# A Monte Carlo Simulation Comparing Parameter Estimates from Multiple Linear Regression and Hierarchical Linear Modeling

**Daniel J. Mundfrom** **Mark R. Schultz**
University of Northern Colorado

In this simulation study, the parameter estimates obtained from hierarchical linear modeling (HLM) and multiple linear regression (MLR) were examined for differences under different values of the intraclass correlation. 15,000 data sets were generated for each of ten different ranges of intraclass correlations. The resulting vectors of parameter estimates from both HLM and MLR were subtracted, averaged across 50 data sets and compared to a null vector of zeros using Hotelling's $T^2$ statistic. Little difference was found between the vectors of parameter estimates in any of the intraclass correlation ranges.

Multiple Linear Regression (MLR) and Hierarchical Linear Modeling (HLM) are two statistical procedures that can be used to model the relationship between a numerical dependent (i.e., response) variable and two or more numerical independent (i.e., predictor) variables. For an algebraic description and comparison of the HLM and MLR models see Mundfrom & Schultz (2001). Although similar in many ways, these two procedures are not identical in how they analyze the data and consequently may not produce the same results on any specific set of data. Raudenbush & Bryk (1986) and Goldstein (1987) have suggested that HLM is useful for data-analytic situations that may not be adequately handled using the general linear model, of which MLR is a specific case. Specifically, HLM is believed to be, and in fact was designed to be, more accurate in situations involving multi-level data, i.e., situations in which data are measured at more than one level.

Multi-level data situations are not uncommon, particularly in educational research. A common example involves data measured at both the student level and also at the teacher and/or the school level. This hierarchical or nested structure does not appear to be adequately modeled using the general linear model framework or more specifically, multiple linear regression. However, Mundfrom and Schultz (2001) found that little difference existed between the predicted values generated using HLM and those obtained from MLR when an appropriate MLR model was utilized. They did find some differences in parameter estimates between the two procedures, although in most cases those differences were small. Although their findings were obtained from comparing a relatively few actual data sets, the results would seem to indicate that MLR may be an appropriate alternative for analyzing multi-level data.

## Purpose

This study was designed to examine more closely differences among the parameter estimates between HLM and MLR. Littel, Milliken, Stroup, & Wolfinger (1996) show examples of analyses using the general linear model produce identical results to ones using an HLM model. Bryk & Raudenbush (1992) on the other hand cite examples in which the analyses using the two procedures produce similar, but different results. One possibility for explaining why in some cases HLM and MLR produce parameter estimates that are the same whereas in other instances these estimates differ could be differing correlational structures in the data. Specifically, perhaps the size of the intraclass correlation could be affecting the parameter estimates.

The intraclass correlation is often described as the proportion of variability in the dependent variable that is explained by the group membership (Montgomery, 1997). In the typical multi-level data structure, one or more characteristics are measured on individuals in each of several groups, and one or more characteristics are also obtained on these same individuals at the group level. That is, each individual in the same group will have the same value for the group level characteristic(s). Hence, differences among the group-level characteristic(s) can account for some of the variation in the responses, and this variation is referred to as the intraclass correlation. Murray (1998) also refers to this quantity as a clustering effect.

It is conceivable that multi-level data in which little or no differences exist among the group-level characteristic(s) (i.e., little or no intraclass correlation) will produce parameter estimates that are very similar, if not identical, when analyzed using HLM and MLR. But when the intraclass correlation is greater, i.e., larger differences among group-level characteristics, the two analyses will produce parameter estimates that exhibit larger differences among them. The purpose of this study is to compare the parameter estimates obtained from HLM with those obtained from MLR on simulated data in which the intraclass correlation is systematically varied from small values to larger ones.

## Method

In this simulation study the outcome of interest was the difference between the vector of parameter estimates obtained when data are analyzed using both MLR and HLM. The independent variable was the size of the intraclass correlation among the groups. Our objective was to use ten different values of the intraclass correlation: 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9. However, because we were unable to generate directly data sets with a given intraclass correlation, we were forced to "back into" these values by generating data with given correlations and then checking the intraclass correlation. Trial and error and the ability to learn while doing allowed us to become more proficient as the process progressed. However, we were unable to fix the intraclass correlation at these specified values and instead were forced to settle for intraclass correlations within a small, specified range. Therefore, the ranges of intraclass correlations examined in this study were as follows: < .05, between .05 and .15, between .15 and .25, and so on through between .85 and .95. Within each of these intraclass correlation ranges, 15,000 data sets were generated.

The HLM model used in each of these data sets was a simple two-level model with one individual-level variable and one group-level variable. This model can be expressed as:

$$Yij = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij},$$

where $Yij$ represents the response for the $i^{\text{th}}$ individual in the $j^{\text{th}}$ group
$\beta_{0j}$ represents the y-intercept of the regression line for the $j^{\text{th}}$ group
$\beta_{1j}$ represents the slope of the regression line for the $j^{\text{th}}$ group
$X_{ij}$ is the measurement on the individual-level variable for the $i^{\text{th}}$ individual in the $j^{\text{th}}$ group,
$r_{ij}$ represents random error associated with the response for the $i^{\text{th}}$ individual in the $j^{\text{th}}$ group, and
$j$ ranges from 1 to $J$, the number of groups in the data set.

In the HLM model, however, the group parameters, $\beta_{0j}$ and $\beta_{1j}$, are not estimated individually from the raw individual data, but instead are estimated from a second-level model using group-level data. This model can be expressed as:

$$\beta_{kj} = \gamma_{k0} + \gamma_{k1} W_j + u_{kj},$$

where $\beta_{kj}$ represents the $k^{\text{th}}$ parameter for the $j^{\text{th}}$ group
$\gamma_{k0}$ represents the $y$-intercept of the regression line for the $k^{\text{th}}$ parameter
$\gamma_{k1}$ represents the slope of the regression line for the $k^{\text{th}}$ parameter
$W_j$ represents the measurement on the group-level variable for the $j^{\text{th}}$ group, and
$u_{kj}$ represents random error associated with the $k^{\text{th}}$ parameter for the $j^{\text{th}}$ group.

In this model, there are two second-level models, one for the $y$-intercept, $\beta_{0j}$, and one for the slope, $\beta_{1j}$:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j} \quad \text{and} \quad \beta_{1j} = \gamma_{10} + \gamma_{11} W_j + u_{1j}.$$

The MLR model examined in this study includes both the individual-level variable and the group-level variable along with their interaction. This model, with two independent variables and their interaction term, can be expressed as:

$$Yij = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{1ij} X_{2ij} + e_{ij},$$

The number of groups was set to ten with the number of individuals in each group allowed to vary, but the mean group size was approximately 50. Each data set contains 500 observations on the response variable, 500 observations on the individual-level variable, and approximately 50 observations in each of the ten groups on the group-level variable. Hence, the parameter estimate vectors for multiple regression contained four values, an intercept, a coefficient for the individual-level variable, a coefficient for the group-level variable, and a coefficient for the "interaction" between the individual-level variable and the group-level variable, $[\beta_0\ \beta_1\ \beta_2\ \beta_3]'$. For the HLM model, the parameter estimate vectors contained the estimates of the second-level fixed effects parameters (often referred to as "gammas"), $[\gamma_{00}\ \gamma_{10}\ \gamma_{01}\ \gamma_{11}]'$.

Once the 15,000 data sets were produced within each intraclass correlation range, each set was analyzed using both a hierarchical linear model and a multiple linear regression model to produce a vector of parameter estimates. All of the analyses were run using SAS, with the MLR analyses performed with PROC GLM and PROC MIXED used for the HLM analyses. Wang (1997) demonstrated the similarities in results between multi-level analyses preformed in SAS PROC MIXED and the HLM software (e.g., Raudenbush, Bryk, and Congdon, 1999). For each data set, the parameter estimate vectors from MLR and HLM were subtracted to obtain a difference vector. Consecutive sets of 50 difference vectors were grouped together to form mean difference vectors and Hotelling's $T^2$ statistic was used to determine the proportion of mean difference vectors that exceeded an $F$-critical value at the 5% significance level.

## Results

The parameter estimate vectors from both HLM and MLR analyses on each of the 15,000 data sets showed remarkable similarity when compared with each other. Once the parameter estimate vectors were obtained and subtracted to form difference vectors, the average difference vectors were compared to a null hypothesis of a zero vector which would be indicative of no difference between HLM and MLR in terms of the parameter estimates. These mean difference vectors were tested using Hotelling's $T^2$ statistic. If this null hypothesis were true, we would expect to find about 5% of the mean difference vectors differing from the zero vector, simply by chance variation. Larger percentages of the mean difference vectors differing from the zero vector would indicate that HLM and MLR produce parameter estimates, which differ significantly from one another.

For each of the ten intraclass correlation ranges, the results are displayed separately in Table 1. Notice that none of the intraclass correlation ranges had $F$-approximations that approach the expected 5% Type I error criteria. Only when the intraclass correlation exceeded 0.65 did the percentage of F's exceeding the 5% critical value surpass even 1%. In fact, these pairs of parameter estimates are so similar to each other that in more than half of the intraclass correlation ranges studied, less than 1% of the mean difference vectors differed significantly from each other. There appears to be little evidence in these data that the size of the intraclass correlation has any influence upon the difference between the parameter estimates from HLM and MLR.

## Discussion

In this study, numerous data sets were generated and analyzed to investigate the effect, if any, that the intraclass correlation has on the parameter estimates of multiple linear regression as compared to those from hierarchical linear modeling. Previous work had indicated that in some cases these estimates were very similar, if not identical, and that in other cases, the parameter estimates from these two procedures were quite different. It was conjectured that in those data that produced different estimates of the parameters, perhaps it was a larger intraclass correlation that could account for these estimates differing. These results would seem to indicate that it is not the size of the intraclass correlation that is responsible for differences in parameter estimates between HLM and MLR. There can be no doubt that in some instances these estimates do indeed differ. Why they differ in some cases and not in others is still a question that needs answering. It would appear, however, that the intraclass correlation could be eliminated from the list of possible explanations for these differences.

**Table 1**. Percentages of Calculated $F$-Values Beyond the 5% $F$-critical value.

| Intraclass Correlation Range | Percentage of $F$'s beyond the 5% $F$ Critical Value |
|---|---|
| 0.00-<0.05 | 0.9 |
| 0.05-<0.15 | 0.3 |
| 0.15-<0.25 | 0.9 |
| 0.25-<0.35 | 0.6 |
| 0.35-<0.45 | 0.0 |
| 0.45-<0.55 | 0.0 |
| 0.55-<0.65 | 0.0 |
| 0.65-<0.75 | 1.3 |
| 0.75-<0.85 | 1.3 |
| 0.85-<0.95 | 1.1 |

It should be noted that the MLR model that was compared to HLM in this study is the model that contains the interaction term between the individual-level variable and the group-level variable. Previous work by Mundfrom and Schultz (2001) indicated that the simpler MLR model without interactions was not an adequate competitor to HLM in terms of similar predicted values or parameter estimates. Also, in this study, only the simplest full multi-level model was investigated. It could still be the case that more complex multi-level models may show more influence due to the size of the intraclass correlation in terms of differences between parameter estimates from HLM and MLR. It should also be noted that HLM has been shown to provide better, more accurate (i.e., larger) estimates of standard errors of parameter estimates than does multiple regression. Indeed, it is this characteristic of HLM that provides its premier advantage in the analysis of multi-level data. This study did not compare standard errors of the two techniques.

This study did not assume a balanced design (i.e., equal sample sizes per group) although further research on the effect of unbalanced designs on the respective parameter estimates may be warranted. Similarly, the role of heterogeneous variances in the comparison of these parameter estimates may also be worthy of further investigation.

Finally, it is not the intent of this study to imply that hierarchical linear modeling is in any way inadequate or inappropriate for use in analyzing data, particularly multi-level data. The intent here is simply to investigate differences and similarities between the results obtained when these two procedures are used with the same multi-level data sets, and to see if reasons can be identified as to why, or in what situations, these procedures produce different outcomes.

## References


Bryk, A. S. & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage Publications.

Goldstein, H. (1987). *Multilevel models in educational and social research*. London: Griffin.

Littel, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). SAS system for mixed models. Cary, NC: SAS Institute, Inc.

Montgomery, D. C. (1997). *Design and analysis of experiments* (4th edition). New York: Wiley.

Mundfrom, D. J. & Schultz, M. R. (2001). A comparison between hierarchical linear modeling and multiple linear regression in selected data sets. *Multiple Linear Regression Viewpoints*, *27*(1), 3-11.

Murray, D. M. (1998). *Design and analysis of group-randomized trials*. New York: Oxford University Press.

Raudenbush, S. W. & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, *59*, 1-17.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (1999). *HLM Version 5*.Chicago IL: Scientific Software International.

Wang, J. (1997). Using SAS PROC MIXED to demystify the Hierarchical Linear Model. *Journal of Experimental Education*, *66*, 84-93.



Send correspondence to: Daniel J Mundrom, Department of Applied Statistics and Research Methods, University of Northern Colorado, Greeley, Colorado 80639. Email: Daniel.Mundfrom@unco.edu.