) and by total covariate effect (bottom panel)

Model	Parameter	Peculiarity	χ^2	df	p Value
ICE					
CDEMM	γ_{const}		3.630	4	.458
	σ_e^2		3.431	4	.488
OMM	γ'_{const}		26.737*	4	.00002
	$\sigma_e'^2$		3.669	4	.453
TCE					
CDEMM	γ_{const}	Different covariate effects	2.179	2	.336
		Equal covariate effects	0.972	2	.615
	σ_e^2	Different covariate effects	0.667	2	.716
		Equal covariate effects	1.001	2	.606
OMM	γ'_{const}	Different covariate effects	26.453*	2	.00000
	$\sigma_e'^2$	Different covariate effects	0.348	2	.840
		Equal covariate effects	2.803	2	.368

ICE, inequality of covariate effects; TCE, total covariate effect; CDEMM, covariate different-effects model; OMM, ordinary multilevel model.

* Significant at the $p < .0001$ level

about inflated Type II errors with respect to the treatment effect. That is, the coverages show that, when the OMM is applied, the standard errors of the treatment effect γ'_{treat} are estimated without bias (lower panel of Table 6 and 7). The standard errors associated with the constant γ'_{const} however, are biased. In the OMM, they are overestimated in all conditions, even when the within-cluster and contextual effects are equal. The estimate is affected by the inequality of the covariate effects [$\chi^2(4) = 30.368, p < .0001$; Table 8]. Post hoc Fisher's exact tests reveal that the overestimation only differs when the condition with equal within-cluster and contextual effects is compared to the condition in which the within-cluster effect is ten times as large as the contextual effect (Fisher exact $p < .0001$; Table 9). When the within-cluster and contextual effects differ, the overestimation is also affected by the total covariate effect [$\chi^2(2) = 26.453, p < .0001$; Table 8]. The overestimation differs between the small and large total covariate effects (Fisher exact $p < .00033$; Table 9). The standard error of the cluster-level variance σ_u^2 in the CDEMM is slightly underestimated when the within-cluster effect is three times as large as the contextual effect, and when the effects are equal (Table 6). The underestimation is the same in both conditions (Fisher exact $p = .171$; Table 9). The standard error of the cluster-level variance σ_u^2 is also underestimated when the covariate effect is small or medium and the within-cluster and contextual effects of the covariate are equal (right-hand panel of Table 7). The underestimation is the same for both conditions (Fisher exact $p = .406$; Table 9). The standard error of the subject-level variance σ_e^2 , however, is overestimated in all conditions (Tables 6 and 7). The overestimation is not affected by the inequality of the covariate effects [$\chi^2(4) = 3.431, p = .488$] or by the total covariate effect [$\chi^2(2) = 0.667, p = .716$, and $\chi^2(2) = 1.001, p = .606$, for different and equal covariate effects, respectively; Table 8].

In the OMM, standard errors associated with the residual variances are often biased as well. The standard errors of the cluster-level variance σ_u^2 are underestimated in two situations: when the within-cluster effect is ten times as small as the contextual effect (Table 6) and when the total covariate effect is small and the within-cluster and contextual effects differ (Table 7). In all other situations, the standard errors associated with the cluster-level variance σ_u^2 are accurately estimated. The standard errors estimates associated with the subject-level variance σ_e^2 are over-

Table 9 Primary and post hoc Fisher's exact tests on the significant coverages, by inequality of the covariate effects (top panel) and by total covariate effect (bottom panel)

Peculiarity	σ_u^2	γ'_{const}
ICE		
1 vs. 0.333		.013
1 vs. 3	.171 ^A	.059
1 vs. 0.1		.00000*
1 vs. 10		.00012
10 vs. 0.333		.026
3 vs. 10		.040
0.1 vs. 3		.005
0.1 vs. 10		.438
0.333 vs. 3		.550
0.333 vs. 10		.145
TCE		
.1 vs. .3	Equal covariate effects	.406 ^A
.1 vs. .3	Different covariate effects	.042
.3 vs. .5		.003
.1 vs. .5		.00000****

ICE, inequality of covariate effects; TCE, total covariate effect.
^A Primary Fisher's exact test. * Significant at the $p < .0001$ level.
 ** Significant at the $p < .00017$ level. *** Significant at the $p < .001$ level.
 **** Significant at the $p < .00033$ level

estimated in all conditions (Table 6 and 7). Table 8 shows that the overestimation does not depend on the inequality of the covariate effects [$\chi^2(4) = 3.669$, $p = .453$], nor on the magnitude of the total covariate effect [$\chi^2(2) = 0.348$, $p = .840$, when the within-cluster and contextual effects differ, and $\chi^2(2) = 2.803$, $p = .368$, when they are equal].

To sum up, the standard errors of the constant and of the first-level variance are biased in both models in all conditions. The standard errors of the cluster-level variance are biased in both models in some conditions. However, the standard errors of the treatment effect are unbiased, whether the different within- and between-cluster effects are ignored or not.

Summary and discussion

The parameter of main interest—that is, the treatment effect, γ_{treat} in the covariate different-effects multilevel model and γ'_{treat} in the ordinary multilevel model—and its standard errors are estimated without bias in all conditions by both models. If a researcher's only interest is in the treatment effect, it is sufficient to apply an ordinary multilevel model and ignore possible unequal within-cluster and contextual effects of a subject-level covariate. However, researchers who are interested in other parameters as well should take notice of the differences between the two models with respect to parameter and standard error estimates.

When the CDEMM is applied—that is, when the model takes into account the possibility of different within-cluster and contextual effects of a first-level covariate—all parameter estimates are unbiased. This does not hold when a model is applied that assumes the within-cluster effect to be equal to the contextual effect when this assumption is violated—that is, when the OMM is applied to data with unequal within-cluster and contextual effects. In this situation, the random parameters—that is, the cluster-level variance $\sigma_u'^2$ and the subject-level variance $\sigma_e'^2$ —are biased in almost all conditions. This may be tolerable, since the random parameters in cluster-randomized trials are usually viewed as nuisance parameters. However, in, for example, school effectiveness research, the random parameters are of particular interest. In these situations, biased random parameters are not tolerable.

With respect to the standard errors, we have shown that the standard errors of the constant γ_{const} and of the variance components σ_u^2 and σ_e^2 are biased in both the OMM and the CDEMM when the models are applied to data with different covariate effects. Again, these biases may be tolerable, since they are associated with either a nuisance parameter (i.e., the constant γ'_{const}) or with nonnormally distributed parameters (i.e., the residual variances σ_u^2 and σ_e^2). Even if variance parameters are of interest, it is not

advised to use the Wald test when evaluating variance components.

In general, it is unknown whether the effect of the first-level covariate differs from or equals the contextual effect. When the effects are equal, both models perform equally well, and when the effects differ, the CDEMM gives better estimates of the variance components. When a researcher is only interested in the treatment parameter, ignoring a possible difference between the within-cluster and contextual effects will do no harm. However, when a researcher is also interested in the variance components of his model, we advise that the covariate different-effects model be used.

Author Note To our deep sorrow, our colleague C.J.M. Maas passed away unexpectedly in February 2010. This study was funded by the Netherlands Organization for Scientific Research (Nederlandse Organisatie voor Wetenschappelijk Onderzoek; NWO). We thank the anonymous reviewer for his comments on an earlier version of the manuscript.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Begg, M. D., & Parides, M. K. (2003). Separation of individual-level and cluster-level covariate effects in regression analysis of correlated data. *Statistics in Medicine*, 22, 2591–2602. doi:10.1002/sim.1524
- Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349–381). San Francisco: Jossey-Bass.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Dwyer, T., & Blizzard, L. (2005). A discussion of some statistical methods for separating within-pair associations from associations among all twins in research on fetal origins of disease. *Pediatric and Perinatal Epidemiology*, 19, 48–53. doi:10.1111/j.1365.3016.2005.00615.x
- Greenland, S. (2002). A review of multilevel theory for ecologic analyses. *Statistics in Medicine*, 21, 389–395. doi:10.1002/sim.1024
- Hedges, L. V. (2007). Correcting a significance test for clustering. *Journal of Educational and Behavioral Science*, 32, 151–179. doi:10.3102/1076998606298040
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York: Routledge.
- Korendijk, E. J. H., Maas, C. J. M., Moerbeek, M., & Van der Heijden, P. G. M. (2008). The influence of misspecification of the heteroscedasticity on multilevel regression parameter and standard error estimates. *Methodology*, 4, 67–72. doi:10.1027/1614-2241.4.2.67
- Kreft, G. G., & de Leeuw, E. D. (1988). The see-saw effect: A multilevel problem? *Quality and Quantity*, 22, 127–137.
- Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modelling. *Methodology*, 1, 86–92. doi:10.1027/1614-1881.1.386

- Mann, V., De Stavol, B. L., & Leon, D. A. (2004). Separating within and between effects in family studies: An application to the study of blood pressure in children. *Statistics in Medicine*, 23, 2745–2756. doi:10.1002/sim.1853
- Moerbeek, M. (2006). Power and money in cluster randomized trials: When is it worth measuring a covariate? *Statistics in Medicine*, 25, 2607–2617. doi:10.1002/sim.2297
- Murray, D. M. (1998). *Design and analysis of group-randomized trials*. Oxford: Oxford University Press.
- Neuhaus, J. M., & Kalbfleisch, J. D. (1998). Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics*, 54, 638–645.
- Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: A comparison of seven methods. *Statistics in Medicine*, 17, 857–872.
- Palta, M., & Seplaki, C. (2002). Causes, problems and benefits of different between and within effects in the analysis of clustered data. *Health Services and Outcomes Research Methodology*, 3, 177–193.
- Rasbash, J., Steele, F., Browne, W., & Prosser, B. (2004). *A user's guide to MLwiN version 2.0*. London: University of London, Multilevel Models Project.
- Raudenbush, S. W., Bryk, S. A., Cheong, Y. F., Congdon, R. T., & Du Toit, M. (2004). *HLM 6: Hierarchical linear and nonlinear modeling [Software]*. Lincolnwood, IL: Scientific Software International.
- Shen, L., Shao, J., Park, S., & Palta, M. (2008). Between- and within-cluster covariate effects and model misspecification in the analysis of clustered data. *Statistica Sinica*, 18, 731–748.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modelling*. London: Sage.
- Skrondal, A. (2000). Design and analysis of Monte Carlo experiments: Attacking the conventional wisdom. *Multivariate Behavioral Research*, 35, 137–167.