

The effect of missing levels of nesting in multilevel analysis

Seho Park¹, Yujin Chung^{2*}

¹Department of Biostatistics and Health Data Science, Indiana University School of Medicine, Indianapolis, IN 46202, USA

²Department of Applied Statistics, Kyonggi University, Suwon 16227, Korea

Multilevel analysis is an appropriate and powerful tool for analyzing hierarchical structure data widely applied from public health to genomic data. In practice, however, we may lose the information on multiple nesting levels in the multilevel analysis since data may fail to capture all levels of hierarchy, or the top or intermediate levels of hierarchy are ignored in the analysis. In this study, we consider a multilevel linear mixed effect model (LMM) with single imputation that can involve all data hierarchy levels in the presence of missing top or intermediate-level clusters. We evaluate and compare the performance of a multilevel LMM with single imputation with other models ignoring the data hierarchy or missing intermediate-level clusters. To this end, we applied a multilevel LMM with single imputation and other models to hierarchically structured cohort data with some intermediate levels missing and to simulated data with various cluster sizes and missing rates of intermediate-level clusters. A thorough simulation study demonstrated that an LMM with single imputation estimates fixed coefficients and variance components of a multilevel model more accurately than other models ignoring data hierarchy or missing clusters in terms of mean squared error and coverage probability. In particular, when models ignoring data hierarchy or missing clusters were applied, the variance components of random effects were overestimated. We observed similar results from the analysis of hierarchically structured cohort data.

Keywords: hierarchical structure data, missing levels of nesting, multilevel model

Introduction

A multilevel model has gained popularity as a practical and essential analysis tool in various fields such as epidemiological research, public health research, and social or educational research [1-6]. A specific study design involving multiple levels or nested structures results in a hierarchical structure and accompanies the application of multilevel modeling [1]. Multilevel data are often obtained as an incomplete dataset with missing components at any level of the data hierarchy. For example, in practice, a nesting of some hospitals and practices, which are also nested within larger health systems, can be ignored or unobserved in the health care study [7].

In practice, researchers in the field ignore a hierarchy within the data and prefer an ordinary least square regression model (OLS) that treats all observations as if they are measured at the same level [8-12]. Alternatively, a two-level model is also commonly adopted by ignoring missing intermediate levels and accounting for only the top and bottom levels of hierarchy in multilevel data. However, the misleading hierarchy by ignoring intermediate or top levels may potentially impact the parameter estimation in multilevel analysis

and has gained attention among researchers in recent research [13-17].

To handle the missing issues, several studies proposed imputation methods to fill in the missing and create a complete dataset [15-17]. However, imputation methods add more uncertainty to the parameter estimations and especially increase the complexity when the imputation is intertwined with the hierarchical structure of the data [15-18]. In particular, it is difficult to handle the missing when a cluster is entirely missing or the cluster size is very small, and especially the missing values are in the explanatory variables. Alternatively, a single imputation can be considered. It is because it handles the practical issues in multilevel imputation, is simple to adapt, and preserves the design hierarchy of the multilevel data. In this study, we consider a multilevel linear mixed effect model (LMM) with a single imputation that was introduced in Sanders' study [19] for handling missing intermediate or top clusters in multilevel data. An LMM has been widely accepted to model the hierarchical structure in multilevel data and can account for the correlation among units nested within clusters. The method uses a single imputation that replaces the missing clusters by its nesting lower-level unit's measurements and considers as if each of the nesting lower-level unit's measurements is a single observation for missing clusters. As a result, a missing intermediate-level cluster is filled in containing only a singleton, resulting in a complete dataset [19]. However, to our best knowledge, no previous studies have examined and investigated the performance of an LMM with a single imputation and directly compared it with other models that ignore missing levels of nesting.

This study aims to evaluate the performance of a model considering all hierarchies, such as an LMM with single imputation, in three-level data and demonstrate that it outperforms other models that do not match the hierarchy of multilevel data. To do so, we compare three models in three-level hierarchical data with missing intermediate-level clusters. The first model is an OLS regression model, which is a single-level model ignoring any hierarchy in the data and treats all observations measured at the same level. Secondly, we consider a two-level LMM considering only level-1 and level-3 and discarding missing intermediate levels. The third model is a three-level LMM with a single imputation that involves all levels of the data hierarchy, so we can assume that the multilevel model matches the design hierarchy in the data by filling in a missing level-2 unit.

In this study, we apply and compare the three models to the Childhood to Adolescence Transition Study (CATS), a motivating case study of this work, in a three-level structure: school (level-3 unit), individual (level-2 unit), and repeated measures observed

per individual (level-1 unit) with missing level-2 and level-3 units [20,21]. We also conduct a thorough simulation study by comparing the three models to examine the impact of ignorance of missing intermediate level on parameter estimation of three-level data analysis when the missing rate and the cluster sizes of level-2 and level-3 clusters vary.

Methods

Three-level linear mixed effect model

In the context of multilevel data analysis, a LMM is considered as an analysis model that accounts for correlation due to the hierarchical structure of the data. To establish background information on the analysis model for hierarchical structure data, especially for a three-level data structure, we develop a general notation and introduce a brief overview of a three-level model.

Let i be the index of the level-3 unit with size L ($i=1, \dots, L$), j be the index of the level-2 unit with size M_i ($j=1, \dots, M_i$), and k be the index of the level-1 unit with size N_{ij} ($k=1, \dots, N_{ij}$). We consider a three-level model, where a level-1 unit is nested within a level-2 unit that is also nested within a level-3 unit. Let y_{ijk} be the response variable for level-1 unit k in the level-2 unit j and level-3 unit i , x_{ijk} be the associated covariate variable observed at level-1, x_{ij} be the associated covariate variable at level-2, and x_i be the associated covariate at level-3. Then a general three-level linear mixed effect model is:

$$\begin{aligned} y_{ijk} &= \beta_{0ij} + \beta_{1ij}x_{ijk} + e_{ijk} \\ \beta_{0ij} &= \alpha_{0i} + \alpha_{01}x_{ij} + v_{ij} \\ \beta_{1ij} &= \alpha_{1i} + \alpha_{11}x_{ij} + v'_{ij} \\ \alpha_{0i} &= \gamma_{00} + \gamma_{01}x_i + u_i \\ \alpha_{1i} &= \gamma_{10} + \gamma_{11}x_i + u'_i, \end{aligned} \tag{1}$$

where $e_{ijk} \sim N(0, \sigma_e^2)$ are residuals at level-1 that are independent and identically distributed (*iid*) and capture within-cluster residual variations, $v_{ij} = (v_{ij}, v'_{ij})$ are *iid* random effects at level-2 and captures between-cluster intercept differences at level-2, $u_i = (u_i, u'_i)$ are *iid* random effects at level-3 and captures between-cluster intercept differences at level-3.

In this study, we assume that only the regression intercept varies across clusters and hence restrict our attention to a three-level random effect model with a random intercept for each level. It is the simplest type of linear mixed model for multilevel study, although we note that an extended model including random slope can be considered. We revisit this in the discussion. Considering a ran-

dom effect model with random intercepts only, $\alpha_{11}=\nu'_{ij}=\gamma_{11}=u'_i=0$. Therefore, the model in Eq. (1) can be rewritten as

$$y_{ijk} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{ij} + \beta_3 x_{ijk} + u_i + v_{ij} + e_{ijk}, \tag{2}$$

where $\beta_0(=\gamma_{00})$ is the intercept, and $\beta_1(=\gamma_{01}), \beta_2(=\alpha_{01})$, and $\beta_3(=\gamma_{10})$ are the slope coefficients for the fixed effects at level-3, level-2, and level-1, respectively. Considering random intercepts only in the model, we assume that level-3 and level-2 residuals, u_i and v_{ij} , respectively, are multivariate normal with zero means and variances of σ_u^2 and σ_v^2 , respectively, and the within-cluster residuals are normally distributed with mean zero and constant variance σ_e^2 .

$$u_i \sim N(0, \sigma_u^2), v_{ij} \sim N(0, \sigma_v^2), e_{ijk} \sim N(0, \sigma_e^2).$$

The covariance structure of the response vector $y=(y_{111}, y_{112}, \dots, y_{L})$, $M_L, N_{LM})$ is a block-diagonal covariance matrix given by

$$Cov(y) = \begin{bmatrix} G_1 + \sigma_u^2 J_{(n1)} & & \\ & \ddots & \\ & & G_L + \sigma_u^2 J_{(nL)} \end{bmatrix}$$

with an i th block-diagonal matrix of $G_i = \sigma_e^2 I_{(ni)} + \sigma_v^2 J_{(ni)}$, where $I_{(ni)}$ is an $n_i \times n_i$ identity matrix, $J_{(ni)}$ is an $n_i \times n_i$ matrix of 1s, and $n_i = M_i \times N_{imi}$ is a size of a block-diagonal matrix associates with i th level-3 cluster. The off-diagonal elements are zero.

Analysis of incomplete multilevel data using single imputation

To handle the missing level-2 units in a three-level hierarchical data, we adopt a single imputation method introduced in Sanders’ study [19] that replaces the missing clusters by its nesting lower-level unit’s measurements and considers the imputed level-2 unit as a cluster containing a single observation. Specifically, it constructs imaginary level-2 clusters for level-1 measurements whose level-2 clusters were missing entirely. As a result, the single imputation method adds level-2 clusters as many as the number of level-1 units directly under the missing level-2 clusters. By applying the method, we can have a complete three-level dataset without any missing clusters at level-2 units, allowing us to apply a three-level model. The single imputation method allows us to preserve a hierarchical data structure in multilevel data analysis.

The motivating case study CATS has a three-level structure with the components of school, individual, and its repeated measures per individual, and some measurements at the individual level (level-2; academic numeracy score at baseline) were partially not

observed and recorded as missing in the dataset. There were 22.4% missing academic numeracy scores at baseline at level-2, and they were filled in by single imputation that impute a corresponding academic numeracy score measured at level-1 to handle the components with missingness. In other words, we treated each level-1 unit within a missing level-2 unit is nested in a level-2 unit that only contains the level-1 unit itself [19].

Childhood to Adolescence Transition Study

Motivated by the case study of CATS, we focus on an analysis of the simulated data mimicking the CATS data [21]. The CATS is a longitudinal cohort study collected through multiple waves, and it collects mental health to investigate the effect of early depressive symptoms on academic outcomes in children from puberty through adolescence. More details about the data collection and study protocol can be found in Mundy et al. [20].

The simulated data mimicking the CATS have a three-level hierarchical structure. Individuals are nested within schools, and repeated measures within individuals were collected. The data consist of demographic, educational, and social outcomes as well as mental health outcomes for the CATS study: depressive symptoms, National Assessment Programme – Literacy and Numeracy results (NAPLAN) academic numeracy score, socio-economic status, age, and sex are collected for the study [21].

The NAPLAN numeracy score is an outcome of interest and is observed at level 2 at wave 1 (potential baseline) and at level 1 in the following waves with missing cases. As we are interested in a case where missingness happens at level-2 only, we discard cases where the NAPLAN numeracy score has missingness at level-1, and the size of data that we used is $N = 2,592$. The data are unbalanced as the size of level-2 differs, and the number of level-1 units within a level-2 unit varies across individual. The dataset has 163 schools; 54 of 163 schools have a single individual with a single measurement per individual, and the other 109 schools have multiple individuals with a different number of repeated measures. There were 1,142 individuals in the data, and 256 of 1,142 (22.4%) NAPLAN numeracy scores at the individual level are missing.

Simulation data

We conducted a simulation study to evaluate the performance of a three-level LMM with single imputation. In the simulation, we considered complete three-level data sets and incomplete data with various missing rates of level-2 clusters (Table 1) and compared the performance of the three-level LMM with single imputation with those of a single-level model and 2-level LMM.

We generated complete data sets without missing subjected to

Table 1. Simulation schemes to generate complete 3-level data and incomplete data with missing in level-2 units**Complete 3-level data sets**

Given level-3 cluster size L ($= 30, 50, 100$) and level-2 cluster size M ($= 10, 30$),

Generate level-1 cluster size $N_{ij} \sim U_{unif}[15,40]$.

Given N_{ij} , generate a data set $(y_{ijk}, x_{ik}, x_{jk}, x_{ijk})$ from the model in Eq. (2).

Incomplete 3-level data sets with missing in the level-2 units

Given each complete data set and missing rate p ($= 10, 25, 50, 75\%$).

Randomly select and discard $\times p_{x_{ij}}$'s from the complete data set.

Then an incomplete data set consists of $M \times p$ observations of $(y_{ijk}, x_{ik}, x_{jk})$ and $M \times (1-p)$ observations of $(y_{ijk}, x_{ik}, x_{jk}, x_{ijk})$.

We note that complete 3-level data sets can be considered as the case of a 0% missing rate. In total, we considered 30 simulation scenarios (three level-3 cluster sizes \times two level-2 cluster sizes \times five missing rates with a range of 0%, 10%, 25%, 50%, and 75%).

the model in Eq. (2), where we assumed the fixed coefficients of $\beta_0 = 0.5, \beta_1 = \beta_2 = \beta_3 = 0.3$, the random effects at level-2 and level-3 of $u_i \sim N(0, \sigma_u^2)$ and $v_{ij} \sim N(0, \sigma_v^2)$ with $\sigma_u^2 = 1.4$ and $\sigma_v^2 = 1$, respectively, and $e_{ijk} \sim N(0, \sigma_e^2)$ with $\sigma_e^2 = 1.4$ for $i = 1, \dots, L, j = 0, 1, \dots, M_{\square}$ and $k = 1, \dots, N_{ij}$. We considered various cluster sizes at each level (Table 1). The number of level-3 clusters varied as $L=30, 50$, and 100 . For each level-3 cluster, we assumed $M=10, 30$ clusters. The number of level-1 clusters within each level-2 cluster was randomly determined as $N_{ij} \sim Unif[15,40]$. For all cases, the covariates were generated as $x_i \sim N(0,1)$, $x_{ij} \sim \exp(1)+1$, and $x_{ijk} \sim N(1, 1.2^2)$. Intraclass correlations were set to 0.398 and 0.602 for level-3 and level-2, respectively. For each case, we generated 500 replicates.

To evaluate the performance of the three-level LMM in the presence of missingness at the level-2 units, we generated missingness from each of the complete 3-level data sets (Table 1). We set the complete data to be missing in M level-2 clusters and its related variables such as x_{ij} . The missing data rate in level-2 clusters was set to 10, 25, 50, and 75% of M level-2 clusters.

Results

Childhood to Adolescence Transition Study

We use three analysis models to determine the effect of ignorance of hidden levels in the hierarchically structured CATS data: (1) a single-level model (ordinary least square regression) that completely ignores the structure of the hierarchy and treats the data collected at level-1 only (M1), (2) a two-level LMM with a random effect for school (level-3) only and ignoring an individual (level-2) random effect (M2), and (3) a three-level LMM with single imputation considering all levels of hierarchy in the CATS study units (M3). The LMMs we fit for M2 and M3 only have random intercepts, and random slopes are not considered in the data analysis. Tables 2 and 3 report estimates of the fixed and ran-

dom effects, respectively, across the three models.

Table 2 presents estimated coefficients of fixed effects in each model, and Table 3 presents the estimated variance components of random effects and residuals. Note that M2 has one random effect at school and M3 has two random effects at school and individual.

Based on a significance level of 5%, depressive symptoms appeared to be significantly meaningful in supporting the effect of mental health on academic outcomes in children from puberty through adolescence across the three models ($p < 0.001$). Age and NAPLAN numeracy scores observed at baseline were meaningful covariates for explaining the relationship with academic outcome in addition to the mental health outcome. That is, when children are young, have fewer depressive symptoms, and higher NAPLAN numeracy scores at baseline, academic outcomes are likely to be higher. Noticeably, the estimated coefficients of the important covariates do not differ much across the three analysis models.

Table 3 presents the estimated variance components of random effects and residuals in the multilevel data across the three analysis models with different hierarchy levels. The estimated residual variance quantifies the variation within repeated measures and decreases when random effects are included in the analysis model (M2 and M3). The estimated variance components of the random effects explain variations across schools and individuals that were not explained by M1, and the sum of the estimated variance components in M3 is larger than in M2. This implies that M3 captures more unexplained variation in the hierarchical structure data, especially the variation among individuals, which cannot be ignorable. Hence, the analysis without accounting for the variance among the unobserved level of nesting units in data analysis resulted in a larger estimation of the residual variance as shown in Table 3. It might lead to an inaccurate estimation of the total variation in the multilevel data.

Table 2. Estimates of coefficients for CATS data inferred by the three models: single-level model (M1), two-level LMM (M2), and three-level LMM (M3)

	M1			M2			M3		
	Coef	SE	p-value	Coef	SE	p-value	Coef	SE	p-value
(Intercept)	2.789	0.225	<0.001	2.744	0.212	<0.001	2.748	0.234	<0.001
Depressive symptom ^a	-0.043	0.013	<0.001	-0.047	0.012	<0.001	-0.048	0.010	<0.001
Age ^a	-0.173	0.023	<0.001	-0.172	0.022	<0.001	-0.173	0.025	<0.001
Sex ^a	0.037	0.041	0.371	0.018	0.038	0.633	0.021	0.044	0.901
SES ^a	0.019	0.021	0.360	0.011	0.019	0.579	0.006	0.022	0.764
NAPLAN at baseline ^b	0.677	0.021	<0.001	0.675	0.019	<0.001	0.676	0.021	<0.001

CATS, Childhood to Adolescence Transition Study; SE, standard error; SES, socio-economic status; NAPLAN, National Assessment Programme – Literacy and Numeracy results.

^aCovariate at level-1.

^bCovariate at level-2.

Table 3. Estimated variance components of random effects and residuals for CATS data inferred by the three models: single-level model (M1), two-level LMM (M2), and three-level LMM (M3)

Variance component	M1	M2	M3
Level 3: school	-	0.216	0.137
Level 2: individual	-	-	0.325
Residual	1.002	0.786	0.426

CATS, Childhood to Adolescence Transition Study; LMM, linear mixed effect model.

Simulation data

Considering 30 scenarios (Table 1), we applied a single-level model (M1) and 2-level LMM (M2) and 3-level LMM with single imputation (M3). M1 and M2 estimate four fixed coefficients, and the standard deviation σ_e^2 of the error term but does not estimate some of the variances of the random effects, while M3 estimates all parameters described in Eq. (2). Since M1 ignores the hierarchical structure of the data, it does not contain the random effect terms of u_i and v_{ij} . M2 considers level-1 and level-3 clusters by assuming that level-1 units are directly connected to level-3 clusters. Therefore, M2 does not include the random effect term of v_{ij} . In summary, M3 only estimates σ_e^2 , and M2 and M3 models estimates σ_u^2 . The estimates of all fixed parameters were compared to the true values of the parameters used to simulate the data. In particular, we calculated the mean squared error (MSE) across 500 replications, and the coverage probability was estimated by the proportion of replications with 95% confidence intervals containing the true value.

We expected that M3 performed well in the presence of unobserved level-2 clusters, whereas M2 would poorly estimate the coefficient β_2 of the level-2 associated covariate with missing clusters since M2 ignored the level-2 clusters and partially incorporated the date structure of level-1 and level-3 clusters. Since the M1 did

not consider the hierarchical data structure, we expected that it poorly estimated β_1 and β_2 .

In general, the M3 performed well and better estimated the fixed coefficients in Eq. (2) than the M1 and M2 models. As expected, in particular, we observed that the M1 and M2 models poorly estimated the coefficient β_2 of the level-2 associated covariate with missing clusters. Figs. 1 and 2 show the MSEs and the coverage probabilities for β_2 , respectively, and compared those values by the three models. The MSEs by the M3 ranged from 0.00039 to 0.00557, while those by the M1 and M2 were 0.0011–0.0639 and 0.0004–0.037, respectively (Fig. 1). The coverage probabilities for β_2 by the M3 were approximately 0.814–0.964 (Fig. 2). The M1 and M2 models did poorly estimate β_2 and their coverage probabilities low as 0.27–0.406 and 0.392–0.498, respectively (Fig. 2). The MSEs for β_2 seemed to increase with the missing rate (Fig. 1); however, the coverage probabilities did not change as the missing rate increased.

We observed MSEs for β_0 , β_1 , and β_3 similar to Fig. 1. The differences in MSEs among the three models were small, but M3 performed the best overall. Moreover, MSEs tended to increase with missing rates. Similar to Fig. 2, the coverage probabilities for β_0 and β_1 were around 0.95 from the M3, while we observed small coverage probabilities from M1 (e.g., 0.06–0.302 for β_1) and comparable coverage probabilities from M2 (e.g., 0.926–0.96 for β_1). The three models yielded similar coverage probabilities for β_3 .

The three models contain different variance components. M3 can estimate the three standard deviations, σ_u^2 , σ_v^2 and σ_e^2 . However, M2 does not estimate the variance σ_v^2 among level-2 clusters, and M1 can estimate the variance σ_e^2 among individuals but not σ_u^2 and σ_v^2 . We expected that the M1 and M2 models could overestimate variance components by failing to explain the 3-level data structure.

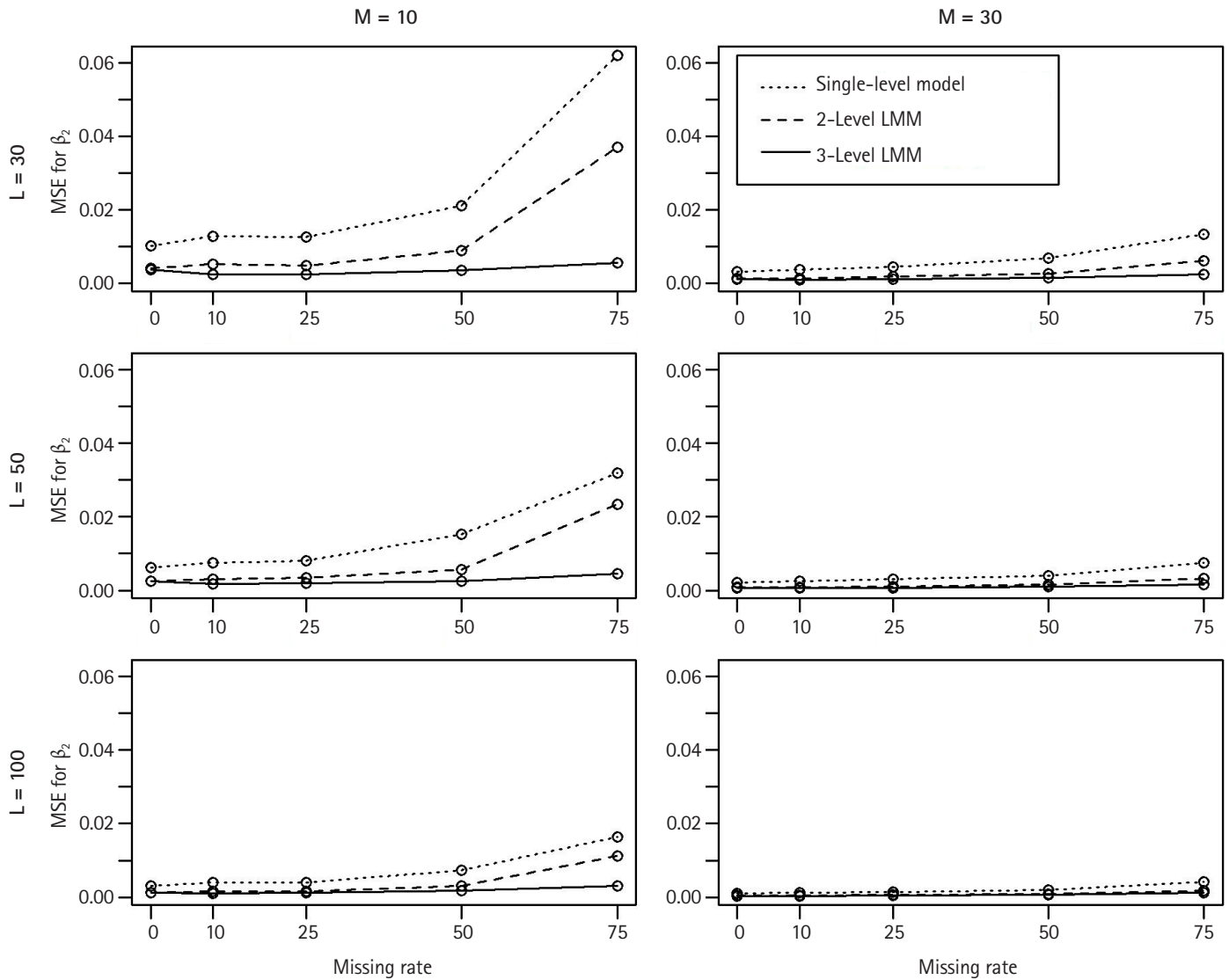


Fig. 1. The MSEs for β_2 , the coefficient of the level-2 associated covariate, when estimated by the single-level model (···), 2-level LMM (---), or 3-level LMM (—). For each plot of L ($=30, 50, 100$) level-3 clusters and M ($=10, 30$) level-2 clusters, each point represents the MSE across 500 simulated data sets over a range of missing rates in level-2 clusters. MSE, mean squared error; LMM, linear mixed effect model.

Figs. 3 and 4 show that the MSEs and coverage probabilities for σ_v^2 from M3 worsened as the missing rate increased for each case of L and M values. At the same missing rate, the larger value of M , the better the coverage. However, more level-3 clusters did not improve MSE and coverage probability. This is because each level-3 cluster contains unobserved level-2 clusters, and hence the number of unobserved level-2 clusters increases as L increases.

We observed that the estimates for σ_u^2 and σ_e^2 from M1 were unbiased with small MSEs and their coverage probabilities were close to the nominal level of 95% (0.918–0.96 for σ_u^2 ; 0.936–0.974 for σ_e^2). However, the coverage probabilities of M1 and M2 models were very low (e.g., 0–0.344 for σ_u^2 and 0–0.006 for σ_e^2 from the

2-level model), and σ_u^2 and σ_e were overestimated by M1 and M2 models (see Fig. 5 for the average estimates for σ_e).

Discussion

In this study, we investigated the effect of missing intermediate level of the hierarchical structure data with varying missing rates, as well as the consequence of ignorance of the level of nesting on the parameter estimation of multilevel analysis. Since missing level-2 units are imputed by the nested level-1 unit’s measurement, the intermediate-level unit has a single observation per unit, and the data can be considered sparse. We compared three models with differ-

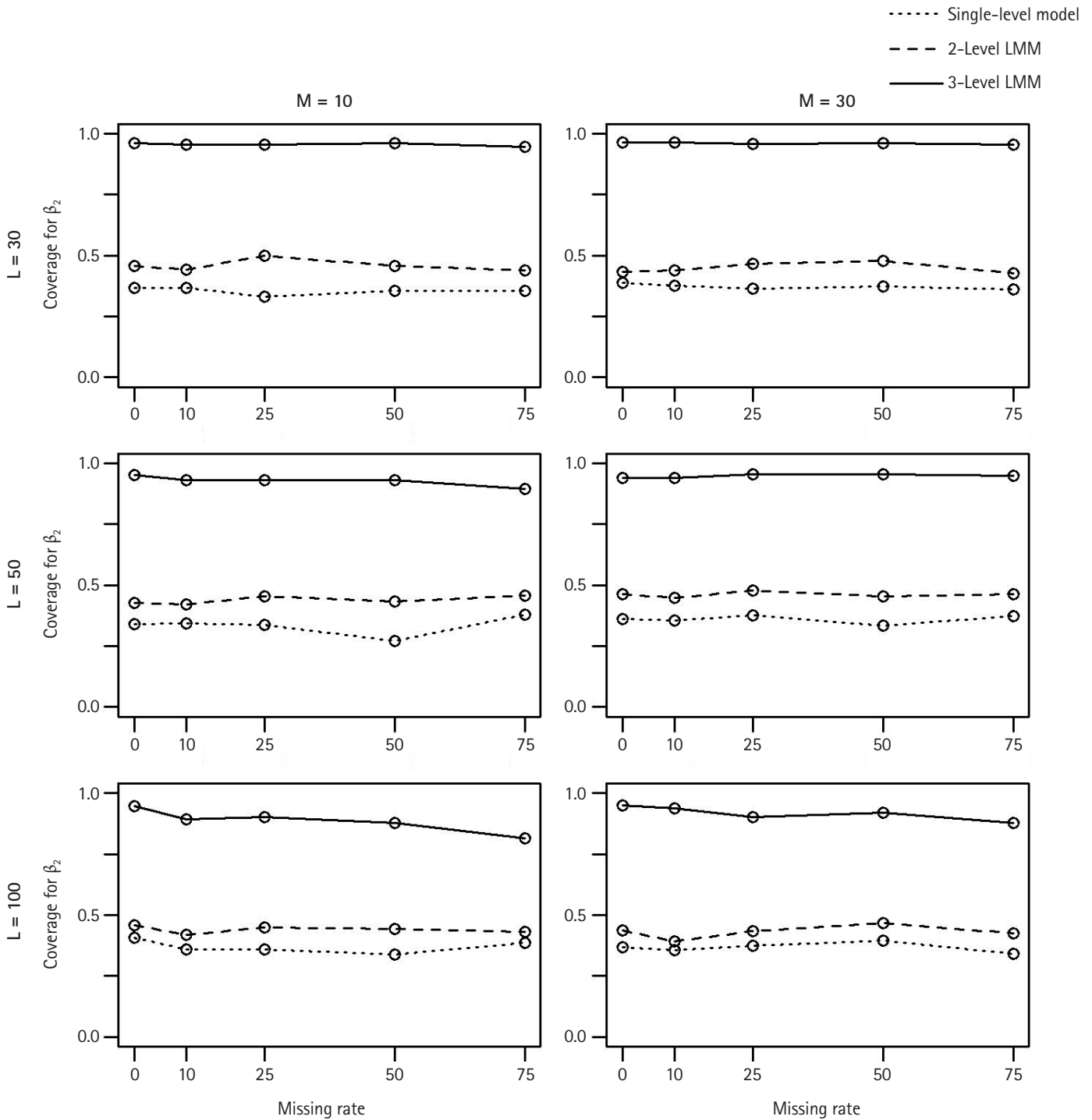


Fig. 2. The coverage probability for β_2 , the coefficient of the 2-level associated covariate, when estimated by the single-level model (···), 2-level LMM (---), or 3-level LMM (-). For each plot of L (=30, 50, 100) level-3 clusters and M (=10, 30) level-2 clusters, each point represents the coverage probability across 500 simulated data sets over a range of missing rates in level-2 clusters. LMM, linear mixed effect model.

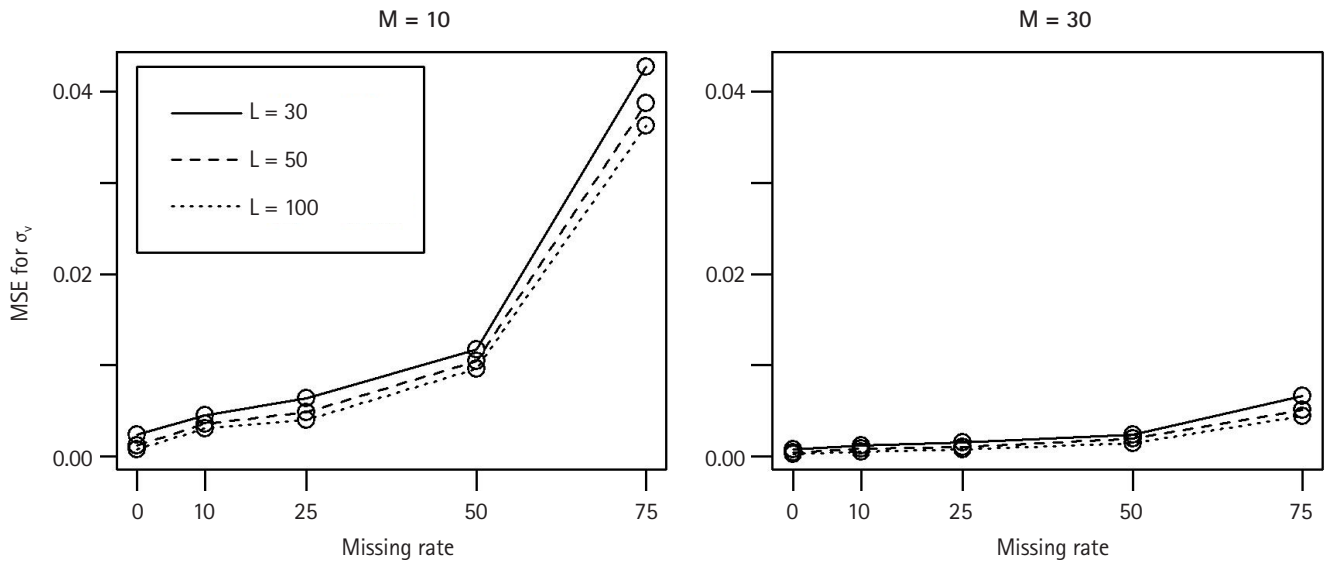


Fig. 3. The MSEs for σ_v from the 3-level LMM (M3) when the level-3 cluster size changes as $L = 30$ (—), $L = 50$ (---), or $L = 100$ (· · ·). For each case of $M = 10$ or 30 level-2 cluster sizes, 500 datasets were simulated over a range of missing rates in level-2 clusters. MSE, mean squared error; LMM, linear mixed effect model.

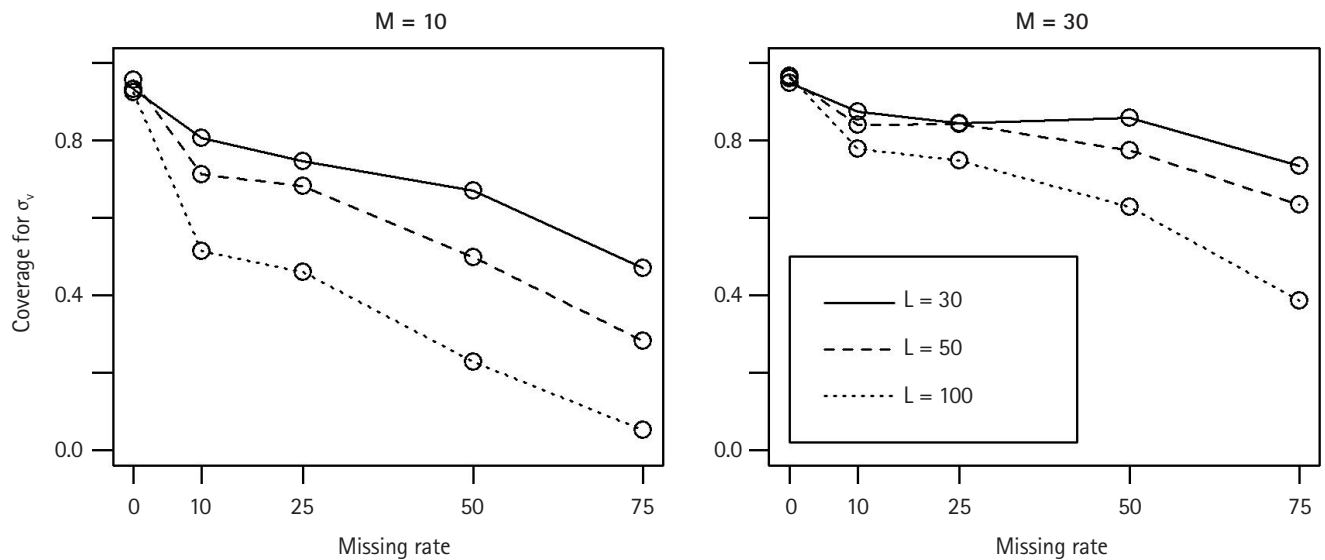


Fig. 4. The coverage probability for σ_v from the 3-level LMM (M3) when the level-3 cluster size changes as $L = 30$ (—), $L = 50$ (---), or $L = 100$ (· · ·). For each case of $M = 10$ or 30 level-2 cluster sizes, 500 datasets were simulated over a range of missing rates in level-2 clusters. LMM, linear mixed effect model.

ent levels of hierarchy across varying missing rates at the intermediate-level and various cluster sizes in the top- and intermediate-level clusters. We observed that the three-level LMM with single imputation showed a better performance compared to the other two models, which ignore higher levels of nesting in terms of MSE and coverage. Moreover, lower-level variance components were overestimated, indicating that variance components are af-

fected by ignorance of the intermediate level.

We considered a random effect model for the 2-level and 3-level linear mixed effect models in the simulation study and data application for brevity, and it can be expanded to consider a random slope model and/or interaction terms. However, it requires caution to generalize the model analysis to be more complex because we have a large number of intermediate-level clusters with a single

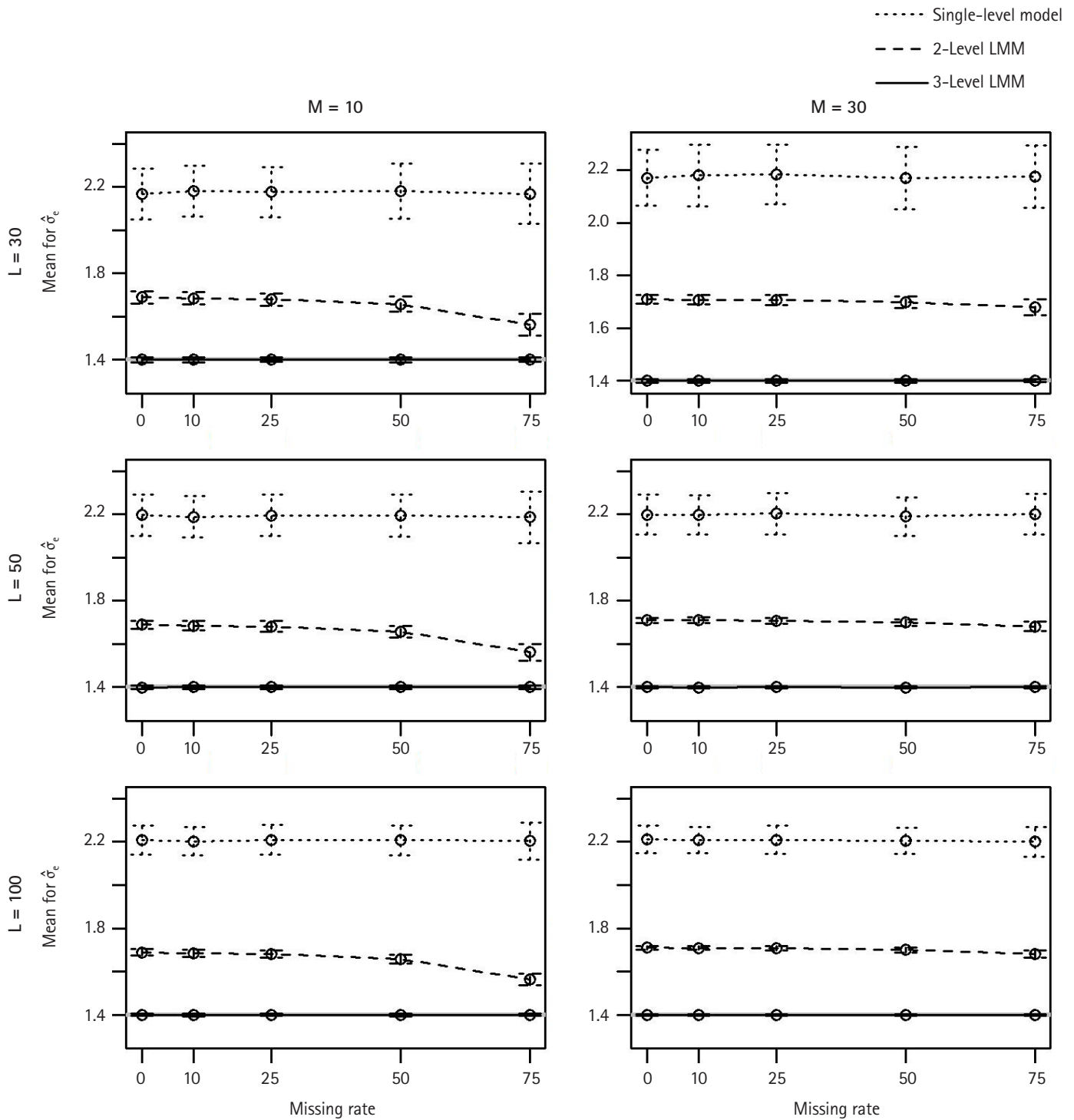


Fig. 5. The average estimates for σ_e when estimated by the single-level model (---), 2-level LMM (---), or 3-level LMM (-). For each plot of L (=30, 50, 100) level-3 clusters and M (=10, 30) level-2 clusters, each point represents the mean of estimates across 500 replicates, and vertical lines indicate the standard errors. The true value σ_e is 1.4. LMM, linear mixed effect model.

observation, and such sparse clusters might lead to biased estimates in random slope estimation [22]. Also, we have incomplete variables for intermediate-level units. An imputation requires careful caution when incomplete explanatory variables with a random slope and/or interaction terms exist in the model because they might lead to invalid estimates [10].

Furthermore, we used a three-level LMM with single imputation that replaces the missing by the nesting lower-level unit's measurements and considering that a single observation is measured per intermediate-level unit. As a future study, other imputation methods for multilevel data can be considered and compared their performances with an approach studied in this work. In practice, it is common that level-1 and level-2 units are both missing, and in such case, the imputation method, such as a multiple imputation handling multiple level's missingness, should be considered.

ORCID

Seho Park: <https://orcid.org/0000-0002-4750-1916>

Yujin Chung: <https://orcid.org/0000-0002-3801-3961>

Authors' Contribution

Conceptualization: SP. Data curation: SP, YC. Formal analysis: SP, YC. Funding acquisition: YC. Methodology: SP. Writing - original draft: SP, YC. Writing - review & editing: SP, YC.

Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2021R1C1C1011250).

References

- Greenland S. Principles of multilevel modelling. *Int J Epidemiol* 2000;29:158-167.
- Bryk AS, Raudenbush SW. Hierarchical Linear Models. Newbury Park: Sage Publications, 1992.
- Searle SR, Casella G, McCulloch CE. Variance Components. New York: John Wiley & Sons Inc., 1992.
- Goldstein H. Multilevel Statistical Models. 2nd ed. London: Edward Arnold, 1995.
- Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian Data Analysis. New York: Chapman and Hall, 1995.
- Kreft IG, de Leeuw J. Introducing Multilevel Modelling. London: Sage Publications, 1998.
- O'Malley AJ, Park S. A novel cluster sampling design that couples multiple surveys to support multiple inferential objectives. *Health Serv Outcomes Res Methodol* 2020;20:85-110.
- Clarke P. When can group level clustering be ignored? Multilevel models versus single-level models with sparse data. *J Epidemiol Community Health* 2008;62:752-758.
- Marston L, Peacock JL, Yu K, Brocklehurst P, Calvert SA, Greenough A, et al. Comparing methods of analysing datasets with small clusters: case studies using four paediatric datasets. *Paediatr Perinat Epidemiol* 2009;23:380-392.
- Moineddin R, Matheson FI, Glazier RH. A simulation study of sample size for multilevel logistic regression models. *BMC Med Res Methodol* 2007;7:34.
- McNeish DM, Stapleton LM. The effect of small sample size on two-level model estimates: a review and illustration. *Educ Psychol Rev* 2016;28:295-314.
- Stegmuller D. How many countries for multilevel modeling? A comparison of frequentist and Bayesian approaches. *Am J Polit Sci* 2013;57:748-761.
- Hill PW, Goldstein H. Multilevel modeling of educational data with cross-classification and missing identification for units. *J Educ Behav Stat* 1998;23:117-128.
- Goldstein H, Carpenter JR. Multilevel multiple imputation. In: Handbook of Missing Data Methodology (Molenberghs G, Fitzmaurice G, Kenward MG, Tsiatis A, Verbeke G, eds.). Boca Raton: CRC Press, 2015. pp. 295-316.
- Ludtke O, Robitzsch A, Grund S. Multiple imputation of missing data in multilevel designs: a comparison of different strategies. *Psychol Methods* 2017;22:141-165.
- Grund S, Ludtke O, Robitzsch A. Multiple imputation of missing data for multilevel models: simulations and recommendations. *Organ Res Methods* 2018;21:111-149.
- Drechsler J. Multiple imputation of multilevel missing data: rigor versus simplicity. *J Educ Behav Stat* 2015;40:69-95.
- Van Buuren S. Multiple imputation of multilevel data. In: Handbook of Advanced Multilevel Analysis (Hox JJ, Roberts JK, eds.). New York: Routledge, 2011. pp. 173-196.
- Sanders EA. Multilevel analysis methods for partially nested cluster randomized trials. Ph.D. Dissertation. Seattle: University of Washington, 2011.

20. Mundy LK, Simmons JG, Allen NB, Viner RM, Bayer JK, Olds T, et al. Study protocol: the Childhood to Adolescence Transition Study (CATS). *BMC Pediatr* 2013;13:160.
21. Wijesuriya R, Moreno-Betancur M, Carlin JB, Lee KJ. Evaluation of approaches for multiple imputation of three-level data. *BMC Med Res Methodol* 2020;20:207.
22. Bell BA, Ferron JM, Kromrey JD. 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*. Alexandria: American Statistical Association, 2008. pp. 1122-1129.