Journal of Modern Applied Statistical Methods

Volume 17 | Issue 1 Article 26

11-14-2018

The Impact of Sample Size in Cross-Classified Multiple Membership Multilevel Models

Hyewon Chung
Chungnam National University, hyewonchung7@gmail.com

Jiseon Kim
jiseonk5@gmail.com

Ryoungsun Park
Wayne State University, ryoungsun.park@wayne.edu

Hyeonjeong Jean
Chungnam National University, gomtw1121@gmail.com

The Impact of Sample Size in Cross-Classified Multiple Membership Multilevel Models

Cover Page Footnote
This work was supported by research fund of Chungnam National University.

Journal of Modern Applied Statistical Methods
May 2018, Vol. 17, No. 1, eP2491
doi: 10.22237/jmasm/1542209860

ISSN 1538 − 9472

The Impact of Sample Size in Cross-Classified Multiple Membership Multilevel Models

Hyewon Chung
Chungnam National University
Daejeon, South Korea

Jiseon Kim
Independent Researcher

Ryoungsun Park
Wayne State University
Detroit, MI

Hyeonjeong Jean
Chungnam National University
Daejeon, South Korea

A simulation study was conducted to examine parameter recovery in a cross-classified multiple membership multilevel model. No substantial relative bias was identified for the fixed effect or level-one variance component estimates. However, the level-two cross-classification multiple membership factor variance components were substantially biased with relatively fewer groups.

Keywords: Cross-classified multiple membership multilevel model, sample size

Introduction

Cross-classified multiple membership random effects modeling, which is an extension of traditional multilevel modeling, is used to handle the complexity of cross-classified multiple membership data structures (Goldstein, 2010). Traditional multilevel models or hierarchical linear models enable researchers to investigate not only the effect of lower-level units but also the effect of higher-level units and the impact of their characteristics on outcome measures (Goldstein, 2010; Raudenbush & Bryk, 2002). Although in practice, most theoretical and empirical studies that employ multilevel models address purely hierarchical data structures, multilevel data often cannot be adequately represented by such structures. A typical example of a more realistic non-pure hierarchy is the data structure that arises in large-scale longitudinal studies that track the same subjects over periods of time.

doi: 10.22237/jmasm/1542209860 | Accepted: February 22, 2018; Published: November 14, 2018.
Correspondence: Hyewon Chung, hyewonchung7@gmail.com

For example, Leckie (2009) conducted a study using cross-classified multiple membership multilevel data to investigate the effect of student mobility on academic achievement. The study used a dataset tracking students' moving between secondary schools as well as between neighborhoods; some students transferred secondary schools and/or neighborhoods (i.e., were members of multiple secondary schools and/or neighborhoods). Consequently, it was necessary to consider two classifications at level two (i.e., secondary schools and neighborhoods) and the students were cross-classified by secondary school and neighborhood. Thus, the dataset had a cross-classified multiple membership data structure. The data employed in Leckie's study revealed that 8% of the students transferred to at least one other secondary school and that 27 % of the students changed residence, with 23% of the students changing neighborhoods over the course of the study. Part of Leckie's study was to compare modeling systems that do or do not account for cross-classification and/or multiple membership of secondary schools and neighborhoods. The model that ignored cross-classification of neighborhoods and multiple membership of secondary schools and neighborhoods (i.e. application of traditional multilevel modeling) resulted in smaller parameter and standard error estimates of the secondary school variance component as compared to the estimates in the model that accounted for cross-classified multiple membership.

Additional examples of cross-classified multiple membership data structures can be found in diverse fields (e.g., Browne, Goldstein, & Rasbash, 2001; Goldstein, Burgess, & McConnell, 2007). For example, in medical research, patients typically consult with doctors and nurses; this results in complex situations in which patients can be cross-classified by doctors and nurses. Multiple membership relations also occur when patients see different doctors and/or nurses on different occasions. Refer to the chapter by Beretvas (2011) for more detailed examples including a series of contingency tables for various pure hierarchical data, cross-classified data and cross-classified multiple membership data structures. In this paper, the parameterization of cross-classified multiple membership random effects models will be explained using the example introduced above of students (level-one) cross-classified by schools (one level-two classification) and neighborhoods (another level-two classification), where some students transferred schools and/or neighborhoods during the period of study.

Two-Level Cross-Classified Multiple Membership Multilevel Model

Unconditional Cross-Classified Multiple Membership Multilevel Model

The unconditional cross-classified multiple membership multilevel model is expressed as follows: at level one

$$y_{i\{j_1\}\{j_2\}} = \beta_{0\{j_1\}\{j_2\}} + e_{i\{j_1\}\{j_2\}} \tag{1}$$

and, at level two,

$$\beta_{0\{j_1\}\{j_2\}} = \gamma_{000} + \sum_{h_1 \in \{j_1\}} W_{ih_1 0} U_{0h_1 0} + \sum_{h_2 \in \{j_2\}} W_{i0h_2} U_{00h_2} \tag{2}$$

where $y_{i\{j_1\}\{j_2\}}$ is the outcome for level-one unit *i* (here, a student). This student is a member of one or more elements of a set $\{j_1\}$ of level-two units of the first type (schools) and of another set $\{j_2\}$ of units corresponding to the other level-two cross-classification factor (neighborhoods). The sum of the weights for each type of level-two unit to level-one unit *i* belongs is equal to one, i.e.

$$\sum_{h_1 \in \{j_1\}} W_{ih_1 0} = 1 \quad \text{and} \quad \sum_{h_2 \in \{j_2\}} W_{i0h_2} = 1$$

(Goldstein, 2010).

Conditional Cross-Classified Multiple Membership Multilevel Model

In the current example, particular characteristics of each student (*X*), school (*S*), and neighborhood (*N*) can be added to the model. The resulting conditional cross-classified multiple membership multilevel model is expressed as follows: at level one

$$Y_{i\{j_1\}\{j_2\}} = \beta_{0\{j_1\}\{j_2\}} + \beta_{1\{j_1\}\{j_2\}} X_{i\{j_1\}\{j_2\}} + e_{i\{j_1\}\{j_2\}} \tag{3}$$

and, at level two,

$$\begin{cases} \beta_{0\{j_1\}\{j_2\}} = \gamma_{000} + \gamma_{010} \sum_{h_1 \in \{j_1\}} W_{ih_1 0} S_{0h_1 0} + \sum_{h_1 \in \{j_1\}} W_{ih_1 0} U_{0h_1 0} + \gamma_{001} \sum_{h_2 \in \{j_2\}} W_{i0h_2} N_{00h_2} \\ \qquad + \sum_{h_2 \in \{j_2\}} W_{i0h_2} U_{00h_2} \\ \beta_{1\{j_1\}\{j_2\}} = \gamma_{100} \end{cases} \tag{4}$$

where the multiple membership classification predictors (S and N) are weighted in the same way as the factor residuals $\left[U_{0\{j_1\}0}\right]$ and $\left[U_{00\{j_2\}}\right]$. In equation (4), the coefficient of the level-one predictor, X, is modeled as fixed, although additional predictors and/or random effects can be added to the model.

In a traditional unconditional multilevel model, the level-one (σ^2) and level-two (τ_{00}) residual variance components are typically used to provide a measure of the degree of dependence of the outcome measure. Specifically, the intra-class correlation coefficient, ρ_{ICC} (Raudenbush & Bryk, 2002), is calculated as follows:

$$\rho_{\text{ICC}} = \frac{\tau_{00}}{\sigma^2 + \tau_{00}}. \tag{5}$$

The higher the value of ρ_{ICC}, the higher the proportion of the variability in the outcome measure that is related to level-two units. For an unconditional cross-classified multiple membership multilevel model, a similar coefficient is used to represent the degree of variability in the outcome measure that is attributable to each level-two classification factor. This coefficient is called the intra-unit correlation coefficient (IUCC), ρ_{IUCC} (Raudenbush & Bryk, 2002), and is calculated as

$$\rho_{\text{IUCC},\{j_1\}} = \frac{\tau_{u0\{j_1\}}}{\sigma^2 + \tau_{u0\{j_1\}} + \tau_{u0\{j_2\}}} \tag{6}$$

for classification factor $\{j_1\}$ and as

$$\rho_{\text{IUCC},\{j_2\}} = \frac{\tau_{u0\{j_2\}}}{\sigma^2 + \tau_{u0\{j_1\}} + \tau_{u0\{j_2\}}} \tag{7}$$

for classification factor $\{j_2\}$. As with ρ_{ICC}, the larger the value of ρ_{IUCC}, the larger the proportion of the variability in the outcome measure that is attributable to the relevant classification.

Sample Size

The appropriateness of sample size has been widely studied in the multilevel modeling literature to determine the minimum desirable sample size. The impact of sample size is more complex in the case of multilevel models because multilevel models involve multiple sample sizes, and researchers need to determine a reasonable sample size for each level (Bell, Ferron, & Kromrey, 2008; Bell, Morgan, Kromrey, & Ferron, 2010). For two-level multilevel modeling analysis, Kreft (1996) recommended the '30/30' rule, which prescribes a minimum of 30 units at each level of the analysis to obtain unbiased estimates of all parameters and their associated standard errors. Hox (1998) recommended 50 groups with a minimum of 20 observations per group when modeling cross-level interactions. Previous researchers have emphasized that a large number of groups is more critical than a large number of observations per group for obtaining accurate estimates (Clarke & Wheaton, 2007; Newsom & Nishishiba, 2002). Although fixed effect estimates are less sensitive to the number of groups, variance component estimates are substantially influenced by the number of groups. Mok (1995) observed that five groups at level two yielded substantially biased variance estimates, whereas Clarke and Wheaton (2007) recommended at least 100 groups with a minimum of ten observations per group to obtain an unbiased estimate of the intercept variance. For cases in which the slope variance is to be estimated, they recommended at least 200 groups with a minimum of 20 observations per group.

Given the increasing frequency with which cross-classified data structures are being encountered in multilevel modeling (Browne et al., 2001; Goldstein et al., 2007; Grady & Beretvas, 2010; Leckie, 2009), the minimum sample requirement for cross-classified multiple membership multilevel model estimation should be assessed. Therefore, this simulation study was conducted to assess the parameter recovery of the fixed effect and random variance components for cross-classified multiple membership data structures under a variety of manipulated conditions.

Methodology

Simulation Study Design

Five factors were manipulated in the simulation study, namely, the average group size (10, 20, and 40), the number of groups (20, 50, and 100), the multiple membership rate (10%, 20%, and 40%), the cross-classification rate (20%, 40%, and 100%), and the IUCC (10%, 20%, and 30%). A completely crossed design, in which three values were investigated for each of the five factors, was employed; thus, 243 ($3 \times 3 \times 3 \times 3 \times 3$) combinations of conditions were obtained.

Average Group Size

Previous simulation studies using multilevel modeling have typically employed 5 as the minimum sample size at level one. In a previous simulation study by Meyers and Beretvas (2006), the group size values were manipulated from 20 to 40. Meyers and Beretvas employed a balanced cross-classified design, indicating that same number of students per school and neighborhood (i.e., equal cell size). To better approximate real-world situations, the current study selected the following three values for the average group size: 10 [5-15], 20 [15-25], and 40 [30-50]. For example, we randomly generated between 5 and 15 students per school and neighborhood for an average sample size of 10.

Number of Groups

For this study, a simple scenario was constructed in which the number of groups was equal for both level-two classification factors (here, school and neighborhood) in each condition. In previous methodological studies that have employed either cross-classified multilevel models or multiple membership multilevel models, the number of groups has ranged from 20 to 100 (Chung & Beretvas, 2012; Meyers & Beretvas, 2006). Thus, three values for the numbers of schools and neighborhoods were investigated: 20, 50, and 100.

Multiple Membership Rate

In the example considered here, the multiple membership rate can be interpreted as the likelihood of a student being mobile. Student mobility ranged from 12% to 38.5% between 2005 and 2010 (Ihrke & Faber, 2012). Three multiple membership rates were examined in this study: 10%, 20%, and 40%. These mobility values were selected because they correspond closely to the values reported in applied research, as summarized in a previous simulation study using multiple membership multilevel modeling (Chung & Beretvas, 2012; Ihrke & Faber, 2012).

Cross-Classification Rate

The cross-classification rate indicates the ratio of number of students who are cross-classified by schools and neighborhood out of the total number of students. In the context of the current study, students (level one) are cross-classified by schools and neighborhood. Based on previous simulation studies using cross-classified multilevel modeling (e.g., Jeong & Kang, 2013; Meyers & Beretvas, 2006), cross-classification rates of 20%, 40%, and 100% were investigated in this study.

Intra-Unit Correlation Coefficient (IUCC)

The values of the IUCC were manipulated based on values found from previous simulation studies employing cross-classified multilevel models (Hox, Moerbeek, & van de Schoot, 2002; Luo, Cappaert & Ning, 2015). Small, medium and large IUCC values of 10%, 20%, and 30%, respectively, were used.

Data Generation

All simulated datasets were generated using MLwiN 2.36 (Rasbash, Steele, Browne & Goldstein, 2016). MLwiN was used to generate 1,000 datasets per combination of conditions following previous methodological studies that have employed either cross-classified multilevel models or multiple membership multilevel models (e.g., Chung & Beretvas, 2012; Meyers & Beretvas, 2006). The data were generated in accordance with a two-level cross-classified multiple membership multilevel model with students at level one and a cross-classification of schools and neighborhoods at level two, along with a condition-dependent multiple membership rate of students attending multiple schools and/or neighborhoods. In addition, one predictor for each level-two classification factor and one level-one predictor were included in the model, matching the conditional model presented in equations (3) and (4). The single-equation formulation of the model is as follows:

$$\begin{aligned} Y_{i(\{j_1\},\{j_2\})} &= \gamma_{000} + \gamma_{100} X_{i(\{j_1\},\{j_2\})} + \gamma_{010} \sum_{h_1 \in \{j_1\}} W_{ih_1} S_{h_1} + \sum_{h_1 \in \{j_1\}} W_{ih_1} U_{0h_1 0} \\ &\quad + \gamma_{001} \sum_{h_2 \in \{j_2\}} W_{ih_2} N_{h_2} + \sum_{h_2 \in \{j_2\}} W_{ih_2} U_{00h_2} + e_{i(\{j_1\},\{j_2\})}. \end{aligned} \tag{8}$$

The generating values for the fixed effects (see equation (8)) were as follows: 100 for γ_{000}, 0.5 for γ_{100}, 0.5 for γ_{010}, and 0.5 for γ_{001}. The values for the level-one predictor and the two level-two predictors (i.e., for X, S, and N) were generated from a standard normal distribution with a mean of 50 and a standard deviation of 10. All generating values were selected based on previous methodological research in which either a cross-classified multilevel model or a multiple membership multilevel model was used (Chung & Beretvas, 2012; Meyers & Beretvas, 2006; Wolff Smith & Beretvas, 2014).

Analyses

Data Analyses

MLwiN 2.36 (Rasbash et al., 2016) and Markov chain Monte Carlo (MCMC) estimation were used to estimate the cross-classified multiple membership multilevel model with default priors. The default non-informative prior was employed for each fixed effect; that is, the prior was proportional to 1, similar to a uniform distribution (Rasbash et al., 2016). Additionally, default inverse gamma distributions (.001, .001) were used as priors for the random effect variance components at both level one and level two. To determine the required number of iterations, the Raftery-Lewis diagnostic (Raftery & Lewis, 1992) was applied. As a pilot study, 50,000 iterations were run, with 10,000 iterations for burn-in of the first datasets generated across the 243 conditions. This process satisfied the minimum number of iterations for the Gibbs sampler as suggested by the Raftery-Lewis diagnostic.

The converged Gibbs sampling output for one simulated dataset of the condition with level-one sample size = 20, level-two sample size = 20, cc% = 20%, and mm% = 10% is presented in Figure 1. The posterior density plots (Figure 1, column a), autocorrelation plots (Figure 1, column b), and trace plots (Figure 1, column c) are presented for all parameters. Posterior density plots are a useful diagnostic for checking the Gibbs sampling convergence. Non-convergence typically manifests as multimodal distributions. The density plots in Figure 1, column a indicate unimodal distributions for all parameters. Meanwhile, the autocorrelation plots in Figure 1, column b indicate values near 0 after 20 or fewer lags. Thus, the values are approximately independent. Finally, the trace plots in Figure 1, column c randomly fluctuate around the mean after the 10,000 iterations of burn-in. This result indicates that convergence was achieved for each parameter after the initial burn-in. The plots for the other conditions in the simulation study are similar and will not be presented.

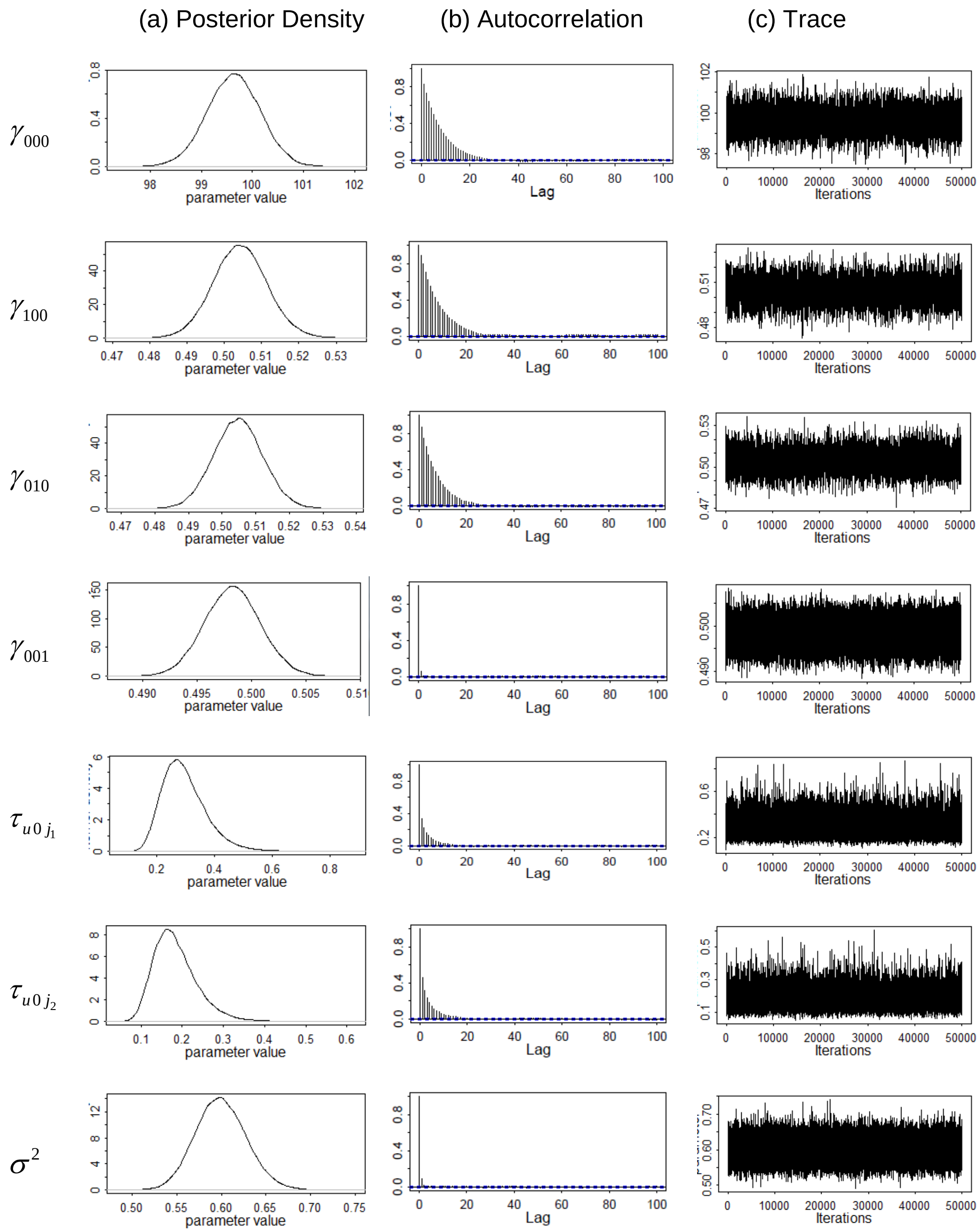


Figure 1. Posterior density, autocorrelation, and trace plots for parameters of one simulated dataset (condition 1)

Relative Parameter Bias

The relative parameter bias was calculated for each fixed and random effect variance component estimate using the following formula:

$$\mathrm{RPB}\left(\hat{\theta}_i\right) = \frac{\bar{\hat{\theta}}_i - \theta_i}{\theta_i} \tag{9}$$

where θ_i is the generating (true) value of the i^{th} parameter and $\bar{\hat{\theta}}_i$ is the average of the estimates for the i^{th} parameter across the 1,000 simulated datasets. For the estimation of each parameter, the relative parameter bias value was considered acceptable if its magnitude was less than .05 (Hoogland & Boomsma, 1998). In addition, analysis of variance (ANOVA) was used to explore the effects of the simulation conditions on the relative bias. Both main effects and 2-way interaction effects were analyzed, with the simulation conditions as the independent variables and the relative bias as the outcome variable. An alpha level of .01 was used as the cutoff for statistical significance. The partial eta squared value (η_p^2) was used to estimate the size of a given effect of an independent predictor. The ANOVA results for the relative bias measures are presented for conditions in which the η_p^2 effect sizes were found to be larger than .01.

Coverage Rate of the 95% Credible Interval

The coverage rates of the 95% credible interval [2.5%, 97.5%] were derived from the quantiles of the 50,000 parameter estimates. For each parameter, the coverage indicator was set to 1 if the true value was included within the credible interval and to 0 if the true value fell outside the credible interval. The coverage rates of the 95% credible interval were computed as the average of the coverage indicators across 1,000 replications for each condition. Then, logistic regression was used to assess the potential impact of the simulated conditions as predictors, with the confidence interval coverage indicator, either zero or one, as the dependent variable.

Root Mean Square Error

The root mean square error (RMSE) was calculated using the equation below.

$$\mathrm{RMSE}\left(\hat{\theta}\right) = \sqrt{\mathrm{MSE}\left(\hat{\theta}\right)} = \sqrt{\mathrm{E}\left(\left(\hat{\theta} - \theta\right)^2\right)}. \tag{10}$$

Table 1. Summary of mean relative parameter biases, coverage rates of the 95% credible interval, and RMSEs for fixed estimates by condition

	Relative parameter bias				Coverage rate of the 95% credible interval				RMSE			
Condition	**Inter**	**Stu**	**Sch**	**Neigh**	**Inter**	**Stu**	**Sch**	**Neigh**	**Inter**	**Stu**	**Sch**	**Neigh**
Average group size												
10	<0.001	<0.001	<0.001	<0.001	0.948	0.953	0.947	0.941	0.668	0.004	0.009	0.009
20	<0.001	<0.001	<0.001	<0.001	0.948	0.952	0.946	0.942	0.591	0.003	0.008	0.008
40	<0.001	<0.001	0.001	<0.001	0.947	0.952	0.946	0.941	0.530	0.002	0.008	0.007
Number of groups												
20	<0.001	<0.001	0.001	<0.001	0.947	0.953	0.943	0.935	0.884	0.004	0.013	0.011
50	<0.001	<0.001	<0.001	<0.001	0.946	0.953	0.946	0.943	0.534	0.003	0.008	0.007
100	<0.001	<0.001	<0.001	<0.001	0.949	0.952	0.950	0.945	0.371	0.002	0.005	0.005
Multiple membership rate (%)												
10	<0.001	<0.001	0.001	<0.001	0.950	0.951	0.947	0.941	0.599	0.003	0.008	0.008
20	<0.001	<0.001	<0.001	<0.001	0.946	0.950	0.944	0.943	0.592	0.003	0.009	0.007
40	<0.001	<0.001	<0.001	<0.001	0.946	0.956	0.948	0.940	0.598	0.003	0.009	0.007
Cross-classification rate (%)												
20	<0.001	<0.001	<0.001	<0.001	0.946	0.953	0.946	0.940	0.609	0.003	0.009	0.008
40	<0.001	<0.001	<0.001	<0.001	0.947	0.953	0.947	0.942	0.583	0.003	0.008	0.007
100	<0.001	<0.001	<0.001	<0.001	0.949	0.951	0.946	0.941	0.597	0.003	0.009	0.008
IUCC (%)												
10	<0.001	<0.001	<0.001	<0.001	0.945	0.952	0.944	0.940	0.519	0.003	0.007	0.007
20	<0.001	<0.001	<0.001	<0.001	0.948	0.954	0.948	0.941	0.610	0.003	0.009	0.008
30	<0.001	<0.001	<0.001	<0.001	0.949	0.952	0.949	0.943	0.666	0.002	0.010	0.008
Mean	<0.001	<0.001	<0.001	<0.001	0.947	0.950	0.945	0.942	0.595	0.003	0.009	0.008

Note: IUCC = Intra-unit correlation coefficient; RMSE = Root mean square error; Inter = Intercept; Stu = Student; Sch = School; Neigh = Neighborhood.

In the presence of parameter estimate bias, the RMSE is a combined measure of the bias and variability of each parameter estimate with respect to the true parameter value.

Results

Fixed Effect Estimates

Relative Parameter Bias

Based on the criterion of Hoogland and Boomsma (1998), no substantial RPB was found in the intercept estimates. Additionally, no substantial parameter estimation bias was detected for the predictor coefficients for the level-one factor (student) or either of the cross-classified multiple membership factors (school and neighborhood) across all conditions. Given the lack of substantial bias found for the fixed effect estimates, ANOVA was not conducted for these results.

Coverage Rate of the 95% Credible Interval

The coverage rates of the 95% credible interval were close to nominal coverage for the intercept, level-one predictor, and level-two predictor estimates across all conditions (see Table 1, coverage rate section). The logistic regression results indicated that the multiple membership rate was significantly related to the coverage rate of the 95% credible interval for the intercept estimates and student-level predictor estimates ($p < .001$). The differences were trivial for the intercept estimates: .950 for 10%, .946 for 20%, and .946 for 40%. The differences were also very small for the student-level predictor estimates: .951 for 10%, .950 for 20%, and .956 for 40%.

In addition, according to the logistic regression results, the number of groups and the IUCC were related to the coverage rates for the school-level predictor estimates (*ps* < 0.001). The differences for different numbers of groups were very small: .943 for 20 groups, .946 for 50 groups, and .950 for 100 groups. Similarly, the differences were very minimal for different IUCC values: .944 for 10%, .948 for 20%, and .949 for 30%.

Root Mean Square Error

With a larger average group size and a larger number of groups, the RMSE decreased for the intercept, level-one predictor, and level-two predictor estimates.

There were no substantial differences in the RMSE results across multiple membership rates, cross-classification rates and IUCC values.

Variance Component Estimates

Relative Parameter Bias

As seen in Table 2, none of the cross-classified multiple membership multilevel model estimates of the level-one (student) variance components was found to be substantially biased across any of the conditions (mean of 0.006, ranging from -0.0005 to 0.028).

However, for one of the cross-classified multiple membership classification factors (school), the variance component estimates were substantially biased for a subset of conditions (mean of 0.059, ranging from –0.013 to 0.159). The ANOVA results revealed that the main overestimation bias for the school variance component was associated with the number of groups: $F(2, 242949) = 1716.11$, $p < .001$, $\eta_p^2 = .014$. The average relative bias was 0.124 for 20 groups, 0.031 for 50 groups, and 0.015 for 100 groups. Increasing the number of level-two groups from 20 to 100 substantially decreased the degree of positive relative bias. Slight overestimation was also noticed for some of the other simulated conditions, including the average group size, multiple membership rate, cross-classification rate, and IUCC (see Table 2, school column of the relative bias section). According to the ANOVA results, no other main effects were found to have a significant and practical impact on the relative parameter bias, and no two-way interaction effects were found to have a noticeable impact on the relative bias.

For the other cross-classified multiple membership classification factor (neighborhood), the variance component estimates were also substantially biased for a subset of conditions (mean of 0.050, ranging from –0.020 to 0.142). The ANOVA results revealed that the main overestimation bias for the neighborhood variance component was again associated with the number of groups: $F(2, 242949) = 691.09$, $p < .001$, $\eta_p^2 = .006$. The average relative bias was 0.091 for 20 groups, 0.036 for 50 groups, and 0.017 for 100 groups. Similar to the results for the school variance component estimates, slight overestimation was also found for some of the other simulated conditions, including the average group size, multiple membership rate, cross-classification rate, and IUCC (see Table 2, neighborhood column of the relative bias section). No other main effects were found to have a significant impact on the relative parameter bias.

Table 2. Summary of mean relative parameter biases, coverage rates of the 95% credible interval and RMSEs for variance component estimates by condition

	Relative parameter bias			Coverage rate of the 95% credible interval			RMSE		
Condition	**Student**	**School**	**Neigh**	**Student**	**School**	**Neigh**	**Student**	**School**	**Neigh**
Average group size									
10	0.013	0.048	0.037	0.949	0.939	0.936	0.047	0.081	0.083
20	0.005	**0.063**	**0.051**	0.949	0.942	0.944	0.032	0.068	0.069
40	0.001	**0.060**	**0.057**	0.952	0.944	0.946	0.021	0.061	0.061
Number of groups									
20	0.012	**0.124**	**0.091**	0.950	0.936	0.938	0.048	0.108	0.108
50	0.005	0.031	0.036	0.948	0.942	0.943	0.031	0.061	0.062
100	0.002	0.015	0.017	0.951	0.947	0.945	0.021	0.041	0.043
Multiple membership rate (%)									
10	0.006	**0.056**	0.050	0.952	0.940	0.939	0.033	0.069	0.070
20	0.006	**0.063**	**0.050**	0.951	0.944	0.945	0.033	0.069	0.069
40	0.007	**0.053**	0.046	0.948	0.941	0.942	0.034	0.071	0.074
Cross-classification rate (%)									
20	0.006	**0.059**	**0.050**	0.949	0.941	0.942	0.033	0.073	0.075
40	0.006	**0.053**	0.046	0.949	0.943	0.942	0.034	0.068	0.068
100	0.006	**0.058**	0.048	0.951	0.941	0.942	0.033	0.069	0.070
IUCC (%)									
10	0.008	0.049	0.036	0.950	0.936	0.938	0.044	0.047	0.048
20	0.007	**0.060**	**0.052**	0.949	0.942	0.941	0.034	0.072	0.073
30	0.005	**0.063**	**0.059**	0.951	0.947	0.946	0.022	0.093	0.093
Mean	0.006	**0.059**	**0.050**	0.951	0.942	0.943	0.033	0.069	0.069

Note: IUCC = Intra-unit correlation coefficient; RMSE = Root mean square error. Highlighted relative parameter bias values exceed the Hoogland and Boomsma (1998) criteria for substantial bias.

Considering the substantial relative biases associated with the cross-classification multiple membership variance component estimates for a small number of groups, Figure 2 presents the relative biases for the cross-classified multiple membership variance component as a function of the number of groups for the different values of each of the other manipulated conditions. As seen in Figure 2, for conditions with at least 50 groups, no substantial relative biases were found.

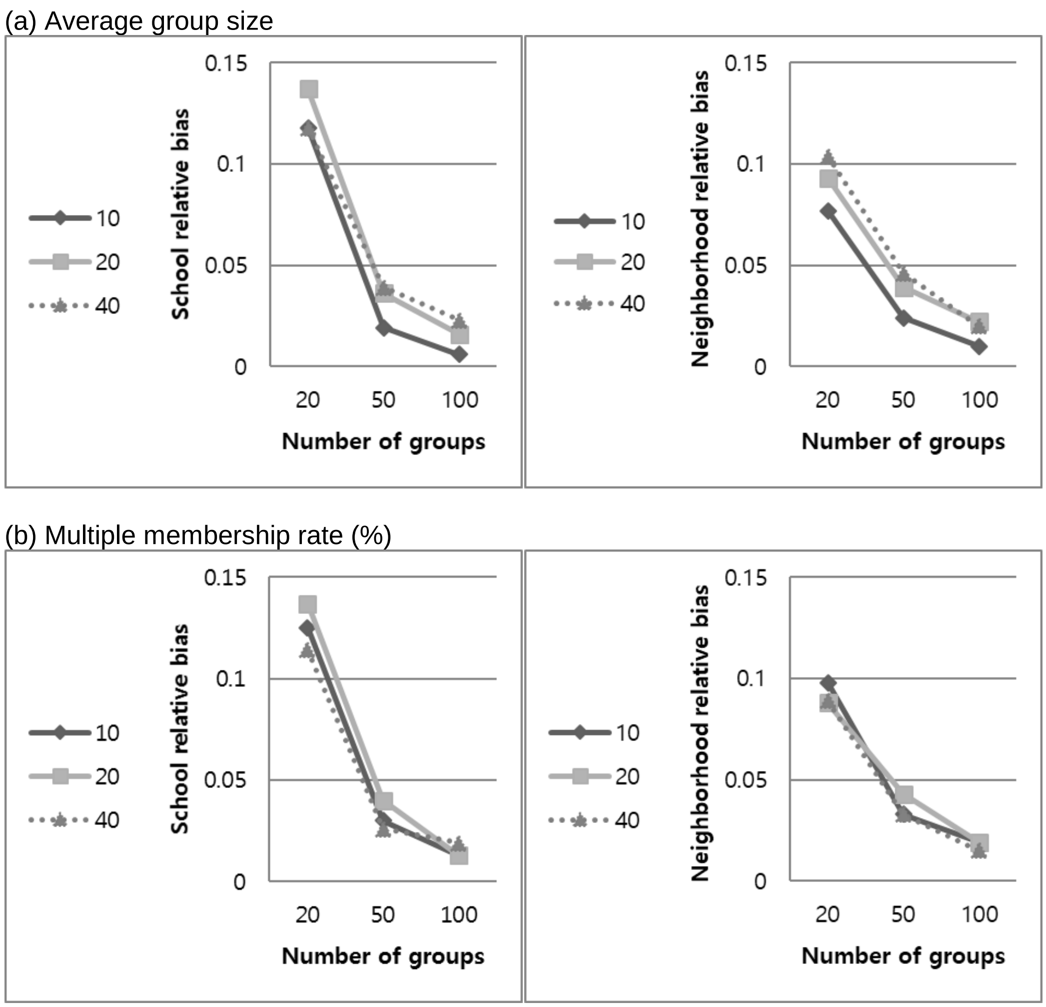
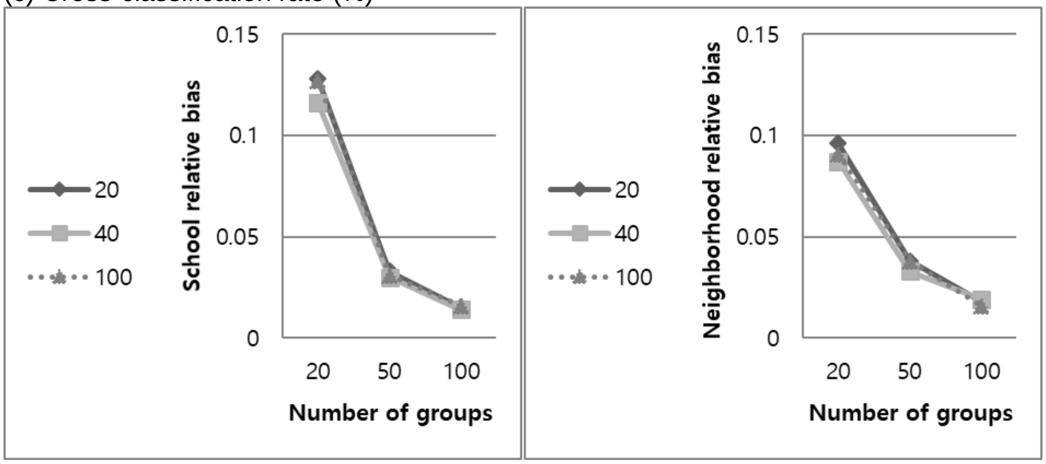


Figure 2. Relative parameter biases of the cross-classified multiple membership variance component estimates for the different simulated conditions

(c) Cross-classification rate (%)



(d) IUCC (%)

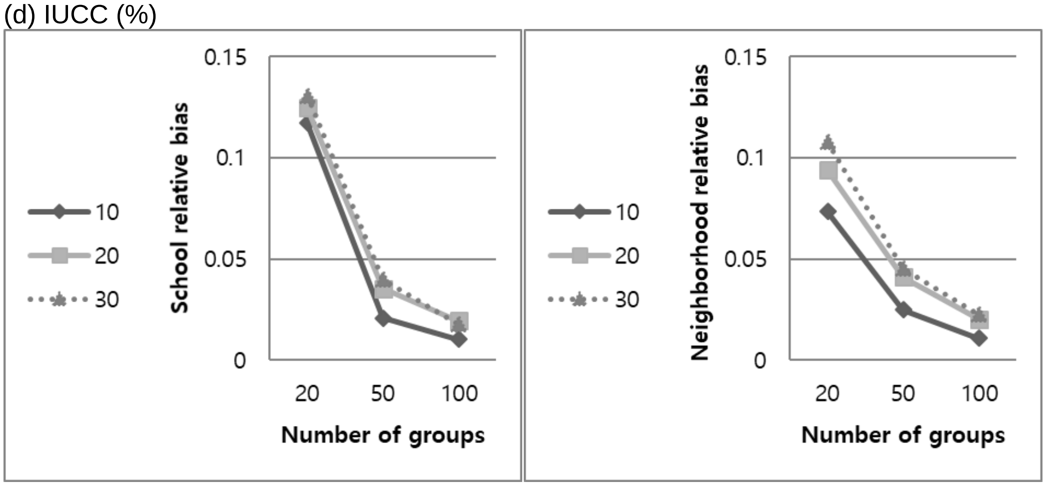


Figure 2 (continued).

Coverage Rate of the 95% Credible Interval

As seen in Table 2, the coverage rates of the 95% credible interval were close to nominal coverage for the level-one (student) and level-two cross-classified multiple membership (school and neighborhood) random effect variance components for all conditions.

The logistic regression results indicated that the average group size and multiple membership rate were significantly related to the coverage rate of the 95% credible interval for the student-level variance component estimates (*ps* < .001). The differences for different average group sizes were very small: .949 for a group size of 10, .949 for a group size of 20, and .952 for a group size of 40. Similarly, the differences were very trivial for different multiple membership rates: .952 for 10%, .951 for 20%, and .948 for 40%.

For the school-level variance component estimates, the regression analysis indicated that the average group size, the number of groups, and the IUCC were significantly related to the coverage rate of the 95% credible interval ($ps < .001$). With a larger average group size, the coverage rate increased: .939 for a group size of 10, .942 for a group size of 20, and .944 for a group size of 40. The coverage rate also increased as the number of groups increased: .936 for 20 groups, .942 for 50 groups, and .947 for 100 groups. The coverage rate increased with a higher IUCC: .936 for 10%, .942 for 20%, and .947 for 30%.

For the neighborhood-level variance component estimates, the logistic regression results showed that the average group size and IUCC were significantly related to the coverage rate of the 95% credible interval ($ps < .001$). With a larger average group size, the coverage rate increased: .936 for a group size of 10, .944 for a group size of 20, and .946 for a group size of 40. The coverage rate also increased with a higher IUCC: .938 for 10%, .941 for 20%, and .946 for 30%.

Root Mean Square Error

The RMSEs associated with the level-one (student) and level-two cross-classified multiple membership (school and neighborhood) random effect variance components were negatively related to the average group size and the number of groups, meaning that the RMSE decreased as the average group size and the number of groups increased. The RMSE also decreased as the IUCC increased for the level-one (student) random effect variance components. By contrast, the RMSE increased as the IUCC increased for the level-two cross-classified multiple membership (school and neighborhood) random effect variance components. There were no substantial differences in the RMSE results across different multiple membership rates and cross-classification rates.

Conclusion

Considering the increasing prevalence of cross-classification data structures in educational and social science research, the effect of sample size on parameter estimation in cross-classified multiple membership multilevel models requires empirical analysis. The current study was designed to address the lack of empirical research regarding the minimal sample requirement by exploring parameter estimates under a variety of conditions. For the conditions examined here, the cross-classified multiple membership multilevel model estimates for the fixed effects and the level-one variance components were not substantially biased. The cross-classified multiple membership multilevel model estimates for the level-two cross-classification multiple membership factor variance components were also unbiased across conditions with at least 50 groups. These results should encourage applied researchers to analyze sufficiently large datasets when using cross-classified multiple membership multilevel models to address cross-classified multiple membership data structures (i.e., at least 50 groups with an average group size of 10). Ultimately, the results of the current study suggest that using a cross-classified multiple membership dataset with fewer groups may lead to inaccurate conclusions.

In general, the RMSE diminished as the number of groups and the average group size increased. With an excessively small number of groups, the coverage rates of the 95% credible interval were slightly less than 5%. On average, however, the coverage rates of the 95% credible interval were close to 5% under the conditions investigated in the current study.

Due to the complexity of cross-classified multiple membership data structures, MCMC estimation is strongly recommended for cross-classified multiple membership multilevel models (Rasbash et al., 2016). With the use of suitable priors, the MCMC estimation procedure can provide more robust and precise parameter estimates than those obtained through maximum likelihood estimation (Browne & Draper, 2006). Thus, future studies should assess the impact of prior distribution selection on the estimation of fixed and random effect variance components in cross-classified multiple membership multilevel models. The findings of this study may be affected by the selection of specific values for each simulated factor. The investigation of additional values will be helpful for generalizing the findings.

References

Bell, B. A., Ferron, J. M., & Kromrey, J. D. (2008). Cluster size in multilevel models: The impact of sparse data structures on point and interval estimates in two-level models. In *JSM proceedings, Section on survey research methods* (pp. 1122-1129). Alexandria, VA: American Statistical Association. Retrieved from http://www.asasrms.org/Proceedings/y2008/Files/300933.pdf

Bell, B. A., Morgan, G. B., Kromrey, J. D., & Ferron, J. M. (2010). The impact of small cluster size on multilevel models: A Monte Carlo examination of two-level models with binary and continuous predictors. In *JSM proceedings,* *Section on survey research methods* (pp. 4057-4067). Alexandria, VA: American Statistical Association. Retrieved from http://www.asasrms.org/Proceedings/y2010/Files/308112_60089.pdf

Beretvas, S. N. (2011). Cross-classified and multiple membership random effects models. In J. Hox & J. K. Roberts (Eds.), *The handbook of advanced multilevel analysis* (pp. 313-334). New York, NY: Routledge.

Browne, W. J., & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis, 1*(3), 473-514. doi: 10.1214/06-ba117

Browne, W. J., Goldstein, H., & Rasbash, J. (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modeling, 1*(2), 103-124. doi: 10.1177/1471082x0100100202

Chung, H., & Beretvas, S. N. (2012). The impact of ignoring multiple membership data structures in multilevel models. *British Journal of Mathematical and Statistical Psychology, 65*(2), 185-200. doi: 10.1111/j.2044-8317.2011.02023.x

Clarke, P., & Wheaton, B. (2007). Addressing data sparseness in contextual population research: Using cluster analysis to create synthetic neighborhoods. *Sociological Methods & Research, 35*(3), 311-351. doi: 10.1177/0049124106292362

Goldstein, H. (2010). *Multilevel statistical models* (4th ed.). New York: Hodder Arnold. doi: 10.1002/9780470973394

Goldstein, H., Burgess, S., & McConnell, B. (2007). Modeling the effect of pupil mobility on school differences in educational achievement. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 170*(4), 941-954. doi: 10.1111/j.1467-985x.2007.00491.x

Grady, M. W. & Beretvas, S. N. (2010). Incorporating student mobility in achievement growth modeling: A cross-classified multiple membership growth curve model. *Multivariate Behavioral Research, 45*(3), 393-419. doi: 10.1080/00273171.2010.483390

Hoogland, J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research, 26*(3), 329-367. doi: 10.1177/0049124198026003003

Hox, J. (1998). Multilevel modeling: When and why. In I. Balderjahn, R. Mathar, & M. Schader (Eds.). *Classification, data analysis, and data highways* (pp. 147-154). Berlin, Germany: Springer. doi: 10.1007/978-3-642-72087-1_17

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum. doi: 10.4324/9781410604118

Ihrke, D. K., & Faber, C. S. (2012). *Geographical mobility: 2005 to 2010* (Report no. P20-567). Washington, DC: United States Census Bureau. Retrieved from https://www.census.gov/content/dam/Census/library/publications/2012/demo/p20-567.pdf

Jeong, S. Y., & Kang, S. J. (2013). Gyochabunlyu dacheungjalyoe daehan mohyeong-ui myeongsehwa olyuga mosuchujeonglyang-ui pyeon-uie michineun yeonghyang: Montekaleullo simyulleisyeon yeongu [The effects of model misspecification on the relative bias of parameter estimators in the analysis of cross-classified multilevel data: A Monte Carlo simulation study]. *Gyoyugpyeonggayeongu, 26*(4), 845-874.

Kreft, I. G. G. (1996). *Are multilevel techniques necessary? An overview, including simulation studies* (Unpublished manuscript). Los Angeles, CA: California State University at Los Angeles.

Leckie, G. (2009). The complexity of school and neighbourhood effects and movements of pupils on school differences in models of educational achievement. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 172*(3), 537-554. doi: 10.1111/j.1467-985x.2008.00577.x

Luo, W., Cappaert, K. J., & Ning, L. (2015). Modelling partially cross-classified multilevel data. *British Journal of Mathematical and Statistical Psychology, 68*(2), 342-362. doi: 10.1111/bmsp.12050

Meyers, J. L., & Beretvas, S. N. (2006). The impact of inappropriate modeling of cross-classified data structures. *Multivariate Behavioral Research, 41*(4), 473-497. doi: 10.1207/s15327906mbr4104_3

Mok, M. (1995). *Sample size requirements for 2-level designs in educational research* (Unpublished manuscript). Sydney, Australia: Macquarie University.

Newsom, J. T., & Nishishiba, M. (2002). *Nonconvergence and sample bias in hierarchical linear modeling of dyadic data* (Unpublished manuscript). Portland, OR: Portland State University.

Raftery, A. E., & Lewis, S. M. (1992). How many iterations in the Gibbs Sampler? In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian Statistics 4* (pp. 763-773). Oxford, UK: Oxford University Press.

Rasbash, J., Steele, F., Browne, W. J., & Goldstein, H. (2016). *A user's guide to MLwiN v2.36*. Bristol, UK: Centre for Multilevel Modelling, University of Bristol. Retrieved from http://www.bristol.ac.uk/cmm/media/software/mlwin/downloads/manuals/2-36/manual-web.pdf

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.

Wolff Smith, L. J., & Beretvas, S. N. (2014). The impact of using incorrect weights with the multiple membership random effects model. *Methodology, 10*(1), 31-42. doi: 10.1027/1614-2241/a000066