Impact of inconsistent imputation models in mediation analysis with clustered data

Ye, Bo* Yucel, Recai†

October 7, 2021

Abstract

The choice of imputation model is often driven by the focus of the post imputation analyses. However, this can be impractical especially when multiple imputation (MI) inference is used in large-scale surveys or in applications with complex data structures. This work investigates the impact of an imputation model on the mediation analysis with clustered data. Specifically, we consider joint and variable-by-variable imputation models leading up to multi-level mediation analysis. We provide theoretical and analytical assessment of the bias under each imputation method. A comprehensive simulation study is conducted to understand the performance of imputation methods in a repetitive sampling framework.

Key Words: multi-level mediation analysis, missing data, multiple imputation, compatible imputation

1 Introduction

Missing data can easily complicate most analyses in many disciplines. Since the seminal work by Dempster et al. (1977) missing data have been subject to many research topics not only in statistics but also in many subject-matter areas. In particular, inference by multiple imputation (MI) has been increasingly the norm in analysis of incomplete data Rubin (1987). Another type of solution makes use Expectation-Maximization (EM) type solutions leading to maximum-likelihood estimation in problem-specific settings. In this work, we focus on how imputation model choice influences multilevel mediation analysis.

There are methods available in simpler problems of cross-sectional data that pertains to joint modeling (JM) and fully conditional specification (FCS) modeling. There is a relatively short history of applying JM and FCS to multilevel data, and less is known about the performance of JM and FCS. Van Buuren (2011) extend the application of FCS to multi-level data based on univariate linear mixed models. Carpenter & Kenward (2013) incorporate cluster-level variable means as covariate to FCS imputation model. Schafer & Yucel (2002) apply JM to multi-level missing data problem based on multivariate linear mixed model.

Some researchers have conducted studies on the application of MI on multilevel dataset. Multilevel imputation methods incorporate the clustering effects into the imputation process. Imputation models that accounts for random effects can appropriately handle multilevel data with missing data problem. Grund et al. (2018) apply MI to multilevel data and summarize the comparison result of imputation models. They suggest that MI provide an effective treatment of missing data in multilevel research. Mistler & Enders (2017) examine four multilevel multiple imputation approaches and conclude that their analytic work and computer simulations show that FCS is more restrictive by imposing implicit equality constraints on functions of the within- and between-cluster covariance matrices.

*Department of Epidemiology and Statistics, School of Public Health, State University of New York at Albany

†Department of Epidemiology and Biostatistics, College of Public Health, Temple University

Multi-level mediation analysis model is also impacted by missing data problem. Krull & MacKinnon (2001) and other researchers introduce multi-level mediation model to appropriately assess mediation effects in of a mediator in mediating the relationship between independent variable and response variable for clustered data. However, limited researches have been conducted to analyze the impact of imputation models on multi-level mediation analysis. In single-level mediation analysis model, Zhang and Wang (2013) introduce and compare four approaches to deal with missing data in single-level mediation analysis including listwise deletion, pairwise deletion, multiple imputation by Rubin (1976), and a two-stage maximum likelihood (TS-ML) method under various missingness mechanisms through simulation studies. They show that MI performed well when missingness mechanism is independent of any of the observed variables (i.e. missing completely at random (MCAR)) or when the missingness mechanism depends only on the observed variables but not on the missing variables (missing at random (MAR)) in the sense defined by Rubin (1976). Milletich II (2018) introduce bayesian bootstrap method to generate posterior inferences for mediation effect and impute missing data using linear and logistic regression models. They incorporate bayesian boostrap (BB) method with MI in the study, and claim it has great performance in imputation. However, they do not extend their research or work to multi-level mediation analysis model. Ye & Yucel (2020) compare the performance of imputation models (JM and FCS) for single-level mediation analysis model, and find JM and FCS are similar in estimating mediation effect. However, the same result is not guaranteed in the multi-level mediation analysis model structure. The multi-level mediation analysis model is constructed by multiple univariate mixed model equations.

Many of the developments so far on MI have been made available to practitioners for multi-level data. A joint modeling approach for clustered data has been developed by Schafer & Yucel (2002) for linear mixed-effect model with an R package, "Pan". Multilevel joint modeling draw imputation from a multivariate mixed model. Van Buuren (2011) develop the R package MICE which implements a wide variety of algorithms under variable-by-variable imputation and provides additional options to draw from the potential conditional predictive distributions. Multilevel variable-by-variable imputation employs multiple draw imputation for one incomplete variable at a time based on univariate mixed model. There are also other programs available in other platforms. Raghunathan, Lepkowski, VanHoewyk, and Solenberger (2001) developed a SAS macro, "IveWare", for the application of variable-by-variable imputation in SAS. In STATA, we have the procedure "ICE" by Royston and White (2011) to implement multiple imputation method for missing data problem.

In this paper, we focus on analytically comparing joint model and variable-by-variable imputation methods in multilevel mediation analysis with missing data problem. This study starts by introducing how to estimate multilevel mediation effects, followed by how imputation method works in this model. Then, the analytical comparison of two methods is provided. Finally, we conduct a simulation study with finite samples to assess and compare the performance of the methods under varying scenarios (different missingness rate, mediation effect size, missingness mechanism and intracluster correlation coefficient (ICC)).

2 Methods

2.1 Notation

Suppose K random variables D = $(D_1, ..., D_K)^T$ are intended to be observed on N subjects with missing values. Subscripts p and q are used to index subjects and variables respectively $(p = 1, ..., N;\ q = 1, ..., K)$. Let d_{pq} denote an p^{th} row and q^{th} column element in an (N*K) matrix. We denote the column q of matrix d by d_q=$(d_{1q}, ..., d_{Nq})^T$. We denote the complete data matrix U = (u_{pq}), the missing data indicator matrix R = (R_{pq}) and ϕ as the unknown parameters in the section of introducing missingness mechanism. In multilevel mediation analysis model, Y is denoted as the response variable, M is denoted as the mediator, and X is denoted as the independent variable. Let y_{ij} denote a value of a random variable Y for subject $j = 1, 2, ..., n_i$ in cluster $i = 1, 2, ..., m$. For example, the independent variable X_{ij} indicates the value for j^{th} subject at i^{th} hospital. Three variables (X_{ij}, M_{ij} and Y_{ij}) are included in the data matrix D. η is the matrix includes M and Y, δ represents mediation effect, β is the matrix of regression coefficients on mediator. For the purpose of missing data imputation, we will have to model all variables regardless of covariates and response as variables to be modeled.

2.2 1-1-1 Multilevel Mediation Analysis Model

In this study, the analysis is based on a specified multilevel mediation analysis model. Multilevel mediation analysis examines the indirect effect (mediation effect) of an independent variable on response variable by a mediator. Different from single-level mediation analysis model, we consider interventions that occurs in cluster, such as hospitals, schools and etc (Ellickson et al. (2003)) in multilevel mediation model. The aim of the multilevel mediation analysis is to determine whether the relationship between the independent variable and the response variable is mediated by the mediator. We are considering a two-level 1-1-1 mediation model example (Krull & MacKinnon (2001)), in which level 1 are nested within level 2. In the $1-1-1$ multilevel mediation model, the independent variable (X_{ij}), mediator (M_{ij}), and response variable (Y_{ij}) are all evaluated at level 1 (individual-level).The ij subscript on each variable refers to an individual j within i group. We focus on $1-1-1$ multilevel mediation analysis model in this study.

We present a typical display from the Equation 1 to Equation 3 to form the mixed model equations for the 1-1-1 multi-level mediation model:

$$Y_{ij} = \alpha_{1i} + cX_{ij} + \epsilon_{1ij}, \tag{1}$$

$$M_{ij} = \alpha_{2i} + aX_{ij} + \epsilon_{2ij}, \tag{2}$$

$$Y_{ij} = \alpha_{3i} + c'X_{ij} + bM_{ij} + \epsilon_{3ij}, \tag{3}$$

where Y_{ij} is a representative of the response variable, and it is a continuous variable in this study. M_{ij} is a representative of the mediator, and it is a continuous variable in this study. X_{ij} refers to the independent variable and it could be either discrete or continuous variable. All three variables, Y_{ij}, M_{ij} and X_{ij} are observed in the individual-level. c, a, c' and b are fixed effects. ϵ_{1ij}, ϵ_{2ij} and ϵ_{3ij} are error terms, and α_{1i}, α_{2i} and α_{3i} are random effects for corresponding models. We assume normal distributions for the error terms and random effects: $\epsilon_{1ij} \sim N(0, \sigma^2_{\epsilon_{1ij}})$, $\epsilon_{2ij} \sim N(0, \sigma^2_{\epsilon_{2ij}})$, $\epsilon_{3ij} \sim N(0, \sigma^2_{\epsilon_{3ij}})$, $\alpha_{1i} \sim N(0, \sigma^2_{\alpha_{1i}})$, $\alpha_{2i} \sim N(0, \sigma^2_{\alpha_{2i}})$ and $\alpha_{3i} \sim N(0, \sigma^2_{\alpha_{3i}})$ for $i = 1, 2, ..., m$ and $j = 1, 2, ..., n_i$. c is the total effect of X on Y. c' is the direct effect of X on Y. $a \times b$ is the indirect effect (Mediation effect) of X on M. Figure 1 and Figure 2 shows a diagram of a $1-1-1$ mediation analysis model. Figure 1 represents the total effect of X on Y. In Figure 2, we partition the total effect into direct effect and indirect effect, as indirect effect goes through variable M.

We have underlying missingness mechanisms assumptions in this study as defined by Rubin (1976) and have been extensively discussed in missing data literature. We discuss three missingness mechanisms in this paper. The first missingness mechanism is missing completely at random (MCAR), which means that missingness probabilities are independent of any observed data; The second missingness mechanism is missing at random (MAR), which means that missingness probabilities may depend on observed data but not on variables subject to missing data and, finally, missing not at random (MNAR) means that probabilities of missingness depend on the missing values and hence they need to be modeled. In this paper, we focus on investigating performance under MCAR and MAR.

2.3 Multiple Imputation

Multiple imputation (MI) is a very useful method to impute missing values. MI is a practice of replacing missing values that are typically drawn from a posterior predictive distribution of missing data. We denote $Y_{ij(obs)}$ as the observed outcome variable Y_{ij}, and denote $Y_{ij(mis)}$ as the unobserved outcome variable Y_{ij}. Y_{ij} includes $Y_{ij(obs)}$ and $Y_{ij(mis)}$. Let ϕ denotes the unknown parameters. MI obtains multiple independent draws from the posterior predictive distribution and generate M complete data sets. The missing values are drawn from its posterior predictive distribution as follows:

$$P(Y_{ij(mis)}|Y_{ij(obs)}) = \int P(Y_{ij(mis)}|Y_{ij(obs)}, \phi)P(\phi|Y_{ij(obs)})d\phi \tag{4}$$

After imputation, we have M complete data sets. We use standard MI combination rule to pool all results by Rubin (1987). Due to the reason that $P(\phi|Y_{ij(obs)})$ is difficult to estimate, we utilize a Markov Chain Monte Carlo (MCMC) method. Specifically, we focus on a Gibbs sampler, which is a bayesian simulation technique that samples from the conditional distributions Van Buuren et al. (2006).

Firstly, we draw ϕ from the conditional distribution of $P(\phi|\alpha, Y)$. This process includes three steps. We start to draw $\alpha^{(t+1)}$ by the following conditional distribution:

$$\alpha^{(t+1)} \sim P(\alpha|Y_{ij(obs)}, Y^t_{ij(mis)}, \phi^{(t)}) \tag{5}$$

Then, we draw $\theta^{(t+1)}$ as follows:

$$\phi^{(t+1)} \sim P(\phi|Y_{ij(obs)}, Y^t_{ij(mis)}, \alpha^{(t+1)}) \tag{6}$$

After that, we draw $Y^{t+1}_{ij(mis)}$ as follows:

$$Y^{(t+1)}_{ij(mis)} \sim P(Y_{ij(mis)}|Y_{ij(mis)}, \phi^{t+1}, \alpha^{(t+1)}) \tag{7}$$

Based on the initial values of $(\alpha,\ \phi,\ Y_{ij(mis)})^{(0)}$, we can use the above steps to finalize a full cycle of Gibbs sampler. And the cycle will keep generating sequences for θ and $Y_{ij(mis)}$, i.e., we have $\{\theta^{(1)}, \theta^{(2)}, \theta^{(3)}, ...\}$ and $\{Y^{(1)}_{ij(mis)}, Y^{(2)}_{ij(mis)}, Y^{(3)}_{ij(mis)}, ...\}$. The samples in the sequences follow a Marcov chain.

This predictive distribution of missing data can be modeled using parametric or non-parametric methods. Multiple imputation can be done completely non-parametrically or parametrically using either joint model or fully conditional specification.

In parametric methods, there are two commonly used modeling strategies. Joint modeling (Carpenter & Kenward (2013); Schafer & Yucel (2002)) and conditional (or sequential or variable-by-variable) (Van Buuren (2011)) imputation models are two widely applied methods for multi-level data with missing data problem.

Multilevel joint modeling draw imputations from a multivariate mixed model (two-level hierarchical model). Specifically, JM employs bayesian theory to generate imputations from two distributions. The missing values are drawn from a multivariate distribution that conditions on the complete variable, random effects and fixed effects. The random effects are drawn from a multivariate distribution with zero means and corresponding variances.

Multilevel fully conditional specification (FCS) imputation employs multiple draw imputation for one incomplete variable at a time based on univariate mixed model. Specifically, the FCS imputation procedure employs similar bayesian procedure as JM. For each variable with missing data problem, imputations are drawn from a univariable distribution that conditions on the all other variables, fixed effects and random effect. The random effect is drawn from a univariate distribution with zero mean and corresponding variance.

3 Analytic comparison of joint versus variable-by-variable imputation model

3.1 Population joint distribution

Consider three variables Y, M, and X distributed as multivariate normal distribution:

$$\begin{bmatrix} Y \\ M \\ X \end{bmatrix} \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{8}$$

where the mean vector ($\boldsymbol{\mu}$) and covariace matrice ($\boldsymbol{\Sigma}$) are:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_Y \\ \mu_M \\ \mu_X \end{bmatrix} \tag{9}$$

$$\begin{aligned} \boldsymbol{\Sigma} &= \boldsymbol{\Sigma_W} + \boldsymbol{\Sigma_B} \\ &= \begin{bmatrix} \sigma^2_{Y(W)} & \sigma_{YM(W)} & \sigma_{YX(W)} \\ \sigma_{MY(W)} & \sigma^2_{M(W)} & \sigma_{MX(W)} \\ \sigma_{XY(W)} & \sigma_{MX(W)} & \sigma^2_{X(W)} \end{bmatrix} + \begin{bmatrix} \sigma^2_{Y(B)} & \sigma_{YM(B)} & \sigma_{YX(B)} \\ \sigma_{MY(B)} & \sigma^2_{M(B)} & \sigma_{MX(B)} \\ \sigma_{XY(B)} & \sigma_{MX(B)} & \sigma^2_{X(B)} \end{bmatrix} \\ &= \begin{bmatrix} \sigma^2_{Y(W)} + \sigma^2_{Y(B)} & \sigma_{YM(W)} + \sigma_{YM(B)} & \sigma_{YX(W)} + \sigma_{YX(B)} \\ \sigma_{MY(W)} + \sigma_{MY(B)} & \sigma^2_{M(W)} + \sigma^2_{M(B)} & \sigma_{MX(W)} + \sigma_{MX(B)} \\ \sigma_{XY(W)} + \sigma_{XY(B)} & \sigma_{MX(W)} + \sigma_{MX(B)} & \sigma^2_{X(W)} + \sigma^2_{X(B)} \end{bmatrix} \end{aligned} \tag{10}$$

where μ_Y, μ_M and μ_X are expectations of Y, M and X, respectively. We partition the variance-covariance matrix $\boldsymbol{\Sigma}$ into within- and between-cluster components ($\boldsymbol{\Sigma_W}$ and $\boldsymbol{\Sigma_B}$) as shown in Equation 10. For example, σ^2_Y and σ_{YX} are variance and covariance of Y, and Y and X. σ^2_Y and σ_{YX} are partitioned into within-cluster variance and covariance (i.e., $\sigma^2_{Y(W)}$ and $\sigma_{YX(W)}$), and between-cluster variance and covariance (i.e., $\sigma^2_{Y(B)}$ and $\sigma_{YX(B)}$).

Then, we build up a joint distribution of Y, M, X and random effects (α_Y and α_M) based on the joint distribution of Y, M and X with partitioned variance-covariance matrix. The between-cluster variance of Y ($\sigma^2_{Y(B)}$) is assumed to be equal to the variance of random effect of Y (α_Y). The between-cluster variance of M ($\sigma^2_{M(B)}$) is assumed to be equal to the random effect of M (α_{M_i}). And we assume that α_Y and α_{M_i} are independent. Therefore, we have the joint distribution of Y, M, X and random effects (α_Y and α_M) as follows:

$$\begin{bmatrix} Y_{ij} \\ M_{ij} \\ X_{ij} \\ \alpha_{Y_i} \\ \alpha_{M_i} \end{bmatrix} \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

where the mean vector and covariace matrice are:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_Y \\ \mu_M \\ \mu_X \\ 0 \\ 0 \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2_{Y(W)} + \sigma^2_{Y(B)} & \sigma_{YM(W)} + \sigma_{YM(B)} & \sigma_{YX(W)} + \sigma_{YX(B)} & \sigma^2_{Y(B)} & \sigma^2_{YM(B)} \\ \sigma_{YM(W)} + \sigma_{YM(B)} & \sigma^2_{M(W)} + \sigma^2_{M(B)} & \sigma_{MX(W)} + \sigma_{MX(B)} & \sigma_{MY(B)} & \sigma^2_{M(B)} \\ \sigma_{XY(W)} + \sigma_{XY(B)} & \sigma_{MX(W)} + \sigma_{MX(B)} & \sigma^2_{X(W)} + \sigma^2_{X(B)} & \sigma_{XY(B)} & \sigma_{MX(B)} \\ \sigma^2_{Y(B)} & \sigma_{MY(B)} & \sigma_{XY(B)} & \sigma^2_{Y(B)} & 0 \\ \sigma_{YM(B)} & \sigma^2_{M(B)} & \sigma_{MX(B)} & 0 & \sigma^2_{M(B)} \end{bmatrix}$$

In the following content, the processes of constructing conditional distributions for JM and FCS are based on the joint distribution of $Y_{ij}, M_{ij}, X_{ij}, \alpha_{Y_i}, \alpha_{M_i}$ as shown above.

3.2 Implied conditional distributions under joint modeling

We first formulate the population joint distribution as a conditional distribution in terms of mixed model parameters. The linear mixed model can be rewritten as a two-level hierarchical model in Equation 11 and 12:

$$\begin{bmatrix} Y_{ij} \\ M_{ij} \end{bmatrix} | X_{ij}, \alpha_Y, \alpha_M \sim N\left(\begin{bmatrix} \alpha_{Y_i} + \beta_{Y|X} X_{ij} \\ \alpha_{M_i} + \beta_{M|X} X_{ij} \end{bmatrix}, \mathbf{\Sigma}_{Y,M|X,\alpha_Y,\alpha_M} \right) \tag{11}$$

$$\begin{bmatrix} \alpha_{Y_i} \\ \alpha_{M_i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{Y(B)} & 0 \\ 0 & \sigma^2_{M(B)} \end{bmatrix} \right) \tag{12}$$

where α_{Y_i} and α_{M_i} are random effects. $\beta_{Y|X}$, $\beta_{M|X}$ are fixed values (typically underlying computational algorithms are MCMC type algorithms which operates on conditional distributions fixing these unknown values at a previous iteration). The prior distribution for $\boldsymbol{\alpha_i}$ is multivariate normal distribution with zero means and variances ($\sigma^2_{\alpha_{Y_i}}$ and $\sigma^2_{\alpha_{M_i}}$). $\mathbf{\Sigma}_{Y,M|X,\alpha_Y,\alpha_M}$ is the variance-covariance matrix for the conditional distribution of $Y, M|X, \alpha_Y, \alpha_M$, and it is calculated as:

$$\mathbf{\Sigma}_{Y,M|X,\alpha_Y,\alpha_M} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21} \tag{13}$$

where $\mathbf{\Sigma}_{11}$, $\mathbf{\Sigma}_{12}$ and $\mathbf{\Sigma}_{22}$ are shown as below:

$$\mathbf{\Sigma}_{11} = \begin{bmatrix} \sigma^2_{Y(W)} + \sigma^2_{Y(B)} & \sigma_{YM(W)} + \sigma_{YM(B)} \\ \sigma_{YM(W)} + \sigma_{YM(B)} & \sigma^2_{M(W)} + \sigma^2_{M(B)} \end{bmatrix},$$

$$\mathbf{\Sigma}_{12} = \begin{bmatrix} \sigma_{YX(W)} + \sigma_{YX(B)} & \sigma^2_{Y(B)} & \sigma_{YM(B)} \\ \sigma_{MX(W)} + \sigma_{MX(B)} & \sigma_{YM(B)} & \sigma^2_{M(B)} \end{bmatrix},$$

$$\mathbf{\Sigma}_{22} = \begin{bmatrix} \sigma^2_{X(W)} + \sigma^2_{X(B)} & \sigma_{XY(B)} & \sigma_{MX(B)} \\ \sigma_{XY(B)} & \sigma^2_{Y(B)} & 0 \\ \sigma_{MX(B)} & 0 & \sigma^2_{M(B)} \end{bmatrix},$$

$$\mathbf{\Sigma}_{21} = \begin{bmatrix} \sigma_{YX(W)} + \sigma_{YX(B)} & \sigma_{MX(W)} + \sigma_{MX(B)} \\ \sigma^2_{Y(B)} & \sigma_{MY(B)} \\ \sigma_{MY(B)} & \sigma^2_{M(B)} \end{bmatrix}$$

There are multiple steps we take to develop the formula of $\mathbf{\Sigma}_{22}^{-1}$. The details of development of estimating $\mathbf{\Sigma}_{22}^{-1}$ and $\mathbf{\Sigma}_{Y,M|X,\alpha_Y,\alpha_M}$ display in Appendix section 8.1.

3.3 Individual distributions under fully conditional specification

In this study, parametric FCS would employ mixed models as foundation to impute missing values in Y and M. The model is specified as a two-level hierarchical model, and the distribution of the incomplete variable Y_{ij} given the complete variables (M_{ij} and X_{ij}) and random effect (α_{Y_i}), and the prior distribution of α_{Y_i} is:

$$Y_{ij} \mid M_{ij}, X_{ij}, \alpha_{Y_i} \sim N(\alpha_{Y_i} + \beta_{Y|X} X_{ij} + \beta_{Y|M} M_{ij}, \sigma^2_{Y|M,X}) \tag{14}$$

$$\alpha_{Y_i} \sim N(0, \sigma^2_{Y(B)}) \tag{15}$$

where Y represents the drawn missing value from its posterior predictive distribution whose parameters are fixed at the previous iteration, and α_{Y_i} is a random effect. $\beta_{Y|X}$, $\beta_{Y|M}$ are fixed values(typically underlying computational algorithms are MCMC type algorithms which operates on conditional distributions fixing these unknown values at a previous iteration). The procedure of imputing missing values of M is specified as a two-level hierarchical model, which includes a conditional distribution of the incomplete variable M given the complete variables (Y_{ij} and X_{ij}) and random effect (α_{M_i}), and the prior distribution of α_{M_i} is:

$$M_{ij} \mid Y_{ij}, X_{ij}, \alpha_{M_i} \sim N(\alpha_{Y_i} + \beta_{M|X} X_{ij} + \beta_{M|Y} Y_{ij}, \sigma^2_{M|Y,X}) \tag{16}$$

$$\alpha_{M_i} \sim N(0, \sigma^2_{M(B)}) \tag{17}$$

where M represents the drawn missing value from its posterior predictive distribution whose parameters are fixed at the previous iteration, and α_{M_i} is a random effect. $\beta_{M|X}$, $\beta_{M|Y}$ are fixed values(typically underlying computational algorithms are MCMC type algorithms which operates on conditional distributions fixing these unknown values at a previous iteration).

3.4 Application in multi-level mediation analysis model

In this section, we assess the impact of imputation models on multi-level mediation effect (δ). The estimation of individual level mediation effect is a more complicated scenario than estimating simpler scenario with i.i.d random samples mediation effect. The analytical development of estimation for cluster level mediation effect can be found at Ye & Yucel (2020). We firstly formulate the MLE estimator of δ, and then we analytically estimate the bias generated by corresponding imputation models.

We implement the imputation models as described in section 3.1, 3.2 and 3.3 with three variables Y, M and X. As we have stated the multi-level mediation analysis model formulas from Equation 1 to Equation 3. By the properties of multi-level mediation analysis model, the direct effect is estimated by the coefficient of X_{ij} on Y_{ij} (c') in Equation 3. δ is estimated by the multiplication of the coefficient of X_{ij} on M_{ij} (a in Equation 2) and the coefficient of M_{ij} on Y_{ij} (b in Equation 3), i.e., $\delta = a \times b$. The total effect is estimated by the coefficient of X_{ij} on Y_{ij} (c in Equation 1). MacKinnon et al.(1995) have shown the algebraic equivalence of the $a \times b$ and $c - c'$ of the three mediation regression equations by normal theory ordinary least squares and maximum likelihood estimation .

For each linear mixed model (LMM) equation in 1, 2 and 3, we rewrite LMM as two-level hierarchical model. We have maximum likelihood estimator (MLE) or weighted least squares estimator (LSE) of $\hat{c}$ in Equation 1 as follows:

$$\hat{c} = (X^T V_1^{-1} X)^{-1} X^T V_1^{-1} Y$$

where we have

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \in \Re^{n\times 1}, \text{where n} = \sum_{i=1}^{m} n_i, X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \in \Re^{n\times 1}, V_1 = I_n G_{\alpha_{1i}} I_n^T + R_1$$

$G_{\alpha_{1i}}$ is a diagonal matrix of the variance of α_{1i} ($\sigma^2_{\alpha_{1i}}$), I_n is an identity matrix of size n and R_1 is a $n \times n$ variance matrix of residues for regression model in Equation 1. Similarly, we have the MLE or weighted LSE of $\hat{a}$ in Equation 2 as follows :

$$\hat{a} = (X^T V_2^{-1} X)^{-1} X^T V_2^{-1} M$$

where we have

$$M = \begin{pmatrix} M_1 \\ \vdots \\ M_n \end{pmatrix} \in \Re^{n\times 1}, V_2 = I_n G_{\alpha_{2i}} I_n^T + R_2$$

$G_{\alpha_{2i}}$ is a diagonal matrix of the variance of α_{2i}, and R_2 is a $n \times n$ variance matrix of residues for regression model in Equation 2. We have the MLE or weighted LSE of $\hat{c}'$ and $\hat{b}$ as follows:

$$(\hat{c}', \hat{b})^T = (X_{[M]}^T V_3^{-1} X_{[M]})^{-1} X_{[M]}^T V_3^{-1} Y$$

where we have

$$X_{[M]} = \begin{pmatrix} X_1 & M_1 \\ \vdots & \vdots \\ X_n & M_n \end{pmatrix} \in \Re^{n\times 2}, V_3 = I_n G_{\alpha_{3i}} I_n^T + R_3$$

$G_{\alpha_{3i}}$ is a diagonal matrix of the variance of α_{3i}, and R_3 is a $n \times n$ variance matrix of residues for regression model in Equation 3.

All parameters in the following content are estimated by MLE. We implement the invariance property of MLE in the following content to estimate the parameters. Therefore, we have the MLE estimator of $\hat{\delta}$ is as follows:

$$\hat{\delta} = \hat{a}\hat{b} = Z(X^T V_2^{-1} X)^{-1} X^T V_2^{-1} M (X_{[M]}^T V_3^{-1} X_{[M]})^{-1} X_{[M]}^T V_3^{-1} Y \tag{18}$$

where Z is a 1×2 matrix, i.e., $Z = (0, 1)$. Computational details of formulating MLE of $\hat{a}$, $\hat{b}$, $\hat{c}$, $\hat{c}'$ and $\hat{\delta}$ for sample data are provided in Appendix 8.2.

3.4.1 Mediation effect estimate under JM and FCS

We first start from a simpler scenario that we only have one incomplete variable, and then we extend to the scenario that we have two incomplete variables under the framework of multi-level mediation analysis model.

When we only have one incomplete variable (Y) and two fully observed variables (M and X) in the dataset $\boldsymbol{D}$, we find that both JM and FCS models employ the same imputation formula (Equation 14 and 15) as the imputation model to generate imputed values for Y, and the association between Y and M is governed by the same parameter, $\beta_{Y|M}$ in Equation 14. The formulated expression of $\hat{\delta}$ is consistent for both joint and variable-by-variable imputation methods. Thus, JM and FCS are consistent in estimating mediation effect when we have only one incomplete variable under the multi-level mediation analysis model framework.

In the second scenario, we suppose that there are two incomplete variables (Y and M) and one fully observed variable (X) in the dataset $\boldsymbol{D}$. We formulate the MLE estimates of mediation effect by employing imputation models' parameters, and calculate the estimates of bias generated by imputation models at the end of this section.

Firstly, we choose FCS as the imputation model, and we restate that FCS employs the conditional distributions of Y_{ij} given M_{ij}, X_{ij} and α_{Yi} with the prior distribution of α_{Y_i}, and M_{ij} given Y_{ij}, X_{ij} and α_{Mi} with the prior distribution of α_{M_i} to impute incomplete values for Y_{ij} and M_{ij}, respectively. The conditional distribution of Y_{ij} given M_{ij}, X_{ij} and α_{Yi}, and prior distribution of $\alpha_{Y_i}$are shown from Equation 14 to 15. We firstly multiply the conditional distribution of Y_{ij} given $M_{ij}, X_{ij}, \alpha_{Y_i}$ with the prior distribution of α_{Y_i}. We find the conditional distribution of Y_{ij} and α_{Y_i} given M_{ij} and X_{ij} as follows (details are shown in Appendix 8.3):

$$f(Y_{ij}, \alpha_{Y_i} | M_{ij}, X_{ij}) = f(Y_{ij} \mid M_{ij}, X_{ij}, \alpha_{Y_i}) \times f(\alpha_{Y_i})$$

Then, we integrate the conditional distribution of Y_{ij} and α_{Y_i} given M_{ij} and X_{ij} regarding α_{Y_i}, and we have the conditional distribution of Y_{ij} given M_{ij} and X_{ij}, which is free of random effect α_{Y_i}. The detailed process of developing the distribution of Y_{ij} given M_{ij} and X_{ij} is shown at Appendix 8.3. The distribution of Y_{ij} given M_{ij} and X_{ij} is shown as follows:

$$f(Y_{ij}|M_{ij}, X_{ij}) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2_{Y|M,X} + \sigma^2_{\alpha_Y}}} e^{\dfrac{-(Y - \beta_{Y|X}X - \beta_{Y|M}M)^2}{2(\sigma^2_{Y|M,X} + \sigma^2_{\alpha_Y})}}$$

In this case, we find that Y_{ij} given M_{ij} and X_{ij} is a normal distribution with mean $\beta_{Y|X}X_{ij} + \beta_{Y|M}M_{ij}$ and variance $\sigma^2_{Y|M,X} + \sigma^2_{\alpha_Y}$. The conditional expectation of Y_{ij} given M_{ij} and X_{ij} ($E(Y|M,X)$) is $\beta_{Y|X}X + \beta_{Y|M}M$, and we find the estimate of $\hat{\beta}_{Y|M}$ ($\hat{b}$ in mediation analysis model) as follows:

$$\hat{b} = \hat{\beta}_{Y|M} = \frac{\hat{\sigma}^2_M \hat{\sigma}_{YX} - \hat{\sigma}_{MX}\hat{\sigma}_{MY}}{\hat{\sigma}^2_M \hat{\sigma}^2_X - \hat{\sigma}^2_{MX}}$$

Next, we determine the estimate of $\hat{a}$, which corresponds to the coefficient of M_{ij} on X_{ij} in multi-level mediation analysis model (Equation 2). We multiply the conditional distribution of M_{ij} given Y_{ij}, X_{ij} and α_{Mi} with the prior distribution of α_{M_i} in Equation 16 and 17. We find the conditional distribution of M_{ij}, α_M given Y_{ij} and X_{ij} as follows (details are shown in Appendix 8.3):

$$f(M_{ij}, \alpha_{M_i}|Y_{ij}, X_{ij}) = f(M_{ij} \mid Y_{ij}, X_{ij}, \alpha_{M_i}) \times f(\alpha_{M_i})$$

Then, we integrate the conditional distribution of M_{ij} and α_{M_i} given Y_{ij} and X_{ij} with regard to α_{M_i}, and we have the conditional distribution of M_{ij} given Y_{ij} and X_{ij}, which is free of α_{M_i}. The details of developing the distribution of M_{ij} given Y_{ij} and X_{ij} is shown at Appendix 8.4. The distribution of M_{ij} given Y_{ij} and X_{ij} is shown as follows:

$$f(M_{ij}|Y_{ij}, X_{ij}) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2_{M|Y,X} + \sigma^2_{\alpha_M}}} e^{\dfrac{-(M_{ij} - \beta_{M|X}X_{ij} - \beta_{M|Y}Y_{ij})^2}{2(\sigma^2_{M|Y,X} + \sigma^2_{\alpha_M})}}$$

We find that the conditional distribution of M_{ij} given Y_{ij} and X_{ij} is a normal distribution with mean $\beta_{M|X}X_{ij} + \beta_{M|Y}Y_{ij}$ and variance $\sigma^2_{M|Y,X} + \sigma^2_{\alpha_M}$. Then, we find the conditional distribution of M_{ij} and Y_{ij} given X_{ij} by multiplying the conditional distribution of M_{ij} given Y_{ij} and X_{ij} with the distribution of Y_{ij}. Y_{ij} is distributed as $Y_{ij} \sim N(\mu_Y, \sigma^2_{Y(W)} + \sigma^2_{Y(B)})$. Next, we find the distribution of M_{ij} given X_{ij} by integrating the conditional distribution of M_{ij} given Y_{ij} and X_{ij} with respect to Y_{ij}. The process is shown as follows:

$$\begin{aligned} f(M_{ij}|X_{ij}) &= \int f(M_{ij}|Y_{ij}, X_{ij}) f(Y_{ij}) dY \\ &= \int f(M_{ij}, Y_{ij}|X_{ij}) dY \end{aligned}$$

We show the derivation details at Appendix 8.5. Therefore, we find the conditional distribution of M_{ij} given X_{ij} as a shifted normal distribution as follows:

$$f(M_{ij}|X_{ij}) = \frac{1}{\sqrt{\pi(\beta^2_{M|Y}\sigma^2_Y + \sigma^2_{M|Y,X} + \sigma^2_{\alpha_M})}} e^{A_2M^2 + B_2M + C_2}$$

where A_2, B_2 and C_2 are functions of σ^2_Y, $\sigma^2_{\alpha_M}$, $\beta_{M|Y}$ and etc. The expressions for A_2, B_2 and C_2 are provided in Appendix 8.5. Therefore, we find that the distribution of M_{ij} given X_{ij} is a shifted normal distribution. We determine the coefficient of M_{ij} on X_{ij} by calculating the expectation of M_{ij} given X_{ij} ($E(M|X)$), and we have the formula of $E(M|X)$ (details of development are shown in Appendix 8.5) as follows:

$$E(M|X) = -\frac{(\beta_{M|X}\sigma^2_{M|Y,X} + \sigma^2_\alpha\beta_{M|X})X + c'}{2\pi A(\beta^2_{M|Y}\sigma^2_Y + \sigma^2_{M|Y,X} + \sigma^2_\alpha)(\sigma^2_{M|X,Y} + \sigma^2_\alpha)\sqrt{2A(\beta^2_{M|X}\sigma^2_Y + \sigma^2_{M|Y,X} + \sigma^2_\alpha)}}$$

where c' is a constant with respect to X. Therefore, we find the estimate of $\hat{a}$, which is the coefficient of X in the formula above, we have

$$\hat{a} = [\frac{\hat{\sigma}^2_Y\hat{\beta}^2_{M|Y}\hat{\beta}_{M|X}}{(\hat{\beta}^2_{M|Y}\hat{\sigma}^2_Y + \hat{\sigma}^2_{M|Y,X} + \hat{\sigma}^2_\alpha)(\hat{\sigma}^2_{M|X,Y} + \hat{\sigma}^2_\alpha)} - \frac{\hat{\beta}_{M|X}}{\hat{\sigma}^2_{M|X,Y} + \hat{\sigma}^2_\alpha}] \times \\ \frac{1}{2\pi\hat{A}\sqrt{2\hat{A}(\hat{\beta}^2_{M|X}\hat{\sigma}^2_Y + \hat{\sigma}^2_{M|Y,X} + \hat{\sigma}^2_\alpha)}}$$

Therefore, the estimate of mediation effect after applying FCS to data with missing values is as follows:

$$\hat{\delta}_{FCS} = \hat{a}\hat{b} = [\frac{\hat{\sigma}_Y^2 \hat{\beta}_{M|Y}^2 \hat{\beta}_{M|X}}{(\hat{\beta}_{M|Y}^2 \hat{\sigma}_Y^2 + \hat{\sigma}_{M|Y,X}^2 + \hat{\sigma}_\alpha^2)(\hat{\sigma}_{M|X,Y}^2 + \hat{\sigma}_\alpha^2)} - \frac{\hat{\beta}_{M|X}}{\hat{\sigma}_{M|X,Y}^2 + \hat{\sigma}_\alpha^2}] \times \frac{\hat{\beta}_{M|X}}{2\pi \hat{A}\sqrt{2\hat{A}(\hat{\beta}_{M|X}^2 \hat{\sigma}_Y^2 + \hat{\sigma}_{M|Y,X}^2 + \hat{\sigma}_\alpha^2)}} \tag{19}$$

Then, we implement the similar strategy to evaluate the $\hat{\delta}_{JM}$ for JM as we did in previous content for FCS. JM draws incomplete variables, Y_{ij} and M_{ij}, together from the joint conditional distribution of Y_{ij} and M_{ij} given X_{ij}, α_Y and α_M with the prior joint distribution of α_Y and α_M in Equation 11 and 12. However, different from FCS, a and b are not independently generated, we cannot estimate δ directly by multiplying a and b. Therefore, we employ the theorem from MacKinnon (1995) to estimate $\hat{\delta}$ based on the following equivalence:

$$\hat{\delta} = \hat{a} \times \hat{b} = \hat{c} - \hat{c}'$$

We employ the following techniques to find the estimation of mediation effect by c and c'. We determine $\hat{c}$ by calculating $E(Y|X)$ and assessing the association between Y and X. We estimate $\hat{c}'$ by assessing the coefficient of X on Y within the expression of $E(Y|M,X)$. We firstly calculate the distribution of $Y_{ij}, M_{ij}, \alpha_{Y_i}, \alpha_{M_i}$ given X_{ij} by multiplying the conditional distribution of Y_{ij}, M_{ij} given $X_{ij}, \alpha_{Y_i}, \alpha_{M_i}$ and the prior distribution of $\alpha_{Y_i}, \alpha_{M_i}$ as follows:

$$\begin{aligned} f(Y_{ij}, M_{ij}, \alpha_{Y_i}, \alpha_{M_i}|X_{ij}) &= f(Y_{ij}, M_{ij}|X_{ij}, \alpha_{Y_i}, \alpha_{M_i}) \times f(\alpha_{Y_i}, \alpha_{M_i}) \\ &= \frac{exp(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma_c}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))}{\sqrt{(2\pi)^2|\boldsymbol{\Sigma_c}|}} \times \frac{exp(\boldsymbol{\alpha}^T \boldsymbol{\Sigma_\alpha}^{-1}\boldsymbol{\alpha})}{\sqrt{(2\pi)^2 \boldsymbol{\Sigma_\alpha}}} \end{aligned}$$

We find the conditional distribution of Y_{ij}, M_{ij} given X_{ij} by integrating conditional distribution of $Y_{ij}, M_{ij}, \alpha_{Y_i}, \alpha_{M_i}$ given X_{ij} regarding α_{Y_i} and α_{M_i} as follows:

$$f(Y_{ij}, M_{ij}|X_{ij}) = \iint f(Y_{ij}, M_{ij}, \alpha_{Y_i}, \alpha_{M_i}|X_{ij}) d\alpha_{Y_i} d\alpha_{M_i}$$

Therefore, we find the formula of $f(Y_{ij}, M_{ij}|X_{ij})$ as follows (details of development are provided at Appendix 8.6):

$$f(Y_{ij}, M_{ij}|X_{ij}) = \frac{\sqrt{(fk-gh)\sigma_Y^2}}{\pi\sqrt{2(k\sigma_Y^2 + fk - gh)\Sigma_C \Sigma_\alpha A_4}} e^{-HY^2+JY+K}$$

where H, J and K are constants with respect to Y. The expressions for H, J and K are provided at Appendix 8.6. Then, we integrate the conditional distribution of Y_{ij} and M_{ij} given X_{ij} with regard to M_{ij}, and we obtain pdf of Y_{ij} given X_{ij} as follows:

$$f(Y_{ij}|X_{ij}) = \int f(Y_{ij}, M_{ij}|X_{ij}) dM$$

Therefore, we have the expression of $f(Y_{ij}|X_{ij})$ as follows (details of development are provided at Appendix 8.7):

$$f(Y_{ij}|X_{ij}) = \frac{\sqrt{\sigma_Y^2(fk-gh)}}{\sqrt{\pi(k\sigma_Y^2 + fk - gh)\Sigma_C \Sigma_\alpha A_4 H}} e^{UM^2+VM+W}$$

where U, V and W are constant to X_{ij}. The conditional distribution of Y_{ij} given X_{ij} is a shifted normal distribution, and we find the expression of E$(Y|X)$ by the property of shifted normal distribution. We have $E(Y|X)$ (details of development are provided at Appendix 8.7) as follows:

$$E(Y|X) = \frac{2\sqrt{2(fk-gh)\sigma_Y^2 H}}{\sqrt{(k\sigma_Y^2 + fk - gh)\Sigma_C\Sigma_\alpha(P^2-4QH)}} \times \frac{2HR-LRX}{P^2-4HQ}$$

c equals to the coefficient of X in the expression of $E(Y|X)$. We find the expression of $\hat{c}$ as follows:

$$\hat{c} = -\frac{2\sqrt{2(\hat{f}\hat{k}-\hat{g}\hat{h})\hat{\sigma}_Y^2 H}}{\sqrt{(\hat{k}\hat{\sigma}_Y^2 + \hat{f}\hat{k} - \hat{g}\hat{h})\hat{\Sigma}_C\hat{\Sigma}_\alpha(\hat{P}^2-4\hat{Q}\hat{H})}} \times \frac{\hat{L}\hat{R}}{\hat{P}^2-4\hat{H}\hat{Q}}$$

We show the details of deriving $\hat{c}$ at Appendix 8.7. Then we find the conditional distribution of Y_{ij} given M_{ij}, X_{ij} using bayes theorem as follows (details of developing the conditional distribution of Y_{ij} given M_{ij}, X_{ij} are given at Appendix 8.8):

$$f(Y_{ij}|M_{ij}, X_{ij}) = \frac{f(Y_{ij}, M_{ij}|X_{ij})}{f(M_{ij}|X_{ij})}$$

We find the expression of $\hat{c}'$ by assessing the coefficient of X_{ij} on Y_{ij} within the expression of $E(Y|M,X)$. We find Y_{ij} given M_{ij} and X_{ij} is a shifted normal distribution, therefore, we have the formula of $E(Y|M,X)$ as follows:

$$E(Y|M,X) = \frac{E\sqrt{(fk-gh)\sigma_Y^2(\beta_{M|Y}^2\sigma_Y^2 + \sigma_{M|Y,X}^2 + \sigma_\alpha^2)}}{\sqrt{(k\sigma_Y^2 + fk - gh)\Sigma_C\Sigma_\alpha A_4}}[(\frac{F}{E}\beta_{M|X} - \beta_{Y|X})X - \frac{FM}{E}]$$

And we find the expression of $\hat{c}'$ as follows:

$$\hat{c}' = \frac{\hat{E}\sqrt{(\hat{f}\hat{k}-\hat{g}\hat{h})\hat{\sigma}_Y^2(\hat{\beta}_{M|Y}^2\hat{\sigma}_Y^2 + \hat{\sigma}_{M|Y,X}^2 + \hat{\sigma}_\alpha^2)}}{\sqrt{(\hat{k}\hat{\sigma}_Y^2 + \hat{f}\hat{k} - \hat{g}\hat{h})\hat{\Sigma}_C\hat{\Sigma}_\alpha\hat{A}_4}}(\frac{\hat{F}}{\hat{E}}\hat{\beta}_{M|X} - \hat{\beta}_{Y|X})$$

Thus, the estimate of mediation effect after applying FCS to data with missing values is as follows:

$$\begin{aligned}\hat{\delta}_{JM} &= \hat{c} - \hat{c}' \\ &= -\frac{2\sqrt{2(\hat{f}\hat{k}-\hat{g}\hat{h})\hat{\sigma}_Y^2 H}}{\sqrt{(\hat{k}\hat{\sigma}_Y^2 + \hat{f}\hat{k} - \hat{g}\hat{h})\hat{\Sigma}_C\hat{\Sigma}_\alpha(\hat{P}^2-4\hat{Q}\hat{H})}} \times \frac{\hat{L}\hat{R}}{\hat{P}^2-4\hat{H}\hat{Q}} - \\ &\quad \frac{\hat{E}\sqrt{(\hat{f}\hat{k}-\hat{g}\hat{h})\hat{\sigma}_Y^2(\hat{\beta}_{M|Y}^2\hat{\sigma}_Y^2 + \hat{\sigma}_{M|Y,X}^2 + \hat{\sigma}_\alpha^2)}}{\sqrt{(\hat{k}\hat{\sigma}_Y^2 + \hat{f}\hat{k} - \hat{g}\hat{h})\hat{\Sigma}_C\hat{\Sigma}_\alpha\hat{A}_4}}(\frac{\hat{F}}{\hat{E}}\hat{\beta}_{M|X} - \hat{\beta}_{Y|X})\end{aligned} \tag{20}$$

θ is denoted as the bias (difference between the expected value of the estimator and true value) of the estimate of mediation effect after applying imputation models, we first estimate θ_{FCS} by calculating the difference between the expectation of δ after implementing FCS and its corresponding true value, and we have

$$\hat{\theta}_{FCS} = E(\hat{\delta}_{FCS}) - \hat{\delta} \tag{21}$$

where $\hat{\delta}_{FCS}$ is shown in Equation 19. Then, we employ the same process to estimate θ_{JM} by implementing the same approach, and we have

$$\hat{\theta}_{JM} = E(\hat{\delta}_{JM}) - \hat{\delta} \tag{22}$$

where $\hat{\delta}_{JM}$ is shown in Equation 20. The true value of $\hat{\delta}$ in Equation 21 and Equation 22 is estimated based on MLE or Weighted LSE estimate of δ in Equation 18 before we apply any imputation models. MLE has the consistency property that it converges to true parameter in probability as $n \to \infty$. Then, we apply Slutsky's Theorem to find the expectation of $E(\hat{\hat{\delta}}_{FCS})$ and $E(\hat{\hat{\delta}}_{JM})$. Based on Slutsky's Theorem, we have $Y_n X_n \to aX$ in distribution, if $X_n \to X$ in distribution and $Y_n \to a$ in probability, where a is a constant.

Both formulas are very complex and include components such as variances, covariance and etc. However, we can factor out $\beta_{M|X}$ (the coefficient of $\mu_{M|X,Y,\alpha_M}$ for X) in the bias estimate for FCS. The relationship between M and X is directly related to the expected bias generated by FCS. As for JM, we fail to easily factor out a common term, and there is no specific term that affect the bias directly.

4 Simulation assessment

We conduct a simulation study to test the performance of different Multiple Imputation (MI) methods on the multi-level mediation analysis. This study compares PAN and MICE package in RStudio version (Version 1.1.463 – 2009-2018 RStudio, Inc). PAN package samples from multivariate linear mixed models for incomplete data, and MICE package imputes continuous two-level data based on univariate linear mixed model for each incomplete data.

4.1 Data generation

Our data generation process first simulates the independent variable X_{ij}, which is generated from an independent normal distribution, i.e., $X_{ij} \sim N(0, \sigma_X^2)$. Next, the mediator M, is simulated from the following mixed-effects formulation:

$$M_{ij} = aX_{ij} + \alpha_{1i} + \epsilon_{1ij},$$

where α_{1_i} is a cluster random-effect assumed to follow a normal distribution, i.e., $\alpha_{1i} \sim N(0, \sigma_{\alpha_1}^2)$ for any i. ϵ_{ij} refers to residual error term and is assumed to be generated from a normal distribution, i.e., $\epsilon_{1ij} \sim N(0, \sigma_{\epsilon_1}^2)$ for any i and j. α_1 and ϵ_{1ij} are assumed to be independent.

Next, we simulated Y_{ij} conditional on M_{ij}, X_{ij} using the following mixed-effects formulation:

$$Y_{ij} = c'X_{ij} + bM_{ij} + \alpha_{3i} + \epsilon_{3ij}$$

where α_{3i} and ϵ_{3ij} refer to random effect and residual error, respectively. α_{3i} and ϵ_{3ij} are assumed to be generated from independent normal distributions, i.e., $\alpha_{3i} \sim N(0, \sigma_{\alpha_3}^2)$ and $\epsilon_{3ij} \sim N(0, \sigma_{\epsilon_3}^2)$. We use this framework to simulate a population of 1,000 observations under 100 clusters. We repeat this process for 100 times to study the performance criteria stated in section 4.2.

In the first simulation experiment, we set δ to 0.01 (i.e., $a = b = 0.1$, $\delta = 0.01$ and $c' = 0.5$) to test the performance of imputation methods with low mediation effect. In the second simulation experiment, we set δ to 0.1521 (i.e., $a = b = 0.39$, $\delta = 0.1521$, $c' = 0.5$) and study the impact of imputation methods on medium mediation effect. In the third simulation experiment, we set δ to 0.3481 (i.e., $a = b = 0.59$, $\delta = 0.3481$, $c' = 0.5$) to examine the compatibility of imputation models on high mediation effect scenario.

The next step of data generation is imposing missing values under MCAR, MAR and MNAR mechanisms (Rubin (1976)). We impose the missing values on the response variable Y and the mediator M under different missingness mechanisms. Let r_Y and r_M denote missingness indicators for Y and M, respectively. Under MCAR mechanism, both r_Y and r_M are simulated from a bernoulli distribution. We set the success probability to 10 percent, 20 percent and 30 percent to study the behavior under these rates of missingness for Y and M. Under MAR mechanism, the probability of missingness of Y and M depend on the fully observed variable X:

$$logit(P(r_Y = 1|x)) = \gamma_0^Y + \gamma_1^Y X,$$
$$logit(P(r_M = 1|x)) = \gamma_0^M + \gamma_1^M X,$$

where the γ (γ_0 and γ_1) for both Y and M are chosen to lead to approximately 10 percent, 20 percent, and 30 percent rates of missingness. Under MNAR mechanism, missingness in Y and M depends on Y and M themselves, respectively. If Y is smaller than its s^{th} percentile, Y is missing, and if M is smaller than its s^{th} percentile, M is missing. We set s to be 10, 20 and 30 to control the missingness rates of Y and M at 10%, 20% and 30%, respectively.

We repeat this simulation process for three times to investigate the impact of different intracluster correlations (ICC), we allow σ_{α_1} to vary 0.5 from to 2, and σ_{ϵ_1} from 2 to 4. We have ICC=$\frac{\sigma_\alpha^2}{\sigma_\alpha^2+\sigma_\epsilon^2}$, and we only report results for the scenario when ICC equals 0.05, 0.1, and 0.2.

4.2 Estimation and evaluation criteria

For each simulation scenario, we create a total of 5 imputations under each MI method. MI theory suggests that three to five imputations could yield excellent results. Our choice of five imputations is considered based on imputation precision. The general choice number of imputations is discussed in Graham et al. (2007), Royston (2005) and Bodner (2008). The estimates and their standard errors are combined using Rubin's rule (Rubin (1976)).

Comparisons of different imputation methods are conducted based on different effect size of mediation effects, which are always focal to mediation analysis. We use the prespecified four different mediation effects (δ) as the "true values" to test the accuracy of different imputation methods. We use four different criteria to assess the consistency and accuracy for each imputation methods. Let $\hat{l}_r$ and $\hat{u}_r$ denote the lower and upper limits of the 95% confidence interval for the mediation effect in the r^{th} imputed data. We also denote $\hat{\delta}_r=\hat{a}_r\hat{b}_r$ as the mediation effect estimate for the r^{th} replication, where $r=1,...,100$.

The first criterion to assess the performance is the coverage rate (CR):

$$CR=\frac{\sum_{r=1}^{100} I_r(\hat{l}_r<\delta<\hat{u}_r)}{100},$$

where I_r is an indicator function such that:

$$I_r=\begin{cases}1 & \delta\in(\hat{l}_r,\hat{u}_r)\\ 0 & \delta\notin(\hat{l}_r,\hat{u}_r).\end{cases}$$

CR evaluates the proportion of confidence intervals that contain the true mediation effect value. In this study, we set the actual coverage of nominal 95% intervals. The actual rate is supposed to be close to nominal rate for good imputation method. Collins et al. (2001) suggest that the performance of imputation method is poor if the CR is below 90 percent for a 95 percent nominal rate. Secondly, we estimate the standardized bias for imputation models. Standardized bias (SB) is calculated as:

$$SB=100\times\frac{\bar{\hat{\delta}}-\delta}{SE},$$

where SE is the standard error of the estimates, and $\bar{\hat{\delta}}$ is the average estimate of the mediation effect. Collins et al. (2001) comment that once the standardized bias exceeds 40%, the bias may have a noticeable negative impact on imputed parameter. In this study, we consider standardized bias as significant when its absolute value is greater than 40%.

The third criterion is mean square error (MSE). The formula of how to calculate MSE is provided as follows:

$$MSE=\frac{1}{100}\sum_{r=1}^{100}(\hat{\delta}-\delta)^2,$$

MSE is a measure for accuracy and precision. MSE is a measure for accuracy and precision. We also calculate average width (AW) as:

$$AW=\frac{\sum_{r=1}^{1000}(\hat{u}_r-\hat{l}_r)}{1000},$$

Among four different criteria to assess the consistency and accuracy for each method under all conditions, CR and SB are the most important criteria to show the performance of imputation methods under multi-level mediation analysis model.

The same evaluation criteria will also be measured for complete-case-only (CC) analysis. For complete-case-only analysis, we exclude all units for which the outcomes variable and mediator (Y and M) are missing and keep the remaining complete observations. Then, we apply the mediation analysis model to obtain the estimate of mediation effect and compare the results with other datasets applied with imputation models.

4.3 Summary of the results

We compare the performance of FCS and JM methods for incomplete dataset under mediation analysis model. From table 1 to table 3, we show the different coverage rate (CR), mean square error (MSE), average width (AW), and standardized bias (SB) for the scenario when we have low mediation effect condition ($a = b = 0.1$, $\delta = ab = 0.01$, $c' = 0.5$). We report the result when ICC is fixed at 0.05, 0.1 and 0.2. From table 4 to 6, we show results for high mediation ($a = b = 0.59$, $\delta = ab = 0.3481$, $c' = 0.5$) condition and ICC is fixed at 0.05, 0.1 and 0.2. We show medium mediation condition ($a = b = 0.39$, $\delta = ab = 0.1521$, $c' = 0.5$) with fixed ICC as 0.05, 0.1 and 0.2 from figure 3 to figure 5. For each table, we compare CRs, MSEs, AWs and SBs of the estimated mediation effect under complete-case-only analysis and imputation models. Different missingness mechanisms (MCAR, MAR and MNAR) are considered with a sample size of 1,000, and variations in the missingness rates (10%, 20% and 30%). As for figures, we focus on comparing SBs with respect to varying joint missingness rate.

Table 1 presents the result under the scenario of low mediation effect. The missingness rates for M and Y are fixed at 10%, 20% and 30% marginally for each variable, and from 10% to 52.2% jointly. CRs and SBs that are above acceptable limits are shown in bold. We find that FCS and JM are consistent in estimating the effect size of mediation effect under MCAR and MAR mechanisms. Under MCAR, CRs range from 92% to 94% for FCS, and from 92% to 96% for JM. Under MAR, CRs range from 90% to 91% for FCS, and from 92% to 96% for JM. However, FCS, JM and CC do not perform well for MNAR mechanism, and we observe low CRs values (below 90%). We find that SBs are below 40% when the missingness mechanism are MCAR and MAR mechanisms (except for the case when we apply CC and have 30% JMR). We find a significant SB (above 40%) when we apply FCS under MNAR mechanism with 19.3%. SBs are all significant when we have 36.8% or higher JMR under MNAR mechanism regardless of imputation model. Table 2 and Table 3 show the result when we increase ICC (0.1 and 0.2).

Table 4, Table 5 and Table 6 display results for High mediation effect with varying ICC (0.05, 0.1 and 0.2). We find that FCS and JM perform well under MCAR and MAR. The compatibility between FCS and JM remains consistent when joint missing rate increases. However, none of the methods perform well under MNAR.

We show the comparison between exact bias and estimated bias by imputation methods in figure 6. We find that the bias generated by imputation methods is very close to the exact bias especially when the missingness mechanism is MCAR. When the missingness mechanism is MAR or MNAR, we find that the estimated bias by imputation methods is larger than the exact bias.

5 Application

We are studying the mediation effect of attendance rate in the association between the Environmental related program (EPA) and students grades under the multi-level mediation analysis model framework. Two datasets were used in the analysis (Building Condition Survey (BCS) and academic data for students in school-level).

New York State (NYS) Building Condition Survey (BCS) in 2015 is mandated by New York State Education Law and regulations that all NYS public schools submit the BCS to New York State Education Department (NYSED) every 5 years to assess the school environment with respect to building condition and students' learning environment. The BCS aims to measure building systems and school environment using a group of licensed engineers or architects through physical inspection. The BCS also includes information on school building age, size, and ratings of the overall building condition, and of the 53 individual building systems including the building envelope, plumbing systems, heating ventilation and air conditioning (HVAC) systems, roofing and implementation of environmental programs. Since 2005, a new section assessing environmental parameters related to comfort and health, including IAQ, cleanliness, acoustics, and lighting has been added. The other dataset comes from NY State Education Department and includes all school-level academic information of students (Regent test results for high school students, English Language Arts, Mathematics ("NY STATE - NEW YORK STATE REPORT CARD [2018 - 19]," n.d.), Science for high school students) in 2016. The city of where school locates is also included. Both datasets contain geographical information of schools, which was used to link the data records.

Previous researchers (Lorraine E. Maxwell and Suzanne L. Schechtman, 2012) have investigated the impact of EPA tools for school (TfS) on students grades and attendance. Also, Maxwell (2016) reported a significant impact of attendance rate on ELA score (direct effect is 0.39 under mediation analysis model) for middle school student at NYC. However, the mediation effect of attendance rate remains unknown from a multi-level mediation model perspective. We utilize "mlma" package by Yu and Li (2021) in R to estimate the mediation effect, indirect effect and percentage mediated for each test at elementary school (Table 7) and high school (Table 8). City is regarded as the cluster level variable. We find mediation effect of attendance rate in mediating the association between EPA tools for school (TfS) and elementary students' academic performance (mediation effect ranges from -0.33 to -0.48). We find mediation effect of attendance rate on the EPA TfS and high school students' academic performance (mediation effect ranges from 0.002 to -0.51).

6 Discussion

The objective of this simulation study is to verify the analytically-derived formula of bias, and determine the compatibility of FCS and JM imputation models in multi-level mediation analysis model. We find that the choice of imputation models does not have an impact on the multi-level mediation analysis model and the bias is verified.

We recommend researchers to apply imputation methods to assess mediation effect for incomplete dataset when the missingness mechanism is assumed as MCAR or MAR. When we apply imputation methods to incomplete dataset with varying mediation effect (low, medium and high mediation effect), we do not find a significant standardized bias when the missingness mechanism is MCAR with varying ICC (from 0.05 to 0.2). Although JM and FCS are performing as well as complete-case-only analysis for MCAR regarding the bias, MSE, coverage rate and etc, we lose sample size, efficiency and power if we complete-case-only analysis. Mediation analysis is mostly dealing with multivatiate conditions, therefore, losing sample size may have negative impact on the model. When the missingness mechanism is MAR, MI inference outperforms complete-case-only analysis regardless of the imputation methods and ICC as expected. CC will bias the whole sample, because the missing value depends on the covariates. None of the methods perform well under MNAR, in particular, CC led to substantial biases. Our simulation based study find that the results generated by imputation models are significantly better than CC under MCAR and MAR, because MI do not necessarily have the assumption for MNAR and we specify the model to impute missing values under MNAR.

This study has several limitations. First, comparison of JM and FCS is based on a pseudo-random simulation study, aimed to mimic a real-world problem. We supplemented this simulation with analytical assessment of the bias induced by each of imputation method. To the best of our knowledge, our work is the first to make such an assessment. Second, we only consider a mediation analysis model in the sense that we have only one mediator. We did not investigate different types of response variables or types of covariates/mediators. These aspects will be subject of our future research topics.

7 Tables and Figures

7.1 Mediation analysis model visualization

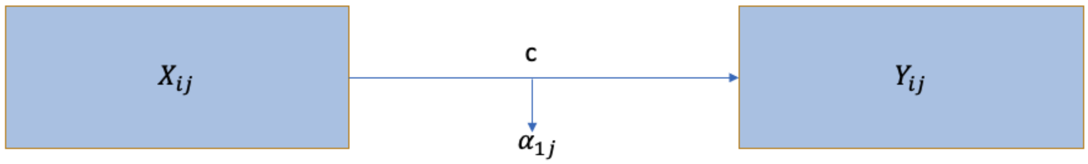

Figure 1: The total effect of the independent variable on the dependent variable

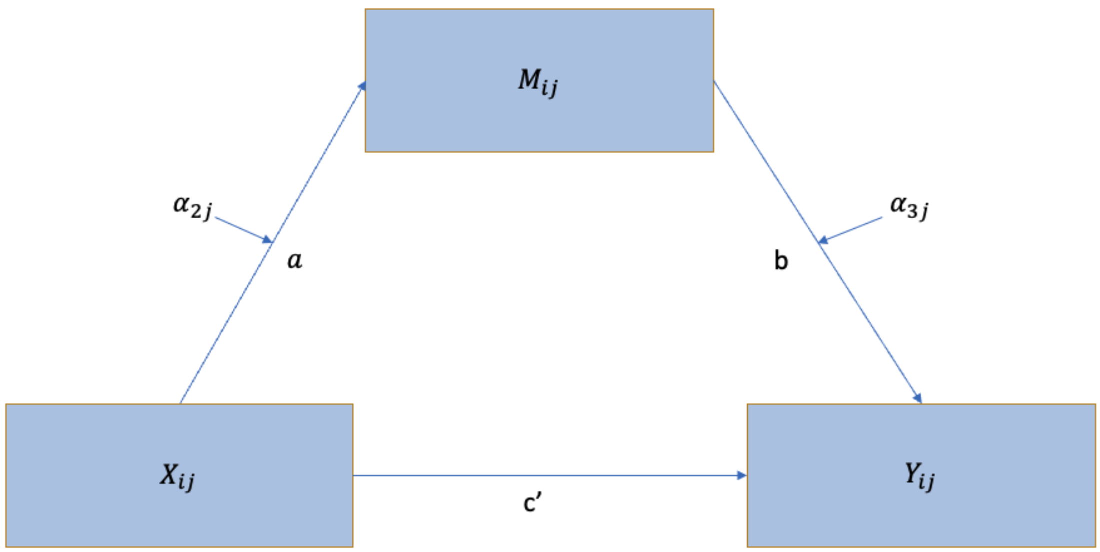

Figure 2: The indirect effect of the independent variable on the dependent variable through the mediator variable

Table 1: **Simulation Results for Low Mediation Effect (δ=0.01, ICC=0.05)**

MM	Imputation Model	JMR	CR	MSE	AW	SB
mcar	FCS_Regresion	0.190	0.94	0.00013	0.033	0.142
	JM		0.90	0.00011	0.022	0.326
	CC		0.96	0.00005	0.033	0.171
	FCS_Regresion	0.361	0.92	0.00009	0.036	0.366
	JM		0.96	0.00011	0.032	0.267
	CC		0.92	0.00007	0.036	0.151
	FCS_Regresion	0.510	0.93	0.00009	0.037	0.369
	JM		0.93	0.00013	0.051	0.100
	CC		0.91	0.00011	0.040	0.266
mar	FCS_Regresion	0.100	0.91	0.00013	0.054	0.111
	JM		0.92	0.00008	0.035	0.281
	CC		0.95	0.00007	0.036	0.391
	FCS_Regresion	0.200	0.91	0.00014	0.061	0.070
	JM		0.96	0.00012	0.055	0.213
	CC		0.92	0.00009	0.042	0.292
	FCS_Regresion	0.300	0.90	0.00010	0.063	0.264
	JM		0.93	0.00015	0.072	0.094
	CC		0.96	0.00014	0.050	**0.460**
mnar	FCS_Regresion	0.193	**0.54**	0.00008	0.020	**0.505**
	JM		**0.86**	0.00006	0.014	0.220
	CC		**0.74**	0.00006	0.023	0.255
	FCS_Regresion	0.368	**0.43**	0.00008	0.019	**0.428**
	JM		**0.73**	0.00008	0.020	**0.429**
	CC		**0.63**	0.00007	0.022	**0.594**
	FCS_Regresion	0.522	**0.43**	0.00009	0.020	**0.526**
	JM		**0.50**	0.00010	0.025	**0.424**
	CC		**0.53**	0.00008	0.021	**0.417**

note. MM: Missingness Mechanism; MR: Missingness Rate; JMR: Joint Missingness Rate; CR: Coverage Rate; MSE: Mean Square Error; AW: Average Width; SB: Standardized Bias; FCS: fully conditional specification; JM: joint modeling; CC: complete-case-only only analysis; mcar: missing completely at random; mar: missing at random; mnar: missing not at random.

Table 2: **Simulation Results for Low Mediation Effect (δ=0.01, ICC=0.1)**

MM	Imputation Model	JMR	CR	MSE	AW	SB
mcar	FCS_Regresion	0.190	0.91	0.00068	0.068	0.129
	JM		0.94	0.00128	0.085	0.225
	CC		0.93	0.00133	0.144	0.300
	FCS_Regresion	0.360	0.91	0.00034	0.083	0.316
	JM		0.97	0.00111	0.138	0.271
	CC		0.94	0.00165	0.162	0.343
	FCS_Regresion	0.510	0.95	0.00027	0.088	0.189
	JM		0.97	0.00160	0.182	0.273
	CC		0.94	0.00208	0.193	0.333
mar	FCS_Regresion	0.100	0.97	0.00063	0.133	0.268
	JM		0.95	0.00184	0.166	0.222
	CC		0.95	0.00174	0.159	0.225
	FCS_Regresion	0.200	0.99	0.00043	0.147	0.381
	JM		0.96	0.00263	0.233	0.286
	CC		0.91	0.00258	0.185	0.381
	FCS_Regresion	0.300	0.96	0.00024	0.140	0.259
	JM		0.97	0.00271	0.312	**0.669**
	CC		**0.87**	0.00266	0.213	**0.664**
mnar	FCS_Regresion	0.193	**0.79**	0.00020	0.045	0.144
	JM		**0.83**	0.00033	0.053	0.343
	CC		**0.85**	0.00037	0.088	0.356
	FCS_Regresion	0.668	**0.81**	0.00013	0.042	**0.454**
	JM		**0.88**	0.00032	0.073	**0.685**
	CC		**0.85**	0.00040	0.090	**0.500**
	FCS_Regresion	0.522	**0.79**	0.00010	0.043	**0.666**
	JM		**0.87**	0.00030	0.091	**0.553**
	CC		**0.85**	0.00044	0.097	**0.425**

note. MM: Missingness Mechanism; MR: Missingness Rate; JMR: Joint Missingness Rate; CR: Coverage Rate; MSE: Mean Square Error; AW: Average Width; SB: Standardized Bias; FCS: fully conditional specification; JM: joint modeling; CC: complete-case-only only analysis; mcar: missing completely at random; mar: missing at random; mnar: missing not at random.

Table 3: **Simulation Results for Low Mediation Effect (δ=0.01, ICC=0.2)**

MM	Imputation Model	JMR	CR	MSE	AW	SB
mcar	FCS_Regresion	0.190	0.94	0.00230	0.129	0.204
	JM		0.93	0.00309	0.141	0.106
	CC		0.94	0.00352	0.243	0.196
	FCS_Regresion	0.359	0.93	0.00136	0.152	0.244
	JM		0.93	0.00394	0.214	0.280
	CC		0.95	0.00444	0.284	0.304
	FCS_Regresion	0.510	0.96	0.00078	0.147	0.341
	JM		0.96	0.00357	0.287	0.300
	CC		0.94	0.00613	0.314	0.218
mar	FCS_Regresion	0.100	0.95	0.00164	0.259	0.265
	JM		0.93	0.00686	0.327	0.282
	CC		0.91	0.00621	0.294	0.288
	FCS_Regresion	0.200	0.97	0.00100	0.254	0.329
	JM		0.97	0.00768	0.449	0.286
	CC		0.94	0.00695	0.345	0.384
	FCS_Regresion	0.300	0.98	0.00070	0.241	0.383
	JM		0.98	0.00922	0.586	0.393
	CC		0.95	0.01025	0.390	**0.511**
mnar	FCS_Regresion	0.193	**0.76**	0.00086	0.070	**0.477**
	JM		**0.88**	0.00170	0.104	**0.479**
	CC		**0.85**	0.00162	0.178	0.320
	FCS_Regresion	0.367	0.93	0.00066	0.087	0.346
	JM		**0.87**	0.00125	0.136	**0.502**
	CC		**0.86**	0.00127	0.165	**0.577**
	FCS_Regresion	0.521	**0.89**	0.00021	0.098	**0.728**
	JM		**0.88**	0.00118	0.161	**0.507**
	CC		**0.86**	0.00159	0.176	**0.513**

note. MM: Missingness Mechanism; MR: Missingness Rate; JMR: Joint Missingness Rate; CR: Coverage Rate; MSE: Mean Square Error; AW: Average Width; SB: Standardized Bias; FCS: fully conditional specification; JM: joint modeling; CC: complete-case-only only analysis; mcar: missing completely at random; mar: missing at random; mnar: missing not at random.

Table 4: **Simulation Results for High Mediation Effect (δ=0.3481, ICC=0.05)**

MM	Imputation Model	JMR	CR	MSE	AW	SB
mcar	FCS_Regresion	0.190	0.93	0.00011	0.034	0.280
	JM		0.97	0.00007	0.021	**0.485**
	CC		0.93	0.00007	0.033	0.268
	FCS_Regresion	0.360	**0.88**	0.00007	0.040	0.332
	JM		0.94	0.00009	0.034	0.391
	CC		0.96	0.00010	0.038	0.346
	FCS_Regresion	0.510	0.94	0.00008	0.043	**0.433**
	JM		0.95	0.00010	0.044	0.352
	CC		0.92	0.00011	0.042	**0.454**
mar	FCS_Regresion	0.100	0.91	0.00014	0.055	0.068
	JM		0.90	0.00010	0.040	0.352
	CC		**0.89**	0.00008	0.037	0.366
	FCS_Regresion	0.200	0.94	0.00012	0.062	0.118
	JM		0.95	0.00013	0.057	0.181
	CC		**0.88**	0.00010	0.042	0.302
	FCS_Regresion	0.300	0.96	0.00010	0.064	0.271
	JM		0.97	0.00019	0.070	0.083
	CC		**0.88**	0.00013	0.088	**0.467**
mnar	FCS_Regresion	0.193	**0.55**	0.00010	0.031	0.264
	JM		**0.59**	0.00006	0.055	0.377
	CC		**0.74**	0.00006	0.053	**0.423**
	FCS_Regresion	0.367	**0.45**	0.00010	0.070	0.243
	JM		**0.56**	0.00007	0.089	**0.582**
	CC		**0.64**	0.00007	0.071	**0.556**
	FCS_Regresion	0.523	**0.45**	0.00010	0.051	**0.693**
	JM		**0.55**	0.00008	0.085	**0.449**
	CC		**0.64**	0.00008	0.092	**0.487**

note. MM: Missingness Mechanism; MR: Missingness Rate; JMR: Joint Missingness Rate; CR: Coverage Rate; MSE: Mean Square Error; AW: Average Width; SB: Standardized Bias; FCS: fully conditional specification; JM: joint modeling; CC: complete-case-only only analysis; mcar: missing completely at random; mar: missing at random; mnar: missing not at random.

Table 5: **Simulation Results for High Mediation Effect (δ=0.3481, ICC=0.1)**

MM	Imputation Model	JMR	CR	MSE	AW	SB
mcar	FCS_Regresion	0.190	0.94	0.00045	0.066	0.160
	JM		0.92	0.00084	0.071	0.179
	CC		0.92	0.00092	0.130	0.139
	FCS_Regresion	0.360	0.95	0.00036	0.084	0.164
	JM		0.94	0.00128	0.120	0.210
	CC		0.92	0.00157	0.150	0.257
	FCS_Regresion	0.510	0.97	0.00019	0.073	0.054
	JM		0.98	0.00111	0.161	0.257
	CC		0.97	0.00153	0.173	0.327
mar	FCS_Regresion	0.100	0.96	0.00045	0.121	0.237
	JM		0.96	0.00138	0.165	0.232
	CC		0.94	0.00137	0.152	0.319
	FCS_Regresion	0.200	0.98	0.00044	0.142	0.292
	JM		0.99	0.00205	0.241	0.335
	CC		0.93	0.00229	0.180	0.324
	FCS_Regresion	0.300	0.99	0.00026	0.138	0.276
	JM		0.98	0.00310	0.316	**0.695**
	CC		0.91	0.00286	0.211	**0.662**
mnar	FCS_Regresion	0.193	**0.84**	0.00023	0.051	0.190
	JM		**0.85**	0.00045	0.060	0.352
	CC		**0.76**	0.00046	0.096	0.381
	FCS_Regresion	0.368	**0.81**	0.00013	0.041	0.369
	JM		0.93	0.00029	0.068	0.347
	CC		**0.78**	0.00032	0.087	**0.475**
	FCS_Regresion	0.522	**0.89**	0.00013	0.042	0.360
	JM		**0.87**	0.00035	0.085	**0.490**
	CC		**0.75**	0.00043	0.091	**0.549**

note. MM: Missingness Mechanism; MR: Missingness Rate; JMR: Joint Missingness Rate; CR: Coverage Rate; MSE: Mean Square Error; AW: Average Width; SB: Standardized Bias; FCS: fully conditional specification; JM: joint modeling; CC: complete-case-only only analysis; mcar: missing completely at random; mar: missing at random; mnar: missing not at random.

Table 6: **Simulation Results for High Mediation Effect (δ=0.3481, ICC=0.2)**

MM	Imputation Model	JMR	CR	MSE	AW	SB
mcar	FCS_Regresion	0.190	0.95	0.00169	0.124	0.279
	JM		0.98	0.00375	0.143	0.278
	CC		0.96	0.00461	0.260	0.194
	FCS_Regresion	0.360	0.98	0.00107	0.134	0.293
	JM		0.92	0.00446	0.244	0.223
	CC		0.95	0.00495	0.287	0.297
	FCS_Regresion	0.509	0.97	0.00052	0.147	0.466
	JM		0.96	0.00687	0.301	0.305
	CC		0.93	0.00803	0.343	0.372
mar	FCS_Regresion	0.100	0.95	0.00168	0.239	0.153
	JM		**0.86**	0.00560	0.297	0.121
	CC		0.91	0.00613	0.286	0.214
	FCS_Regresion	0.200	0.98	0.00115	0.244	0.218
	JM		0.93	0.00879	0.424	0.347
	CC		0.93	0.00813	0.331	0.389
	FCS_Regresion	0.300	0.99	0.00082	0.222	0.393
	JM		0.91	0.00990	0.605	0.290
	CC		0.94	0.00936	0.387	**0.489**
mnar	FCS_Regresion	0.193	**0.87**	0.00083	0.079	0.253
	JM		**0.79**	0.00170	0.090	0.269
	CC		**0.65**	0.00183	0.169	0.302
	FCS_Regresion	0.368	**0.87**	0.00030	0.088	0.260
	JM		**0.85**	0.00110	0.424	**0.529**
	CC		**0.77**	0.00116	0.556	**0.543**
	FCS_Regresion	0.522	**0.83**	0.00023	0.679	**0.669**
	JM		**0.88**	0.00093	0.551	**0.528**
	CC		**0.79**	0.00098	0.564	**0.529**

note. MM: Missingness Mechanism; MR: Missingness Rate; JMR: Joint Missingness Rate; CR: Coverage Rate; MSE: Mean Square Error; AW: Average Width; SB: Standardized Bias; FCS: fully conditional specification; JM: joint modeling; CC: complete-case-only only analysis; mcar: missing completely at random; mar: missing at random; mnar: missing not at random.

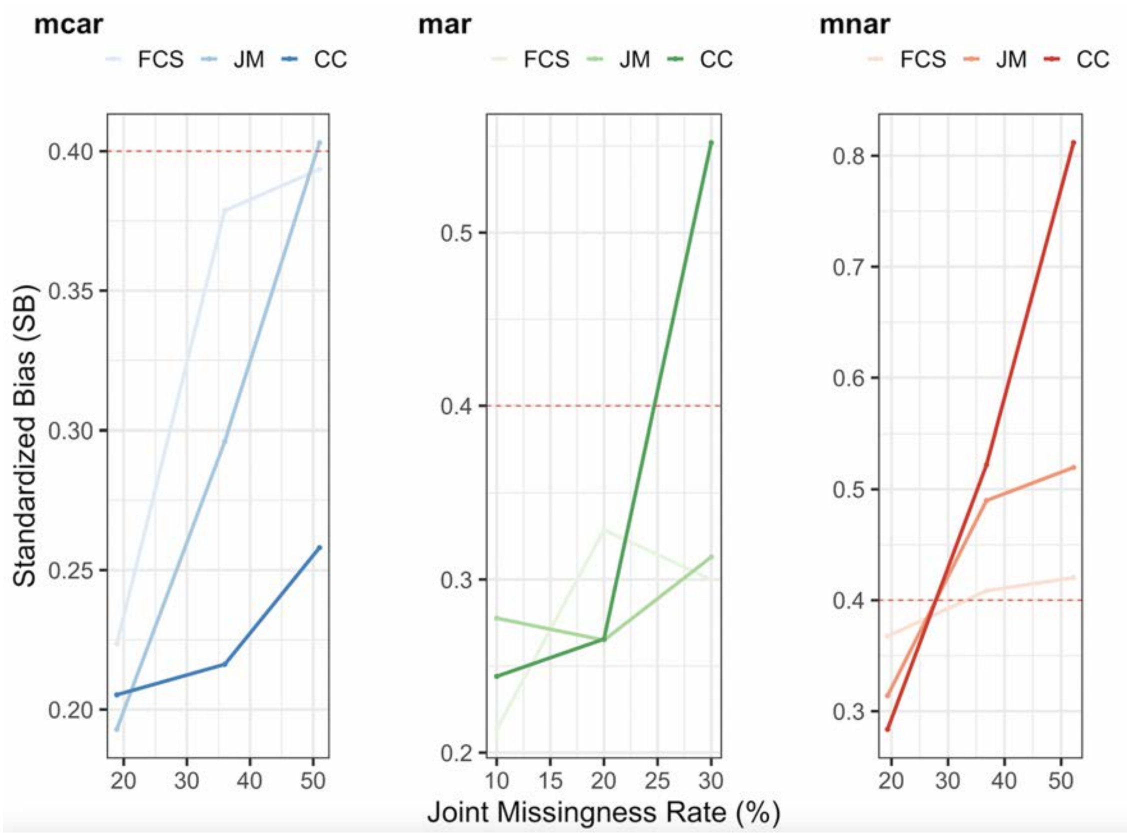

Figure 3: Estimated standardized bias versus joint missingness rate under different missingness mechanisms (ICC=0.05, medium effect size of mediation effect)

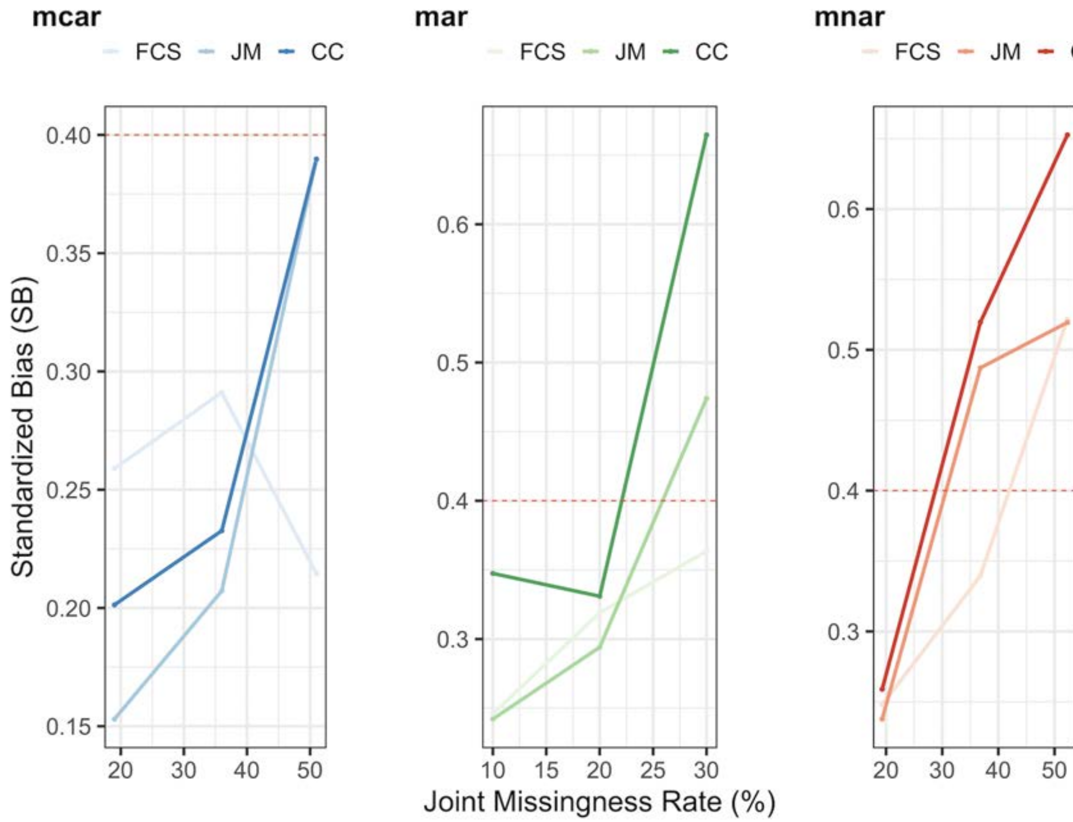

Figure 4: Estimated standardized bias versus joint missingness rate under different missingness mechanisms (ICC=0.1, medium effect size of mediation effect)

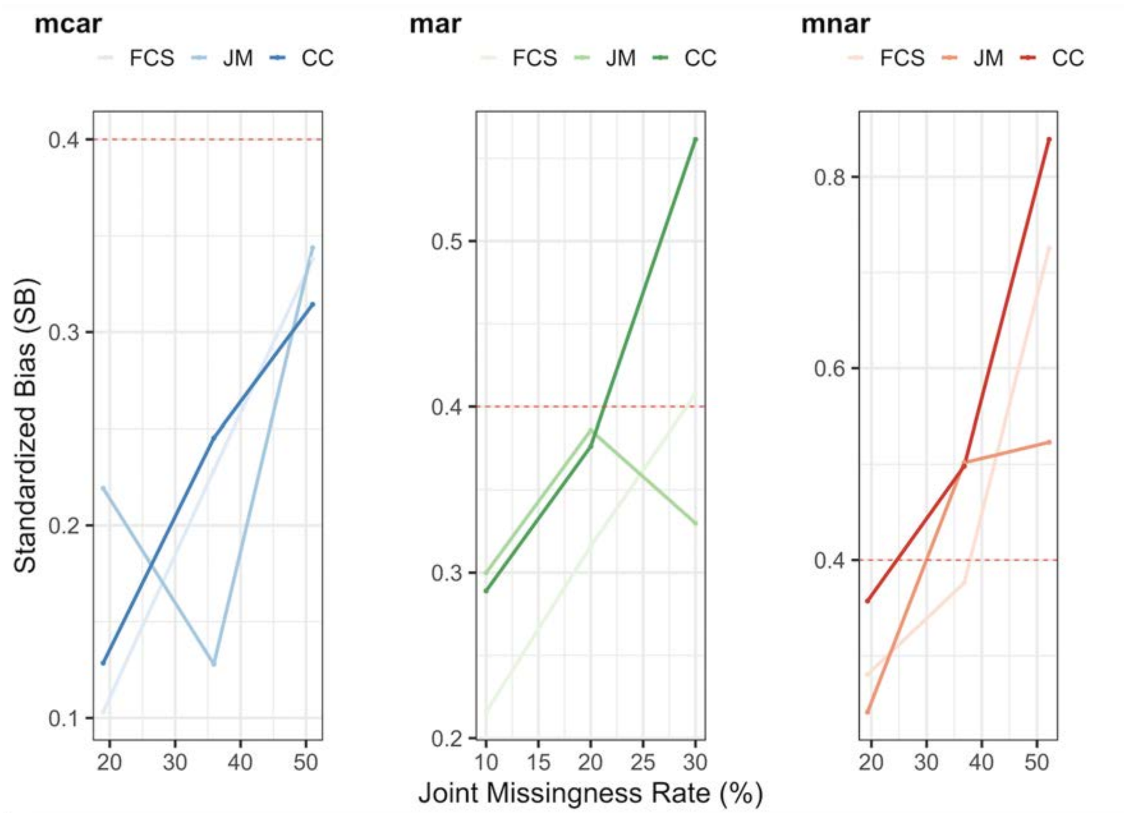

Figure 5: Estimated standardized bias versus joint missingness rate under different missingness mechanisms (ICC=0.2, medium effect size of mediation effect)

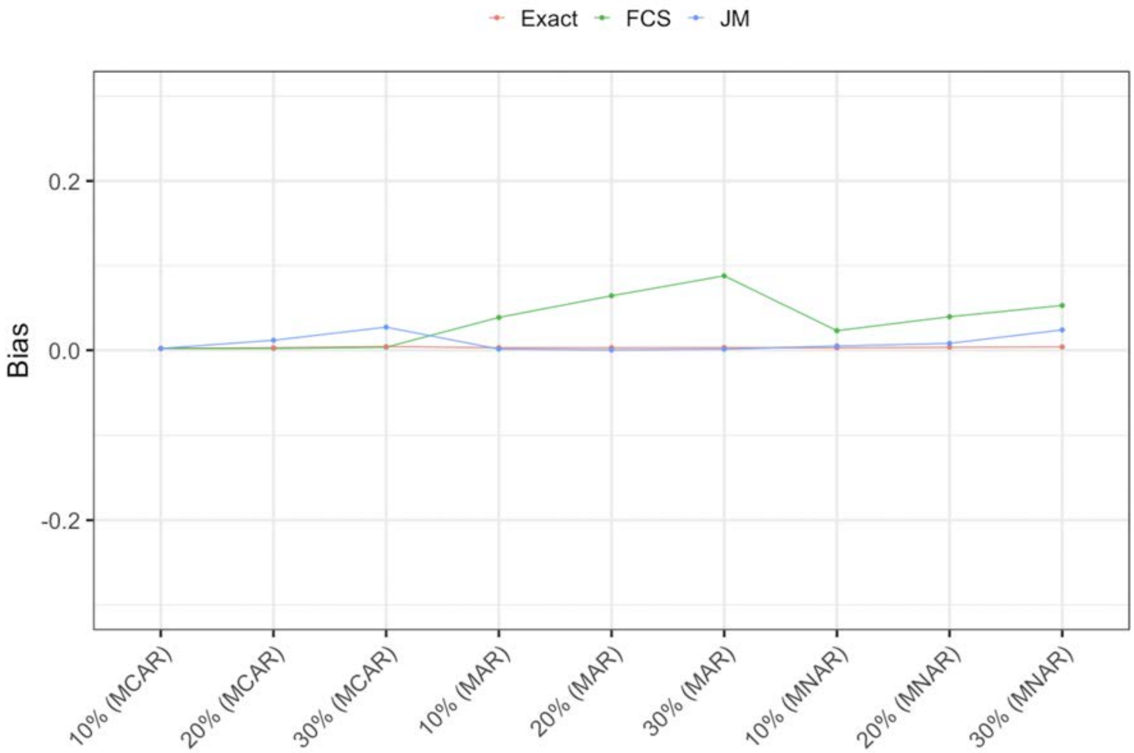

Figure 6: Comparison between exact bias and estimated bias by imputation methods

Table 7: **Mediation effect of attendance rate for Elementary school students**

Test	Mediation effect	Direct	Total
ELA	-0.329	-1.465	-1.794
Math	-0.446	0.383	-0.063
Science	-0.483	0.735	0.253

note. ELA: English Language Arts

Table 8: **Mediation effect of attendance rate for high school students**

Test	Mediation	Direct	Total
Algebra I	0.002	-3.325	-3.323
Chemistry	-0.121	-2.063	-2.185
Earth Science	-0.022	-4.143	-4.165
English	-0.209	-1.708	-1.917
Living Environment	-0.335	-2.652	-2.987
Geometry	-0.251	-2.646	-2.897
Global history	-0.510	-7.332	-7.842
Physics	-0.044	-1.560	-1.604
Trigonometry	-0.158	-6.766	-6.924
U.S. History	-0.178	-0.526	-0.705

8 Appendix

8.1 Derivation for $\boldsymbol{\Sigma}_{22}^{-1}$, $\boldsymbol{\Sigma}_{Y_{ij},M_{ij}|X_{ij},\alpha_Y,\alpha_M}$

We take four steps to calculate the $\boldsymbol{\Sigma}_{22}^{-1}$. We restate $\boldsymbol{\Sigma}_{22}$ as follows:

$$\boldsymbol{\Sigma}_{22} = \begin{bmatrix} \sigma^2_{X(W)} + \sigma^2_{X(B)} & \sigma_{XY(B)} & \sigma_{MX(B)} \\ \sigma_{XY(B)} & \sigma^2_{Y(B)} & 0 \\ \sigma_{MX(B)} & 0 & \sigma^2_{M(B)} \end{bmatrix} \tag{23}$$

In the first step, we calculate the matrix of minors of $\boldsymbol{\Sigma}_{22}$, and we have the matrix of minors for each cell as follows:

$$\begin{bmatrix} \sigma^2_{Y(B)}\sigma^2_{M(B)} & \sigma_{XY(B)}\sigma^2_{M(B)} & -\sigma^2_{Y(B)}\sigma_{MX(B)} \\ \sigma_{XY(B)}\sigma^2_{M(B)} & (\sigma^2_{X(B)} + \sigma^2_{X(W)})\sigma^2_{M(B)} - \sigma^2_{MX(B)} & -\sigma_{XY(B)}\sigma_{MX(B)} \\ -\sigma^2_{Y(B)}\sigma_{MX(B)} & -\sigma_{XY(B)}\sigma_{MX(B)} & (\sigma^2_{X(B)} + \sigma^2_{X(B)})\sigma^2_{Y(B)} - \sigma^2_{XY(B)} \end{bmatrix}$$

Then, we employ the second step to turn the matrix of minors into the Matrix of Cofactors as follows:

$$\begin{bmatrix} \sigma^2_{Y(B)}\sigma^2_{M(B)} & -\sigma_{XY(B)}\sigma^2_{M(B)} & -\sigma^2_{Y(B)}\sigma_{MX(B)} \\ -\sigma_{XY(B)}\sigma^2_{M(B)} & (\sigma^2_{X(B)} + \sigma^2_{X(W)})\sigma^2_{M(B)} - \sigma^2_{MX(B)} & \sigma_{XY(B)}\sigma_{MX(B)} \\ -\sigma^2_{Y(B)}\sigma_{MX(B)} & \sigma_{XY(B)}\sigma_{MX(B)} & (\sigma^2_{X(B)} + \sigma^2_{X(B)})\sigma^2_{Y(B)} - \sigma^2_{XY(B)} \end{bmatrix}$$

The third step is to transpose the matrix. We have a symmetric matrix and it remains the same. The fourth step is to multiple the matrix by corresponding $\frac{1}{|\Sigma|}$. The determinant $|\Sigma|$ is calculated as

$$|\Sigma| = -\sigma^2_{MX(B)}\sigma^2_{Y(B)} + \sigma^2_{M(B)}(\sigma^2_{X(W)}\sigma^2_{Y(B)} + \sigma^2_{X(B)}\sigma^2_{Y(B)} - \sigma^2_{XY(B)}) \tag{24}$$

Therefore, we have the $|\Sigma_{22}^{-1}|$ as follows:

$$\Sigma_{22}^{-1} = \frac{1}{|\Sigma|} \times \begin{bmatrix} \sigma^2_{Y(B)}\sigma^2_{M(B)} & -\sigma_{XY(B)}\sigma^2_{M(B)} & -\sigma^2_{Y(B)}\sigma_{MX(B)} \\ -\sigma_{XY(B)}\sigma^2_{M(B)} & (\sigma^2_{X(B)} + \sigma^2_{X(W)})\sigma^2_{M(B)} - \sigma^2_{MX(B)} & \sigma_{XY(B)}\sigma_{MX(B)} \\ -\sigma^2_{Y(B)}\sigma_{MX(B)} & \sigma_{XY(B)}\sigma_{MX(B)} & (\sigma^2_{X(B)} + \sigma^2_{X(B)})\sigma^2_{Y(B)} - \sigma^2_{XY(B)} \end{bmatrix}$$

After we calculate the Σ_{22}^{-1}, we employ the following formula of calculating conditional variance:

$$\boldsymbol{\Sigma}_{Y,M|X,\alpha_Y,\alpha_M} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \tag{25}$$

As we have shown the Σ_{11}, Σ_{12}, Σ_{22}^{-1} and Σ_{21}, we have $\boldsymbol{\Sigma}_{Y,M|X,\alpha_Y,\alpha_M}$ as follow:

$$\boldsymbol{\Sigma}_{Y,M|X,\alpha_Y,\alpha_M} = \begin{bmatrix} f & g \\ h & k \end{bmatrix} \tag{26}$$

where

$$f = \sigma^2_{Y(W)} + \sigma^2_{Y(B)} - \frac{1}{|\Sigma|}[\sigma_{YX(W)}\sigma^2_{Y(B)}\sigma^2_{M(B)}\sigma_{XY(W)} + \sigma^2_{Y(B)}\sigma^2_{M(B)}(\sigma^2_{X(W)} + \sigma^2_{X(B)})$$
$$- 2\sigma_{YM(B)}\sigma_{MX(B)}\sigma^2_{Y(B)}\sigma_{XY(W)} - \sigma^4_{Y(B)}\sigma^2_{MX(B)} - \sigma^2_{YX(B)}\sigma^2_{M(B)}\sigma^2_{Y(B)}$$
$$+ \sigma^2_{YM(B)}\sigma^2_{Y(B)}(\sigma^2_{X(W)} + \sigma^2_{X(B)}] - \sigma^2_{YM(B)}\sigma^2_{XY(B)}$$
$$g = \sigma_{YM(W)} + \sigma_{YM(B)} + \sigma_{YX(W)}\sigma^2_{Y(B)}\sigma^2_{M(B)}\sigma_{MX(W)} - \sigma_{YM(B)}\sigma_{MX(B)}\sigma^2_{Y(B)}\sigma_{MX(W)}$$
$$- \sigma_{YM(B)}\sigma^2_{MX(B)}\sigma^2_{Y(B)} - \sigma_{YX(W)}\sigma_{XY(B)}\sigma^2_{M(B)}\sigma_{MY(B)} - \sigma^2_{YX(B)}\sigma^2_{M(B)}\sigma_{MY(B)}$$
$$\sigma^2_{MY(B)}\sigma_{MX(B)}\sigma_{XY(B)} + \sigma^2_{M(B)}\sigma_{YM(B)}\sigma^2_{Y(B)}\sigma^2_{X(W)} + \sigma^2_{M(B)}\sigma_{YM(B)}\sigma^2_{Y(B)}\sigma^2_{X(B)}$$
$$- \sigma^2_{M(B)}\sigma_{YM(B)}\sigma^2_{XY(B)}$$
$$h = \sigma_{YM(W)} + \sigma_{YM(B)} + \sigma_{MX(W)}\sigma^2_{Y(B)}\sigma^2_{M(B)}\sigma_{XY(W)} - \sigma_{YM(B)}\sigma_{XY(B)}\sigma^2_{M(B)}\sigma_{XY(W)}$$
$$- \sigma_{YM(B)}\sigma_{XY(B)})\sigma^2_{M(B)}\sigma_{XY(B)} + \sigma^2_{Y(B)}\sigma_{YM(B)}\sigma^2_{X(W)}\sigma^2_{M(B)} + \sigma^2_{Y(B)}\sigma_{YM(B)}\sigma^2_{X(B)}\sigma^2_{M(B)}$$
$$- \sigma^2_{Y(B)}\sigma_{YM(B)}\sigma^2_{MX(B)} - \sigma_{YM(B)}\sigma_{MX(W)}\sigma^2_{Y(B)}\sigma_{MX(B)} - \sigma_{YM(B)}\sigma_{MX(B)}\sigma^2_{Y(B)}\sigma_{MX(B)}$$
$$+ \sigma^2_{YM(B)}\sigma_{XY(B)}\sigma_{MX(B)} + \sigma_{YM(B)}\sigma^2_{M(B)}\sigma^2_{Y(B)} + \sigma_{YM(B)}\sigma^2_{X(W)} + \sigma_{YM(B)}\sigma^2_{X(B)}$$
$$- \sigma_{YM(B)}\sigma^2_{XY(B)}\sigma^2_{M(B)}$$
$$k = \sigma^2_{M(W)} + \sigma^2_{M(B)} + \sigma_{MX(W)}\sigma^2_{Y(B)}\sigma^2_{M(B)}\sigma_{MX(W)} - \sigma_{YM(B)}\sigma_{XY(B)}\sigma^2_{M(B)}\sigma_{MX(W)}$$
$$- \sigma_{MY(B)}\sigma_{MX(W)}\sigma_{XY(B)}\sigma_{M(B)} + \sigma^2_{MY(B)}\sigma^2_{X(W)}\sigma^2_{M(B)} + \sigma^2_{MY(B)}\sigma^2_{X(B)}\sigma^2_{M(B)}$$
$$- \sigma^2_{MY(B)}\sigma^2_{MX(B)} - \sigma^2_{M(B)}\sigma^2_{MX(B)}\sigma^2_{Y(B)} + \sigma^4_{M(B)}\sigma^2_{Y(B)} + \sigma^2_{M(B)}\sigma^2_{X(W)} + \sigma^2_{M(B)}\sigma^2_{X(B)}$$
$$- \sigma^2_{M(B)}\sigma^2_{XY(B)}\sigma^2_{M(B)} \tag{27}$$

8.2 Derivation of MLE estimator for $\hat{c}$, $\hat{a}$, $\hat{b}$ and $\hat{\delta}$

We have the mixed model as follows:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}_c + \boldsymbol{\alpha}_1 + \boldsymbol{\epsilon}_1 \tag{28}$$

We rewrite the mixed model as $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}_c + \boldsymbol{\epsilon}^*_1$, where $\boldsymbol{\epsilon}^*_1 = \boldsymbol{\alpha}_1 + \boldsymbol{\epsilon}_1$, and $\boldsymbol{\epsilon}^*_1 \sim N(\boldsymbol{0}, \boldsymbol{V})$, with $\boldsymbol{V} = \boldsymbol{G} + \boldsymbol{R}$. We have $\boldsymbol{V} = \boldsymbol{V}^{\frac{1}{2}}(\boldsymbol{V}^{\frac{1}{2}})^T$. We multiply $\boldsymbol{V}^{-\frac{1}{2}}$ on the both side of $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}_c + \boldsymbol{\epsilon}^*_1$, and we have:

$$\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{Y} = \boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{\epsilon}^*_1 \tag{29}$$

where

$$\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{\epsilon}^*_1 \sim N(\boldsymbol{0}, \boldsymbol{I}_n) \tag{30}$$

Therefore, we generate the equation for $\hat{\boldsymbol{\beta}}$ as follows:

$$\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{Y} = \boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{X}\hat{\boldsymbol{\beta}} \tag{31}$$

We multiply $\boldsymbol{X}^T(\boldsymbol{V}^{-\frac{1}{2}})^T$ on both sides of the equation above, and we have:

$$\boldsymbol{X}^T(\boldsymbol{V}^{-\frac{1}{2}})^T\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{Y} = \boldsymbol{X}^T(\boldsymbol{V}^{-\frac{1}{2}})^T\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{X}\hat{\boldsymbol{\beta}} \tag{32}$$

Thus, we find MLE or weighted LSE of $\hat{\beta}$ by multiplying $\boldsymbol{X}^T(\boldsymbol{V}^{-\frac{1}{2}})^T\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{X}$ on both sides of the Equation above, and we have

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{Y} \tag{33}$$

$\boldsymbol{X}$ is a $n \times 1$ matrix, and $\hat{\boldsymbol{\beta}}$ in this case is a scalar, and we have $\hat{c} = \hat{\boldsymbol{\beta}}$. Similar procedures are employed to estimate $\hat{a}$ and $\hat{b}$ based on different mixed models as follows:

$$\boldsymbol{M} = \boldsymbol{X}\boldsymbol{\beta}_{\boldsymbol{a}} + \boldsymbol{\alpha}_2 + \boldsymbol{\epsilon}_2 \tag{34}$$

$$\boldsymbol{Y} = \boldsymbol{X}'\boldsymbol{\beta}_{bc} + \boldsymbol{\alpha}_3 + \boldsymbol{\epsilon}_3 \tag{35}$$

8.3 Derivation of $f(Y|M,X)$ and $E(Y|M,X)$ for FCS

We have the distribution of Y_{ij} given α_{Y_i}, M_{ij} and X_{ij}, and the corresponding prior distribution of α_{Y_i}. We find the distribution of Y_{ij}, α_{Y_i} given M_{ij}, X_{ij} as follows:

$$\begin{aligned} f(Y_{ij}, \alpha_{Y_i}|M_{ij}, X_{ij}) &= f(Y_{ij} \mid M_{ij}, X_{ij}, \alpha_{Y_i}) \times f(\alpha_{Y_i}) \\ &= \frac{1}{\sqrt{2\pi}\sigma_{Y|M,X}} \times e^{\frac{-(Y_{ij}-\alpha_{Y_i}-\beta_{Y|X}X_{ij}-\beta_{Y|M}M_{ij})^2}{2\sigma^2_{Y|M,X}}} \times \\ &\quad \frac{1}{\sqrt{2\pi}\sigma_{\alpha_{Y_i}}} \times e^{\frac{-\alpha^2_{Y_i}}{2\sigma^2_{\alpha_{Y_i}}}} \end{aligned} \tag{36}$$

We reformulate the distribution of Y_{ij}, given $\alpha_{Y_i}, M_{ij}X_{ij}$ into a function of α_{Y_i}, and we have:

$$\begin{aligned} f(Y_{ij}, \alpha_{Y_i}|M_{ij}, X_{ij}) &= \frac{e^{-(\frac{1}{2\sigma^2_{Y|M,X}}+\frac{1}{2\sigma^2_{\alpha_{Y_i}}})(\alpha^2_{Y_i}-\frac{\beta_{Y|X}X_{ij}+\beta_{Y|M}M_{ij}-Y_{ij}}{\frac{1}{2}+\frac{\sigma^2_{Y|M,X}}{2\sigma^2_{\alpha_{Y_i}}}}\alpha_{Y_i})}}{2\pi\sigma_{Y|M,X}\sigma_{\alpha_{Y_i}}} \times \\ &\quad \frac{1}{2\pi\sigma_{Y|M,X}\sigma_{\alpha_{Y_i}}} e^{-\frac{(\beta_{Y|X}X_{ij}+\beta_{Y|M}M_{ij}-Y_{ij})^2}{2\sigma^2_{Y|M,X}}} \\ &= \frac{1}{2\pi\sigma_{Y|M,X}\sigma_{\alpha_{Y_i}}} e^{-(\frac{1}{2\sigma^2_{Y|M,X}}+\frac{1}{2\sigma^2_{\alpha_{Y_i}}})[\alpha_{Y_i}-\frac{\sigma^2_{\alpha_{Y_i}}(\beta_{Y|X}X+\beta_{Y|M}M-Y)}{\sigma^2_{\alpha_{Y_i}}+\sigma^2_{Y|M,X}}]^2+G} \end{aligned} \tag{37}$$

where

$$G = \frac{(\beta_{Y|X}X+\beta_{Y|M}M-Y)^2}{2\sigma^2_{Y|M,X}+\frac{2\sigma^4_{Y|M,X}}{\sigma^2_{\alpha_{Y_i}}}} - \frac{(\beta_{Y|X}X+\beta_{Y|M}M-Y)^2}{2\sigma^2_{Y|M,X}} \tag{38}$$

Then, we develop the conditional distribution of Y given M, X by integral the distribution $f(Y_{ij}, \alpha_{Y_i}|M_{ij}, X_{ij})$ with regard to α_{Y_i}, and we have $f(Y|M,X)$ as follows:

$$\begin{aligned} f(Y|M,X) &= \int f(Y_{ij}, \alpha_{Y_i}|M_{ij}, X_{ij}) d\alpha_{Y_i} \\ &= \frac{1}{\sqrt{2\pi}\sigma_{Y|M,X}\sigma_{\alpha_{Y_i}}} \times \sqrt{\frac{\sigma^2_{Y|M,X}\sigma^2_{\alpha_{Y_i}}}{\sigma^2_{Y|M,X}+\sigma^2_{\alpha_{Y_i}}}} e^{\frac{(\beta_{Y|X}X+\beta_{Y|M}M-Y)^2}{2\sigma^2_{Y|M,X}+\frac{2\sigma^4_{Y|M,X}}{\sigma^2_{\alpha_{Y_i}}}}} \times \\ &\quad e^{-\frac{(\beta_{Y|X}X+\beta_{Y|M}M-Y)^2}{2\sigma^2_{Y|M,X}}} \\ &= \frac{1}{\sqrt{2\pi(\sigma^2_{Y|M,X}+\sigma^2_{\alpha_{Y_i}})}} e^{-\frac{(Y-\beta_{Y|X}X-\beta_{Y|M}M)^2}{2(\sigma^2_{\alpha_{Y_i}}+\sigma^2_{Y|M,X})}} \end{aligned} \tag{39}$$

Therefore, we find that $f(Y|M,X)$ is a normal distribution with mean $\beta_{Y|X}X+\beta_{Y|M}M$ (i.e., $E(Y|M,X) = \beta_{Y|X}X + \beta_{Y|M}M$) and variance $\sigma^2_{\alpha_{Y_i}} + \sigma^2_{Y|M,X}$.

8.4 Derivation of $f(M|Y,X)$ for FCS

We have the distribution of M_{ij} given α_{M_i}, Y_{ij} and X_{ij} and the corresponding prior distribution of α_{M_i}. We find the distribution of M_{ij}, α_{M_i} given Y_{ij} and X_{ij} as follows:

$$\begin{aligned} f(M_{ij}, \alpha_{M_i}|Y_{ij}, X_{ij}) &= f(M_{ij} \mid Y_{ij}, X_{ij}, \alpha_{M_i}) \times f(\alpha_{M_i}) \\ &= \frac{1}{\sqrt{2\pi}\sigma_{M|Y,X}} \times e^{\dfrac{-(M_{ij} - \alpha_{M_i} - \beta_{M|X}X_{ij} - \beta_{M|Y}Y_{ij})^2}{2\sigma^2_{M|Y,X}}} \times \\ &\quad \frac{1}{\sqrt{2\pi}\sigma_{\alpha_{M_i}}} \times e^{\dfrac{-\alpha^2_{M_i}}{2\sigma^2_{\alpha_{M_i}}}} \end{aligned} \tag{40}$$

Similarly as we did for the distribution of M_{ij} given α_{M_i}, Y_{ij} and X_{ij}, we reconstruct the distribution of M_{ij}, α_{M_i} given Y_{ij} and X_{ij} into a function of α_{M_i}, and we have:

$$\begin{aligned} f(M_{ij}, \alpha_{M_i}|Y_{ij}, X_{ij}) &= \frac{e^{-(\frac{1}{2\sigma^2_{M|Y,X}} + \frac{1}{2\sigma^2_{\alpha_{M_i}}})(\alpha^2_{M_i} - \frac{\beta_{MX}X_{ij} + \beta_{M|Y}Y_{ij} - M_{ij}}{\frac{1}{2} + \frac{\sigma^2_{M|Y,X}}{2\sigma^2_{\alpha_{M_i}}}}\alpha_{M_i})}}{2\pi\sigma_{M|Y,X}\sigma_{\alpha_{M_i}}} \times \\ &\quad \frac{1}{2\pi\sigma_{M|Y,X}\sigma_{\alpha_{M_i}}} e^{-\dfrac{(\beta_{M|X}X_{ij} + \beta_{M|Y}Y_{ij} - M_{ij})^2}{2\sigma^2_{M|Y,X}}} \\ &= \frac{1}{2\pi\sigma_{M|Y,X}\sigma_{\alpha_{M_i}}} e^{-(\frac{1}{2\sigma^2_{M|Y,X}} + \frac{1}{2\sigma^2_{\alpha_{M_i}}})[\alpha_{M_i} - \frac{\sigma^2_{\alpha_{M_i}}(\beta_{M|X}X + \beta_{M|Y}Y - M)}{\sigma^2_{\alpha_{M_i}} + \sigma^2_{M|Y,X}}]^2 + G_1} \end{aligned} \tag{41}$$

where

$$G_1 = \frac{(\beta_{M|X}X + \beta_{M|Y}Y - M)^2}{2\sigma^2_{M|Y,X} + \dfrac{2\sigma^4_{M|Y,X}}{\sigma^2_{\alpha_{M_i}}}} - \frac{(\beta_{M|X}X + \beta_{M|Y}Y - M)^2}{2\sigma^2_{M|Y,X}} \tag{42}$$

Then, we develop the conditional distribution of M given Y, X by integral the distribution $f(M_{ij}, \alpha_{M_i}|Y_{ij}, X_{ij})$ with regard to α_{M_i}, and we have $f(M|Y,X)$ as follows:

$$\begin{aligned} f(M|Y,X) &= \int f(M_{ij}, \alpha_{M_i}|Y_{ij}, X_{ij}) d\alpha_{M_i} \\ &= \frac{1}{\sqrt{2\pi}\sigma_{M|Y,X}\sigma_{\alpha_{M_i}}} \times \sqrt{\frac{\sigma^2_{M|Y,X}\sigma^2_{\alpha_{M_i}}}{\sigma^2_{M|Y,X} + \sigma^2_{\alpha_{M_i}}}} e^{\dfrac{(\beta_{M|X}X + \beta_{M|Y}Y - M)^2}{2\sigma^2_{M|Y,X} + \frac{2\sigma^4_{M|Y,X}}{\sigma^2_{\alpha_{M_i}}}}} \times \\ &\quad e^{-\dfrac{(\beta_{M|X}X + \beta_{M|Y}Y - M)^2}{2\sigma^2_{M|Y,X}}} \\ &= \frac{1}{\sqrt{2\pi(\sigma^2_{M|Y,X} + \sigma^2_{\alpha_{M_i}})}} e^{-\dfrac{(M - \beta_{M|X}X - \beta_{M|Y}Y)^2}{2(\sigma^2_{\alpha_{M_i}} + \sigma^2_{M|Y,X})}} \end{aligned} \tag{43}$$

Therefore, we find that $f(M|Y,X)$ is a normal distribution with mean $\beta_{M|X}X + \beta_{M|Y}Y$ (i.e., $E(M|Y,X) = \beta_{M|X}X + \beta_{M|Y}Y$) and variance $\sigma^2_{\alpha_{M_i}} + \sigma^2_{M|Y,X}$.

8.5 Derivation of $f(M|X)$ and $E(M|X)$ for FCS

Firstly, we find the expression of the conditional distribution of M given Y and X as follows:

$$\begin{aligned} f(M,Y|X) &= f(M|Y,X)f(Y) \\ &= \frac{1}{\sqrt{2\pi(\sigma^2_{M|Y,X}+\sigma^2_{\alpha_M})}} e^{-\frac{(M-\beta_{M|X}X-\beta_{M|Y}Y)^2}{2(\sigma^2_{M|Y,X}+\sigma^2_{\alpha})}} \times \\ &\quad \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{(Y-\mu_Y)^2}{2\sigma^2_Y}} \end{aligned} \tag{44}$$

Then, we reformulate the distribution of Y, M given X as a function of Y, we have:

$$f(M,Y|X) = \frac{1}{2\pi\sigma_Y\sqrt{\sigma^2_{M|Y,X}+\sigma^2_{\alpha_M}}} e^{-A_1(Y-B_1)^2} + C_1 \tag{45}$$

where

$$\begin{aligned} A_1 &= \frac{\beta^2_{M|Y}}{2\sigma^2_{M|Y,X}+2\sigma^2_{\alpha_M}} + \frac{1}{2\sigma^2_Y} \\ B_1 &= \frac{\mu_Y(\sigma^2_{M|Y,X}+\sigma^2_{\alpha}) - \beta_{M|Y}(\beta_{M|X}X-M)\sigma^2_Y}{\beta^2_{M|Y}\sigma^2_Y+\sigma^2_{M|Y,X}+\sigma^2_{\alpha_M}} \\ C_1 &= \frac{\sigma^2_Y(\sigma^2_{M|Y,X}+\sigma^2_{\alpha_M})}{2(\beta^2_{M|Y}\sigma^2_Y+\sigma^2_{M|Y,X}+\sigma^2_{\alpha_M})}\left(\frac{\mu_Y}{\sigma^2_Y} - \frac{\beta_{M|X}(\beta_{M|X}-M)}{\sigma^2_{M|Y,X}+\sigma^2_{\alpha_M}}\right)^2 - \\ &\quad \frac{(\beta_{M|X}X-M)^2}{2(\sigma^2_{M|Y,X}+\sigma^2_{\alpha_M})} - \frac{\mu^2_Y}{2\sigma^2_Y} \end{aligned} \tag{46}$$

Therefore, we integrate the $f(M,Y|X)$ with respect to Y, and we have the distribution of M given X as follows:

$$\begin{aligned} f(M|X) &= \int f(M,Y|X)dY \\ &= \frac{1}{\sqrt{\pi(\beta^2_{M|Y}\sigma^2_Y+\sigma^2_{M|Y,X}+\sigma^2_{\alpha_M})}} e^{A_2M^2+B_2M+C_2} \end{aligned} \tag{47}$$

where

$$\begin{aligned} A_2 &= \frac{\beta^2_{M|Y}\sigma^2_Y}{2(\beta^2_{M|Y}\sigma^2_Y+\sigma^2_{M|Y,X}+\sigma^2_{\alpha_M})(\sigma^2_{M|Y,X}+\sigma^2_{\alpha_M})} - \frac{1}{2(\sigma^2_{M|Y,X}+\sigma^2_{\alpha_M})} \\ B_2 &= \frac{\mu_Y\beta_{M|Y}}{\beta^2_{M|Y}\sigma^2_Y+\sigma^2_{M|Y,X}+\sigma^2_{\alpha_M}} - \frac{\sigma^2_Y\beta^2_{M|Y}\beta_{M|X}X}{(\beta^2_{M|Y}\sigma^2_Y+\sigma^2_{M|Y,X}+\sigma^2_{\alpha_M})(\sigma^2_{M|Y,X}+\sigma^2_{\alpha_M})} + \\ &\quad \frac{\beta_{M|X}X}{\sigma^2_{M|Y,X}+\sigma^2_{\alpha_M}} \\ C_2 &= -\frac{\mu^2_Y}{2\sigma^2_Y} - \frac{\beta^2_{M|X}X^2}{2(\sigma^2_{M|Y,X}+\sigma^2_{\alpha_M})} + \frac{\mu^2_Y(\sigma^2_{M|Y,X}+\sigma^2_{\alpha_M})}{2\sigma^2_Y(\beta^2_{M|Y}\sigma^2_Y+\sigma^2_{M|Y,X}+\sigma^2_{\alpha_M})} - \\ &\quad \frac{\mu_Y\beta_{M|Y}\beta_{M|X}X}{\beta^2_{M|X}\sigma^2_Y+\sigma^2_{M|Y,X}+\sigma^2_{\alpha_M}} + \frac{\sigma^2_Y\beta^2_{M|Y}\beta^2_{M|X}X^2}{2(\beta_{M|Y}\sigma^2_Y+\sigma^2_{\alpha_M}+\sigma^2_{M|Y,X})(\sigma^2_{M|Y,X}+\sigma^2_{\alpha_M})} \end{aligned} \tag{48}$$

8.6 Derivation of $f(Y,M|X)$ for JM

We find the distribution of $Y_{ij}, M_{ij}, \alpha_{Y_i}, \alpha_{M_i}$ given X_{ij} by multiplying the conditional distribution of Y_{ij}, M_{ij} given $X_{ij}, \alpha_{Y_i}, \alpha_{M_i}$ and the prior distribution of $\alpha_{Y_i}, \alpha_{M_i}$ as follows:

$$\begin{aligned} f(Y_{ij}, M_{ij}, \alpha_{Y_i}, \alpha_{M_i}|X_{ij}) &= f(Y_{ij}, M_{ij}|X_{ij}, \alpha_{Y_i}, \alpha_{M_i}) \times f(\alpha_{Y_i}, \alpha_{M_i}) \\ &= \frac{1}{4\pi^2\sqrt{\mathbf{\Sigma}_C\mathbf{\Sigma}_\alpha}} e^{-\frac{k(Y-\alpha_Y-\beta_{Y|X}X)^2+f(M-\alpha_M-\beta_{M|X}X)^2}{2(fk-gh)}} \times \\ & e^{-\frac{(g+h)(M-\alpha_M-\beta_{M|X}X)(Y-\alpha_Y-\beta_{Y|X}X)}{2(fk-gh)}} \times \\ & e^{\frac{1}{2\sigma_Y^2\sigma_M^2}(\sigma_M^2\alpha_Y^2+\sigma_Y^2\alpha_M^2)} \end{aligned} \tag{49}$$

In the above derivation, we have the expression of $\mathbf{\Sigma}_C$ ($\mathbf{\Sigma}_{Y_{ij},M_{ij}|\alpha_{Y_i},\alpha_{M_i},X_{ij}}$ in Equation 13, we use $\mathbf{\Sigma}_C$ to simplify the notation), and the expression of $\mathbf{\Sigma}_\alpha$ is shown in Equation 12. We find the conditional distribution of Y, M given X by integrating conditional distribution of $Y_{ij}, M_{ij}, \alpha_{Y_i}, \alpha_{M_i}$ given X_{ij} regarding α_{Y_i} and α_{M_i} to obtain the conditional distribution of Y, M given X ($f(Y,M|X)$) as follows:

$$f(Y_{ij}, M_{ij}|X_{ij}) = \iint f(Y_{ij}, M_{ij}, \alpha_{Y_i}, \alpha_{M_i}|X_{ij}) d\alpha_{Y_i} d\alpha_{M_i} \tag{50}$$

We firstly reformulate the $f(Y_{ij}, M_{ij}, \alpha_{Y_i}, \alpha_{M_i}|X_{ij})$ into a function of α_Y, therefore, we have:

$$f(Y_{ij}, M_{ij}, \alpha_{Y_i}, \alpha_{M_i}|X_{ij}) = \frac{1}{4\pi^2\sqrt{\mathbf{\Sigma}_C\mathbf{\Sigma}_\alpha}} e^{A_3\alpha_Y^2+B_3\alpha_Y+C_3} \tag{51}$$

where

$$\begin{aligned} A_3 &= -\frac{k}{2(fk-gh)} - \frac{1}{2\sigma_Y^2} \\ B_3 &= -\frac{k(\beta_{Y|X}X-Y)+(g+h)(M-\alpha_M-\alpha_{M|X}X)}{fk-gh} \\ C_3 &= -\frac{k(\beta_{Y|X}X-Y)^2-(g+h)(M-\alpha_M-\beta_{M|X}X)(Y-\beta_{Y|X}X)}{2(fk-gh)} - \\ & \frac{f(M-\alpha_M-\beta_{M|X}X)^2}{2(fk-gh)} \end{aligned} \tag{52}$$

Then we find the expression of $f(Y_{ij}, M_{ij}, \alpha_M|X_{ij})$ and reformulate it into a function of α_M as follows:

$$f(Y_{ij}, M_{ij}, \alpha_M|X) = \frac{\sqrt{(fk-gh)\sigma_Y^2}}{2\pi\sqrt{\pi(k\sigma_Y^2+fk-gh)\mathbf{\Sigma}_C\mathbf{\Sigma}_\alpha}} e^{A_4\alpha_M^2+B_4\alpha_M+C_4} \tag{53}$$

where

$$
\begin{aligned}
A_4 &= \frac{(g+h)^2(fk-gh)\sigma_Y^2}{2(k\sigma_Y^2+fk-gh)(fk-gh)^2} \\
B_4 &= \frac{(g+h)[k(\beta_{Y|X}X-Y)+(g+h)(M-\beta_{M|X}X)]^2\sigma_Y^2(fk-gh)}{(k\sigma_Y^2+fk-gh)(fk-gh)^2}+ \\
&\frac{(g+h)(Y-\beta_{Y|X}X)-2a(M-\beta_{M|x}X)}{2(fk-gh)} \\
C_4 &= \frac{[k(\beta_{Y|X}X-Y)+(g+h)(M-\beta_{M|X}X)]^2\sigma_Y^2(fk-gh)}{2(k\sigma_Y^2+fk-gh)(fk-gh)}- \\
&\frac{k[\beta_{Y|X}X-Y]^2-(g+h)(M-\beta_{M|X}X)(Y-\beta_{Y|X}X)+f(M-\beta_{M|X}X)^2}{2(fk-gh)}
\end{aligned} \tag{54}
$$

Then, we find the expression of $f(Y,M|X)$ by integrating the conditional distribution $f(Y,M,\alpha_M|X)$ with regard to α_M. We show $f(Y,M|X)$ as a function of Y as follows:

$$
\begin{aligned}
f(Y_{ij},M_{ij}|X_{ij}) &= \int f(Y_{ij},M_{ij},\alpha_M|X_{ij})d\alpha_M \\
&= \frac{\sqrt{(fk-gh)\sigma_Y^2}}{\pi\sqrt{2(k\sigma_Y^2+fk-gh)\Sigma_C\Sigma_\alpha A_4}}e^{-HY^2+JY+K}
\end{aligned} \tag{55}
$$

where

$$
\begin{aligned}
H &= [\frac{(g+h)k\sigma_Y^2}{(k\sigma_Y^2+fk-gh)(fk-gh)}-\frac{g+h}{2(fk-gh)}]^2-\frac{k^2\sigma_Y^2}{2k\sigma_Y^2+2(fk-gh)}+ \\
&\frac{k}{2(fk-gh)} \\
J &= [\frac{(g+h)k\sigma_Y^2}{(k\sigma_Y^2+fk-gh)(fk-gh)}-\frac{g+h}{2(fk-gh)}]^2 2\beta_{Y|X}X-\frac{2d^2\sigma_Y^2\beta_{Y|X}X}{2k\sigma_Y^2+2(fk-gh)} \\
&+\frac{2\beta_{Y|X}Xd}{2(fk-gh)}-[\frac{k(g+h)\sigma_Y^2}{k\sigma_Y^2+2(fk-gh)}+\frac{g+h}{2(fk-gh)}](M-\beta_{M|X}X)+ \\
&2(\frac{(g+h)k\sigma_Y^2}{(k\sigma_Y^2+fk-gh)(fk-gh)}-\frac{g+h}{2(fk-gh)})(\frac{(g+h)^2\sigma_Y^2}{(k\sigma_Y^2+fk-gh)(fk-gh)}-\frac{f}{(fk-gh)})\times \\
&(M-\beta_{M|X}X) \\
K &= [\frac{k(g+h)\sigma_Y^2}{k\sigma_Y^2+2(fk-gh)}+\frac{g+h}{2(fk-gh)}](M-\beta_{M|X}X)\beta_{Y|X}X- \\
&2(\frac{(g+h)k\sigma_Y^2}{(k\sigma_Y^2+fk-gh)(fk-gh)}-\frac{g+h}{2(fk-gh)})(\frac{(g+h)^2\sigma_Y^2}{(k\sigma_Y^2+fk-gh)(fk-gh)}-\frac{f}{fk-gh})\times \\
&(M-\beta_{M|X}X)\beta_{Y|X}X-(M-\beta_{M|X}X)^2\times \\
&[(\frac{(g+h)^2\sigma_Y^2}{(k\sigma_Y^2+fk-gh)(fk-gh)}-\frac{f}{fk-gh})^2-\frac{2(g+h)^2\sigma_Y^2}{k\sigma_Y^2+fk-gh}+\frac{f}{2(fk-gh)}]
\end{aligned} \tag{56}
$$

8.7 Derivation of $f(Y|X)$, $E(Y|X)$ and $f(M|X)$ for JM

In order to find the expression of $f(Y|X)$, we reformulate $f(Y,M|X)$ as a function of M as follows:

$$
f(Y_{ij},M_{ij}|X_{ij}) = \frac{\sqrt{(fk-gh)\sigma_Y^2}}{\pi\sqrt{2(k\sigma_Y^2+fk-gh)\Sigma_C\Sigma_\alpha A_4}}e^{-H_1M^2+J_1M+K_1} \tag{57}
$$

where

$$\begin{aligned}
H_1 &= \frac{(g+h)\sigma_Y^2}{2(k\sigma_Y^2 + fk - gh)} + \frac{f}{2(fk-gh)} \\
J_1 &= \frac{(g+h)[k(\beta_{Y|X}X - Y) - (g+h)\beta_{M|X}X]^2}{(k\sigma_Y^2 + fk - gh)} - \frac{(g+h)(Y - \beta_{Y|X}X)}{2(fk-gh)} - \frac{f\beta_{Y|X}X}{fk-gh} \\
K_1 &= \frac{[k(\beta_{Y|X}X - Y) - (g+h)\beta_{M|X}X]^2\sigma_Y^2}{2(k\sigma_Y^2 + fk - gh)} - \frac{k(\beta_{Y|X} - Y)^2 + (g+h)\beta_{M|X}X(Y - \beta_{Y|X}X)}{2(fk-gh)}
\end{aligned} \tag{58}$$

We find the expression of $f(Y_{ij}|X_{ij})$ by integrating $f(Y_{ij}, M_{ij}|X_{ij})$ with respect to M_{ij}, and we show $f(Y_{ij}|X_{ij})$ as a function of Y_{ij} as follows:

$$\begin{aligned}
f(Y_{ij}|X_{ij}) &= \int f(Y_{ij}, M_{ij}|X_{ij})dM \\
&= \frac{\sqrt{2\sigma_Y^2(fk-gh)}}{\sqrt{\pi(k\sigma_Y^2 + fk - gh)\Sigma_C\Sigma_\alpha A_4 H_1}} e^{K_1 - \frac{J_1^2}{4H_1}}
\end{aligned} \tag{59}$$

We find the expression of $f(M_{ij}|X_{ij})$ by integrating $f(Y_{ij}, M_{ij}|X_{ij})$ with respect to Y_{ij}, and we show $f(M_{ij}|X_{ij})$ as a function of M_{ij} as follows:

$$\begin{aligned}
f(M_{ij}|X_{ij}) &= \int f(Y_{ij}, M_{ij}|X_{ij})dY \\
&= \frac{\sqrt{\sigma_Y^2(fk-gh)}}{\sqrt{\pi(k\sigma_Y^2 + fk - gh)\Sigma_C\Sigma_\alpha A_4 H}} e^{UM^2 + VM + W} \\
&= \frac{\sqrt{\sigma_Y^2(fk-gh)}}{\sqrt{\pi(k\sigma_Y^2 + fk - gh)\Sigma_C\Sigma_\alpha A_4 H}} e^{U(M + \frac{V}{2U})^2 + W - \frac{V^2}{4U}}
\end{aligned} \tag{60}$$

where

$$\begin{aligned}
U &= Q^2 - R \\
V &= (2\beta_{Y|X}P - Q\beta_{M|X})QX - Q\beta_{Y|X}X + 2R\beta_{M|X}X \\
W &= [(2\beta_{Y|X}P - Q\beta_{M|X})X]^2 + Q\beta_{M|X}X^2\beta_{Y|X} - R\beta_{M|X}^2X^2 \\
P &= [\frac{(g+h)k\sigma_Y^2}{(k\sigma_Y^2 + fk - gh)(fk-gh)} - \frac{g+h}{2(fk-gh)}]^2 - \frac{2k^2\sigma_Y^2}{2k\sigma_Y^2 + 2(fk-gh)} + \frac{k}{2(fk-gh)} \\
Q &= [2(\frac{(g+h)k\sigma_Y^2}{(k\sigma_Y^2 + fk - gh)(fk-gh)} - \frac{g+h}{2(fk-gh)})(\frac{(g+h)^2\sigma_Y^2}{(k\sigma_Y^2 + fk - gh)(fk-gh)} - \frac{f}{fk-gh}) \\
&\quad - \frac{k(g+h)\sigma_Y^2}{k\sigma_Y^2 + fk - gh} - \frac{g+h}{2(ad-bc)}]^2 \\
R &= [\frac{(g+h)^2\sigma_Y^2}{(k\sigma_Y^2 + fk - gh)(afk-gh)} + \frac{f}{(fk-gh)}]^2 - \frac{f}{2(fk-gh)} - \frac{2(g+h)^2\sigma_Y^2}{d\sigma_Y^2 + fk - gh}
\end{aligned} \tag{61}$$

Therefore, we find M given X is a shifted normal distribution, and we find the expression for $E(M|X)$ as follows:

$$E(M|X) = \frac{-V\sqrt{\sigma_Y^2(fk-gh)}}{\sqrt{2U(k\sigma_Y^2 + fk - gh)\Sigma_C\Sigma_\alpha A_4 H}} e^{W - \frac{V^2}{2U}} \tag{62}$$

8.8 Derivation of $f(Y|M,X)$ and $E(Y|M,X)$for JM

We apply bayes' theorem to generate the formula of $f(Y|M,X)$ as follows:

$$
\begin{aligned}
f(Y_{ij}|M_{ij},X_{ij}) &= \frac{f(Y_{ij},M_{ij}|X_{ij})}{f(M_{ij}|X_{ij})} \\
&= \frac{1}{\sqrt{2\pi H}} e^{-H(Y-\frac{J}{2H})^2+\frac{J^2}{4H}+K-UM^2-VM-W}
\end{aligned}
\tag{63}
$$

Therefore, the conditional distribution of Y given M,X is a shifted normal distribution, and we have the expression of $E(Y|M,X)$ as follows:

$$
E(Y|M,X) = \frac{J}{2H} e^{\frac{J^2}{4H}+K-UM^2-VM-W} \tag{64}
$$

References

Bodner, T. E. (2008), 'What improves with increased missing data imputations?', *Structural Equation Modeling* **15**(4), 651–675.

Carpenter, J. R. & Kenward, M. G. (2013), 'Multiple Imputation and its Application'.

Collins, L. M., Schafer, J. L. & Kam, C. M. (2001), 'A comparison of inclusive and restrictive strategies in modern missing data procedures', *Psychological Methods* **6**(3), 330–351.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977), *Maximum Likelihood from Incomplete Data Via the EM Algorithm* , Vol. 39.

Ellickson, P. L., McCaffrey, D. F., Ghosh-Dastidar, B. & Longshore, D. L. (2003), 'New Inroads in Preventing Adolescent Drug Use: Results from a Large-Scale Trial of Project ALERT in Middle Schools', *American Journal of Public Health* **93**(11), 1830–1836.

Graham, J. W., Olchowski, A. E. & Gilreath, T. D. (2007), 'How many imputations are really needed? Some practical clarifications of multiple imputation theory', *Prevention Science* **8**(3), 206–213.

Grund, S., Lüdtke, O. & Robitzsch, A. (2018), 'Multiple Imputation of Missing Data for Multilevel Models: Simulations and Recommendations', *Organizational Research Methods* **21**(1), 111–149.

Krull, J. & MacKinnon, D. (2001), 'Multilevel Modeling of Individual and Group Level Mediated', *Multivariate Behavioral Research* **36**(2), 227–248.

Milletich II, R. J. (2018), 'Multiple imputation of missing data in structural equation models with mediators and moderators using gradient boosted machine learning', *Dissertation Abstracts International: Section B: The Sciences and Engineering* **78**(7-B(E)).

Mistler, S. A. & Enders, C. K. (2017), 'A Comparison of Joint Model and Fully Conditional Specification Imputation for Multilevel Missing Data', *Journal of Educational and Behavioral Statistics* **42**(4), 432–466.

Royston, P. (2005), 'Multiple imputation of missing values: Update of ice', *Stata Journal* **5**(4), 527–536.

Rubin, D. B. (1976), 'Biometrika Trust Inference and Missing Data Author (s): Donald B . Rubin Published by : Oxford University Press on behalf of Biometrika Trust Stable URL : http://www.jstor.org/stable/2335739 Accessed : 12-06-2016 21 : 34 UTC', *Biometrika* **63**(3), 581–592.

Rubin, D. B. (1987), 'Multiple Imputation for Nonresponse in Surveys', *Journal of Marketing Research* **26**(4), 485.

Schafer, J. L. & Yucel, R. M. (2002), 'Computational strategies for multivariate linear mixed-effects models with missing values', *Journal of Computational and Graphical Statistics* **11**(2), 437–457.

Van Buuren, S. (2011), 'Multiple Imputation of Multilevel Data', *Handbook of Advanced Multilevel Analysis* (11237), 173–196.

Van Buuren, S., Brand, J. P., Groothuis-Oudshoorn, C. G. & Rubin, D. B. (2006), 'Fully conditional specification in multivariate imputation', *Journal of Statistical Computation and Simulation* **76**(12), 1049–1064.

Ye, B. & Yucel, R. M. (2020), *Impact of inconsistent imputation models in mediation analysis.*, American Statistical Association., Alexandria, VA.