Title
Addressing Common Covariate Modeling Issues in Latent Class Analysis Models

Author
Visser, Marieke Adriana

Publication Date
2022

UNIVERSITY OF CALIFORNIA, MERCED

Addressing Common Covariate Modeling Issues in Latent Class Analysis Models

A dissertation submitted in partial satisfaction of the requirements
for the degree of Doctor of Philosophy

in

Psychological Sciences

by

Marieke A. Visser

2022

Committee members:
Professor Sarah Depaoli, Chair
Professor Ren Liu
Professor Fan Jia

Signature Page

The dissertation of Marieke A. Visser is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

(Professor Sarah Depaoli, Chair)

(Professor Fan Jia)

(Professor Ren Liu)

University of California, Merced

2022

To my family

Table of Contents

Online Supplemental Material:

Supporting R and M*plus* code can be found with the following Dropbox link: https://www.dropbox.com/sh/35c79gmhh5q50qt/AADcwrhxbrJKnIhHaGY6CxG9a?dl=0.

List of Figures

List of Tables

Acknowledgments

Countless people helped me during my five years at UC Merced. It seems appropriate to take the time to thank some of the people most influential during my graduate studies.

First and foremost, I would like to thank my academic advisor, Sarah Depaoli. My father always emphasized the importance of picking the right Ph.D. advisor, and I am confident Sarah was the right person for me. Sarah was a wonderful mentor who not only taught me the nuts and bolts of methodological research but also taught me much about professionalism, work-life balance, storytelling, time management, and so much more. My future feels wide open and full of possibilities. I cannot possibly thank Sarah enough for the opportunities she gave me and the mentorship she provided along the way.

I also want to thank my dissertation committee: Ren Liu and Fan Jia. You were so generous with your time and expertise. Your feedback has consistently improved the quality of my work, and I appreciate your taking such care when reviewing my research. Thank you for being such supportive figures during my time at UC Merced.

My journey at UC Merced would have been much more difficult were it not for the lifelong friendships I formed in my lab. I would especially like to thank my lab mom, Sonja, who was always ready to serve as an encouraging voice and resource to the graduate students who came behind her. I knew when I met Sonja during visitation that she would play an important role in my graduate studies, but I couldn't imagine how much she would teach me along the way. Another person I would like to thank is Luca, who wore many hats for me: BFF, work buddy, roommate, sidekick, and coding guru. I am truly lucky to have gotten the chance to befriend Luca, and I greatly benefited from his (and Al's) support along the way. Finally, I would like to thank the latest edition of the Depaoli lab, Ihnwhi. It felt like coming full circle when I got to pass along some of my knowledge, as Sonja once did for me.

Next, I want to thank my friends who had nothing to do with academia. Your unwavering support and absolute confidence that I would finish this degree kept me steadfast when things inevitably did not go my way. Despite being far away, I always knew my closest friends (Ashley, Shelby, and Kirstin) were rooting for me. My visits with each of you were bright spots during the last five years.

Lastly, I thank my parents, Sent and Mary Visser, and my brother, Dirk. They have been supportive and encouraging figures throughout my graduate studies. I am fortunate to have parents who were so ready to help me achieve my goal of completing a Ph.D. program. Thank you for the opportunities you've provided and for loving me unconditionally. I hope I have made you proud!

VITA

EDUCATION

2018 – Present	**Ph.D. in Psychological Sciences** University of California, Merced, CA Dissertation: Addressing Common Covariate Modeling Issues in Latent Class Analysis Models Advisor: Dr. Sarah Depaoli
2016 – 2018	**Master of Arts in Psychological Research** Texas State University, San Marcos, TX Advisor: Dr. Reiko Graham
2012 – 2016	**Bachelor of Arts in Psychology** Southwestern University, Georgetown, TX Advisor: Dr. Erin Crockett

SELECT PUBLICATIONS

Visser, M., & Depaoli, S. (2022). A Guide to Detecting and Modeling Local Dependence in Latent Class Analysis Models. *Structural Equation Modeling: A Multidisciplinary Journal.* doi: 10.1080/10705511.2022.2033622

Depaoli, S., Winter, S. D., & **Visser, M.** (2020). The Importance of Prior Sensitivity Analysis in Bayesian Statistics: Demonstrations Using an Interactive Shiny App. *Frontiers in Psychology*, *11*(November), 1–18. doi: 10.3389/fpsyg.2020.608045

Abstract of Dissertation

Addressing Common Covariate Modeling Issues in Latent Class Analysis Models

by
Marieke A. Visser

Doctor of Philosophy in Quantitative Methods, Measurement, and Statistics

University of California, Merced, 2022
Professor Sarah Depaoli, Chair

This dissertation is organized into two studies investigating common modeling issues that occur when including a covariate in a latent class analysis (LCA) model. When estimating a conditional LCA model, applied researchers must make decisions about the estimation strategy (one-step vs. three-step), the handling of incomplete covariate data, and specification of covariate relationships. Study 1 examined the performance of different methods for handling incomplete covariate data when using a three-step approach to estimation. The simulation results found that Bayesian estimation with informative normal priors correctly centered on the regression coefficient population values produced the most consistent and accurate regression coefficient estimates, regardless of the covariate distribution, strength, and missing data pattern. However, informative priors centered on the wrong population values produced some of the most biased regression coefficients. In most modeling conditions, full information maximum likelihood (FIML) and multiple imputation (MI) still worked well. When estimating a conditional LCA model, applied researchers must also make decisions about how to specify the covariate relations with the LCA measurement model. Specifically, applied researcher must decide if the covariate only has an indirect effect on the observed indicators via the latent class variable or if the covariate is related to one or more of the observed indicator variables. The goal of Study 2 was to explore the utility of using small-variance priors to help evaluate covariate relationships. Specifically, small-variance normal priors centered on zero were specified for the direct effects between the covariate and the latent class indicators for a series of population models with varying covariate relationships. Results from the Study 2 simulation indicate small-variance priors can be a useful tool for detecting covariate misspecifications, depending on the number of direct effects, sample size, and class sizes. Overall, findings from Study 1 and Study 2 highlight how Bayesian estimation can be especially helpful for handling common modeling issues in conditional LCA models.

Chapter 1: Overview of Dissertation

Latent class analysis (LCA; Goodman, 1974; Lazarfeld & Henry, 1968) is a popular method for creating measurement models in the social sciences. For many LCA applications, applied researchers are interested in using the LCA measurement model as part of a larger structural equation model (SEM) that includes covariates (Brewsaugh, Masyn, & Salloum, 2018; Herman, Prewett, Eddy, Savala, & Reinke, 2020; Nasiopoulou, Williams, Sheridan, & Hansen, 2019). When and how to include these covariates has become a central issue in the methodological literature, where two different estimation approaches have been introduced for handling covariables: a one-step and a three-step approach. The one-step approach simultaneously estimates the LCA measurement model and regresses the latent class variable on the covariate (Vermunt, 2010). In contrast, the three-step approach uses a stepwise estimation approach in which the LCA measurement model is established independently of the covariate before regressing the latent class variable on the covariate (Bakk, Tekle, & Vermunt, 2013; Vermunt, 2010). Despite recent methodological advances in estimating conditional LCA models, there are still gaps in the literature for how to handle common covariate modeling issues (e.g., incomplete covariates, covariate misspecifications). This dissertation will address these gaps.

The dissertation is organized into two different studies. The first study is entitled "Addressing Missing Data in Latent Class Analysis When Using Three-Step Approach." This study aims to provide recommendations for how best to deal with incomplete covariates when using a three-step approach. Currently, mixture modeling statistical software defaults to listwise deletion of incomplete covariates. However, several alternative techniques are available, including full information maximum likelihood (FIML), switching to Bayesian estimation in the third step, and multiple imputation (MI). Study 1 investigates the performance of these methods under different covariate missing conditions (e.g., proportion missing, missing mechanism) and covariate distributions (e.g., standard normal, binomial). Results from this study will be valuable to applied researchers seeking to address incomplete covariates while using a three-step approach to estimation.

The second dissertation study is entitled "Using Small-Variance Priors to Detect Covariate Misspecifications in Latent Class Analysis Models." The aim of this study is to illustrate how informative, small-variance priors can be used to detect covariate misspecifications in conditional LCA models. The most common way of including a covariate is to regress the latent class variable on the covariate. Still, there are many ways the covariate could be related to the measurement model (e.g., the covariate could be directly related to one or more of the class indicators; Nylund-Gibson & Masyn, 2016). One method that may be effective for detecting these types of covariate misspecifications is a restrictive prior strategy, which involves the use of informative, small-variance priors on parameters that are typically constrained to zero (i.e., direct effects from the covariate to class indicators). Study 2 aims to provide a new analytical framework for methodological and applied researchers seeking to detect direct effects in conditional LCA models.

Chapter 2: General Methodology

Mixture modeling has become an increasingly popular analytical method for describing and explaining heterogeneity in an unobserved population. Mixture modeling is used to describe the overall population distribution of a set of class indicators using a finite mixture of unobserved subpopulations. Each subpopulation (i.e., latent class) has its own multivariate distribution of the observed indicator variables. Mixture modeling techniques can be applied to both cross-sectional data (e.g., latent class analysis and latent profile analysis) and longitudinal data (e.g., latent transition analysis and growth mixture modeling). In all mixture modeling applications, at least one multinomial latent class variable is estimated, which divides the unobserved population into a finite number of mutually exclusive and exhaustive latent classes. The aim of these analyses is to identify substantively meaningful groups of individuals who responded similarly to the indicator variables (Muthén, 2004).

The most basic finite mixture model is the cross-sectional latent class analysis (LCA) model (Lazarsfeld & Henry, 1968; Goodman 1974; McCutcheon, 1987; Magidson & Vermunt, 2004), which traditionally uses binary latent class indicator variables. The LCA model has been extended to include class indicators of various scales: ordinal, interval, or ratio. In addition, LCA can be applied to indicator variables of all the same scale or of mixed scales (e.g., binary class indicators and ratio class indicators can be used in the same analysis). However, the parameterization of the LCA model becomes increasingly complex with the addition of mixed scale class indicators.

LCA has been used in a variety of substantive settings, including the classification of adolescent smoking subgroups (Henry & Muthén, 2010), alcohol dependence subgroups (Moss, Chen, & Yi, 2007), internet gambling subgroups (Llyod et al., 2010), adolescent obesity subgroups (Huh et al., 2011), and peer victimization subgroups (Nylund, Bellmore, Nishina, & Graham, 2007). As another example, Quirk, Nylund-Gibson, and Furlong used the LCA measurement model to help explain the heterogeneity kindergarten readiness based on a set of readiness indicators that had previously been identified. In the following sections, the parameterization of the unconditional LCA model will be discussed. Next, the LCA model will be extended to include covariates (i.e., predictor variables). Finally, different approaches to estimating the conditional LCA model will be presented.

2.1 Latent Class Analysis

In the LCA model, there are two types of parameters of interest: measurement and structural parameters. The measurement parameters describe the relationship between the observed latent class indicators and the latent class variables (i.e., the class-specific item endorsements probabilities, which are the distribution of the binary class indicators conditional on the latent class variable). In contrast, the structural parameters describe the multinomial distribution of the latent class variable (i.e., the proportion of cases in each latent class). The parameterization of the LCA model with binary indicators is detailed below, with notation first presented in Nylund-Gibson and Masyn (2016) and Masyn (2017).

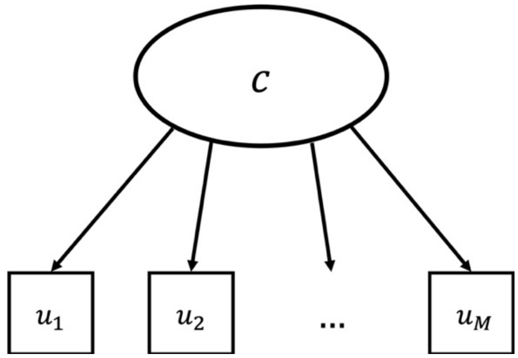


Figure 1. The unconditional LCA model with M binary latent class indicators. The latent class indicators are represented with $u_1, u_2, \ldots, u_M$, and c represents the underlying multinomial latent class variable.

Figure 1 provides a visual representation of the unconditional LCA model. In Figure 1, each latent class indicator is observed on n individuals with u_{mi} representing individual i's response to class indicator m. The latent class variable has K classes where $c_i = k$ when individual i belongs to Class k. The latent classes are mutually exclusive; therefore, individual i can only be assigned to one of K classes. The relationship between the observed class indicator variables and the latent class variable can be formulated with:

$$\Pr(u_{1i}, u_{2i}, \ldots, u_{Mi}) = \sum_{k=1}^{K} [\pi_k \cdot \Pr(u_{1i}, u_{2i}, \ldots, u_{Mi} | c_i = k)], \tag{1}$$

where π_k is a structural parameter representing the prevalence of individuals in Class k (i.e., class proportions). Considering the latent classes are mutually exclusive, $\sum \pi_k = 1$.

The measurement model for the latent class variable can be parameterized as the relationship between the observed class indicators $u_1, u_2, \ldots, u_M$ and the latent class variable c, which can be formulated with:

$$\Pr(u_m = 1 | c = k) = \frac{1}{1 + \exp(\tau_{mk})}, \tag{2}$$

where τ_{mk} is the negative log odds of endorsing class indicator u_m given membership to latent class k. In other words, τ_{mk} is equal to $-\text{logit}(\text{E}[u_m | c = k])$. The class-specific item response probabilities suggest how likely an individual is to endorse a particular item given latent class membership.

For the structural model, the unconditional distribution of the multinomial latent class variable, c, can be parameterized with a multinomial logistic regression formulation. Specifically, the π_k parameters can be defined as intercepts on the inverse multinomial logit scale, such that:

$$\pi_k = \Pr(c = k) = \frac{\exp(\gamma_k)}{\sum_{j=1}^{K} \exp(\gamma_j)}. \quad (3)$$

The γ_k represents the log odds of membership in class k, given membership in either class k or K. For identification purposes, γ_{0K} is constrained to 0.

The LCA model assumes local independence, which suggests the M binary latent class indicators are uncorrelated conditional on class membership. In other words, latent class membership fully explains any correlations between observed class indicators. Software capable of LCA imposes the local independence assumption by default. By making the local independence assumption, Equation (1) is further simplified to:

$$\Pr(u_{1i}, u_{2i}, \ldots, u_{Mi}) = \sum_{k=1}^{K} \left[\pi_k \cdot \left(\prod_{m=1}^{M} \Pr\,(u_{mi} | c_i = k) \right) \right]. \quad (4)$$

Violating the local independence assumption can impact parameter estimates and model fit indices (Albert & Dodd, 2004; Asparouhov & Muthén, 2011; Lee et al., 2020; Torranc-Rynard & Walter, 1998; Vacek, 1985). In applied settings, the local independence assumption should be evaluated because it is possible to relax the assumption (i.e., allowing residual correlations between two or more latent class indicators in one or more latent classes), if necessary. For a detailed explanation of how to evaluate and relax the local independence assumption in popular mixture modeling software, see Visser and Depaoli (2022).

When estimating an LCA model, users must select the number of latent classes in the population. Applied users often lack prior knowledge about the number of latent classes. A statistical analysis procedure (i.e., class enumeration) can aid applied users in selecting the number of classes. During class enumeration, an iterative procedure is used to estimate several LCA models with a different number of specified classes. The best-fitting LCA model is then selected based on model fit and comparison indices, see Nylund, Asparohov, and Muthén (2007) for a detailed explanation of class enumeration.

2.2 Latent Class Analysis with Covariates

In many practical applications of LCA, the LCA measurement model is used as part of a larger structural equation model (SEM). These models often include observed explanatory variables (i.e., covariates, predictors, independent variables, external variables, or concomitant variables[1]) that predict the latent class variable. For example, Quirk et al., (2013) extended their LCA model for kindergarten readiness to include several predictors (e.g., student's prior preschool experiences, age, language skills, and gender). The addition of these covariates allows researchers to explore research questions about why an individual was assigned to a particular latent class. A visual example of the latent class model with a covariate (i.e., conditional LCA model) can be seen in Figure 2. The covariate, x_1, can be categorical (Clogg, 1985; Goodman, 1974; Haberman, 1979; Hagenaars, 1990; Hagenaars, 1993; Vermunt, 1997) or continuous (Bandeen-Roche, Zeger, & Rathouz, 1997; Dayton & Macready, 1988; Kamakura Wedel, & Agrawal, 1994; Yamaguchi, 2000).

[1] A concomitant variable is a variable that is not the focus of the study, but the variable may influence variables of interest to the study (e.g., the dependent variable).

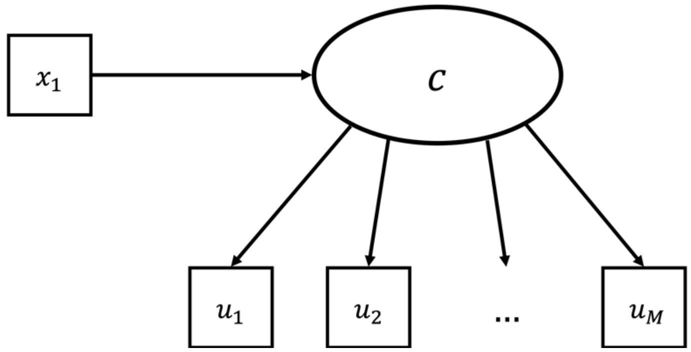


Figure 2. The conditional LCA model with M binary latent class indicators and a single covariate, x_1. The latent class indicators are represented with $u_1, u_2, \ldots, u_M$, and c represents the underlying multinomial latent class variable.

To include a covariate, the latent class model is combined with the latent class regression model into a joint model, which is typically estimated with the maximum-likelihood (ML) estimator. This approach is often referred to as the *one-step approach* in the methodological literature because the measurement model and structural model (i.e., the logistic regression in which the latent classes are related to the covariates) are simultaneously estimated in a single step (Asparouhov & Muthén, 2014; Bandeen-Roche et al., 1997; Dayton & Macready, 1998; Vermunt, 2010). More specifically, the latent class variable is regressed on the covariate using multinomial logistic regression parametrization (Nylund-Gibson & Masyn, 2016). Using notation first presented in Nylund-Gibson & Masyn (2016), the relationship between the LCA model and covariate x_i can be expressed as a multinomial logistic regression model:

$$\Pr(c_i = k|x_i) = \frac{\exp(\gamma_{0k} + \gamma_{1k}\, x_i)}{\sum_{j=1}^{K} \exp(\gamma_{0j} + \gamma_{1j}\, x_i)}, \tag{5}$$

where $\gamma_{0K} = \gamma_{1K} = 0$ for model identification. In Equation (5), the latent class indicator variables are considered independent of the covariate conditional on class membership. Therefore, Equation (4) can be adapted to include covariate x_i such that

$$\Pr(u_{1i}, u_{2i}, \ldots, u_{Mi}|x_i) = \sum_{k=1}^{K} \left[\Pr(c_i = k|x_i) \cdot \left(\prod_{m=1}^{M} \Pr\,(u_{mi}|c_i = k)\right)\right]. \tag{6}$$

When K number of classes have correctly been identified, the exclusion of x_i from the model has no impact on the point estimates (i.e., τ_{mk}) for each class indicator (i.e., $u_1, u_2, \ldots, u_M$). In other words, latent class membership will depend on x_i, but the class indicator responses should only depend on class membership. Thus, the covariate only has an indirect effect on the latent class indicators via the latent class variable.

The one-step approach may appear straightforward enough, but several drawbacks have been noted in the methodological literature. Specifically, Vermunt (2010) notes that the one-step approach is impractical when using many covariates, as is typical for exploratory studies. With each additional covariate, the LCA measurement model and the structural model must be estimated again. In addition, Vermunt (2010) highlights the model-building issues surrounding the inclusion of covariates. Applied users must decide whether to pick the number of latent classes before or after including covariates. Although covariates can help aid class enumeration when properly specified (Li & Hser, 2011; Lubke & Muthén, 2007; Muthén, 2002), applied researchers are unlikely to properly specify the covariates in the model without prior knowledge. Misspecifying the covariate relationships with the LCA measurement model can impact the class enumeration procedure, resulting in an over-extraction of the number of classes (Nylund-Gibson & Masyn, 2016). Therefore, several methodological studies suggest the number of latent classes should be established prior to including covariates (Collins & Lanza, 2010; Masyn, 2013; Petras & Masyn, 2010). Vermunt (2010) also notes that applied researchers do not find the joint model to be intuitive because they often wish to introduce covariates after classifying individuals. In addition, the applied researcher who establishes the latent class measurement model may not be the same researcher who is building the structural model.

To address some of the drawbacks of the one-step approach, methodologists have proposed a stepwise approach to estimation (Vermunt, 2010; Asparouhov & Muthén, 2014), where the latent class model and the relationship between the latent class variable and covariate are independently evaluated. By using a stepwise approach, the measurement model and structural models are decoupled, which can resolve many of the issues with the one-step approach. In the following section, the procedures for using a typical stepwise approach will be discussed in detail.

2.2.1 Stepwise Approaches

The conditional LCA methodological literature has been in flux in recent years. More recently, methodologists have recommended the use of a stepwise approach when building SEMs that include auxiliary variables (e.g., covariates, distal outcomes). When using a stepwise approach to estimation, the LCA measurement model is established prior to the inclusion of auxiliary variables. The general procedure for implementing a typical stepwise approach is as follows:

1. The LCA measurement model is built. During this step, applied researchers must decide how many class indicators should be included in the LCA measurement model and how many classes should be specified. In addition, the local independence assumption should be evaluated during this step.
2. Using the parameter estimates from the LCA measurement model in Step 1, cases are assigned to the different latent classes based on their posterior membership probabilities.
3. The standard multinomial logistic regression is estimated, using the class membership assignment from Step 2 as an observed indicator of the latent class variable.

The described stepwise approach consistently underestimates the relationship between the covariate and latent class (Bolck, Croon, & Hagenaars, 2004). As the classification error in Step 2 increases, the relationship between the covariate and the latent class variable is attenuated. In response to these findings, several new methods have been proposed to address the measurement error issue in Step 2. One such method is the *maximum likelihood (ML) three-step approach*, which was developed by Vermunt (2010) and expanded upon by Bolck et al. (2004).

The three-step approach is suitable for exploring the relationship between the LCA measurement model and covariate(s). We follow Asparouhov and Muthén (2014) in describing the procedure for the ML three-step approach. First, the unconditional LCA model is estimated using only the observed class indicators. In the second step, the most likely class variable, M, is created using the latent class posterior distribution, which was produced during the estimation of the unconditional LCA model. The most likely class variable is a nominal variable and for each observation, M is set to the class for which $\Pr(c = k|\boldsymbol{u})$ is the largest (Asparouhov & Muthén, 2014); $\boldsymbol{u}$ represents the latent class indicators and c is the latent class variable. The classification uncertainty rate for M can be calculated with the following equation:

$$p_{c_1,c_2} = \Pr(c = c_2|M = c_1) = \frac{1}{M_{c_1}} \sum_{M_i=c_1} \Pr(c_i = c_2|\boldsymbol{u}_i), \qquad (7)$$

where M_{c_1} is the number of cases assigned to class c_1 by the most likely class variable M, M_i is the most likely class variable for the ith observation, c_i is the true latent class variable for the ith observation, and $\boldsymbol{u}_i$ represents the class indicator variables for the ith observation. The probability $\Pr(c_i = c_2|\boldsymbol{u}_i)$ can be computed with the estimated unconditional LCA model from the first step.[2] After calculating the classification uncertainty with Equation (7), it is possible to calculate the classification measurement error with:

$$q_{c_1,c_2} = P(M = c_1|c = c_2) = \frac{p_{c_1,c_2} N_{c_1}}{\sum_c p_{c,c_2} N_c}, \qquad (8)$$

where N_c is the number of observations classified in class c by the most likely class variable M. In this way, the most likely class variable can be treated as an imperfect measurement of c with measurement error q_{c_1,c_2}. The measurement error can then be transformed into logits using $\log(q_{c_1,c_2} / q_{K,c_2})$, where K is used as a reference class. In the third step, the latent class variable is regressed on x while taking into account the measurement error. Specifically, the most likely class variable M is used as a single, nominal class indicator of the latent class variable c. The logits are used as fixed parameter values that describe the direct relationship between the latent class variable and the most likely class variable. The multinomial regression of c on predictor x is freely estimated. A visual representation of the three-step approach can be seen in Figure 3, which was adapted from Asparouhov and Muthén (2014).

[2] The conditional probabilities for the class assignment given true latent class membership are automatically computed by M*plus* when estimating an LCA model. These conditional probabilities can be found in the Results section under the title "Classification Probabilities for the Most Likely Latent Class Membership (Row) by Latent Class (Column)." See Asparouhov and Muthén (2014) and Vermunt (2010) for more details on how to compute the conditional probabilities for the ML three-step approach. In the Vermunt (2010) article, the ML three-step approach is referred to as Modal ML.

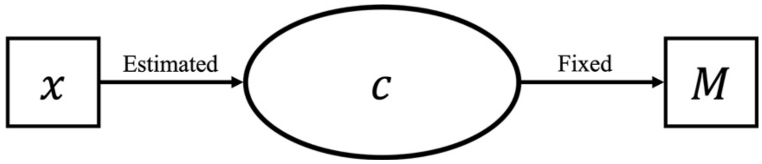


Figure 3. A visual of the ML three-step approach. The latent class variable c is regressed on covariate x. The most likely class variable, M, is used as a single class indicator of the latent class variable c. The relationship between M and c is fixed, and the relationship between x and c is freely estimated.

Ample simulation research has explored the performance of the ML three-step approach under different modeling conditions (e.g., Asparouhov & Muthén, 2014; Bakk, Tekle, & Vermunt, 2013; Nylund-Gibson et al., 2019; Vermunt, 2010). Results from these simulation studies suggest the three-step approach can produce unbiased parameter estimates if the LCA measurement model has sufficient class separation. In mixture modeling, class separation refers to how distinct the latent classes are from one another. When class separation is poor, it can be trickier to properly assign cases to latent classes, resulting in increased measurement error in the latent class variable.

Statistical software capable of mixture modeling (e.g., M*plus,* Latent GOLD) has largely automated the three-step approach, allowing applied researchers to implement the procedure much more easily. Although this automation can be helpful, the automation limits the user's ability to adjust how the model is estimated. Specifically, the automation limits the user's ability to address missing data and the estimator is limited to ML. When "manually" implementing the three-step approach instead, the user has a greater ability to adjust how the model is estimated. For example, the user can address the missing data in the covariate and the user can switch to a Bayesian estimation framework in the third step.

2.3 Bayesian Conditional Latent Class Models

The previous section discussed estimation strategies available for the conditional LCA models in the frequentist framework (e.g., one-step approach, three-step approach). An alternative method for estimating LCA models is to use Bayesian estimation. In recent years, the Bayesian estimation framework has become increasingly popular as statistical software has made it more accessible to applied researchers (van de Schoot, Winter, Zondervan-Zwijnenburg, Ryan, & Depaoli, 2017). The primary distinction between the frequentist and Bayesian estimation is the addition of prior distributions in the model. The prior distributions (or priors) represent what a parameter in the model should look like based on a prior belief about the relationship. For every parameter estimated in the model, it is possible to specify a prior distribution that describes these prior beliefs. These prior distributions are incorporated into the estimation process and can provide information about the parameters in the model.

Depending on the certainty of a researcher's prior information, it is possible to specify a prior with varying degrees of informativeness. For example, if a researcher has very specific knowledge about the parameter, the researcher can specify a narrower prior. However, if the researcher is uncertain about what a parameter looks like, the researcher can set a less informative prior. The degree of informativeness about a prior can be set with hyperparameters, the parameters that compromise a probability distribution. When a prior is narrow and contains specific information about the parameter, it is considered an informative prior. A wider, less informative prior is called a noninformative (or diffuse) prior. The information (or lack of information) specified within the prior is incorporated with the data during the estimation process. In mixture models, the ability to incorporate accurate prior knowledge about class-specific parameters and class proportions can greatly improve estimation (Depaoli, 2013, 2014; Lu, Zhang, & Lubke, 2011). Bayesian estimation offers several advantages when addressing model estimation issues, such as model assumption violations (Asparouhov & Muthén, 2011; Bauer, 2007), convergence to local maxima (Hipp & Bauer, 2006), and inaccurate parameter estimates (Depaoli, 2013).

One important concept in latent class modeling is how separated the latent classes are from one another at the population level. When the latent classes are difficult to distinguish from one another (i.e., poor class separation), it can be less clear which latent class a particular case belongs to. In addition, it may not be obvious how many latent classes are present in the population. Estimating the class-specific parameters can be much more difficult when class separation is poor. When using Bayesian estimation, it is possible to incorporate prior knowledge about the latent classes, which can be a helpful tool for accurately estimating class-specific parameters. In contrast, in the frequentist estimation framework, one of the only viable options for overcoming these estimation challenges is to collect a much a larger sample (Depaoli, 2013, 2014; Lu et al., 2011).

Another source of estimation issues in latent class models is the relative size of the latent classes. When a latent class is small relative to the other latent classes (e.g., Class 1 = 18% vs. Class 2 = 82%), parameters specific to the minority class are much more difficult to estimate (Depaoli, 2013, 2014; Lu et al., 2011; Tueller & Lubke, 2010). Bayesian estimation can be a useful tool for incorporating prior knowledge about the relative size of the latent classes in the model. By incorporating accurate prior distributions about the relative size of the latent classes, the model is better able to identify and accurately estimate small latent classes (Depaoli et al., 2017). In the following section, the prior distributions relevant to class proportions in unconditional and conditional LCA models will be discussed in detail.

2.3.1 Class Proportion Prior Specifications for Latent Class Models

In an unconditional LCA model, the prior distribution for the class proportions typically follows the Dirichlet distribution. When assigning a Dirichlet (D) prior, the class proportions (π_1,…,π_c) for latent class c can be modeled such that:

$$(\pi_1,\ldots,\pi_c) \sim \mathrm{D}(\delta_1,\ldots,\delta_c). \tag{9}$$

The class proportions $\sum \pi_k = 1$, and each of the δ elements represent the hyperparameters, which control how uniform the distribution is. Depending on the statistical software being used, the δ elements can represent either the number of cases or proportion of cases that will be added to each latent class according to the prior. In the M*plus* statistical software (Muthén & Muthén, 1998-2017), the Dirichlet prior is placed on the class proportion threshold, and the δ elements indicate the number of cases that should be added to each latent class according to the prior.

The least informative Dirichlet prior for 2-class model would be $\mathrm{D}(1,1)$, which only has a single case representing each class and provides no information about the proportion of cases in each class. In contrast, a more informative version of the Dirichlet prior would assign the δ elements based on prior knowledge about the size of the latent classes relative to the total number of participants. For example, assume a dataset has 100 participants and a researcher believes the class proportions should be an 82%-18% split between two classes. In this example, the informative Dirichlet prior could be set to $\mathrm{D}(82,18)$ to reflect the expected number of cases in each class. Past methodological research on growth mixture models (GMM) suggests informative, accurate priors on the class proportions can help aid estimation when the class proportions are unequal (Depaoli, 2012; Depaoli, 2013). Unequal class proportions can be challenging in mixture models, especially when one of the latent classes is much smaller relative to the other classes. In these situations, a more diffuse Dirichlet prior with equal δ elements can exacerbate the problem by making the latent classes appear more equal in size than they are in the population (Depaoli et al., 2017). In other words, when there are unequal sizes in the latent classes, and especially when the overall sample size is small, a "diffuse" Dirichlet prior (e.g., $\mathrm{D}(10,10)$) has the potential to act as an informative prior that forces classes to be equal in size (Depaoli, 2013).

When including a covariate in a latent class measurement model, the latent class variable is regressed on the covariate using multinomial logistic regression parametrization. As a result, the latent class proportions are now a function of the intercept and slope. In contrast to the unconditional LCA model, it is no longer possible to specify a Dirichlet prior directly on the latent class proportions. Instead, prior knowledge about the relationship between the covariate and the latent class variable can be incorporated by specifying a prior distribution on the intercept and slope the logistic regression.

The methodological research about the use of informative priors on the multinomial logistic regression parameters is very limited. Previous applications of Bayesian conditional LCA models have placed a normally distributed prior, $\mathrm{N}(\mu, \sigma^2)$, on the logistic regression intercept and slope terms (Garrett & Zeger, 2000; Neelon, Swamy, Burgette, & Miranda, 2011). More specifically, the mean hyperparameters (μ) were assigned a value of 0, and the variance hyperparameters (σ^2) were assigned a value of $9/4$. These values were selected because 95% of the distribution falls between approximately $\pm$ 3.00, which encompass the range of expected values for the intercept and slope of the multinomial logistic regression. These prior specifications result in a weakly informative prior that limits parameter space for the intercept and slope to more plausible values.

2.3.2 Convergence

Bayesian estimation is implemented with the Markov chain Monte Carlo estimation algorithm (MCMC). Although there are several different diagnostics available to assess parameter convergence within the MCMC algorithm, the potential scale reduction factor (PSRF, or $\hat{R}$; Brooks & Gelman, 1998; Gelman & Rubin, 1992a, 1992b; Vehtari, Gelman, Simpson, Carpenter, & Bürkner, 2020) is one of the most common. The PSRF is a diagnostic based on the analysis of variance of parallel MCMC chains with different starting values. More specifically, the PSRF looks at the ratio of the overestimate of the target distribution (i.e., between-chain variance) and the underestimate of the target distribution (i.e., within-chain variance). Using notation first presented in Muthén and Asparouhov (2012), we can define the within- and between-chain variation as,

$$\bar{\pi}_{.j} = \frac{1}{n}\sum_{i=1}^{n}\pi_{ij}, \tag{10}$$

$$\bar{\pi}_{..} = \frac{1}{m}\sum_{j=1}^{m}\pi_{.j}, \tag{11}$$

$$W = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n}\sum_{i=1}^{n}(\pi_{ij} - \pi_{.j})^2, \tag{12}$$

$$B = \frac{1}{m-1}\sum_{j=1}^{m}(\pi_{.j} - \bar{\pi}_{..})^2. \tag{13}$$

In Equations (10)-(13), there are n iterations in m chains, where π_{ij} represents the value of parameter π in iteration i of chain j. Using the within- and between-chain variation, the PSRF can be calculated with,

$$PSRF = \sqrt{\frac{W + B}{W}}. \tag{14}$$

Convergence can then be assessed by interpreting the PSRF value. If the PSRF is close to 1, there is evidence to suggest convergence has been reached because the between-chain variance is equal to the within-chain variance (implying that the chains overlap and have converged together). If the PSRF is greater than 1.01, concerns about nonconvergence are warranted. In some situations, increasing the length of the MCMC chains can produce a more favorable PSRF value.

2.3.3 Label Switching

One well known issue that commonly arises when estimating finite mixture models is *label switching*. More specifically, label switching occurs when the ordering of the latent classes arbitrarily changes because the order of the classes is not typically defined within the mixture model. When working with a single sample, label switching may not be an issue because the model will converge on a single solution. In simulation studies, the same analysis is repeated across many samples, which allows the ordering of the latent classes to change across replications. If the ordering of the latent classes does not remain stable across replications, then the statistics used to summarize parameter estimates across all replications are suspect (Chung, Loken, & Schafer, 2004; Tueller, Drotar, & Lubke, 2011). To avoid label switching, one popular solution is to introduce inequality constraints to the mixture model, which order the latent classes in a specific way across replications.

Bayesian estimation introduces two additional situations in which label switching can occur (Celeux, Hurn, & Robert, 2000; Chung et al., 2004; Frühwirth-Schnatter, 2001). Bayesian estimation via the MCMC algorithm allows for between-chain and within-chain label switching. Between-chain label switching occurs when multiple chains are used to estimate model parameters. Considering the order of the latent classes are arbitrary, each chain may be estimating the parameter for a different latent class. When the chains are averaged together for a posterior estimate, the results are meaningless because each chain is estimating a parameter for a different latent class. To prevent between-chain label switching, a single MCMC chain should be used to estimate model parameters in mixture models. Within-chain label switching occurs when the latent class labels are arbitrarily switched between latent classes within a single MCMC chain (Jasra, Holmes, & Stephens, 2005). Within-chain label switching is most evident when assessing convergence with trace plots. Specifically, the trace plots reveal dramatic jumps in the parameter estimate within a single chain, which suggest the MCMC algorithm has switched between latent classes. Within-chain label switching can be resolved by implementing inequality constraints on model parameter(s) that most readily distinguish the latent classes (Depaoli et al., 2016; Lu et al., 2011).

2.4 Common Modeling Issues in Conditional Latent Class Analysis Models

The purpose of this dissertation is to address two common issues that appear when including predictor variables in LCA models. In Study 1, we explore the utility of different methods for addressing incomplete covariate data when using the ML three-step approach. Results from this study will help applied researchers in make informed decisions on how to handle incomplete covariate data. In Study 2, we explore the performance of small-variance priors in detecting direct effects between the covariate and latent class indicators. Detecting direct effects between the covariate and latent class indicator can be tricky. Study 2 results may give credence to the use of small-variance priors as another method available for detecting non-zero direct effects. Taken together, these studies will help aid applied researchers using conditional LCA models in their own research and highlight potential benefits of Bayesian estimation.

Chapter 3: Study 1 – Addressing Missing Data in Latent Class Analysis When Using Three-Step Approach

3.1 Introduction

Missing data occurs in most empirical datasets, regardless of efforts by substantive researchers to minimize missingness. Ignoring missing values can be problematic for two key reasons. First, partially complete cases could be substantively different from complete cases, and these differences need to be adjusted to prevent bias in parameter estimates. Second, partially complete cases are often removed from statistical analyses. The reduction in sample size can lead to estimation problems and poor statistical power (Collins & Lanza, 2010). Therefore, methodologists strongly recommend applied researchers address missing data in their analyses.

The methodological literature has identified three missing-data mechanisms (Little & Rubin, 2002; Rubin, 1976; Schafer, 1997; Schafer & Graham, 2002). Data can be missing not at random (MNAR), missing completely at random (MCAR), or missing at random (MAR). Data are considered MNAR when the probability of having incomplete data on variable Y is related to the values of Y itself, despite controlling for observed variables (Enders, 2010). When the probability of incomplete data on Y depends on another observed variable in the analysis model and not the values of Y itself, the data are considered MAR). Data is considered MCAR when the probability of incomplete Y data is not related to another observed variable or the values of Y itself. MNAR is the most problematic because there is a systematic reason for the missing cases, but there is no way to account for that reason. The parameter estimates cannot be adjusted for the unknown variable and will be biased. MCAR and MAR are considered ignorable missingness because modern missing data techniques are available to handles these types of missing data (Collins & Lanza, 2010).

In latent class models, missing cases in the class indicator variables can be easily addressed if the missing data mechanism is MCAR or MAR (Collins & Lanza, 2010; Kolb & Dayton, 1996). Software packages capable of fitting LCA models are typically equipped to address missing latent class indicators rigorously. SEM software (e.g., M*plus*) often defaults to model-based missing data procedures without additional syntax from the user. More specifically, M*plus* computes the parameter estimates with all the available information (i.e., using the MAR assumption). These software defaults allow missing class indicators to be addressed without much consideration from applied researchers. Unfortunately, including a grouping variable (i.e., covariate, predictor, independent variable) can complicate the situation. When estimating an LCA model that includes covariates, the user must decide on the estimation strategy (one-step vs. stepwise) and develop their own analysis plan for addressing incomplete covariates.

More recently, methodologists have begun to recommend a stepwise approach to estimating conditional LCA models (Vermunt, 2010; Asparouhov & Muthén, 2014). The primary reason for this recommendation is that applied researchers often wish to establish the LCA measurement model independently of the covariate(s). One way to achieve this aim is to use the ML three-step approach (Vermunt, 2010). The ML three-step approach uses the following general steps:

1. The best-fitting unconditional LCA model is identified.

2. A most likely class variable, M, is created using the latent class posterior distribution from Step 1. The conditional probabilities for the class assignment given true latent class membership are also computed. These computed quantities will be used as the estimated classification errors in class assignment.
3. The structural model is estimated, allowing for the inclusion of predictors of the latent class variable. M is used as a single, nominal indicator of the latent class variable. The computed quantities are used as fixed parameter values that describe the direct relationship between the latent class variable and M.

The analysis process for the ML three-step approach can be very cumbersome. Statistical software has helped make the method more accessible to applied researchers by automating many or all the estimation steps. Unfortunately, this automation can make for an inflexible modeling experience in which users cannot adapt the model specification. The inflexibility of the automatic method can be especially problematic when addressing incomplete covariates because the software defaults to listwise deletion, which deletes cases with any incomplete covariate data in the analysis.

Statistical software defaults to listwise deletion for incomplete covariates (and not for incomplete latent class indicators) because covariates are exogenous variables in the larger SEM. The outcome (i.e., the latent class variable) is conditional on the covariate, which has no distributional assumptions. Therefore, the covariate is assumed to be fixed and fully observed in a conditional mixture model (Sterba, 2014). The listwise deletion of exogenous covariates is not unique to conditional mixture models; conditional non-mixture models often specify covariates as exogenous variables (Sterba, 2014). The unique issue here is the common practice of using an automated stepwise approach that does not allow users to adapt model specifications to address the incomplete covariates. When users rely on the automatic ML three-step approach, all the available data are used when estimating the unconditional LCA model in the first step, but individuals with incomplete covariates will be removed from the analysis in the third step. Applied users may be caught off-guard by the reduced sample size caused by the listwise deletion of cases with incomplete covariates. A different method is preferred because listwise deletion decreases statistical power and biases parameter estimates under the missing-at-random (MAR) assumption (Little, 1992; Little & Zhang, 2011).

3.1.1 Methods for Addressing Incomplete Covariates in LCA Models

Three alternative methods have been identified to address incomplete covariates when using a manual ML three-step approach. The first method analyzes the data from individuals with complete and partially complete data together, and the model estimates are adjusted based on the information provided. The second method uses Bayesian estimation to internally impute incomplete covariate values using all available information. The third method imputes plausible values in place of missing values across many datasets, analyzes each dataset, and pools the parameter estimates from each analysis. Each of these methods will be described in detail in the following subsections.

3.1.1.1 FIML.

One state-of-the-art method for addressing missing data is *full information maximum likelihood* (FIML), which utilizes ML estimation to handle missing values. ML estimation involves the repeated audition of different combinations of population parameter values until a specific combination of values obtains the highest log-likelihood, which then represents the best fit of the model to data (Enders, 2010). FIML can address the missing values by altering the individual log-likelihood combinations to account for an unequal number of observations across participants. In addition, adjustments are made to the standard error (SE) computations. These computations often require the use of an iterative optimization algorithm such as the expectation-maximization (EM) algorithm. For a detailed explanation of calculating the missing data log-likelihood and SEs, see Enders (2010). In the methodological literature, FIML has earned state-of-the-art status because it yields unbiased parameter estimates under MCAR and MAR data (Schafer & Graham, 2002).

When using the ML estimator, statistical software often automatically addresses missing values in endogenous variables. For example, the ML three-step approach automatically handles missing latent class indicators via FIML during the first estimation step. To address incomplete covariates with FIML, a relatively straightforward programming trick is required. In the third step, the user must specify model parameters specific to the covariates (e.g., means, variances, and covariances) in addition to the conditional LCA model. By estimating covariate parameters, the software treats the covariates as dependent variables and applies a normality assumption. The EM algorithm will then address the missing values by maximizing the joint likelihood (Sterba, 2014). Significantly, this programming trick does not change the interpretation of model parameters (e.g., the regression coefficient for "c on x" maintains the same meaning). Past simulation research suggests the EM algorithm effectively addresses incomplete covariates when using a one-step approach to estimation (Sterba, 2014). One drawback of this approach is that the EM algorithm can be computationally intensive when addressing missing data for several covariates in a single model. For each additional covariate with missing data, increasingly heavy numerical integration computations are required, which increases how long the analysis takes (Asparouhov & Muthén, 2021). Although the heavy numerical integration issue has primarily been discussed anecdotally (Asparouhov & Muthén, 2021), it does suggest there is reason to explore alternative strategies that may yield unbiased parameter estimates in a shorter time.

3.1.1.2 Bayesian Third Step.

An alternative strategy for addressing missing data is to use Bayesian estimation. In contrast to the ML estimator, the Bayesian estimator handles missing values with an internal imputation process (Asparouhov & Muthén, 2021). In M*plus*, Bayesian estimation is implemented with a MCMC estimation algorithm. There are multiple samplers available for use with MCMC methods, but we focus here on the Gibbs sampler (Gelman et al., 2014). The Gibbs sampler iteratively generates a sequence of model parameters, latent variables, and missing observations, which can be used to construct the posterior distribution upon convergence (Asparouhov & Muthén, 2010). The Bayesian estimator is considered a full-information estimator and typically produces similar results to the ML estimator with missing data (i.e., FIML; Asparouhov & Muthén, 2021). When using the ML three-step approach, the user can switch the estimator in the third step (i.e., *Bayesian third step*) and estimate parameters related to the incomplete covariate (e.g., means, variances, and covariances). The missing values are then modeled and imputed internally using an unrestricted model (Asparouhov & Muthén, 2021). Correlating the covariates is helpful because an observed covariate may help impute the missing values in the other covariate.

Perhaps the most significant advantage of using the Bayesian third step is that the method allows for the specification of priors on parameters of substantive interest (e.g., regression coefficients). Past methodological research suggests the stepwise approaches yield biased structural parameter estimates when the latent classes are not very distinct (Bakk & Vermunt, 2016; Nylund-Gibson, Grimm, & Masyn, 2019; Vermunt, 2010). Depending on the prior specification for the covariate effect, Bayesian estimation may improve the accuracy of the parameter estimate, especially under poor class separation. Although no previous simulation studies have explored the potential benefits of a Bayesian third step, previous mixture modeling research suggests Bayesian estimation can aid estimation for latent class models (Depaoli, 2013, 2014; Lu et al., 2011); therefore, it follows that Bayesian estimation may be helpful in this situation. Another advantage of the Bayesian estimator is computation speed, which can sometimes be much quicker than the numerical integration required by the EM algorithm (Asparouhov & Muthén, 2021). This is most evident when the analysis has many predictor variables with missing values, which requires heavier numerical integration. Despite these potential advantages, the Bayesian third step does have disadvantages in certain modeling situations. One critical limitation of the Bayesian third step in M*plus* is that the covariates are assumed to be normally distributed (Asparouhov & Muthén, 2021). In other words, the Bayesian third step may produce biased results when applied to a categorical covariate.

3.1.1.3 Multiple Imputation.

A final option for addressing missing data is another state-of-art method called *multiple imputation* (MI). MI consists of three distinct steps: the imputation phase, the analysis phase, and the pooling phase. During the imputation phase, m copies of the original dataset are created. Each copy contains plausible values for the missing values in the original dataset. During the analysis step, each of the m datasets is analyzed using the same statistical procedure that would have been performed had the original dataset been complete. Considering the analysis phase results in m parameter estimates and SEs, the results need to be pooled together into a single set of results during the pooling phase. For a detailed explanation of each of the three MI steps, see Enders (2010). In the context of the ML three-step approach, MI can either be performed before or after the first estimation step (Asaprouhov & Muthén, 2021). Using MI prior to the first estimation step may be useful if the researchers wish to include direct effects between the covariate and latent class indicators. In contrast, using MI after the first estimation step presents a more simplified strategy for addressing incomplete covariate data.

One factor that can further improve the accuracy of MI parameter estimates is the addition of auxiliary variables (AVs). AVs are ancillary variables that are not part of the primary analysis but are possibly correlated with missingness or the incomplete analysis model variables (Enders, 2010; Schafer, 1997). Although AVs are not of substantive interest, they can increase power and reduce bias in parameter estimates (Enders, 2010). The omission of an important AV can convert an analysis from MAR to MNAR because a variable is not explaining the missingness in the analysis. Even in modeling situations that are firmly MCAR or MAR, the addition of AVs can improve power and reduce bias with almost no downside (Collins, Schafer, & Kam, 2001; Graham, 2003; Schafer & Graham, 2002); therefore, methodologists strongly recommend an inclusive analysis strategy. MI can be implemented following the first analysis step when using the ML three-step approach. After estimating the unconditional LCA model, the user can impute the incomplete covariate values by specifying an imputation model for the incomplete covariates, which should include variables that are possibly correlated with the missingness (e.g., covariates, latent class indicators, AVs). After the imputation phase, each complete dataset is analyzed with a conditional LCA model. The parameter estimates are then pooled. Statistical software packages have automated much of the MI process, but the user still needs to make important decisions concerning the number of datasets and how convergence will be assessed. In addition, users should include AVs that can help predict missingness during the imputation phase (Enders, 2010).

Perhaps the biggest advantage of MI is the ability to specify the distribution of the incomplete covariate during the imputation phase. Both FIML and the Bayesian third step impose normality assumptions on the covariate, whereas MI allows the user to specify the covariate distribution. In practical settings with more trivial the missing data, the assumption violation introduced by FIML and the Bayesian third step may not be as problematic (Asparouhov & Muthén, 2021). One issue that should be considered is the past methodological work, which strongly recommends against the MI of covariates in mixture models because it can lead to substantial parameter bias (Enders & Gottschall, 2011). MI produces biased parameter estimates because covariate relations can vary across classes, making it highly unlikely that the imputation model will be specified correctly in applied settings (Enders & Gottschall, 2011; Sterba, 2014; Sterba, 2017). Although there is good reason to be wary of MI in mixture modeling, there are situations with non-normal covariates (e.g., binary covariate) that may benefit.

3.1.2 Overview of the Current Study

There is some evidence suggesting each of the described alternative methods could effectively address incomplete covariates. The aim of the present simulation study is to investigate the performance of four methods (listwise deletion, FIML, Bayesian third step, and MI) for handling incomplete covariates in LCA models. To achieve this aim, we varied the missing data mechanism (MAR and MCAR), percentage of missing data (15%, 35%, and 55%), covariate distribution (binomial and standard normal), and the strength of the covariate effect (weak, moderate, and strong). The simulation study consisted of 288 cells, and each cell had 500 replications. Previous studies using conditional LCA models have found 500 replications to be sufficient (Di Mari & Bakk, 2018; Janssen et al., 2019; Kim et al., 2016; Nylund-Gibson & Masyn, 2016).

3.2 Design

To investigate the performance of each of the missing data methods (i.e., listwise deletion, FIML, Bayesian third step, and MI), datasets were generated in M*plus* using a conditional LCA model specification. The population model for this study was divided into two parts: the measurement part of the model and the structural part of the model. The measurement part of the model consisted of the unconditional latent class model. The structural part of the model related the covariates to the latent class variable. A visual representation of the population model is displayed in Figure 4, where $u_1 - u_5$ are the binary observed class indicators, c is the categorical latent class variable, and x_1 and x_2 are the observed covariates predicting the latent class variable. The observed variables, x_1 and x_2, are predictors of the latent class variable. The regression coefficient for the effect of x_1 on c is labeled, γ_{11}, and the regression coefficient for the effect of x_2 on c is labeled, γ_{21}.

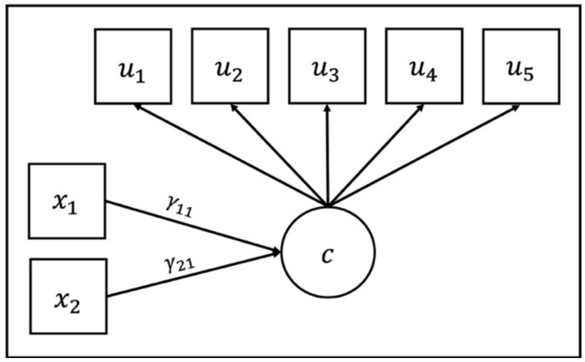


Figure 4. The population model for Study 2.

3.2.1 Measurement Model

The measurement model consisted of two classes ($K = 2$) of equal size ($\pi_1 = 0.5, \pi_2 = 0.5$) with five binary class indicator variables. All population models had moderate class separation, which was achieved by setting the item thresholds in Class 1 to $\tau = -1.25$ and the item thresholds in Class 2 to $\tau = 1.25$. Similar population values have been used in other simulation studies with conditional LCA models (Nylund-Gibson & Masyn, 2016; Masyn, 2013). These item threshold specifications corresponded to a conditional item-response probability (for all items) of 0.78 for Class 1 and 0.22 for Class 2. The sample size was fixed at $n = 500$ across conditions.[3]

3.2.2 Structural Model

The structural model of interest was a binomial logistic regression in which the latent class variable, c, was regressed on two covariates, x_1 and x_2. The x_1 variable was fully observed (i.e., no missing data) and followed a standard normal distribution. The effect of x_1 on c was held constant at $\gamma_{11} = -1$ across conditions. In contrast, x_2 varied across conditions. The structural model varied on the following factors: the strength of the effect of x_2 on c, the distribution of x_2, percentage of missing data on x_2, and the missing data mechanism.

3.2.2.1 Strength of the Covariate Effect.

The strength of the covariate effect of x_2 on c was set to be either weak ($\gamma_{21} = 0.5$), moderate ($\gamma_{21} = 1$), or strong ($\gamma_{21} = 1.5$). These regression coefficient specifications correspond to an odds ratio of 1.65, 2.72, and 4.48, respectively. Previous simulation studies on conditional LCA models have used similar regression coefficient values (Bakk, Obserski, & Verumnt, 2014; Nylund-Gibson & Masyn, 2016; Vermunt & Magidson, 2021). Manipulating the strength of the regression coefficient may be an important factor for illustrating the pitfalls of the automatic ML three-step approach, which results in listwise deletion. Even strong effects can be lost to reduced statistical power from listwise deletion.

3.2.2.2 Distribution of the Covariate.

One factor that is likely to impact the performance of the methods for addressing incomplete covariates is the distribution of covariate x_2. Therefore, x_2 will either follow a

---

[3] Sterba et al., (2014) found $n = 500$ to be a common sample size in mixture model studies in the social sciences. In addition, Nylund-Gibson and Masyn (2016) uses $n = 500$ for a smaller sample size condition in their simulation study with an LCA model with a continuous covariate.

standard normal distribution (like x_1) or a binomial distribution. For conditions with a binary x_2, the logit threshold was set to 0, which equates to a response probability of 0.5. To our knowledge, no previous simulation study has explored missing binary covariates in conditional LCA models, but it is likely missing binary covariates will pose a greater estimation challenge (Asparouhov & Muthén, 2021).

3.2.3 Missing Data Conditions

Missing data were generated on x_2 according to six different missing conditions. Specifically, missing data were generated under two different missing mechanisms: MAR and MCAR. In addition, the percentage of missing data on x_2 was generated as 15%, 35%, or 55%. These missing data conditions are in line with the limited methodological research on incomplete covariates in conditional mixture models (Sterba, 2014). Past research suggests listwise deletion can produce unbiased estimates under MCAR assumptions, but estimation efficiency and statistical power are reduced (Little, 1992; Little & Zhang, 2011; Skrondal & Rabe-Hesketh, 2014). In contrast, listwise deletion produces biased estimates under MAR assumptions. One way to illustrate the consequences of using an automatic ML three-step approach (i.e., listwise deletion) is to include MAR and MCAR missing mechanisms. The missing percentages selected (15% and 35%) represent realistic missing data situations in practice (Enders & Bandalos, 2001; Merkle, 2011; Wothke, 2000), and the 55% condition represents a worst-case scenario.

3.2.3.1 Missing Data Generation*.*

Missing data were generated in M*plus* using the MISSING option in the Montecarlo command, which allows the user to specify a logistic regression model to generate the missing data for one or more variables (i.e., x_2). More specifically, the logistic regression model was used to derive the intercept (α) and slope (β) parameters from the regression of a missing data indicator (R) on one of the latent class indicators (u_1):

$$p(R = 1|u_1) = \frac{\exp(\alpha + \beta u_1)}{1 + \exp(\alpha + \beta u_1)}. \tag{15}$$

In Equation (15), the missing data indicator (R) is a binary dependent variable that is scored as 0 for not missing and scored as 1 for missing on the dependent variables in the data generation model (i.e., x_2). The latent class indicator u_1 is considered a covariate (or predictor) of missingness in the model. Depending on the values selected for the intercept and slope of the logistic regression model, the probability of missingness, $p(R = 1|u_1)$, changes.

For conditions with MCAR missing data on x_2, the slope of u_1 was set to 0, suggesting the missingness was unrelated. Depending on the desired percent missing on x_2 (i.e., 15%, 35%, and 55%), different values for the logistic regression intercept were used (i.e., -1.734, -0.619, and 0.201) to achieve the desired probability of missingness (i.e., 0.15, 0.35, and 0.55). For conditions with MAR missing data, the slope of u_1 was set to 1.48. The slope selected produced a squared correlation of 0.40, which is in line with Enders and Mansolf (2018). The selected slope value indicates a moderately strong relationship between the cause of missingness (u_1) and the underlying latent probability of missing data. In addition to setting the slope to 1.48, different values for the logistic regression intercept were used (i.e., -2.66, -1.445, and -0.51) to achieve the desired probability of missingness (i.e., 0.15, 0.35, and 0.55). Table 1 summarizes the population values used to generate missing data conditions.

Table 1. The population values for generating missing data conditions.

	α		β	
Missing Assumption	MAR	MCAR	MAR	MCAR
Missing %				
15	-2.66	-1.734	1.48	0
35	-1.445	-0.619	1.48	0
55	-0.51	0.201	1.48	0

3.2.4 Analysis Models

In accordance with the manual ML three-step approach procedure in M*plus*, each generated dataset was analyzed with a 2-class unconditional LCA model.[4] Model constraints were applied to prevent label-switching across replications. After completing the first estimation step, the measurement part of the model needed to be combined with the structural part of the model as part of a larger SEM while addressing the missingness on the covariate x_2. Four methods were implemented to address the incomplete covariates: listwise deletion, FIML, Bayesian third step, and MI. The first three methods were implemented during the third estimation step, whereas MI was implemented prior to the third estimation step. All analyses were performed in M*plus*, and the process was automated using the M*plus* Automation package (Hallquist & Wiley, 2018) in R (R Core Team, 2019). In the following sections, the implementation of each of the methods will be described.

3.2.4.1 Listwise Deletion.

The listwise deletion method only required the specification of a conditional LCA model with a single latent class indicator, M, and two covariates, x_1 and x_2. With this model specification, covariates were considered exogenous variables in the SEM and were listwise deleted from the analysis during the third estimation step. Notably, this model specification is equivalent to using the automatic ML three-step approach.

3.2.4.2 FIML.

To address missing data in the third step with FIML, a conditional LCA model with a single latent class indicator, M, and two covariates, x_1 and x_2 was specified. In addition, the means, variances, and covariance of x_1 and x_2 were specified. By also estimating parameters specific to the covariates, x_1 and x_2 were treated as dependent variables by the M*plus* software. All missingness on x_2 was automatically addressed during model estimation by maximizing the joint likelihood, and an assumption is made that x_2 is normally

[4] The first step of the manual ML three-step approach requires the user to estimate an unconditional LCA. The most likely class variable, M, is saved using the SAVEDATA option and including the statements "FILE=Step3.dat" and "SAVE=CPROB". [4] In addition, x_1 and x_2 are saved in the new dataset by including these variables in the auxiliary statement. The new dataset contains $u_1 - u_5$, x_1, x_2, the individual posterior probabilities for each latent class, and M. During the third estimation step, the new dataset is used to specify the structural model "c on x_1 x_2", and M is fixed in each class using the logits from the "Logits for the Classification Probabilities the Most Likely Latent Class Membership (Column) by Latent Class (Row)" section of the output from the first step. These logits are used to take into account measurement error in M.

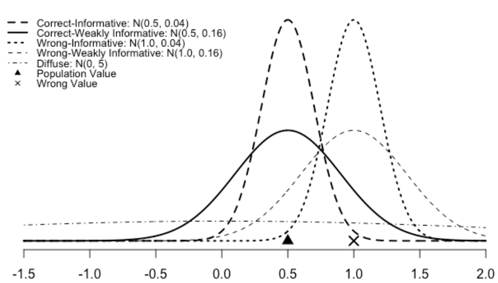
distributed. Numerical integration was required to estimate the joint likelihood; therefore, the Monte Carlo integration algorithm was applied.

3.2.4.3 Bayesian Third Step.

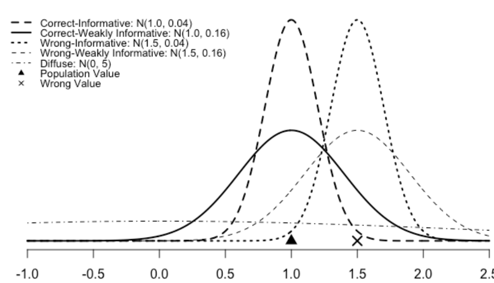
To implement Bayesian estimation in the third step, the estimator was switched from ML to Bayesian. In addition, a conditional LCA model with a single latent class indicator, M, and two covariates, x_1 and x_2 was specified. As was the case for FIML, the means, variances, and covariance of x_1 and x_2 were also specified in the model. By switching estimators and estimating parameters specific to the covariates (i.e., means, variances, covariances), the missing values on x_2 were imputed internally, and an assumption was made that x_2 was normally distributed.

To investigate the impact of prior specifications (or misspecifications), five prior conditions were considered for coefficients, γ_{11} and γ_{21}. The five prior specifications included the default prior in M*plus* for regression coefficients, *N*(0,5), which is considered *diffuse*. In addition, priors correctly centered on the population value (i.e., correct priors) and priors that are centered on a wrong value (i.e., wrong priors) were also considered. For both the correct and wrong prior conditions, the degree of informativeness was varied (i.e., informative vs. weakly informative). The combination of these prior means and variances produced four additional prior conditions: correct-informative, correct-weakly informative, wrong-informative, and wrong-weakly informative. Figure 5 provides a visual of the five prior conditions utilized for both γ_{11} and γ_{21}. Notably, the prior specifications change for γ_{21} depending on the strength of the covariate effect (e.g., 0.5, 1.0, or 1.5) with correct priors centered on 0.5, 1.0, or 1.5 and wrong priors centered on 1.0, 1.5, or 2.0, respectively. The variance hyperparameter was set to 0.04 in informative conditions and 0.16 in weakly-informative conditions. Figure 5 Panels A, B, and C provide the prior specifications for each γ_{21} condition, respectively. In contrast, γ_{11} is held constant at -1.0 across conditions, and all five prior conditions can be seen in Figure 2 Panel D. A single MCMC chain was utilized for parameter estimation to prevent between-chain label switching. The number of iterations was set to 30,000 for all analyses, and the first 15,000 iterations were discarded as burn-in. Convergence was assessed via careful examination of trace plots and autocorrelation plots and monitoring the PSRF.

Panel A ($\gamma_{21} = 0.5$)



Panel B ($\gamma_{21} = 1.0$)



Panel C ($\gamma_{21} = 1.5$) **Panel D ($\gamma_{11} = -1.0$)**

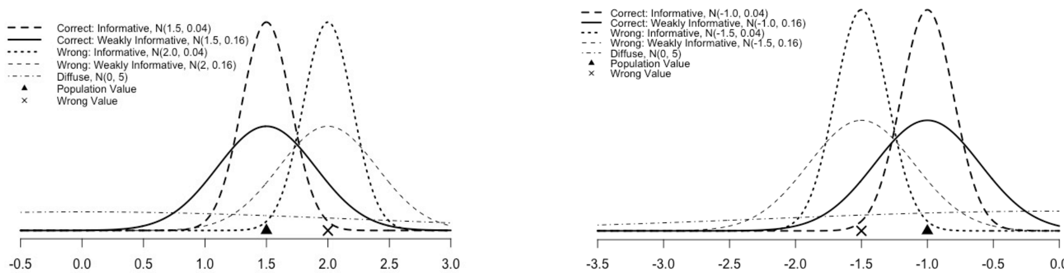


Figure 5. Prior conditions for the regression coefficients in the Bayesian third step. Panels A, B, and C show the prior specifications for γ_{21}, which have a different population value depending on the condition. Panel D shows the prior specification for γ_{11}, which is held constant across conditions.

3.2.4.4 Multiple Imputation.

The procedure for implementing MI was quite different from the previous three methods discussed. A manual ML three-step approach was still required, but missing data were imputed after the first estimation instead of proceeding directly to the final estimation step. The missing values on x_2 were imputed using the "Data Imputation" option in M*plus*, which helps automate the MI process. To aid the imputation process, x_1 and $u_1 - u_5$ were included in the imputation model. Depending on the distribution of x_2 (e.g., normal or binomial) in the condition, x_2 was either imputed as a normally distributed variable or a categorical variable. In other words, the imputation model was never misspecified. The number of imputations per replication was set to 20, which is in line with previous simulation studies using MI (Enders & Mansolf, 2018; Vera & Enders, 2021). Chain convergence was assessed using the M*plus* default criteria (e.g., the PSRF is close to 1 for each parameter). After the imputation step was complete, the structural model of interest (i.e., "c on x_1 x_2") was estimated using the ML estimator for each imputed dataset, which had no missing data. Parameter estimates were then averaged across the imputed datasets.

3.3 Results

The primary results of interest in this study pertain to the following parameters: γ_{11}, γ_{21}, γ_{01}, which represent the regression coefficients (i.e., "c on x_1" and "c on x_2") and the intercept of the binomial logistic regression, respectively. Bias for each parameter was calculated by subtracting the population value from the mean parameter estimate in the cell. In addition, the mean square error (MSE) was calculated by adding the variance of the estimates across the replications to the squared parameter bias.

3.3.1 Convergence

To prevent within-chain label switching across replications of the simulation study, a model constraint was included on the latent class indicator u_1 during the first estimation step such that the values adhered to the following order: Class 2 > Class 1. Overall, each cell in the simulation with a continuous covariate converged without issue and a set of stable estimates for the model parameters was obtained, regardless of the estimation strategy in the third step. In conditions using ML estimation (implemented via the EM algorithm) in the third step, M*plus* defaults were utilized, which specifies 20 sets of random starts in the initial stage and 4 final stage optimizations. In each replication, the log-likelihood was replicated indicating convergence. In conditions with Bayesian estimation in the third estimation step, convergence was assessed using PSRF. If PSRF values were less than 1.01, a replication was considered converged. According to this criterion, all replications using Bayesian estimation converged.

3.3.2 How to Read the Tables

Tables 2-4 display the bias and MSE results for population models with a continuous x_2 covariate and the regression coefficient γ_{21} set to 0.5, 1.0, and 1.5, respectively. Tables 5-7 display the bias and MSE results for population model models with a binary x_2 covariate and the regression coefficient γ_{21} set to 0.5, 1.0, and 1.5, respectively. Each method of addressing missing data in in x_2 are listed on the left side of the tables (i.e., Listwise Deletion, FIML, MI, and Bayes). For conditions that utilize Bayesian estimation, five prior specifications were included (i.e., Bayes-Correct Informative, Bayes-Correct Weakly Informative, Bayes-Wrong Informative, Bayes-Wrong Weakly Informative, and Bayes-Diffuse). The six different patterns of missing data are presented at the top of the table. Specifically, there is a combination of missing mechanism (MAR vs. MCAR) and missing percentage (15%, 35%, vs. 55%). For each parameter of interest (i.e., $\gamma_{11}, \gamma_{21}, \gamma_{01}$) the bias and MSE was calculated in each condition. To help illustrate the pattern of results, conditions with ± 0.1 bias are bolded in Tables 2-7.

Table 2. The bias and MSE in the regression coefficients for c on x_1 (i.e., γ_{11}), c on x_2 (i.e., γ_{21}), and the intercept (i.e., γ_{01}), under different missing data conditions for $\gamma_{21} = 0.5$ and a continuous x_2.

Missing Mechanism			MAR						MCAR					
Missing Percentage			15%		35%		55%		15%		35%		55%	
Approach	Parameter	Pop. Value	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
Listwise Deletion	γ_{11}	-1	-.015	.029	-.036	.046	-.044	.060	-.021	.031	-.032	.041	-.037	.066
	γ_{21}	0.5	.008	.023	.013	.032	.023	.044	.010	.022	.020	.029	.019	.046
	γ_{01}	0	**-.137**	.048	**-.371**	.179	**-.610**	.425	.001	.026	.005	.036	-.015	.043
FIML	γ_{11}	-1	-.010	.025	-.022	.028	-.010	.026	-.020	.024	-.022	.027	-.020	.030
	γ_{21}	0.5	.008	.023	.009	.030	.010	.041	.009	.022	.017	.028	.013	.044
	γ_{01}	0	.011	.026	-.004	.029	-.002	.024	-.001	.024	.007	.025	-.009	.027
MI	γ_{11}	-1	-.007	.024	-.016	.027	.000	.025	-.017	.023	-.015	.025	-.011	.028
	γ_{21}	0.5	-.006	.021	-.012	.026	-.037	.034	-.004	.021	-.013	.022	-.034	.035
	γ_{01}	0	.011	.025	-.005	.029	-.004	.024	-.001	.023	.009	.025	-.004	.026
Bayes-Correct Informative	γ_{11}	-1	-.010	.009	-.017	.010	-.010	.009	-.017	.009	-.017	.010	-.015	.010
	γ_{21}	0.5	.007	.010	.005	.010	.006	.010	.008	.009	.012	.009	.007	.010
	γ_{01}	0	.012	.026	-.002	.029	.002	.024	-.001	.024	.008	.025	-.005	.026
Bayes-Correct Weakly Informative	γ_{11}	-1	-.024	.019	-.036	.021	-.026	.020	-.034	.019	-.036	.021	-.035	.023
	γ_{21}	0.5	.015	.019	.015	.022	.019	.026	.017	.018	.025	.021	.024	.029
	γ_{01}	0	.012	.026	-.003	.030	.000	.025	-.001	.024	.008	.026	-.005	.027
Bayes-Wrong Informative	γ_{11}	-1	**-.260**	.079	**-.276**	.089	**-.282**	.091	**-.269**	.084	**-.277**	.089	**-.288**	.096
	γ_{21}	0.5	**.229**	.065	**.270**	.086	**.333**	.123	**.232**	.066	**.278**	.089	**.329**	.121
	γ_{01}	0	.011	.034	-.010	.041	-.012	.035	.000	.032	.008	.034	-.005	.037
Bayes-Wrong Weakly Informative	γ_{11}	-1	**-.113**	.035	**-.132**	.042	**-.132**	.041	**-.125**	.037	**-.132**	.041	**-.142**	.047
	γ_{21}	0.5	.092	.030	**.116**	.040	**.160**	.058	.095	.029	**.125**	.040	**.161**	.061
	γ_{01}	0	.011	.029	-.006	.034	-.005	.029	-.001	.027	.009	.029	-.006	.031
Bayes-Diffuse	γ_{11}	-1	-.031	.027	-.047	.032	-.039	.030	-.042	.027	-.047	.030	-.051	.036
	γ_{21}	0.5	.019	.025	.023	.033	.033	.048	.022	.024	.035	.031	.041	.052
	γ_{01}	0	.012	.027	-.004	.031	-.002	.026	-.001	.025	.009	.026	-.006	.028

Table 3. The bias and MSE in the regression coefficients for c on x_1 (i.e., γ_{11}), c on x_2 (i.e., γ_{21}), and the intercept (i.e., γ_{01}), under different missing data conditions for $\gamma_{21} = 1$ and a continuous x_2.

Missing Mechanism			MAR						MCAR					
Missing Percentage			15%		35%		55%		15%		35%		55%	
Approach	Parameter	Pop. Value	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
Listwise Deletion	γ_{11}	-1	-.032	.037	-.024	.049	-.056	.078	-.023	.032	-.036	.046	-.047	.075
	γ_{21}	1	.028	.037	.039	.050	.068	.086	.023	.032	.041	.053	.055	.082
	γ_{01}	0	**-.146**	.054	**-.368**	.176	**-.623**	.448	.006	.030	.004	.038	-.018	.048
FIML	γ_{11}	-1	-.028	.032	-.018	.032	-.014	.035	-.023	.026	-.028	.033	-.034	.043
	γ_{21}	1	.027	.037	.033	.049	.040	.072	.022	.032	.037	.053	.049	.081
	γ_{01}	0	.002	.028	.004	.029	.000	.031	.003	.027	.001	.031	-.014	.031
MI	γ_{11}	-1	-.016	.029	.011	.027	.028	.028	-.010	.024	.001	.028	.012	.031
	γ_{21}	1	-.009	.032	-.048	.037	-.082	.052	-.013	.028	-.043	.038	-.080	.050
	γ_{01}	0	.001	.027	.003	.028	-.005	.029	.003	.026	.007	.029	-.007	.028
Bayes-Correct Informative	γ_{11}	-1	-.018	.010	-.011	.009	-.007	.009	-.016	.008	-.017	.010	-.016	.010
	γ_{21}	1	.015	.011	.017	.010	.016	.010	.015	.009	.018	.011	.017	.010
	γ_{01}	0	.003	.028	.008	.029	.006	.030	.003	.027	.005	.030	-.008	.030
Bayes-Correct Weakly Informative	γ_{11}	-1	-.040	.023	-.032	.022	-.028	.022	-.036	.019	-.040	.023	-.042	.026
	γ_{21}	1	.038	.026	.044	.030	.051	.037	.036	.023	.048	.033	.054	.037
	γ_{01}	0	.002	.029	.007	.030	.003	.031	.004	.028	.005	.032	-.008	.031
Bayes-Wrong Informative	γ_{11}	-1	**-.304**	.105	**-.311**	.108	**-.324**	.115	**-.304**	.103	**-.316**	.111	**-.333**	.123
	γ_{21}	1	**.319**	.115	**.363**	.143	**.416**	.184	**.319**	.113	**.362**	.144	**.411**	.180
	γ_{01}	0	-.002	.040	-.001	.042	-.013	.046	.006	.039	.004	.044	-.010	.045
Bayes-Wrong Weakly Informative	γ_{11}	-1	**-.157**	.052	**-.161**	.053	**-.178**	.060	**-.153**	.046	**-.169**	.056	**-.193**	.069
	γ_{21}	1	**.165**	.059	**.203**	.077	**.267**	.117	**.163**	.054	**.207**	.082	**.264**	.112
	γ_{01}	0	.001	.033	.003	.036	-.007	.039	.005	.032	.005	.037	-.009	.039
Bayes-Diffuse	γ_{11}	-1	-.056	.037	-.051	.039	-.060	.046	-.050	.030	-.061	.039	-.080	.057
	γ_{21}	1	.055	.043	.070	.059	**.105**	.102	.051	.037	.077	.065	**.112**	.109
	γ_{01}	0	.002	.030	.006	.031	-.001	.034	.004	.028	.005	.033	-.009	.034

Table 4. The bias and MSE in the regression coefficients for c on x_1 (i.e., γ_{11}), c on x_2 (i.e., γ_{21}), and the intercept (i.e., γ_{01}), under different missing data conditions for $\gamma_{21} = 1.5$ and a continuous x_2.

Missing Mechanism			MAR						MCAR					
Missing Percentage			15%		35%		55%		15%		35%		55%	
Approach	Parameter	Pop. Value	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
Listwise Deletion	γ_{11}	-1	-.016	.037	-.033	.061	-.044	.092	-.020	.035	-.047	.055	-.058	.083
	γ_{21}	1.5	.027	.052	.052	.087	.069	.125	.037	.056	.072	.102	.083	.151
	γ_{01}	0	**-.137**	.051	**-.399**	.204	**-.610**	.438	.011	.036	.003	.047	-.025	.057
FIML	γ_{11}	-1	-.017	.033	-.025	.047	-.022	.052	-.020	.029	-.035	.041	-.040	.055
	γ_{21}	1.5	.026	.051	.041	.087	.038	.119	.035	.056	.064	.099	.069	.143
	γ_{01}	0	.018	.029	-.011	.035	.021	.038	.007	.033	.003	.039	-.020	.040
MI	γ_{11}	-1	.007	.030	.033	.034	.061	.038	.007	.026	.028	.031	.052	.036
	γ_{21}	1.5	-.042	.043	**-.113**	.061	**-.193**	.095	-.035	.045	-.098	.063	**-.184**	.093
	γ_{01}	0	.015	.028	-.014	.032	.008	.031	.007	.031	.011	.035	-.009	.033
Bayes-Correct Informative	γ_{11}	-1	-.012	.009	-.012	.010	-.012	.010	-.012	.008	-.014	.010	-.016	.010
	γ_{21}	1.5	.017	.009	.018	.009	.028	.009	.020	.009	.024	.011	.020	.009
	γ_{01}	0	.019	.029	-.005	.034	-.005	.034	.007	.033	.009	.039	-.012	.038
Bayes-Correct Weakly Informative	γ_{11}	-1	-.032	.022	-.035	.026	-.035	.027	-.034	.020	-.041	.025	-.044	.028
	γ_{21}	1.5	.049	.030	.058	.036	.061	.042	.056	.032	.072	.045	.071	.047
	γ_{01}	0	.018	.030	-.007	.036	.023	.039	.008	.034	.009	.040	-.012	.040
Bayes-Wrong Informative	γ_{11}	-1	**-.320**	.113	**-.334**	.123	**-.352**	.134	**-.322**	.114	**-.337**	.125	**-.356**	.138
	γ_{21}	1.5	**.395**	.165	**.430**	.194	**.468**	.226	**.399**	.170	**.435**	.200	**.468**	.228
	γ_{01}	0	.018	.042	-.020	.051	.010	.056	.010	.048	.009	.057	-.016	.058
Bayes-Wrong Weakly Informative	γ_{11}	-1	**-.174**	.056	**-.196**	.070	**-.221**	.081	**-.178**	.056	**-.203**	.071	**-.232**	.087
	γ_{21}	1.5	**.239**	.092	**.292**	.128	**.356**	.175	**.250**	.099	**.305**	.145	**.365**	.187
	γ_{01}	0	.018	.036	-.016	.044	.013	.050	.009	.041	.008	.049	-.015	.052
Bayes-Diffuse	γ_{11}	-1	-.050	.039	-.070	.060	-.090	.073	-.054	.035	-.080	.054	**-.107**	.080
	γ_{21}	1.5	.077	.064	**.116**	.119	**.166**	.191	.089	.070	**.143**	.139	**.193**	.226
	γ_{01}	0	.018	.031	-.011	.038	.019	.044	.008	.035	.008	.043	-.014	.046

Table 5. The bias and MSE in the regression coefficients for c on x_1 (i.e., γ_{11}), c on x_2 (i.e., γ_{21}), and the intercept (i.e., γ_{01}), under different missing data conditions for $\gamma_{21} = 0.5$ and a binary x_2.

Missing Mechanism			MAR						MCAR					
Missing Percentage			15%		35%		55%		15%		35%		55%	
Approach	Parameter	Pop. Value	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
Listwise Deletion	γ_{11}	-1	-.022	.030	-.013	.039	-.029	.068	-.017	.030	-.027	.039	-.033	.063
	γ_{21}	0.5	-.001	.068	-.009	.088	.009	.145	.008	.081	-.002	.092	.009	.148
	γ_{01}	0	**-.156**	.073	**-.368**	.195	**-.600**	.451	.003	.049	.001	.055	-.013	.078
FIML	γ_{11}	-1	-.019	.024	-.012	.026	-.010	.028	-.016	.025	-.024	.025	-.015	.024
	γ_{21}	0.5	.000	.069	-.005	.087	.006	.136	.008	.081	-.004	.091	.001	.144
	γ_{01}	0	-.006	.045	-.002	.047	-.001	.062	-.001	.046	.004	.048	-.007	.058
MI	γ_{11}	-1	-.018	.024	-.012	.025	-.009	.027	-.015	.025	-.022	.024	-.013	.024
	γ_{21}	0.5	-.009	.064	-.032	.081	-.037	.126	-.005	.076	-.035	.080	-.046	.123
	γ_{01}	0	-.002	.043	.010	.046	.017	.059	.006	.045	.020	.046	.018	.051
Bayes-Correct Informative	γ_{11}	-1	-.015	.009	-.010	.010	-.007	.010	-.012	.009	-.017	.009	-.009	.009
	γ_{21}	0.5	.000	.009	-.002	.008	.000	.007	.003	.010	-.001	.008	.001	.007
	γ_{01}	0	-.006	.030	-.001	.026	.005	.028	.001	.028	.004	.028	-.005	.025
Bayes-Correct Weakly Informative	γ_{11}	-1	-.030	.019	-.024	.020	-.020	.021	-.027	.019	-.034	.019	-.024	.018
	γ_{21}	0.5	.003	.033	-.002	.035	.005	.040	.010	.039	.002	.036	.006	.042
	γ_{01}	0	-.005	.037	.000	.034	.005	.037	.001	.036	.005	.035	-.006	.034
Bayes-Wrong Informative	γ_{11}	-1	**-.245**	.071	**-.240**	.070	**-.240**	.071	**-.241**	.070	**-.248**	.073	**-.243**	.070
	γ_{21}	0.5	**.352**	.133	**.381**	.153	**.423**	.186	**.354**	.136	**.381**	.154	**.420**	.184
	γ_{01}	0	**-.156**	.062	**-.168**	.061	-**.182**	.069	**-.145**	.055	**-.159**	.060	**-.187**	.067
Bayes-Wrong Weakly Informative	γ_{11}	-1	**-.114**	.035	**-.110**	.035	**-.111**	.037	**-.111**	.034	**-.121**	.036	**-.114**	.035
	γ_{21}	0.5	**.176**	.066	**.205**	.079	**.266**	.113	**.181**	.074	**.207**	.082	**.262**	.114
	γ_{01}	0	-.082	.046	-.093	.045	**-.116**	.054	-.074	.044	-.087	.045	**-.121**	.051
Bayes-Diffuse	γ_{11}	-1	-.039	.027	-.033	.028	-.034	.030	-.036	.028	-.045	.027	-.039	.027
	γ_{21}	0.5	.002	.069	-.005	.087	.006	.140	.011	.082	.001	.091	.009	.146
	γ_{01}	0	-.003	.046	.003	.047	.005	.063	.002	.047	.007	.048	-.005	.058

Table 6. The bias and MSE in the regression coefficients for c on x_1 (i.e., γ_{11}), c on x_2 (i.e., γ_{21}), and the intercept (i.e., γ_{01}), under different missing data conditions for $\gamma_{21} = 1$ and a binary x_2.

Missing Mechanism			MAR						MCAR					
Missing Percentage			15%		35%		55%		15%		35%		55%	
Approach	Parameter	Pop. Value	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
Listwise Deletion	γ_{11}	-1	-.004	.030	-.003	.037	-.026	.067	-.020	.032	-.019	.040	-.031	.069
	γ_{21}	1	-.015	.085	-.011	.124	.001	.160	.008	.092	-.011	.102	.028	.173
	γ_{01}	0	**-.133**	.060	**-.357**	.191	**-.611**	.470	.004	.051	.002	.055	-.011	.081
FIML	γ_{11}	-1	-.001	.025	-.016	.024	-.013	.031	-.019	.028	-.015	.025	-.011	.027
	γ_{21}	1	-.006	.087	.006	.128	.006	.158	.007	.092	-.015	.101	.014	.167
	γ_{01}	0	.013	.041	.007	.052	-.004	.062	.002	.047	.004	.048	-.005	.061
MI	γ_{11}	-1	.002	.025	-.011	.023	-.004	.029	-.015	.027	-.008	.024	.000	.025
	γ_{21}	1	-.032	.081	-.049	.110	-.076	.141	-.019	.085	-.074	.091	-.080	.139
	γ_{01}	0	.023	.040	.031	.050	.028	.060	.014	.046	.032	.046	.040	.055
Bayes-Correct Informative	γ_{11}	-1	-.003	.009	-.012	.008	-.008	.011	-.013	.009	-.013	.009	-.007	.009
	γ_{21}	1	-.002	.009	.000	.009	.001	.007	.002	.010	-.004	.008	.004	.007
	γ_{01}	0	.013	.028	.013	.030	.003	.030	.004	.031	.002	.030	.002	.027
Bayes-Correct Weakly Informative	γ_{11}	-1	-.014	.018	-.027	.017	-.023	.023	-.028	.020	-.027	.019	-.020	.019
	γ_{21}	1	.002	.038	.008	.044	.011	.041	.011	.040	-.001	.036	.018	.044
	γ_{01}	0	.015	.033	.013	.037	.003	.038	.005	.037	.005	.036	.001	.036
Bayes-Wrong Informative	γ_{11}	-1	**-.252**	.075	**-.265**	.080	**-.264**	.083	**-.262**	.081	**-.265**	.081	**-.262**	.080
	γ_{21}	1	**.377**	.152	**.409**	.176	**.445**	.205	**.381**	.155	**.404**	.171	**.446**	.206
	γ_{01}	0	**-.112**	.047	**-.130**	.055	**-.161**	.066	**-.121**	.053	**-.138**	.058	**-.153**	.058
Bayes-Wrong Weakly Informative	γ_{11}	-1	**-.113**	.035	**-.131**	.038	**-.133**	.045	**-.128**	.040	**-.130**	.039	**-.130**	.040
	γ_{21}	1	**.199**	.082	**.242**	.108	**.300**	.135	**.209**	.087	**.232**	.094	**.304**	.139
	γ_{01}	0	-.057	.039	-.077	.046	**-.114**	.056	-.066	.045	-.083	.046	**-.110**	.051
Bayes-Diffuse	γ_{11}	-1	-.021	.027	-.039	.027	-.039	.035	-.039	.031	-.038	.028	-.036	.031
	γ_{21}	1	.000	.087	.011	.129	.018	.162	.014	.093	-.005	.101	.032	.169
	γ_{01}	0	.019	.042	.015	.052	.004	.062	.008	.048	.011	.048	.000	.061

Table 7. The bias and MSE in the regression coefficients for c on x_1 (i.e., γ_{11}), c on x_2 (i.e., γ_{21}), and the intercept (i.e., γ_{01}), under different missing data conditions for $\gamma_{21} = 1.5$ and a binary x_2.

Missing Mechanism			MAR						MCAR					
Missing Percentage			15%		35%		55%		15%		35%		55%	
Approach	Parameter	Pop. Value	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
Listwise Deletion	γ_{11}	-1	-.002	.033	-.003	.047	.004	.065	-.014	.035	-.015	.043	-.023	.072
	γ_{21}	1.5	.006	.097	-.036	.122	-.047	.171	.003	.113	.006	.135	.046	.215
	γ_{01}	0	**-.138**	.069	**-.348**	.178	**-.572**	.425	.010	.053	.006	.057	-.007	.085
FIML	γ_{11}	-1	-.002	.028	-.004	.033	.000	.033	-.011	.031	-.009	.030	-.002	.032
	γ_{21}	1.5	.026	.101	.004	.131	.001	.189	.000	.112	-.001	.130	.022	.202
	γ_{01}	0	.007	.048	.009	.048	.001	.071	.007	.049	.004	.051	-.009	.063
MI	γ_{11}	-1	.004	.027	.010	.031	.019	.030	-.005	.030	.006	.027	.020	.028
	γ_{21}	1.5	-.023	.092	-.097	.109	**-.137**	.163	-.039	.103	-.098	.115	**-.124**	.164
	γ_{01}	0	.024	.047	.046	.046	.050	.070	.023	.048	.044	.050	.053	.058
Bayes-Correct Informative	γ_{11}	-1	-.003	.009	-.004	.010	-.003	.010	-.009	.009	-.009	.009	-.001	.009
	γ_{21}	1.5	.009	.008	.001	.007	.002	.006	.000	.009	.000	.007	.007	.007
	γ_{01}	0	.017	.033	.014	.031	.008	.037	.008	.033	.008	.034	.003	.031
Bayes-Correct Weakly Informative	γ_{11}	-1	-.015	.019	-.016	.022	-.014	.022	-.023	.021	-.022	.020	-.013	.020
	γ_{21}	1.5	.028	.039	.012	.039	.015	.040	.011	.043	.012	.040	.029	.043
	γ_{01}	0	.014	.039	.015	.036	.008	.045	.010	.040	.009	.039	.000	.038
Bayes-Wrong Informative	γ_{11}	-1	**-.272**	.085	**-.276**	.089	**-.280**	.090	**-.277**	.088	**-.281**	.090	**-.278**	.088
	γ_{21}	1.5	**.420**	.185	**.436**	.198	**.467**	.224	**.410**	.177	**.436**	.197	**.471**	.228
	γ_{01}	0	-.083	.049	**-.102**	.050	**-.127**	.064	-.091	.050	**-.105**	.054	**-.123**	.055
Bayes-Wrong Weakly Informative	γ_{11}	-1	**-.132**	.041	**-.139**	.047	**-.146**	.048	**-.139**	.045	**-.145**	.045	**-.144**	.046
	γ_{21}	1.5	**.263**	.111	**.284**	.124	**.341**	.160	**.244**	.107	**.282**	.125	**.351**	.170
	γ_{01}	0	-.052	.045	-.068	.044	**-.100**	.060	-.055	.049	-.072	.048	**-.101**	.053
Bayes-Diffuse	γ_{11}	-1	-.024	.030	-.028	.037	-.029	.037	-.033	.034	-.034	.033	-.031	.036
	γ_{21}	1.5	.037	.102	.017	.136	.025	.194	.013	.114	.015	.134	.053	.209
	γ_{01}	0	.016	.048	.020	.048	.012	.071	.016	.050	.014	.051	.000	.063

3.3.3 Continuous Covariate with Missing Data

3.3.3.1 Bias

Tables 2-4 provide the bias for the regression coefficients, γ_{11} and γ_{21}, and the intercept, γ_{01}. Across all three levels of covariate strength (i.e., 0.5, 1.0, 1.5), several important patterns of results emerged. Listwise Deletion produced unbiased regression coefficient estimates, regardless of condition. However, the intercept, γ_{01}, was consistently biased when the missing data mechanism was MAR. Using FIML to address the missing x_2 data resulted in unbiased regression coefficients and intercepts, regardless of the missing data pattern and the strength of the regression coefficient. Similarly, MI produced unbiased parameter estimates for γ_{11} and γ_{01} estimates, across all conditions. When the γ_{21} regression coefficient strength was weak or moderate (i.e., $\gamma_{21} = 0.5$ and $\gamma_{21} = 1.0$), MI also produced unbiased γ_{21} estimates. However, there were several conditions with a strong regression coefficient (i.e., $\gamma_{21} = 1.5$) that had biased γ_{21} estimates. As evidenced by Table 4, the γ_{21} regression coefficient was biased when using MI in conditions with 55% missing data. In addition, γ_{21} was biased when using MI when 35% missing data when the missing data mechanism was MAR.

When using Bayesian estimation in the third step, prior specification impacted results for the regression coefficients, γ_{11} and γ_{21}. Informative and weakly informative priors centered on the correct regression coefficient population value produced unbiased parameter estimates, regardless of the missing data pattern and regression coefficient strength. In contrast, informative and weakly informative priors centered on the wrong value biased the regression coefficient estimates. Diffuse priors always resulted in unbiased regression coefficients when the missing data percentage was 15%. However, diffuse priors biased γ_{21} parameter estimates when the percentage of missing data increased (i.e., 35% and 55%) and the strength of the regression coefficient increased (1.0 and 1.5), regardless of the missing data mechanism. The only exception to this trend was when the missing percentage was 35% and $\gamma_{21} = 1.0$. Notably, γ_{11} was also biased when the missing percentage was 55% and $\gamma_{21} = 1.5$.

3.3.3.2 Mean Square Error

Tables 2-4 display the MSE for the regression coefficients, γ_{11} and γ_{21}, and the intercept, γ_{01}. The pattern of MSE results were similar across all three levels of covariate strength (i.e., 0.5, 1.0, 1.5); however, the MSE values tended to be higher when $\gamma_{21} = 1.5$. Regardless of the condition, similar MSE values were obtained for the regression coefficients, γ_{11} and γ_{21}, when using Listwise Deletion, FIML, and MI. Across these three methods for addressing missing x_2 data, the MSE tended to increase as the percentage of missing data increase from 15% to 55%. Despite the similarities in the regression coefficient results, Listwise Deletion had inflated MSE values for the intercept, γ_{01}, when the missing data mechanism was MAR. In contrast, FIML and MI had lower MSE values for the intercept.

When using Bayesian estimation, the MSE for the regression coefficients were largely dependent on the prior specifications. Across all conditions with a continuous x_2, addressing missing data with a Bayesian third step using informative priors correctly centered on the population value resulted in the lowest MSE in the regression coefficients, γ_{11} and γ_{21}. In contrast, the highest MSE values for the regression coefficients were seen in conditions with informative priors centered on the wrong population value, regardless of the regression coefficient strength and missing data pattern. Weakly informative priors correctly centered on the population value were comparable to FIML and MI, whereas diffuse priors and weakly informative priors centered on the wrong value tended to have higher MSE than FIML and MI. For the regression coefficients, diffuse priors produced comparable MSE values to FIML and MI when $\gamma_{21} = 0.5$, but the MSE values were higher than FIML and MI when $\gamma_{21} = 1.0$ and $\gamma_{21} = 1.5$. Overall, the Bayesian estimator has the potential to produce the lowest and highest MSE for γ_{11} and γ_{21}, depending on the prior specifications.

3.3.4 Categorical Covariate with Missing Data

3.3.4.1 Bias

Tables 5-7 provide the bias for the regression coefficients, γ_{11} and γ_{21}, and the intercept, γ_{01}. The patterns of bias in the binary x_2 conditions were similar to the patterns seen in the continuous x_2 conditions. Regardless of the simulation condition, Listwise Deletion produced unbiased parameter estimates for the regression coefficients. However, Listwise Deletion severely biased the intercept parameter when the missing data mechanism was MAR. As the percentage of missing data increased, the intercept became increasingly biased when using Listwise Deletion. Both FIML and MI produced unbiased γ_{11} and γ_{01} estimates, across simulation condition. Despite these similarities, FIML produced unbiased γ_{21} estimates, whereas MI had some conditions with biased γ_{21} parameters. As seen in Table 7, MI introduced bias to the γ_{21} parameter when the missing data percentage was 55% and $\gamma_{21} = 1.5$, regardless of the missing data mechanism.

The patterns of bias in the binary x_2 conditions were highly influenced by the prior specifications when using the Bayesian third step. Informative and weakly informative priors centered on the correct population value produced unbiased estimates for the regression coefficients and intercept, regardless of the simulation condition. Diffuse priors also produced unbiased parameter estimates. In contrast, informative and weakly informative priors centered on the wrong value biased the regression coefficients, γ_{11} and γ_{21}. Notably, using priors centered on the wrong value for the regression coefficients also biased the intercept, γ_{01}, when the missing percentage was 55%.

3.3.4.2 Mean Square Error

Tables 5-7 provide the MSE for the regression coefficients, γ_{11} and γ_{21}, and the intercept, γ_{01}. Listwise Deletion, MI, and FIML tended to have similar MSE values for the regression coefficients in conditions with a binary x_2, regardless of the covariate strength (i.e., 0.5, 1.0, 1.5) and missing data pattern. In contrast, Listwise Deletion had some the highest MSE values for the intercept. When using Listwise Deletion, the combination of a higher percentage of missing x_2 data and the MAR missing data mechanism resulted in the highest MSE values for the intercept, γ_{01}.

When using Bayesian estimation, the MSE values for the regression coefficients was highly dependent on the prior specifications. The lowest MSE values for the regression coefficients were seen in conditions using informative priors that were correctly centered on the population values. The MSE was much lower in these conditions than FIML and MI. However, informative priors centered on the wrong population values tended to inflate the MSE for the regression coefficients. Weakly-informative priors correctly centered on the population value also produced lower MSE values for the regression coefficients when compared to FIML and MI. In contrast, weakly informative priors centered on the wrong population value tended to have similar MSE values for the regression coefficients when compared to FIML and MI. When using diffuse priors for the regression coefficients and intercept, Bayesian estimation produced MSE values that were comparable to FIML and MI. Thus, the advantages of Bayesian estimator were lost when using more diffuse priors.

3.3.5 Additional Points for Discussion

Regardless of whether the x_2 variable was continuous or binary, Bayesian estimation with correct informative prior specifications produced unbiased estimates and the lowest MSE values for the regression coefficients, γ_{11} and γ_{21}. When the x_2 variable was continuous, the MSE values for the intercept, γ_{01}, tended to be equivalent for FIML, MI, and Bayesian estimation with informative priors correctly centered on the population value. In contrast, when the x_2 variable was binary, Bayesian estimation with informative priors correctly centered on the population value had the lowest MSE values for the intercept. One possible (and likely) explanation for the Bayesian third step not always having the lowest MSE values for the intercept is that a diffuse prior was used on the intercept for Bayesian analysis models. If a more informative prior correctly centered on the intercept population value was used, it is likely Bayesian estimation would produce the lowest MSE values for the intercept, regardless of condition. When using the ML three-step approach, the user would have information about the intercept from class enumeration in the first estimation step that could be used to help specify a prior in the third estimation step.

3.4 Discussion

The primary aim of this simulation study was to explore the performance of available methods for addressing incomplete covariate data when using the ML three-step approach. To accomplish this aim, we generated data with different missing data mechanisms (MAR and MCAR), percentages of missing data (15%, 35%, and 55%), covariate distributions (binomial and standard normal), and strengths of the covariate effect (weak, moderate, and strong). Next, we analyzed the datasets with four different methods for addressing the incomplete covariate data: Listwise Deletion, FIML, MI, and Bayesian estimation. When using Bayesian estimation, a variety of prior specifications were considered. No prior simulation study has compared these methods for addressing incomplete covariate data when using the ML three-step approach.

Statistical software such as M*plus* defaults to Listwise Deletion when using the ML three-step approach. Previous methodological research suggests Listwise Deletion is a poor method for addressing missing data because it reduces sample size and power (Little, 1992; Little & Zhang, 2011). In addition, Listwise Deletion can bias parameter estimates when the missing data mechanism MAR (Little, 1992; Little & Zhang, 2011; Rabe-Hesketh & Skrondal, 2014). The results from the current study illustrate this point; Listwise deletion introduced bias into the intercept, γ_{01}, when the missing data mechanism was MAR. This is especially problematic because bias in the intercept suggests the latent class measurement model is not remaining intact during the third estimation step. For this reason, Listwise Deletion is the worst option for addressing incomplete covariate data when using the ML three-step approach.

In contrast to Listwise Deletion, FIML consistently produced unbiased parameter estimates, regardless of the missing data mechanism and the percentage of missing data. Despite misspecifying the categorical x_2 as continuous, FIML still produced unbiased parameter estimates for the regression coefficients. For this reason, we would recommend using FIML in modeling situation with a single covariate with missing data. If the user includes several covariates with missing data in the model, FIML may not be the best option because computation time increases with each additional covariate, see Asparouhov and Muthén (2021) for more details. Another important factor to consider when using FIML is the variability in the estimator. Although FIML provided unbiased parameter estimates, there was more variability in the estimator than Bayesian estimation with informative and weakly-informative priors. This trend was especially evident in conditions with a categorical x_2 variable.

The MI results were somewhat surprising. MI produced unbiased parameter estimates in all conditions with a covariate strength of $\gamma_{21} = 0.5$ and $\gamma_{21} = 1.0$. However, MI underestimated the γ_{21} regression coefficient in conditions with high percentage of missing data when the covariate strength was $\gamma_{21} = 1.5$. In conditions with a continuous x_2, MI underestimated γ_{21} when 55% of the x_2 data was missing and when 35% of the x_2 data was MAR. In conditions with a binary x_2, MI underestimated γ_{21} when 55% of the x_2 data was missing. One possible explanation for why MI produced biased regression coefficients in conditions with higher percentages of missing data is the number of imputations used. Specifically, we set the number imputations across conditions to 20, which is considered standard in simulation research (Enders & Mansolf, 2018; Vera & Enders, 2021). However, some past methodological research suggests the optimal number of imputations should reflect the percentage of missing data (e.g., 55 imputations for 55% missingness; Von Hippel, 2009; White, Royston, & Wood, 2011). Future methodological research should consider whether a greater number of imputations would improve the performance of MI. Another factor to consider is the variability in the estimator. When the x_2 variable was categorical, MI had much greater variability in the estimator than Bayesian estimation with informative and weakly-informative priors. Based on these findings, MI may not be the best choice when there is a high percentage of incomplete covariate data. Alternative methods (e.g., FIML and Bayesian estimation) provide unbiased γ_{21} estimates under similar conditions.

Bayesian estimation has the potential to be the best or worst method for addressing incomplete covariate data in the third step, depending on the prior specification. When using informative priors correctly centered on the population value of the regression coefficient, the parameter estimates were consistently unbiased and the variability in the estimator was very low, regardless of the strength of the covariate, distribution of the covariate, and the missing data pattern. Similarly, weakly-informative priors correctly centered on the population value produced unbiased parameter estimates, across conditions. However, when using informative or weakly informative priors that are centered on the wrong value, we see some of the highest levels of bias in the regression coefficients. For this reason, we would recommend the use of informative and weakly-informative priors in the third estimation step if *a priori* knowledge about the relationship between the covariate and latent class variable is available. When using Bayesian estimation in applied settings, it is important to always perform a prior sensitivity analysis. For an example of how to implement a prior sensitivity analysis, see Depaoli, Winter, and Visser (2020).

Considering wrong priors can have a dramatic impact on model results, applied users may be tempted to use diffuse priors for the regression coefficients in the third estimation step. When using diffuse priors for all parameters, Bayesian estimation tended to produce unbiased estimates in most conditions. The primary exception to this trend is when the incomplete covariate data is continuous, the strength of the covariate is either 1 or 1.5, and the missing percentage is high 55%. The reason these conditions tended to be trickier for diffuse priors is that the diffuse prior is centered on zero, whereas the population value was higher (i.e., 1 or 1.5). Although diffuse priors may be appealing to applied researchers who do not have much knowledge about the relationship between the covariate and the latent class variable, most of the advantages of Bayesian estimation disappear when using diffuse priors. An applied researcher would be as well off using FIML or MI in these situations. Regardless of the prior specifications on the regression coefficients, all Bayesian estimation conditions had a diffuse prior on the latent class intercept. It is likely the Bayesian estimator would have lower variability in the intercept if more informative priors were specified. In applied settings, the user can incorporate prior information about the class proportions by specifying more informative priors on the latent class intercept and regression coefficients.

The current study was not without limitations. The simulation study only explored the performance of the methods for addressing incomplete covariate data in conditions with moderate class separation and equal class proportions. These factors can have a dramatic impact on the performance of the ML three-step approach and may also influence the performance of the methods for addressing missing data. Bayesian estimation may be especially helpful in situations with poor class separation or a minority latent class (Depaoli, 2012; Depaoli, 2013; Kim, 2014; Lu et al., 2011; Nylund et al., 2007; Tueller & Lubke, 2010). Future research should examine the possible benefits of the Bayesian third step in these modeling situations. The current study was also limited to the conditional LCA model. These results may not be applicable for addressing incomplete covariate data in GMMs and mixture CFA models. MI worked well for many conditions in the simulation study, but previous research suggests MI does not typically work well in mixture models (Enders & Gottschall, 2011; Sterba, 2014; Sterba, 2017).

Overall, results from the current study suggest a variety of methods can be used to address the incomplete covariate data when using the ML three-step approach. Based on our findings, we would recommend the use of Bayesian estimation when it is possible to use informative or weakly-informative priors on the regression coefficients. In situations where informative priors are not possible, FIML works well when the missing data is limited to a single covariate, and MI works well when the missing data percentage is limited to 15% or 35%. Listwise deletion should be avoided whenever possible.

Chapter 4: Study 2 – Using Small-Variance Priors to Detect Covariate Misspecifications in Latent Class Analysis Models

4.1 Introduction

LCA is a modeling technique that allows for the identification of an underlying categorical latent variable from a set of observed latent class indicators. The LCA model is considered a measurement model because the categorical latent variable represents the different subgroups identified in the population. Recent methodological advances have primarily focused on expanding the LCA measurement model to include external variables (e.g., predictors, distal outcomes) as part of a larger structural model. Traditionally, LCA models with external variables were estimated using a one-step approach with the ML estimator (McCutcheon, 1987; Vermunt, 2010). The primary issue with the one-step approach is that each alternation to the external variables in the model results in the reidentification of the model, which can change the interpretation of the latent classes (Asparouhov & Muthén, 2014; Bakk & Kuha, 2018; Bolck, Croon, & Hagenaars, 2004; Vermunt, 2010). An alternative strategy to the one-step approach is to use a stepwise approach to estimation, where the LCA measurement model is established independently of the structural model (Asparouhov & Muthén, 2014; Vermunt, 2010).

Regardless of the estimation strategy being pursued (i.e., one-step vs. stepwise approach), the inclusion of external variables in the LCA model presents an opportunity for model misspecifications. For example, one of the most common external variables to include in LCA is a latent class-predicting covariate. By including a latent class-predicting covariate, researchers can explore variables that explain the clustering in the LCA measurement model. One way of adding a covariate to an LCA model is to regress the latent class variable on the covariate (Nylund-Gibson & Masyn, 2016). By specifying the conditional LCA model this way, an assumption is being made that the relationship between the covariate and observed latent class indicators is fully explained by the indirect effect of the covariate on the latent class variable. In other words, the direct effect of the covariate on each class indicator would be fixed to zero during model estimation. Unfortunately, this assumption does not always hold true in practice. Often, there are direct covariate effects on the observed class indicator variables in applied settings.

4.1.1 Covariate Misspecifications

Several previous methodological studies have explored the consequences of misspecifying covariate effects in LCA models (Collins & Lanza, 2010; Masyn, 2013; Nylund-Gibson et al., 2016; Petras & Masyn, 2010). For example, one simulation study found that the misspecification of covariate effects leads to the over-extraction of the number of latent classes during class enumeration (Nylund-Gibson et al., 2016). For this reason, a growing body of methodological research that suggests class enumeration should take place before modeling external variables (Collins & Lanza, 2010; Nylund-Gibson et al., 2016; Masyn, 2013; Petras & Masyn, 2010). Although covariates may provide additional information that aids class enumeration (Li & Hser, 2011; Lubke & Muthén, 2007; Muthén, 2002), covariates should only be included during class enumeration when the covariate relationships are known *a priori* (Petras & Masyn, 2010). Researchers are unlikely to know the covariate relationships in most applied settings in advance.

A seemingly well-fitting unconditional LCA measurement model can become a poorly fitting conditional LCA model when the covariate is misspecified, regardless of the estimation approach used. When using the one-step approach, ignoring direct effects between the covariate and latent class indicators can bias structural parameters (Janssen et al., 2019). Covariate misspecifications bias structural parameters even more under poor class separation (Janssen et al., 2019). In theory, the stepwise approach to estimation should safeguard the measurement model against covariate misspecifications. However, excluding direct effects still biases structural parameters even when using a stepwise approach to analysis (Asparohouv & Muthén, 2014; Janssen et al., 2019). In other words, failure to model direct effects between the covariate and latent class indicators can compromise the LCA model. Notably, recent methodological advances have been made on modeling direct covariate effects when using a stepwise approach. More specifically, Vermunt and Magidson (2021) demonstrate how the ML three-step approach can be modified to include direct effects during the first estimation step. The direct effects must first be detected to be accommodated in the procedure described in Vermunt and Magidson (2021). The following section discusses techniques available for detecting direct covariate effects in LCA models.

4.1.2 Methods for Detecting Covariate Misspecifications

The methodological research on detecting direct effects in conditional LCA models is limited to one previous simulation study (Janssen et al., 2019). The authors illustrated how both residual and fit statistics could be used to identify direct effects in the ML estimation framework. Residual statistics can test for potentially problematic restrictions in the model. For example, residual statistics could be used to test whether the correlation between a pair of class indicators should be freed. In addition, residual statistics can be used to determine whether the path between the covariate and a class indicator should be freed. For a detailed explanation of implementing residual statistics in M*plus* and Latent GOLD, see Visser and Depaoli (2022). Alternatively, an inferential method (e.g., Wald Test) could also be used to test whether the inclusion of a direct effect improves model fit. According to Janssen et al., (2019), the effectiveness of the residual and fit statistics at detecting direct effects largely depends on the number of direct effects present, the size of the direct effects, and whether the direct effects were class-specific. Although the detection methods in Janssen et al. (2019) were promising, the methods struggled to detect multiple direct effects.

Another study proposed the use of multiple indicator multiple cause (MIMIC) modeling procedures that are commonly used in the item-response theory (IRT) framework to detect differential item functioning (DIF) in conditional LCA models (Masyn, 2017). In the context of LCA models with a class-predicting covariate, DIF occurs in two main ways (Nylund-Gibson et al., 2016). First, the latent class indicator could function differently for individuals with different covariate values. Second, the probability of endorsing the latent class indicator could correspond to a particular difference in the covariate for an individual in a specific latent class. The second type of DIF would occur when there is a class-varying direct effect. An abundance of past methodological work suggests the MIMIC modeling procedure to detecting DIF in factor analysis models is effective (Finch, 2005; Muthén, 1985; Wang, Shih, & Yang, 2009; Willse & Goodman, 2008; Woods, 2009; Woods & Grimm, 2011). Masyn (2017) expanded on this methodological work to develop a similar MIMIC modeling procedure for LCA. The proposed LCA MIMIC modeling procedure is an iterative procedure that could be used in conjunction with a stepwise approach to estimation; see Masyn (2017) for a detailed explanation of how to implement the proposed LCA MIMIC modeling procedure. This new LCA MIMIC modeling procedure's performance has yet to be evaluated in a simulation study. However, Masyn (2017) points out that there may be

Type I error inflation due to the iterative testing procedure used to test each possible direct effect.

4.1.3 Small-Variance Priors

An alternative strategy for detecting non-zero direct effects in conditional LCA models that may be effective for a broader range of situations is Bayesian structural equation modeling (BSEM). BSEM allows for a very flexible modeling experience, where it is possible to relax model assumptions in restrictive models. Ample methodological research has demonstrated the value of using BSEM to relax model assumptions for confirmatory factor analysis (CFA) models (Muthén & Asparouhov, 2012; Stromeyer et al., 2015; Xiao, Liu, & Hau, 2019). For example, BSEM has been used to relax model assumptions about cross-loadings in CFA, residual correlations in CFA, and measurement non-invariance in MIMIC modeling (Muthén & Asparouhov, 2012). In addition, BSEM has been used to relax assumptions about measurement invariance in CFA models (Hoijtink & van de Schoot, 2018; Sedding & Leitgöb, 2018; Winter & Depaoli, 2020). The Bayesian methodology allows us to relax model assumptions by implementing *approximate-zero priors*, which are near-zero priors centered on zero with very small variances. By applying approximate-zero priors to parameters that are typically constrained to zero in the ML estimation framework, a less restrictive version of the model can be estimated, improving model fit and interpretation (Depaoli, 2021).

Approximate-zero priors can also help detect non-zero direct effects. For example, BSEM has handled direct effects between a covariate and factor indicator variables in CFA models (Muthén & Asparouhov, 2012). In traditional CFA models, the direct effects between covariates and the factor indicators are typically fixed to zero, introducing problematic model misspecifications. Using small-variance, normal priors centered at zero for all direct effects between the covariate and the factor indicators, it is possible to relax this strict assumption and say the direct effects are approximately zero. The near-zero priors will allow some "wiggle" room surrounding the cross-loadings, but the substantive meaning remains (e.g., the cross-loadings are minor and unimportant). If a cross-loading is truly non-zero, the data should conflict with the near-zero priors, resulting in a non-zero estimate for the cross-loading. The approximate-zero strategy can improve model fit and aid researchers in detecting non-zero cross-loadings (Muthén & Asparouhov, 2012).

Previous applications of approximate-zero priors in LCA models have been limited to relaxing assumptions about local independence between latent class indicators (Asparouhov & Muthén, 2011; Lee et al., 2020). More specifically, near-zero priors can help detect non-zero residual correlations between latent class indicators in LCA models. For a detailed guide for detecting and modeling conditional dependence in LCA, see Visser and Depaoli (2022). The current study seeks to expand approximate-zero priors' application to include direct effects between a continuous covariate and latent class indicators. Considering the approximate-zero strategy is effective for direct effects in CFA models (Muthén & Asparouhov, 2012), we anticipate similar benefits will be seen in LCA models.

4.2 Design

Random samples were generated from six different population models with known covariate relationships to investigate the utility of the approximate-zero strategy. This study aimed to illustrate how small-variance priors can effectively identify and model non-zero direct effects that commonly occur in applied settings. All data were generated using M*plus.* All population models were generated with two latent classes ($K = 2$), five binary class indicator variables ($u_1 - u_5$), and one standard normal covariate (x). For each population model, sample size (*N* = 500 and *N* = 1,000) and class proportions (equal and unequal) were varied. Each generated dataset was analyzed with several Bayesian conditional LCA models with different prior specifications. To explore the impact of prior specifications (or misspecifications), we included different prior conditions on γ_1 (i.e., x predicting c) and the direct effects (i.e., x predicting $u_1 - u_5$). The prior specifications for each population are discussed in detail below. There were 324 cells in this simulation study, and each cell had 500 replications. Previous studies using conditional LCA models have found 500 replications to be sufficient (Di Mari & Bakk, 2018; Janssen et al., 2019; Kim et al., 2016; Nylund-Gibson & Masyn, 2016).

4.2.1 Class Proportions

Data were generated according to two different class proportion conditions (equal vs. unequal). For the equal class proportion condition, the classes had a 50%-50% split ($\pi_1 = \pi_2 = 0.5$). For the unequal class proportion condition, the classes had an 82%-18% split ($\pi_1 = 0.82, \pi_2 = 0.18$). Considering unequal class sizes pose a greater estimation challenge in mixture models (Kim, 2014; Lubke & Tueller, 2010; Nylund, Asparouhov, & Muthén, 2007), this factor could be important for illustrating the value of informative prior specifications for the binomial logistic regression slope. The population values selected were equivalent to a previous simulation study that used conditional LCA models (Nylund-Gibson & Masyn, 2016).

4.2.2 Sample Size

Two levels of sample size were considered in the current study (*N* = 500 vs. *N* = 1,000). The *N* = 500 is a common sample size for mixture models in applied settings (Sterba et al., 2014), whereas *N* = 1,000 represents an ideal scenario. These sample sizes have previously been used in a simulation study using conditional LCA models (Nylund-Gibson & Masyn, 2016).

4.2.3 Population Models

Datasets were generated from different LCA population models with known covariate relations that commonly occur in applied research. As seen in Figure 6, LCA population models with six known covariate effect specifications for x were considered (population models 1-6, labeled P1-P6). Across the population models, there were two pathways through which the covariate x could influence an observed latent class indicator. The first pathway is indirect via the latent class variable c, as seen in the P1-P4 population models. The second pathway is a direct pathway that bypasses c, as seen in P2-P6 population models. Some population models have indirect and direct effects, whereas others have only indirect or direct effects.

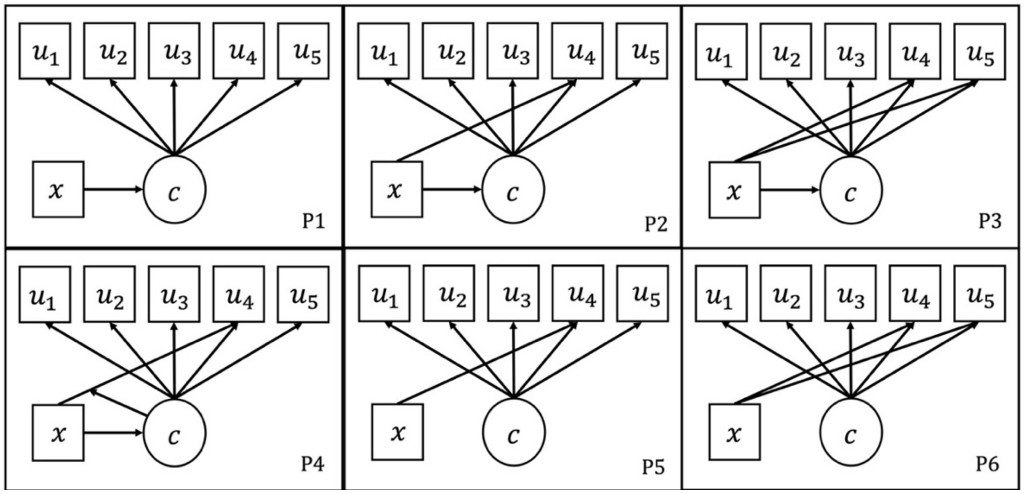


Figure 6. Conditional LCA population models in the current study (P1-P6). The paths from x to c represent the indirect effects of the covariate on the class indicators (i.e., u_1-u_5) via the latent class variable, c. Each path from x to an observed class indicator (e.g., u_4) represents a direct effect. In P4, the arrow pointing from c to the direct effect path indicates a class-varying direct effect.

4.2.3.1 Indirect Effects

P1, the first covariate effect specification in Figure 6, represents the most common conditional LCA model, with only an indirect effect on the latent class indicators via the latent class variable c (Nylund-Gibson & Masyn, 2016). The relationship between x and c can be represented with a single regression path (i.e., γ_1). In the P2, P3, and P4 population models, x was also specified to have an indirect effect on the latent class indicators via the latent class variable c. For population models P1-P4, the effect of x on c (i.e., specified as "c on x" in M*plus* language) was fixed to $\gamma_1 = 1$, which corresponds to an odds ratio of 2.72 for membership in Class 1 compared to Class 2 for a 1 standard deviation (SD) increase in x. The fifth and sixth covariate effect specifications in Figure 6, P5 and P6, differ from the other population models because there was no relationship between x and c (i.e., $\gamma_1 = 0$).

4.2.3.2 Direct Effects

The P2-P6 population models also have direct effects between x and one or more latent class indicators. Expressly, P2 and P5 were specified to have a single direct effect between x and u_4, whereas P3 and P6 were specified to have direct effects between x and u_4 as well as x and u_5. P4 was very similar to the P2 population model with a single direct effect between x and u_4. However, in the P4 covariate specification, the path between x and u_4 was class-varying. The class-varying effect is represented in Figure 6 with the arrow pointing from c to the path between x and u_4. All the population models (excluding P1) have a direct effect of x on u_4, which can be represented with the regression coefficient β_4. Across population models, $\beta_4 = 1$, which corresponds to an odds ratio of 2.72 for item endorsement of u_4 for a 1 SD increase in x, given latent class membership. In population model P4, the class-varying direct effect of x on u_4 was specified such that the Class 1 direct effect was fixed to $\beta_{41} = 0.5$, and the Class 2 direct effect was fixed to $\beta_{42} = 1.5$.

Population models P3 and P6 also had a direct effect of x on u_5, which was set to $\beta_5 = 1$. Notably, both P3 and P6 represent a violation of the local independence assumption because covariate x was a shared antecedent of u_4 and u_5 above and beyond the latent class variable.

4.2.4 Analysis Models

This study assumes that the 2-class unconditional LCA model has already been correctly selected during class enumeration. Our primary aim is to accurately model the covariate relations with the measurement model (e.g., latent class variable and class indicators). For each replication from each population model in Figure 4, we estimated a series of LCA analysis models with different covariate effects and prior specifications. We varied the prior specifications for γ_1 and the direct effects (e.g., β_1-β_5). The levels for the γ_1 prior specifications and the direct effect prior specifications depended on the population model used for data generation. All analyses were performed in M*plus* using the Bayesian estimator with a single MCMC chain per parameter. Each analysis model used 30,000 iterations, and the first 15,000 iterations were discarded as burn-in. Convergence was assessed by carefully examining trace plots and autocorrelation plots and monitoring the PSRF. The M*plus* default priors were utilized for all other parameters in the analysis models. More specifically, the default prior specifications implement a normal prior $\sim N(0,5)$ for the individual item thresholds and the intercept of the latent class variable.

4.2.4.1 γ_1 Prior Specifications

Depending on the population model used for data generation, there may or may not be an indirect effect. For population models with an indirect effect (e.g., P1-P4), three levels of normal priors (informative-correct, informative-wrong, and diffuse) were considered for γ_1. Figure 7 provides a visual representation of the three levels of priors used for γ_1 in P1-P4. The *informative-correct* conditions assigned a normal prior $\sim N(1,0.04)$ for γ_1, which is an informative prior correctly centered on 1. The *informative-wrong* conditions assigned a normal prior $\sim N(0.5,0.04)$ for γ_1, which is an informative prior incorrectly centered on 0.5. The *diffuse* conditions assigned a $\sim N(0,5)$ for γ_1, which is the M*plus* default prior.

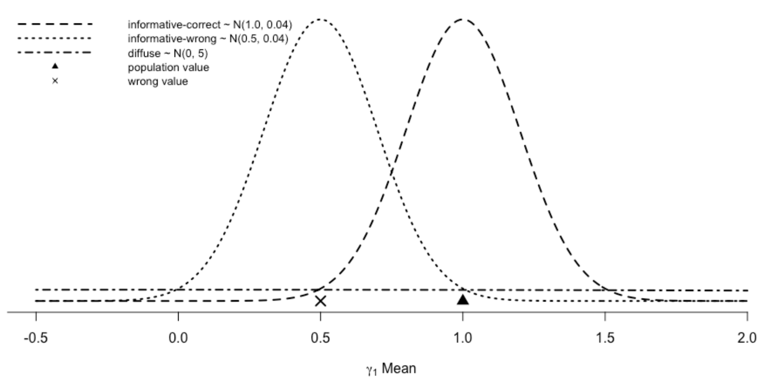


Figure 7. The prior specification levels (informative-correct, informative-wrong, and diffuse) for γ_1 in population models P1-P4.

For population models without an indirect effect (i.e., P5 and P6), we explored the impact of misspecifying γ_1 in the analysis model. More specifically, we considered three different specifications for γ_1 (not specified, informative misspecification, and diffuse misspecification). The γ_1 parameter was not included in the analysis model in the *not specified* conditions, and no prior was assigned. For the γ_1 misspecification conditions, the normal prior $\sim N(0.5, 0.04)$ was assigned for the *informative misspecification* condition and the M*plus* default prior was used for the $\sim N(0,5)$ *diffuse misspecification* condition.

4.2.4.2 Direct Effect Prior Specifications

All population models were analyzed with two levels of small-variance priors (overall and class-specific) for the direct effects of $u_1 - u_5$ on x, which can be represented with regression coefficients β_1-β_5. For the small-variance *overall* prior level, β_1-β_5 were assigned a normal prior $\sim N(0.0, 0.0025)$. For the small-variance class-specific prior level, the direct effects were estimated as class-varying effects (i.e., Class 1 = β_{11}-β_{51}; Class 2 = β_{12}-β_{52}), and each regression coefficient was assigned a normal prior $\sim N(0.0, 0.0025)$. Conditions utilizing the small-variance overall (SVO) priors will likely have fewer estimation problems (Nylund-Gibson & Masyn, 2016); however, the small-variance class-specific (SVCS) priors are essential for detecting class-varying effects such as β_{41} and β_{42} in P4. The small-variance prior conditions aim to determine if the approximate-zero strategy effectively detects non-zero direct effects. We would expect the regression coefficients that are truly zero in the population model to be estimated close to zero and the regression coefficients that are non-zero in each population model to "escape" the restrictive small-variance prior (especially when sample sizes are relatively higher). In addition to the two levels of small-variance priors, the P2-P6 population models had three additional levels of direct effect priors (informative-correct, informative-wrong, and diffuse). These three additional levels illustrate how a more parsimonious model can be estimated after detecting the non-zero direct effects with the small-variance priors. For these three levels of priors, only truly non-zero direct effects were included in the analysis models.

Population models P2, P3, P5, and P6 included the direct effect β_4, and population models P3 and P6 also had direct effect β_5. The *informative-correct* prior level assigned $\sim N(1.0, 0.04)$ to each non-zero direct effect. In contrast, *informative-wrong* prior level assigned $\sim N(0.5, 0.04)$ to each non-zero direct effect. The informative-correct prior is likely to be effective because it will be centered on the population value; however, this level of accuracy in the prior is unlikely to occur in practice. Therefore, we included an informative-wrong prior condition where the prior is centered on a wrong value (i.e., 0.5). The wrong value represents a weaker direct effect, which may be tempting because the estimated direct effect will be pulled towards zero under a small-variance prior. It is important to understand the impact of wrong informative prior specifications because, in most applications, the researcher is likely to be off from the "truth" (i.e., population value). The proposed wrong prior level mimics a situation that is likely to occur in applied research. In cases where the researcher does not want to include informative priors (e.g., no previous knowledge available, exploratory analysis), a common practice is to use the default prior in the software. To assess the viability of this estimation strategy, we included the direct effect *diffuse* prior level, which is the M*plus* default prior $\sim N(0,5)$. Figure 8 provides a visual representation of the five levels of direct effect priors (SVO, SVCS, informative-correct, informative-wrong, and diffuse) for population models P2, P3, P5, and P6.

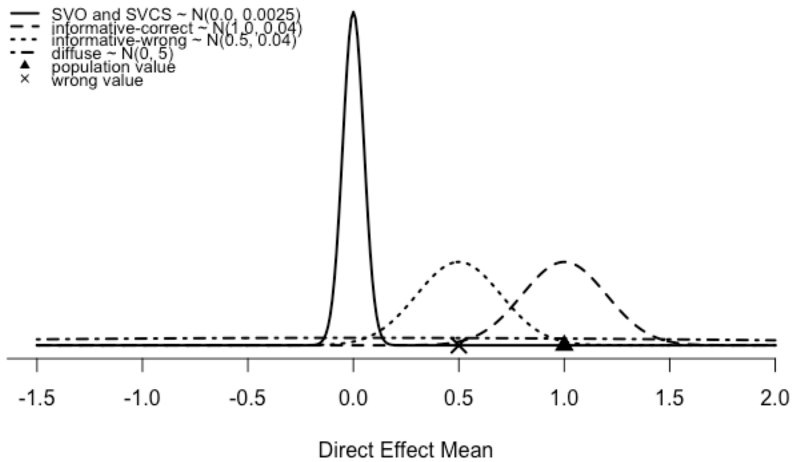


Figure 8. The prior specification levels (SVO, SVCS, informative-correct, informative-wrong, and diffuse) for the direct effects (e.g., β_4) in population models P2, P3, P5, and P6.

The P4 population model is unique because it has a class-varying direct effect ($\beta_{41} = 0.5$ and $\beta_{42} = 1.5$). To adjust for the class-varying direct effect in the analysis models, the *informative-correct* prior level was adapted to estimate the class-specific direct effect of x on u_4. The Class 1 (C1) regression coefficient β_{41} was assigned $\sim N(0.5,0.04)$, and the Class 2 (C2) regression coefficient was assigned $\sim N(1.5,0.04)$. Figure 9 illustrates the adapted informative-correct prior condition.

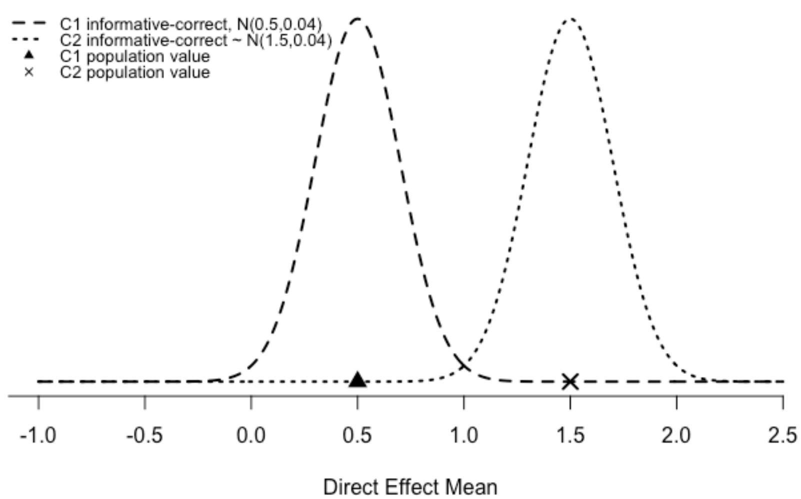


Figure 9. The direct effect prior specifications for the informative-correct conditions (C1 informative-correct, C2 informative-correct) in population model P4.

4.3 Results

4.3.1 Convergence

To prevent within-chain label switching across replications of the simulation study, a model constraint was included on the latent class indicator u_1 such that the values adhered to the following order: Class 2 > Class 1. A single MCMC chain was utilized for parameter estimation to prevent between-chain label switching. The number of iterations was set to 30,000 for all analyses, and the first 15,000 iterations were discarded as burn-in. Each cell in the simulation study converged without issue and a set of stable estimates for the model parameters was obtained. Convergence for each replication was assessed using PSRF. If PSRF values were less than 1.01, a replication was considered converged. According to this criterion, all replications converged as expected.

4.3.2 Detecting Direct Effects with Small-Variance Priors

In this section, we examine the power of small-variance priors to detect non-zero direct effects between the covariate and the latent class indicators. If a non-zero direct effect escapes the small-variance prior that is centered on zero, there is evidence to suggest a direct effect may need to be included when estimating the relationship between x and c. If the direct effect is truly zero, it should be held to "approximately zero" by the small-variance prior. Small-variance priors were specified on the $\beta_1 - \beta_5$ parameters for each of the six population models (P1-P6). P1 results were included to illustrate whether an inflated Type I error is likely to occur when there is no direct effect in the population model. P2 and P3 results explored whether the performance of the small-variance priors was impacted by the number of direct effects (one vs. two). P4 results were included to determine whether a class-varying direct effect can be detected with small-variance priors. P5 and P6 were included to examine the impact of misspecifiying γ_1 when detecting direct effects with small-variance priors. Factors such as sample size and class sizes could impact the power to detect the direct effects when using small-variance prior.

Results from P1-P6 are presented in Tables 8-13, respectively; see Figure 6 for a visual representation of the population models. For each population model, two types of small-variance priors (SVO vs. SVCS) were applied to the direct effects of $u_3 - u_5$ on x, which can be denoted as $\beta_3 - \beta_5$. The SVO prior represents a small-variance prior applied to the overall direct effect, whereas the SVCS prior is a class-specific small-variance prior. In addition to the small-variance priors for the direct effect, three different levels of priors (informative-correct, informative-wrong, and diffuse) were applied to γ_1.

Results presented in Tables 8-13 include the average parameter estimates for three of the direct effects for each latent class (i.e., $\beta_3 - \beta_5$) and γ_1. In addition, the tables display the percentage of replications in which the null hypothesis (e.g., $\beta_5 = 0$) is rejected. When the population value is non-zero, this percentage represents an estimate of power for a single parameter (i.e., the probability of rejecting the null hypothesis when it is false). A cut-off of 80% was used as the standard for power. When the population value is truly zero, this percentage represents an estimate of the Type I error (i.e., the probability of rejecting the null hypothesis when it is true). We would expect a Type I error (i.e., false positive) rate of 5% or less when a population parameter equals zero.

Table 8. The average parameter estimates and % of significant coefficients for P1 parameters when using two levels of small variance priors (SVO and SVCS) on β_3-β_5 and three levels of priors (diffuse, informative-correct, and informative-wrong) on γ_1.

Sample Size		$N = 500$				$N = 1{,}000$				$N = 500$				$N = 1{,}000$			
Class Prop.		50%/50%		82%/18%		50%/50%		82%/18%		50%/50%		82%/18%		50%/50%		82%/18%	
Parameter	Pop. Value	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig
		(SVO) β_1-β_5~N(0,0.0025) & γ_1~N(0,5)								(SVCS) β_1-β_5~N(0,0.0025) & γ_1~N(0,5)							
$C1$: β_3	0	.000	0	.001	0	.000	0	.000	0	.001	0	-.005	0	.001	0	-.009	0
β_4	0	.000	0	.001	0	-.001	0	-.001	0	.000	0	-.005	0	-.001	0	-.010	0
β_5	0	.001	0	.001	0	.001	0	.000	0	.000	0	-.005	0	-.001	0	-.010	0
$C2$: β_3	0	.000	0	.001	0	.000	0	.000	0	-.002	0	.006	0	-.001	0	.013	0
β_4	0	.000	0	.001	0	-.001	0	-.001	0	-.001	0	.008	0	.000	0	.014	0
β_5	0	.001	0	.001	0	.001	0	.000	0	-.001	0	.007	0	.002	0	.016	0
γ_1	1	1.020	100	1.011	100	1.008	100	1.003	100	1.012	100	1.011	100	1.008	100	1.002	100
		(SVO) β_1-β_5~N(0,0.0025) & γ_1~N(1,0.04)								(SVCS) β_1-β_5~N(0,0.0025) & γ_1~N(1,0.04)							
$C1$: β_3	0	.000	0	.001	0	.000	0	.000	0	.001	0	-.005	0	.001	0	-.009	0
β_4	0	.000	0	.001	0	-.001	0	-.001	0	.000	0	-.005	0	-.001	0	-.010	1
β_5	0	.001	0	.001	0	.001	0	.000	0	.000	0	-.005	0	.000	0	-.010	0
$C2$: β_3	0	.000	0	.001	0	.000	0	.000	0	-.002	0	.007	0	-.001	0	.013	0
β_4	0	.000	0	.001	0	-.001	0	-.001	0	-.001	0	.008	0	.000	0	.014	0
β_5	0	.001	0	.001	0	.001	0	.000	0	.001	0	.007	0	.002	0	.014	0
γ_1	1	1.012	100	1.005	100	1.006	100	1.003	100	1.013	100	1.005	100	1.006	100	1.001	100
		(SVO) β_1-β_5~N(0,0.0025) & γ_1~N(0.5,0.04)								(SVCS) β_1-β_5~N(0,0.0025) & γ_1~N(0.5,0.04)							
$C1$: β_3	0	.003	0	.004	0	.003	0	.003	1	.004	0	-.003	0	.003	0	-.007	0
β_4	0	.003	0	.004	0	.002	0	.003	0	.002	0	-.003	0	.001	0	-.008	0
β_5	0	.005	0	.004	0	.004	1	.003	0	.002	0	-.003	0	.001	0	-.008	0
$C2$: β_3	0	.003	0	.004	0	.003	0	.003	1	.000	0	.008	0	.001	0	.014	0
β_4	0	.003	0	.004	0	.002	0	.003	0	.001	0	.009	0	.002	0	.015	0
β_5	0	.005	0	.004	0	.004	1	.003	0	.003	0	.009	0	.004	0	.016	0
γ_1	1	.872	100	.832	100	.922	100	.895	100	.873	100	.833	100	.923	100	.895	100

Table 9. The average parameter estimates and % of significant coefficients for P2 parameters when using two levels of small variance priors (SVO and SVCS) on β_3-β_5 and three levels of priors (diffuse, informative-correct, and informative-wrong) on γ_1.

Sample Size		$N = 500$				$N = 1{,}000$				$N = 500$				$N = 1{,}000$			
Class Prop.		50%/50%		82%/18%		50%/50%		82%/18%		50%/50%		82%/18%		50%/50%		82%/18%	
Parameter	Pop. Value	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig
		(SVO) β_1-β_5~N(0,0.0025) & γ_1~N(0,5)								(SVCS) β_1-β_5~N(0,0.0025) & γ_1~N(0,5)							
$C1$: β_3	0	-.020	0	-.016	0	-.026	2	-.020	1	-.013	0	-.022	0	-.020	0	-.032	3
β_4	1	.196	100	.219	100	.320	100	.351	100	.110	80	.185	100	.197	100	.303	100
β_5	0	-.020	0	-.014	0	-.024	3	-.017	1	-.012	0	-.020	0	-.018	0	-.029	3
$C2$: β_3	0	-.020	0	-.016	0	-.026	2	-.020	1	-.013	0	.004	0	-.020	0	.009	0
β_4	1	.196	100	.219	100	.320	100	.351	100	.111	82	.050	1	.197	100	.098	54
β_5	0	-.020	0	-.014	0	-.024	3	-.017	1	-.013	0	.004	0	-.019	0	.010	0
γ_1	1	1.183	100	1.192	100	1.142	100	1.148	100	1.209	100	1.228	100	1.174	100	1.190	100
		(SVO) β_1-β_5~N(0,0.0025) & γ_1~N(1,0.04)								(SVCS) β_1-β_5~N(0,0.0025) & γ_1~N(1,0.04)							
$C1$: β_3	0	-.019	0	-.014	0	-.025	2	-.019	1	-.012	0	-.021	0	-.019	0	-.031	3
β_4	1	.198	100	.222	100	.322	100	.353	100	.112	82	.188	100	.198	100	.305	100
β_5	0	-.018	0	-.013	0	-.023	3	-.016	2	-.011	0	-.019	0	-.018	0	-.028	3
$C2$: β_3	0	-.019	0	-.014	0	-.025	2	-.019	1	-.013	0	.005	0	-.020	0	.009	0
β_4	1	.198	100	.222	100	.322	100	.353	100	.113	86	.052	1	.199	100	.099	56
β_5	0	-.018	0	-.013	0	-.023	3	-.016	2	-.012	0	.005	0	-.019	0	.010	0
γ_1	1	1.120	100	1.109	100	1.113	100	1.110	100	1.136	100	1.129	100	1.139	100	1.141	100
		(SVO) β_1-β_5~N(0,0.0025) & γ_1~N(0.5,0.04)								(SVCS) β_1-β_5~N(0,0.0025) & γ_1~N(0.5,0.04)							
$C1$: β_3	0	-.015	0	-.011	0	-.021	2	-.015	1	-.010	0	-.018	0	-.017	0	-.028	2
β_4	1	.203	100	.227	100	.326	100	.357	100	.112	88	.192	100	.202	100	.308	100
β_5	0	-.014	0	-.009	0	-.020	2	-.013	1	-.009	0	-.016	0	-.016	0	-.026	2
$C2$: β_3	0	-.015	0	-.011	0	-.021	2	-.015	1	-.011	0	.006	0	-.017	0	.011	0
β_4	1	.203	100	.227	100	.326	100	.357	100	.116	89	.054	1	.202	100	.102	63
β_5	0	-.014	0	-.009	0	-.020	2	-.013	1	-.011	0	.006	0	-.017	0	.011	0
γ_1	1	.965	100	.923	100	1.019	100	.993	100	.980	100	.941	100	1.043	100	1.022	100

Table 10. The average parameter estimates and % of significant coefficients for P3 parameters when using two levels of small variance priors (SVO and SVCS) on β_3-β_5 and three levels of priors (diffuse, informative-correct, and informative-wrong) on γ_1.

Sample Size		$N = 500$				$N = 1{,}000$				$N = 500$				$N = 1{,}000$			
Class Prop.		50%/50%		82%/18%		50%/50%		82%/18%		50%/50%		82%/18%		50%/50%		82%/18%	
Parameter	Pop. Value	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig
		(SVO) β_1-β_5~N(0,0.0025) & γ_1~N(0,5)								(SVCS) β_1-β_5~N(0,0.0025) & γ_1~N(0,5)							
$C1{:}\,\beta_3$	0	-.042	3	-.038	2	-.059	27	-.052	18	-.025	0	-.033	0	-.042	3	-.057	17
β_4	1	.145	99	.160	98	.255	100	.292	100	.077	18	.116	78	.143	99	.222	100
β_5	1	.146	99	.159	98	.257	100	.293	100	.076	17	.114	71	.143	99	.222	100
$C2{:}\,\beta_3$	0	-.042	3	-.038	2	-.059	27	-.052	18	-.026	0	-.010	0	-.042	3	-.009	0
β_4	1	.145	99	.160	98	.255	100	.292	100	.078	20	.049	0	.144	98	.096	52
β_5	1	.146	99	.159	98	.257	100	.293	100	.080	25	.050	0	.146	99	.097	56
γ_1	1	1.511	100	1.673	100	1.421	100	1.472	100	1.570	100	1.764	100	1.52	100	1.593	100
		(SVO) β_1-β_5~N(0,0.0025) & γ_1~N(1,0.04)								(SVCS) β_1-β_5~N(0,0.0025) & γ_1~N(1,0.04)							
$C1{:}\,\beta_3$	0	-.040	2	-.035	1	-.057	23	-.049	16	-.024	0	-.034	0	-.041	3	-.056	17
β_4	1	.153	100	.176	100	.263	100	.302	100	.081	26	.131	90	.148	99	.237	100
β_5	1	.154	100	.175	100	.264	100	.303	100	.081	25	.129	89	.148	99	.237	100
$C2{:}\,\beta_3$	0	-.040	2	-.035	1	-.057	23	-.049	16	-.025	0	-.006	0	-.041	3	-.004	0
β_4	1	.153	100	.176	100	.263	100	.302	100	.082	26	.052	0	.149	100	.098	55
β_5	1	.154	100	.175	100	.264	100	.303	100	.084	34	.053	0	.152	100	.099	59
γ_1	1	1.303	100	1.327	100	1.313	100	1.316	100	1.338	100	1.367	100	1.384	100	1.382	100
		(SVO) β_1-β_5~N(0,0.0025) & γ_1~N(0.5,0.04)								(SVCS) β_1-β_5~N(0,0.0025) & γ_1~N(0.5,0.04)							
$C1{:}\,\beta_3$	0	-.037	1	-.032	1	-.053	21	-.045	11	-.023	0	-.033	0	-.039	2	-.055	16
β_4	1	.161	100	.186	100	.270	100	.311	100	.085	35	.142	94	.153	99	.248	100
β_5	1	.162	100	.185	100	.272	100	.312	100	.085	32	.140	93	.153	100	.248	100
$C2{:}\,\beta_3$	0	-.037	1	-.032	1	-.053	21	-.045	11	-.023	0	-.004	0	-.039	2	.000	0
β_4	1	.161	100	.186	100	.270	100	.311	100	.086	36	.053	0	.154	100	.098	54
β_5	1	.162	100	.185	100	.272	100	.312	100	.088	41	.054	0	.156	100	.099	59
γ_1	1	1.125	100	1.111	100	1.197	100	1.174	100	1.161	100	1.147	100	1.266	100	1.233	100

Table 11. The average parameter estimates and % of significant coefficients for P4 parameters when using two levels of small variance priors (SVO and SVCS) on β_3-β_5 and three levels of priors (diffuse, informative-correct, and informative-wrong) on γ_1.

Sample Size		$N = 500$				$N = 1{,}000$				$N = 500$				$N = 1{,}000$			
Class Prop.		50%/50%		82%/18%		50%/50%		82%/18%		50%/50%		82%/18%		50%/50%		82%/18%	
Parameter	Pop. Value	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig
		(SVO) β_1-β_5~$N(0,0.0025)$ & γ_1~$N(0,5)$								(SVCS) β_1-β_5~$N(0,0.0025)$ & γ_1~$N(0,5)$							
$C1$: β_3	0	-.019	0	-.013	0	-.024	2	-.018	1	-.009	0	-.018	0	-.013	0	-.026	2
β_4	0.5	.184	100	.149	100	.299	100	.237	100	.075	17	.111	86	.133	99	.179	100
β_5	0	-.018	0	-.012	0	-.022	2	-.015	1	-.009	0	-.016	0	-.012	0	-.024	2
$C2$: β_3	0	-.019	0	-.013	0	-.024	2	-.018	1	-.014	0	.003	0	-.022	0	.005	0
β_4	1.5	.184	100	.149	100	.299	100	.237	100	.134	94	.052	1	.238	100	.102	59
β_5	0	-.018	0	-.012	0	-.022	2	-.015	1	-.014	0	.003	0	-.022	0	.006	0
γ_1	1	1.172	100	1.183	100	1.131	100	1.150	100	1.191	100	1.210	100	1.154	100	1.183	100
		(SVO) β_1-β_5~$N(0,0.0025)$ & γ_1~$N(1,0.04)$								(SVCS) β_1-β_5~$N(0,0.0025)$ & γ_1~$N(1,0.04)$							
$C1$: β_3	0	-.017	0	-.012	0	-.023	2	-.017	1	-.009	0	-.017	0	-.013	0	-.025	2
β_4	0.5	.186	100	.151	100	.300	100	.239	100	.076	19	.112	88	.134	99	.181	100
β_5	0	-.016	0	-.011	0	-.021	1	-.014	1	-.008	0	-.015	0	-.011	0	-.023	2
$C2$: β_3	0	-.017	0	-.012	0	-.023	2	-.017	1	-.014	0	.003	0	-.022	0	.005	0
β_4	1.5	.186	100	.151	100	.300	100	.239	100	.136	95	.054	1	.240	100	.103	63
β_5	0	-.016	0	-.011	0	-.021	1	-.014	1	-.014	0	.003	0	-.021	0	.006	0
γ_1	1	1.111	100	1.105	100	1.104	100	1.112	100	1.124	100	1.120	100	1.123	100	1.135	100
		(SVO) β_1-β_5~$N(0,0.0025)$ & γ_1~$N(0.5,0.04)$								(SVCS) β_1-β_5~$N(0,0.0025)$ & γ_1~$N(0.5,0.04)$							
$C1$: β_3	0	-.014	0	-.009	0	-.019	1	-.014	1	-.007	0	-.015	0	-.011	0	-.023	1
β_4	0.5	.191	100	.156	100	.305	100	.243	100	.078	23	.115	90	.136	99	.183	100
β_5	0	-.013	0	-.008	0	-.017	1	-.011	1	-.006	0	-.013	0	-.009	0	-.021	1
$C2$: β_3	0	-.014	0	-.009	0	-.019	1	-.014	1	-.012	0	.004	0	-.020	0	.006	0
β_4	1.5	.191	100	.156	100	.305	100	.243	100	.141	97	.058	1	.244	100	.107	69
β_5	0	-.013	0	-.008	0	-.017	1	-.011	1	-.012	0	.004	0	-.019	0	.007	0
γ_1	1	.957	100	.919	100	1.010	100	.995	100	.969	100	.933	100	1.028	100	1.016	100

Table 12. The average parameter estimates and % of significant coefficients for P5 parameters when using two levels of small variance priors (SVO and SVCS) on β_3-β_5 with three levels of prior specifications (i.e., no prior, informative misspecification, and diffuse misspecification) for γ_1.

Sample Size		$N = 500$				$N = 1{,}000$				$N = 500$				$N = 1{,}000$			
Class Prop.		50%/50%		82%/18%		50%/50%		82%/18%		50%/50%		82%/18%		50%/50%		82%/18%	
Parameter	Pop. Value	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig
		(SVO) β_1-β_5~N(0,0.0025)								(SVCS) β_1-β_5~N(0,0.0025)							
$C1$: β_3	0	-.016	0	-.012	0	-.020	2	-.014	1	-.011	0	-.011	0	-.016	0	-.015	1
β_4	1	.275	100	.278	100	.412	100	.416	100	.167	100	.243	100	.275	100	.374	100
β_5	0	-.015	1	-.011	0	-.018	3	-.011	2	-.009	0	-.009	0	-.014	0	-.012	1
$C2$: β_3	0	-.016	0	-.012	0	-.020	2	-.014	1	-.011	0	-.004	0	-.016	0	-.006	0
β_4	1	.275	100	.278	100	.412	100	.416	100	.169	100	.071	8	.276	100	.130	99
β_5	0	-.015	1	-.011	0	-.018	3	-.011	2	-.012	0	-.005	0	-.016	0	-.006	0
		(SVO) β_1-β_5~N(0,0.0025) & γ_1~N(1,0.04)								(SVCS) β_1-β_5~N(0,0.0025) & γ_1~N(1,0.04)							
$C1$: β_3	0	-.024	0	-.019	0	-.028	4	-.022	2	-.016	0	-.017	0	-.021	0	-.022	2
β_4	1	.270	100	.273	100	.407	100	.411	100	.164	100	.241	100	.272	100	.371	100
β_5	0	-.022	1	-.017	1	-.027	5	-.019	4	-.013	0	-.015	0	-.019	0	-.019	2
$C2$: β_3	0	-.024	0	-.019	0	-.028	4	-.022	2	-.015	0	-.006	0	-.022	0	-.009	0
β_4	1	.270	100	.273	100	.407	100	.411	100	.166	100	.066	2	.273	100	.124	97
β_5	0	-.022	1	-.017	1	-.027	5	-.019	4	-.016	0	-.007	0	-.022	0	-.010	0
γ_1	0	.166	47	.227	52	.118	42	.159	45	.172	50	.238	56	.127	48	.176	52
		(SVO) β_1-β_5~N(0,0.0025) & γ_1~N(0,5)								(SVCS) β_1-β_5~N(0,0.0025) & γ_1~N(0,5)							
$C1$: β_3	0	-.020	0	-.015	0	-.025	3	-.018	1	-.013	0	-.014	0	-.019	0	-.019	1
β_4	1	.273	100	.276	100	.409	100	.413	100	.165	100	.243	100	.273	100	.373	100
β_5	0	-.019	1	-.014	1	-.023	3	-.015	2	-.011	0	-.012	0	-.017	0	-.016	2
$C2$: β_3	0	-.020	0	-.015	0	-.025	3	-.018	1	-.013	0	-.005	0	-.019	0	-.008	0
β_4	1	.273	100	.276	100	.409	100	.413	100	.168	100	.067	3	.274	100	.125	97
β_5	0	-.019	1	-.014	1	-.023	3	-.015	2	-.014	0	-.006	0	-.019	0	-.008	0
γ_1	0	.081	14	.013	15	.069	17	.082	15	-.090	17	.118	16	.080	22	.102	21

Table 13. The average parameter estimates and % of significant coefficients for P6 parameters when using two levels of small variance priors (SVO and SVCS) on β_3-β_5 with three levels of prior specifications (i.e., no prior, informative misspecification, and diffuse misspecification) for γ_1.

Sample Size		$N = 500$				$N = 1{,}000$				$N = 500$				$N = 1{,}000$			
Class Prop.		50%/50%		82%/18%		50%/50%		82%/18%		50%/50%		82%/18%		50%/50%		82%/18%	
Parameter	Pop. Value	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig	Est.	% Sig
		(SVO) β_1-β_5~N(0,0.0025)								(SVCS) β_1-β_5~N(0,0.0025)							
$C1$: β_3	0	-.047	6	-.033	2	-.058	28	-.038	10	-.033	0	-.033	2	-.047	5	-.044	12
β_4	1	.257	100	.266	100	.389	100	.402	100	.155	100	.234	100	.257	100	.362	100
β_5	1	.257	100	.266	100	.391	100	.403	100	.154	99	.233	100	.258	100	.363	100
$C2$: β_3	0	-.047	6	-.033	2	-.058	28	-.038	10	-.031	0	-.009	0	-.046	5	-.012	0
β_4	1	.257	100	.266	100	.389	100	.402	100	.153	100	.061	1	.256	100	.111	80
β_5	1	.257	100	.266	100	.391	100	.403	100	.155	100	.062	3	.258	100	.111	79
		(SVO) β_1-β_5~N(0,0.0025) & γ_1~N(1,0.04)								(SVCS) β_1-β_5~N(0,0.0025) & γ_1~N(1,0.04)							
$C1$: β_3	0	-.071	32	-.064	27	-.087	70	-.066	40	-.049	3	-.050	6	-.069	27	-.071	44
β_4	1	.237	100	.226	99	.369	100	.382	100	.138	94	.172	75	.240	100	.329	99
β_5	1	.238	100	.225	99	.371	100	.383	100	.138	94	.172	77	.240	100	.330	99
$C2$: β_3	0	-.071	32	-.064	27	-.087	70	-.066	40	-.049	4	-.032	4	-.069	25	-.025	4
β_4	1	.237	100	.226	99	.369	100	.382	100	.137	95	.060	3	.239	100	.104	62
β_5	1	.238	100	.225	99	.371	100	.383	100	.138	95	.061	3	.241	100	.105	62
γ_1	0	.345	90	.533	92	.272	96	.342	93	.395	93	.625	94	.320	99	.439	97
		(SVO) β_1-β_5~N(0,0.0025) & γ_1~N(0,5)								(SVCS) β_1-β_5~N(0,0.0025) & γ_1~N(0,5)							
$C1$: β_3	0	-.068	25	-.067	31	-.083	66	-.063	35	-.048	2	-.047	5	-.067	23	-.069	39
β_4	1	.234	100	.206	90	.372	100	.381	100	.139	93	.142	56	.241	100	.317	96
β_5	1	.240	100	.207	91	.374	100	.382	100	.129	93	.142	55	.242	100	.318	96
$C2$: β_3	0	-.068	25	-.067	31	-.083	66	-.063	35	-.047	3	-.041	6	-.067	23	-.030	8
β_4	1	.234	100	.206	90	.372	100	.381	100	.137	95	.058	1	.240	100	.104	64
β_5	1	.240	100	.207	91	.374	100	.382	100	.139	94	.058	2	.242	100	.105	64
γ_1	0	.295	71	.683	70	.235	85	.311	71	.364	79	.891	77	.290	93	.474	86

4.3.2.1 P1 – No Direct Effects

Table 8 displays the mean coefficient estimate and percentage of replications with a significant coefficient in each condition in population model P1, which has an indirect effect and no direct effects. The left side of the table contains the results for SVO priors, and the right side of the table has results for SVCS. Table 8 includes two different sample size conditions ($N = 500$ vs. $N = 1{,}000$) and two different class sizes (equal vs. unequal). The aim of including the P1 population model was to demonstrate that small-variance priors do not result in an inflated Type I error rate for the truly zero direct effects. As expected, the small-variance priors (SVO and SVCS) consistently produced approximately zero estimates for the truly zero direct effects, regardless of the sample size, class sizes, and prior specification. The number of false positives was acceptable (<5%) for the truly zero direct effects, regardless of condition. The γ_1 regression coefficient was unbiased in conditions with an informative-correct prior or diffuse prior on γ_1. In contrast, the informative-wrong prior on γ_1 resulted in underestimation on the direct effect.

4.3.2.2 P2 – One Direct Effect

Table 9 provides the mean coefficient estimate and percentage of replications with a significant coefficient in each condition in population model P2, which has an indirect effect and one direct effect β_4. The left side of the table contains the results for SVO priors, and the right side of the table has results for SVCS. Table 9 includes two different sample size conditions ($N = 500$ vs. $N = 1{,}000$) and two different class sizes (equal vs. unequal). The P2 population model represents an ideal modeling situation with a single direct effect.

When using SVO priors, the non-zero direct effect β_4 was consistently detected as non-zero, regardless of the sample size, class size, and the γ_1 prior. Although β_4 was underestimated, there was enough power to overcome the restrictive SVO prior, as evidenced by 100% of the replications having a significant coefficient. The truly zero direct effects, β_3 and β_5, had an acceptable level of false positives (<5%). When using the SVCS priors, the situation was more complicated. In conditions with equal class sizes, both latent classes had an acceptable level of power to detect β_4. In conditions with unequal class sizes, the non-zero direct effect was consistently detected in the majority class (C1), but there was inadequate power (<80%) to detect the direct effect in the minority class (C2). Across small-variance prior conditions (SVO vs. SVCS), there was adequate power to detect γ_1.

4.3.2.3 P3 – Two Direct Effects

Table 10 provides the mean coefficient estimate and percentage of replications with a significant coefficient in each condition in population model P3, which has an indirect effect and two direct effects, β_4 and β_5. The left side of the table contains the results for SVO priors, and the right side of the table has results for SVCS. Table 10 includes two different sample size conditions ($N = 500$ vs. $N = 1{,}000$) and two different class sizes (equal vs. unequal). The P3 population model includes a local independence assumption violation because the latent class indicators u_4 and u_5 are related via the covariate x.

The local independence assumption violation introduces complications when using small-variance priors. When using the SVO priors, there was enough power to detect the non-zero direct effects (β_4 and β_5), regardless of sample size, class sizes, and $\gamma_1$prior specification. However, in conditions with $N = 1{,}000$, there was an inflated Type I error rate (>5%) for the truly zero direct effects. As the sample size increased, the SVO prior struggled to hold truly zero direct effects to approximately zero when there was a local independence assumption violation. According to the right side of Table 10, the SVCS priors further complicated results. When $N = 1{,}000$ with equal class sizes, there was enough power to detect the non-zero direct effects, and there was an acceptable number of false positives (<5%) for the truly zero direct effects. When class sizes are unequal, there was enough power (>80%) to detect the non-zero direct effects in the majority class (C1). The only exception to this trend was when $N = 500$, and a diffuse prior was used for the γ_1. A small sample size $N = 500$ with equal class sizes had limited power to detect the non-zero direct effects when using SVCS priors.

Across SVO and SVCS priors, there was enough power to detect γ_1, regardless of condition and prior specification. However, γ_1 was grossly overestimated in conditions with diffuse and informative-correct priors on γ_1, whereas the informative-wrong prior tended to produce a less biased estimate. Overall, the γ_1 parameter estimates illustrate the impact a local independence assumption violation can have on conditional LCA models. When latent class indicators are related to one another via a covariate, the relationship between the covariate and the latent class variable can be inflated.

4.3.2.4 P4 – Class-Varying Direct Effect

Table 11 provides the mean coefficient estimate and percentage of replications with a significant coefficient in each condition in population model P4, which has an indirect effect and a class-varying direct effect, β_4. The left side of the table contains the results for SVO priors, and the right side of the table contains results for SVCS. Table 11 includes two different sample size conditions ($N = 500$ vs. $N = 1{,}000$) and two different class sizes (equal vs. unequal). The P4 population model represents a tricky modeling situation in which the direct effect in stronger in C2 ($\beta_4 = 1.5$) than in C1 ($\beta_4 = 0.5$).

The SVO priors were effective at detecting the presence of the non-zero direct effect β_4, regardless of sample size, class size, and γ_1 priors. Although the SVO prior is not capable, of detecting the class-varying aspect of the direct effect, it was effective at flagging the non-zero direct effect β_4 while constraining truly zero direct effects (i.e., β_3 and β_5) to be approximately zero. There truly zero direct effects rarely lead to false positives when using the SVO prior. The overall pattern of results for the SVO prior was akin to what was seen in Table 9 with the P2 population model, which also had a single direct effect.

The SVCS priors are important for identifying the class-varying aspect of the direct effect because the direct effect parameters are no longer held constant across classes. As seen on the right side of Table 11, the SVCS priors complicated the situation. The only condition in which the SVCS prior had enough power to detect the class-varying direct effect in each latent class was when $N = 1{,}000$ and the classes were equal in size. In contrast, when $N = 1{,}000$ with unequal class sizes, there was only enough power to detect the direct effect in the majority class (C1) but not the minority class (C2). When $N = 500$ with equal class sizes, there was enough power to detect the direct effect in C2 but not C1 because the direct effect is stronger in C2 than C1. When $N = 500$ with unequal class sizes, there was enough power to detect the direct effect in the majority class (C1) but not the minority class (C2).

4.3.2.5 P5 – No Indirect Effect with One Direct Effect

Table 12 provides the mean coefficient estimate and percentage of replications with a significant coefficient in each condition in population model P5, which has a single direct effect β_4 and no indirect effect. The left side of the table contains the results for SVO priors, and the right side of the table contains results for SVCS. In Table 12, the γ_1 regression coefficient is either correctly not specified (top panel), misspecified with an informative prior (middle panel), or misspecified with a diffuse prior (bottom panel). Table 12 includes two sample size conditions ($N = 500$ vs. $N = 1{,}000$) and two class sizes (equal vs. unequal). The P5 population model is a situation where a single latent class indicator u_4 is related to the covariate x, but the latent class variable c is unrelated to the covariate. Often researchers are unaware of the covariate relationships a priori; therefore, the P5 population model represents a realistic situation in which the indirect effect does not exist.

The SVO priors were effective at detecting the non-zero direct effect β_4, regardless of the γ_1 specification, sample size, and class proportions. The SVCS priors were also able to effectively detect the non-zero direct effect β_4 in most conditions. The only exception was in conditions with $N = 500$ and unequal class sizes, where there was not enough power to detect the direct effect in the minority class (C2). Across small-variance prior conditions (SVO and SVCS), the truly zero direct effects were held to approximately zero and had an acceptable level of false positive (<5%). Notably, in conditions with a misspecified γ_1 there was an inflated Type I error rate (>5%) for γ_1. The misspecification with an informative prior on γ_1 resulted in a higher number of false positives compared to the misspecification with a diffuse prior.

4.3.2.6 P6 – No Indirect Effect with Two Direct Effects

Table 13 provides the mean coefficient estimate and percentage of replications with a significant coefficient in each condition in population model P6, which has two direct effects (β_4 and β_5) and no indirect effect. The left side of the table contains the results for SVO priors, and the right side of the table contains results for SVCS. In Table 13, γ_1 is either correctly not specified (top panel), misspecified with an informative prior (middle panel), or misspecified with a diffuse prior (bottom panel). Table 13 includes two different sample size conditions ($N = 500$ vs. $N = 1{,}000$) and two different class sizes (equal vs. unequal). The P6 population model represents a very challenging modeling situation in which a local independence assumption violation is present without an indirect effect. Specifically, the covariate x is related to two latent class indicators, u_4 and u_5, but is unrelated to the latent class variable c.

There was enough power to detect the non-zero direct effects with SVO priors, regardless of the γ_1 specification (or misspecification), sample size, and class proportions. However, the SVO prior did not always hold the truly zero direct effects to be approximately zero. Specifically, there was inflated number of false positives in most conditions. The only condition with an acceptable level of false positives (<5%) for the truly zero direct effect had a sample size of $N = 500$, unequal class sizes, and γ_1 was not specified.

When using SVCS priors, there was enough power to detect non-zero direct effects when there were equal class sizes. When the class sizes are unequal, there was not enough power to detect the direct effect in the minority class (C2). The only condition with unequal class sizes that could detect the direct effect in the minority class had a sample size of $N = 1{,}000$ and γ_1 was not specified. Notably, there was an unacceptable level of false positives (>5%) for the truly zero direct effects when the sample size was $N = 1{,}000$. Additionally, when the sample size was $N = 500$ with unequal class sizes, conditions with a misspecified γ_1 also resulted in a higher level of false positives.

When using small-variance priors, the local independence assumption violation results in a dramatic inflation in the Type I error rate for γ_1. Misspecifying γ_1 resulted in a high number of false positives for γ_1, regardless of whether an informative or diffuse prior was utilized. The P6 population model results highlight how unreliable the γ_1 parameter estimate is when using small-variance priors in the presence of a local independence assumption violation.

4.3.3 Modeling Direct Effects with Bayesian Estimation

After the non-zero direct effect(s) have been detected, a more parsimonious conditional LCA model can be specified that only includes the non-zero direct effects (instead of all possible direct effects) and γ_1. In this section of results, we explore how robust the direct effect and γ_1 parameter estimates are to different prior specifications, using the datasets generated from the P2 population model.[5] Specifically, three levels of priors were specified on the direct effect (informative-correct, informative-wrong, and diffuse) and three levels of priors were used on γ_1 (informative-correct, informative-wrong, and diffuse) for each condition. Figure 10 provides the relative bias in the β_4 direct effect, whereas Figure 11 displays the relative bias in γ_1.

Figure 10 reveals the direct effect parameter estimate was robust to different prior specification on the direct effect and γ_1. Regardless of sample size, class size, and prior specifications, minimal bias was produced for the β_4 parameter estimate. In contrast, Figure 11 reveals the γ_1 parameter estimates was affected by γ_1 prior specification, but not the direct prior specification. When using an informative-wrong prior on the γ_1 regression coefficient, the γ_1 parameter tended to be underestimated. However, the informative-correct prior and the diffuse prior resulted in an unbiased γ_1, regardless of sample size, class size, and the prior specification on the direct effect. Overall, Figures 10 and 11 demonstrate that the conditional LCA model results are relatively robust to different prior specifications.

4.3.4 Modeling Class-Varying Direct Effects

In some conditional LCA modeling situations, the relationship between the covariate and a latent class indicator can be class-varying. To assess the impact of prior specification on model parameter estimates, we analyzed the data generated from the P4 population model with a variety of priors on the direct effect (informative-correct, informative-wrong, and diffuse) and γ_1 (informative-correct, informative-wrong, and diffuse). The informative-correct prior on the direct effect was the only condition that allowed for the estimation of a class-varying direct effect, whereas the diffuse and informative-wrong prior constrained the direct effect to be equal across classes. Considering γ_1 is typically of greater substantive interest, it is important to understand the impact of direct effect prior specification on the γ_1 parameter estimate.

Figures 12 and 13 display the relative bias in the β_{41} and β_{42} parameters, respectively. Unsurprisingly, specifying informative-wrong and diffuse priors biased the β_{41} and β_{42} parameters. The β_{41} tended to be overestimated, especially in conditions with equal

[5] The pattern of results for the P3 population model were very similar to the pattern of results for the P2 population model; therefore, the P3 population model results are stored in Appendix A.

class sizes. In contrast, the β_{42} parameter was underestimated, regardless of class sizes. The direct effect results were robust to the prior specification of γ_1. Figure 14 provides the relative bias in the γ_1 parameter. When an informative-wrong prior is specified on γ_1, the γ_1 was biased in conditions with a sample size of $N = 500$, but not conditions with $N = 1{,}000$. The γ_1 parameter had minimal bias when a diffuse or informative-correct prior was specified on γ_1, regardless of sample size and class size. Notably, the direct effect prior specification did not bias γ_1.

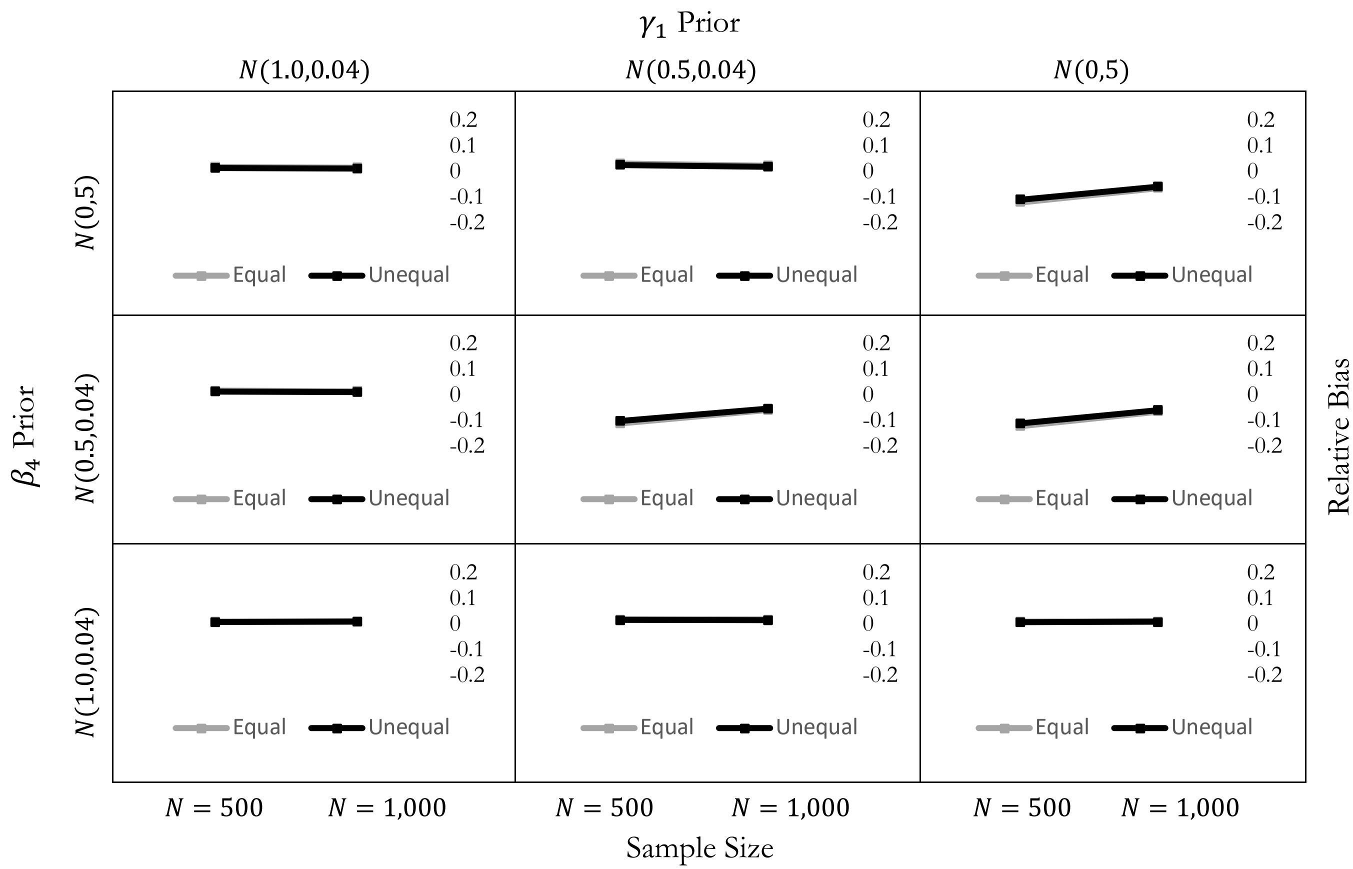


Figure 10. The relative bias in β_4 from P2 when using three levels of priors on γ_1 and three levels of priors on the direct effect.

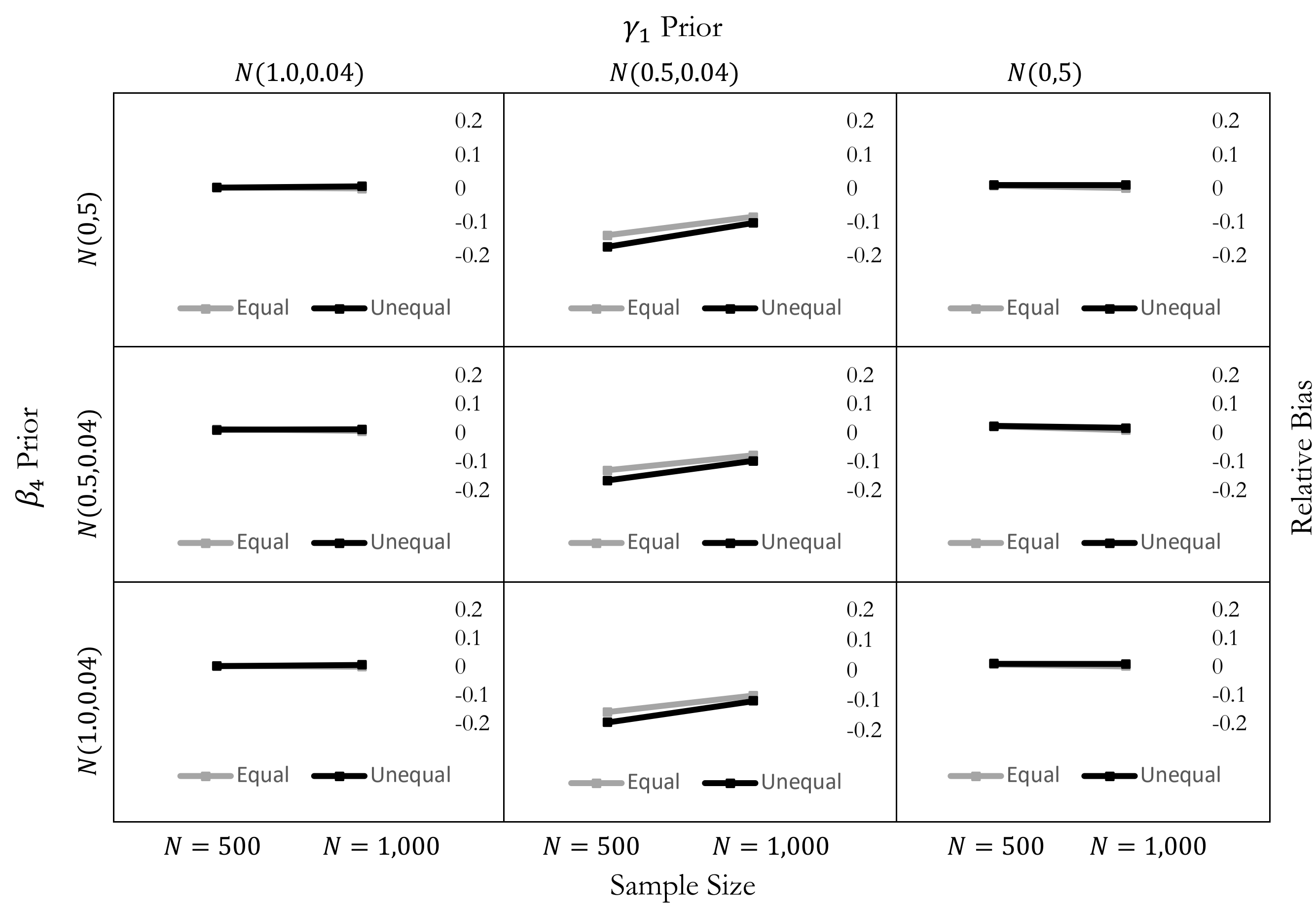


Figure 11. The relative bias in γ_1 from P2 when using three levels of priors on γ_1 and three levels of priors on the direct effect.

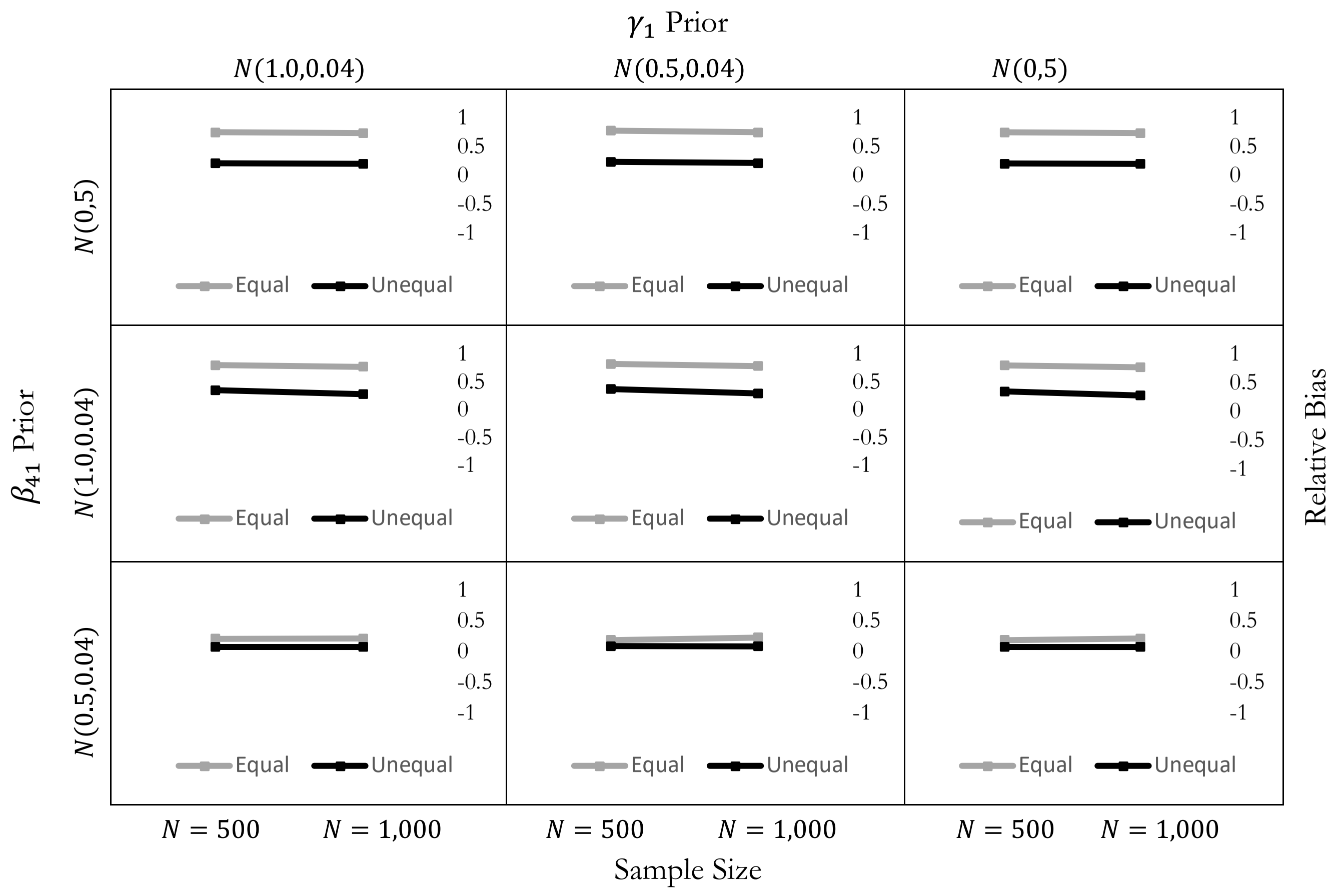


Figure 12. The relative bias in β_{41} from P4 when using three levels of priors on γ_1 and three levels of priors on the direct effect.

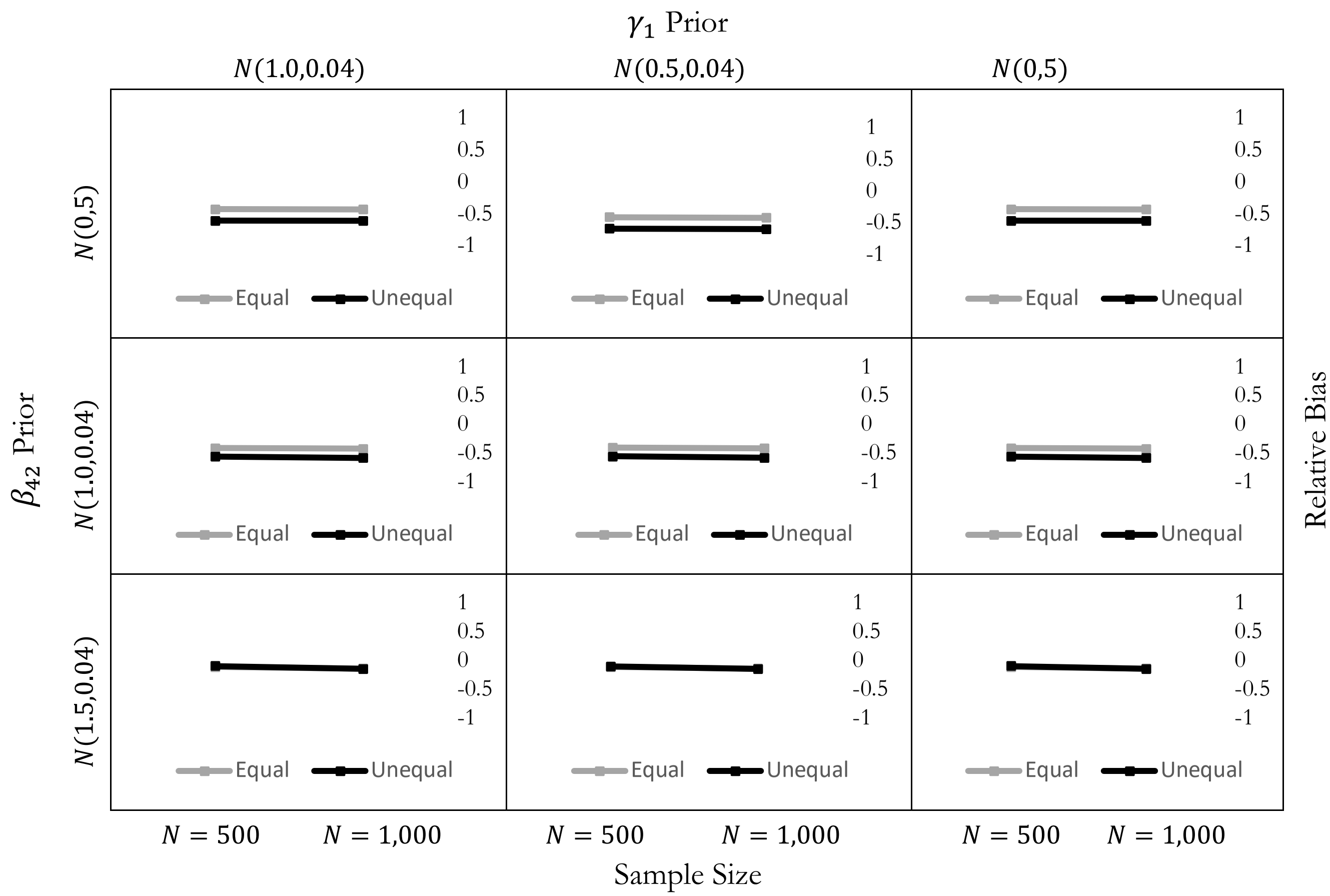


Figure 13. The relative bias in β_{42} from P4 when using three levels of priors on γ_1 and three levels of priors on the direct effect.

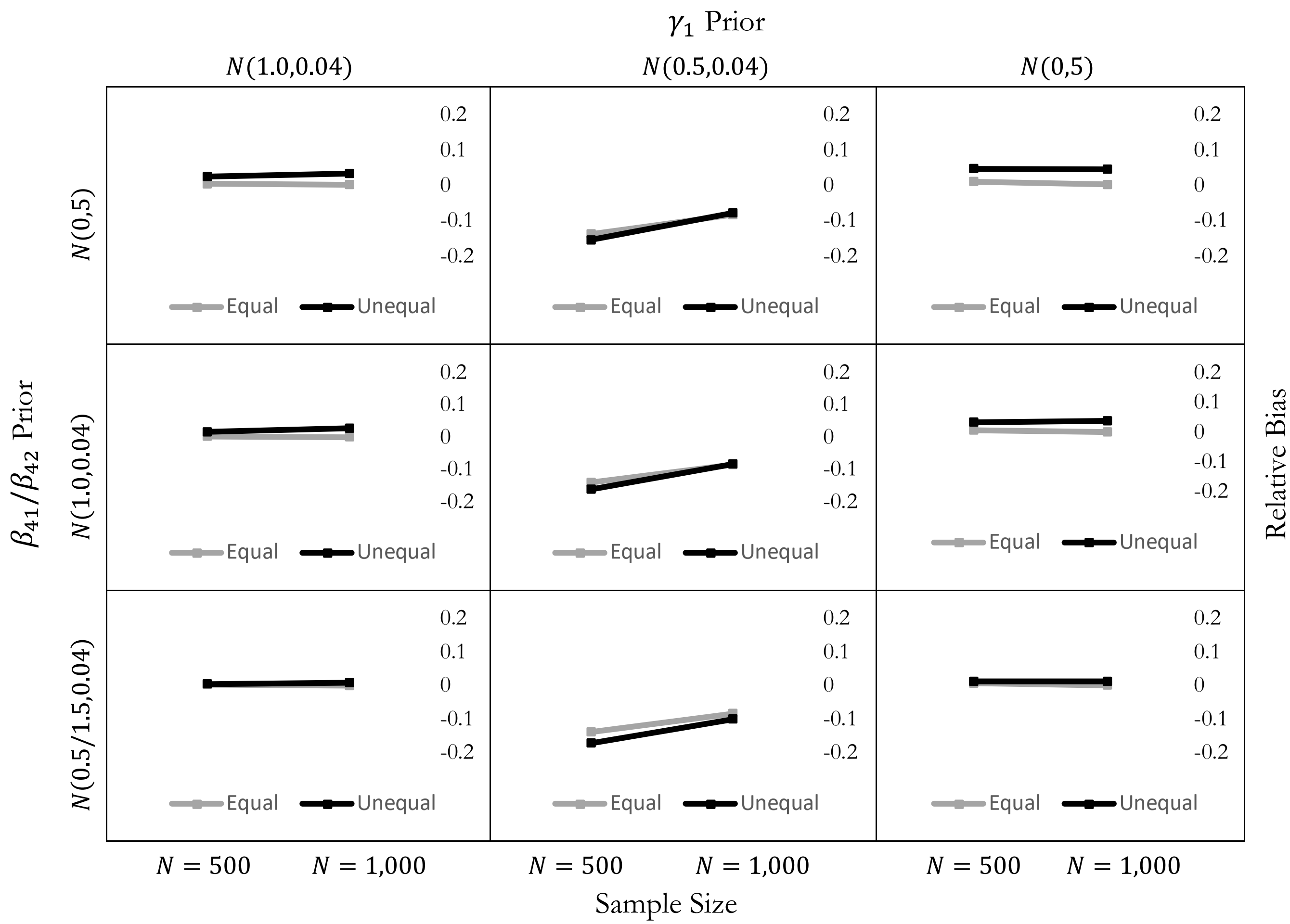


Figure 14. The relative bias in γ_1 from P4 when using three levels of priors on γ_1 and three levels of priors on the direct effect.

4.3.5 Misspecification of γ_1

Applied researchers often assume the presence of an indirect effect when including a covariate variable in the LCA model. In the P5 population model, γ_1 was fixed to zero and there was single direct effect. To explore the impact of misspecifying γ_1 on parameter estimates under different prior conditions, we analyzed the data generated from the P5 population model with three levels of priors on the direct effects (informative-correct, informative-wrong, and diffuse) and two levels of priors on the misspecified γ_1 (informative misspecificiation, diffuse misspecification).[6] Table 14 displays the parameter bias and percentage of replications with a significant coefficient under two sample size conditions ($N = 500$ vs. $N = 1{,}000$) two class size conditions (equal vs. unequal), and six combinations of prior specifications.

Table 14 shows the results for the P5 population model, which has a single direct effect and no indirect effect. The direct effect β_4 was unbiased in all conditions, regardless of sample size, class size, and prior specification. When the truly zero γ_1 was misspecified with informative priors, γ_1 was typically overestimated in the $N = 500$ conditions. Despite γ_1 often being unbiased, there was alarming number of false positives. The γ_1 parameter tended to have an inflated Type I error rate (>5%). The most problematic conditions utilized an informative prior on the misspecified γ_1. When comparing sample size ($N = 500$ vs. $N = 1{,}000$), conditions with a smaller sample size produced more false positives. The combination of a smaller sample size and unequal class sizes produced the greatest number of false positives.

4.4 Discussion

The primary goal of this study was to explore the performance of Bayesian SEM when modeling direct effects in conditional LCA models. The use of small-variance priors to detect non-zero direct effects between covariates and latent class indicators is a novel application of Bayesian SEM. In the conditions we investigated, small-variance priors on the overall direct effects had the power to detect non-zero direct effects in all population models, regardless of sample size and class sizes. Notably, this includes conditions with a single direct effect, two direct effects (i.e., local independence assumption violation), a class-varying direct effect, and a misspecified γ_1. However, the small-variance priors tended to produce a high number of false positives for truly zero direct effects in conditions with a local independence assumption violation, especially when the sample size was large. In addition, conditions with local independence assumption violations and no indirect effect (i.e., population model P6) tended to produce a high number of false positives for γ_1 when it was misspecified. These findings illustrate how problematic local independence assumption violations can be in LCA models.

[6] The P6 population model, which has no indirect effect and two direct effects (i.e., local independence assumption violation), had a very similar pattern of results; therefore, the results from the P6 population model are in Appendix A.

Table 14. The relative bias and % of significant coefficients for P5 parameters when using three levels of priors (diffuse, informative-correct, and informative-wrong) on β_4 and two levels of prior misspecifications (informative misspecification, diffuse misspecification) on γ_1.

Sample Size		$N = 500$				$N = 1{,}000$				$N = 500$				$N = 1{,}000$			
Class Prop.		50%/50%		82%/18%		50%/50%		82%/18%		50%/50%		82%/18%		50%/50%		82%/18%	
Parameter	Pop. Value	Bias	% Sig	Bias	% Sig	Bias	% Sig	Bias	% Sig	Bias	% Sig	Bias	% Sig	Bias	% Sig	Bias	% Sig
		β_4-~N(1,0.04) & γ_1~N(1,0.04)								β_4-~N(1,0.04) & γ_1~N(0,5)							
β_4	1	.007	100	.002	100	.009	100	.005	100	.012	100	.007	100	.013	100	.009	100
γ_1	0	.098	18	.153	25	.055	13	.089	16	.002	5	.008	7	.002	6	.004	5
		β_4-~N(0.5,0.04) & γ_1~N(1,0.04)								β_4-~N(0.5,0.04) & γ_1~N(0,5)							
β_4	1	-.096	100	-.098	100	-.049	100	-.052	100	-.091	100	-.092	100	.046	100	.048	100
γ_1	0	.103	19	.159	28	.058	13	.093	17	.009	6	.016	7	.006	6	.009	5
		β_4-~N(0,5) & γ_1~N(1,0.04)								β_4-~N(0,5) & γ_1~N(0,5)							
β_4	1	.011	100	.004	100	.011	100	.006	100	.018	100	.012	100	.015	100	.010	100
γ_1	0	.098	18	.097	25	.055	13	.089	15	.002	6	.008	7	.002	6	.004	6

In addition to exploring the performance of small-variance priors on the overall direct effects, we also examined the power of small-variance priors on the class-specific direct effects. The aim of a small-variance prior on the class-specific direct effect is to explore the possibility of a class-varying direct effect. Class-specific direct effects are much more difficult to estimate. The primary finding was that the overall sample size and the class size had to be large to detect a class-varying direct effect with small-variance priors. In situations where the class-varying direct effect is not of substantive interest to applied researchers and the sample size is limited, only estimating the overall direct effect may be a better strategy.

The small-variance prior simulations were not without limitations. First, the small-variance prior conditions set the variance hyperparameter to 0.0025, across conditions. Simulation results may be different when a wider (or narrower) prior is specified on the direct effects. A second limitation of the study is how we generated the data for each population model. To examine the performance of small-variance priors on direct effects, we generated data from population models with a relatively strong direct effect between the covariate and latent class variable. The strength of the direct effect would impact the power available to detect the non-zero direct effects. In addition, the covariate was normally distributed with no missing data, which could impact results. Future simulation studies should consider a wider variety of small-variance prior specifications and population models.

Another avenue for future methodological research is model fit. Methodologists should investigate the power of PPP to detect covariate misspecifications in conditional LCA models with small-variance priors. Past research on CFA models suggests PPP lacks power to detect model misspecifications unless the small-variance priors were very restrictive, and the sample size and misspecification are large (Jorgensen, Garnier-Villareal, Pornprasermanit, & Lee, 2019). Methodologists should also compare the performance of Bayesian SEM with other methods available for detecting direct effects such as the LCA MIMIC modeling procedure proposed by Masyn (2017) and the residual and fit statistics discussed in Janssen et al., (2019). Each of these procedures for detecting direct effects have their own limitations and should be explored via simulation research.

In addition to examining small-variance priors, this study also explored how robust the conditional LCA model results are to different combinations of prior specifications (informative-correct, informative-wrong, and diffuse) on β_4 and γ_1. Specifically, we examined the bias in the β_4 and γ_1 parameters. Regardless of the prior specification, the parameter estimates for β_4 and γ_1 were robust. The γ_1 regression coefficient was most impacted by the prior specification on γ_1. Despite these findings, applied researchers should use a prior sensitivity to explore different combinations of priors on the direct effects and γ_1 when modeling a conditional LCA model. For an example of how to implement a prior sensitivity analysis, see Depaoli et al., (2020).

In some modeling situations, the direct effect is not equal in both classes. Based on the small-variance prior results from this study, we know that detecting a class-varying direct effect that is unknown *a priori* is difficult. In addition to a larger sample size requirement, the number of cases assigned to each latent class and the strength of the direct effect in each class impacts our power to detect the direct effect. Considering how challenging class-varying direct effects can be, it was important to understand the impact of misspecifying the class-varying direct effect as an overall direct effect instead. As expected, this misspecification biased the class-specific direct effect parameter estimates. However, the misspecification of the direct effect did not impact the γ_1 parameter estimates. Therefore, it may not be necessary for applied researchers with limited sample sizes or class sizes to consider the class-varying component of the direct effect, if the primary focus of the research is γ_1 . The accuracy of the γ_1 parameter estimates was most influenced by the γ_1 prior specification.

The effect of x on c is often the primary interest of applied researchers using conditional LCA models. The effect of x on c allows researchers to answer questions about why a case was assigned to a particular latent class. Often researchers explore the relationship between demographic variables and the latent class assignment. In some situations, applied researchers may inadvertently assume a covariate is related to the latent class variable (c) when the covariate is only related to a latent class indicator (u_m). This would be a misspecification of γ_1. When using an informative prior on γ_1, the Type I error rate for γ_1 is inflated. Smaller sample sizes and unequal class sizes tended to increase the probability of a false positive, but the Type I error remained inflated in all conditions. A more diffuse prior distribution held the Type I error rate to 0.05-0.07, which is still somewhat inflated. A prior sensitivity analysis on γ_1 would be helpful in this situation because it would demonstrate how influential the prior specification is on γ_1.

Chapter 5: Discussion

The intent of this dissertation was to address gaps in the methodological literature on handling common covariate modeling issues (e.g., incomplete covariates, covariate misspecifications) in LCA models. The findings from this dissertation were used to make recommendations to applied researchers. Study 1 examined the performance of different methods available for addressing incomplete covariate data when using the ML three-step approach. Results from Study 1 allowed us to make recommendations on which method to use, depending on the modeling conditions (e.g., sample size, covariate distribution, strength of covariate effect). Study 2 explored the utility of using small-variance priors to identify non-zero direct effect between the covariate and latent class indicator. Results from Study 2 suggest small-variance priors can be a useful tool for detecting covariate misspecifications, depending on the number of direct effects, sample size, and class sizes. Overall, findings from Study 1 and Study 2 highlight how Bayesian estimation can be especially helpful for handling common modeling issues in conditional LCA models.

Study 1 results highlighted the potential benefits of using Bayesian estimation in the third step of ML three-step approach. Specifically, Bayesian estimation with informative and weakly-informative normal priors correctly centered on the regression coefficient population values resulted in unbiased parameter estimates, regardless of the distribution of the covariate, missing data patterns, and the strength of the covariate effect. The advantages of using Bayesian estimation were most easily seen when the distribution of the covariate was categorical. Alternative strategies (i.e., FIML and MI) had greater variability in the estimator than Bayesian estimation with informative priors. However, the potential advantages of Bayesian estimation were diminished or lost entirely when using diffuse priors on the regression coefficients. If applied researchers lack the knowledge to specify tighter priors on the regression coefficients, then FIML or MI work as well as Bayesian estimation. One pitfall of using Bayesian estimation was the possibility of specifying inaccurate priors on the regression coefficients. In conditions with inaccurate priors (informative and weakly informative), the regression coefficients were biased, which illustrates the importance of using accurate priors in the third estimation step. An important element of using Bayesian estimation in the third step is the elicitation of accurate priors.

One limitation of Study 1 is that it only explored methods for handling incomplete covariate data in conditional LCA models. There is reason to believe the findings in this dissertation may not be applicable to all finite mixture models. Specifically, the MI method of handling incomplete covariate data worked well in most conditions. However, past methodological research suggests MI can bias parameter estimates in other types of fine mixture models (e.g., factor mixture model) because of the class-varying effect of the covariate (Enders & Gottschall, 2011). One explanation for this outcome is that conditional LCA models cannot have a class-varying covariate effect because this would result in a non-recursive model, see Asparouhov (2016) for further details. In situations where a class-varying covariate effect is possible (e.g., factor mixture model), applied researchers may be better off using alternative strategies for addressing missing data such as FIML or Bayesian estimation.

Study 2 estimated the conditional LCA model using a one-step approach with Bayesian estimation. The primary aim of Study 2 was to examine the performance of small-variance normal priors in detecting non-zero direct effects between the covariate and the latent class indicator variables. A variety of modeling conditions were explored including situations with a single direct effect, two direct effects (i.e., local independence assumption violation), a class-varying direct effect, and a misspecified effect of x on c. In all conditions investigated, small-variance priors on the overall direct effects had the power to detect non-zero direct effects, regardless of sample size and class sizes. However, there were an alarming number of false positives for the truly zero direct effect when there was a local independence assumption violation. Small-variance priors on the class-specific direct effects only had the power to detect a class-varying direct effect when the sample size and class sizes were large. Future research should consider different variances for the small-variance priors to see if the number of false positives can be reduced. In Study 2, the variance of the small-variance priors was set to 0.0025 across conditions, but slightly tighter or looser priors may impact the number of false positives and quality of results. Based on the results from Study 2, small-variance priors could be used as an additional tool for detecting non-zero direct effects. Other tools currently available include the LCA MIMIC modeling procedure proposed by Masyn (2017) and the residual and fit statistics discussed in Janssen et al., (2019). After applied researchers have identified non-zero direct effects, it is possible to incorporate these direct effects into a more parsimonious model using either a one-step or three-step approach to estimation. For recommendations on how to incorporate direct effects into the LCA model when using a three-step approach, see Vermunt and Magidson (2021).

One factor that could have a significant impact on the results in both studies is the quality of the measurement model. A high-quality measurement model will have distinct, meaningful latent classes that can be easily identified. In applied settings, poor quality measurement models are a common issue. Previous methodological research suggests the separation of the latent classes has a significant impact on model results, regardless of whether the one- or three-step approach to estimation was utilized (Asparohouv & Muthén, 2014; Janssen et al., 2019). For the purposes of this dissertation, the measurement models in both studies were set to have moderate class separation. Future research should explore the possible advantages of using Bayesian estimation to address common covariate modeling issues when the latent classes are poorly separated.

Several avenues are available for future methodological research involving the inclusion of auxiliary variables in mixture models. Presently, there has been relatively limited research on the misspecification of auxiliary variables in mixture models (Collins & Lanza, 2010; Masyn, 2013; Nylund-Gibson et al., 2016; Petras & Masyn, 2010). Much of the past research has focused on covariate misspecifications, which has the potential to impact the latent class measurement model. However, a misspecified distal outcome may also bias measurement and structural parameters in the conditional LCA model. Therefore, methods of detecting distal outcome misspecification should be further explored. Another avenue of research that has not yet been fully explored is how missing data impacts our ability to detect model misspecifications. In this dissertation, Study 1 focused on handling missing data when using the ML three-step approach, whereas Study 2 focused on identifying covariate misspecifications using small-variance priors. One situation that was not considered in these studies is how missing data and the methods used to address it may impact our ability to detect covariate misspecifications. Missing data is a tricky issue that should be further considered in future methodological research involving mixture models.

Overall, this dissertation provides evidence that Bayesian estimation can help aid applied researchers in addressing common modeling issues in conditional LCA models. However, Bayesian estimation is not without limitations. In Study 1, using inaccurate priors on the regression coefficients resulted in biased parameter estimates, which highlights the importance of identifying accurate priors in applied settings. In Study 2, Bayesian estimation could not resolve the issues that occur when the local independence assumption has been violated, which further suggests conditional dependence is an especially tricky issue to solve. Even when a covariate has been identified (and modeled) that explains the relationship between a pair of latent class indicators, there are still problems in LCA models with conditional dependence. Applied researchers should seek to use Bayesian estimation as a tool for addressing modeling issues, but it is important to keep in mind that there are some modeling situations that Bayesian estimation will not be able to resolve.

References

Albert P. S. & Dodd, L. E. (2004). A cautionary note on the robustness of latent class models for estimating diagnostic error without a gold standard. Biometrics, 60, 427-435. doi: 10.1111/j.0006-341X.2004.00187.x

Asparouhov, T. (2016). Bidirectional association. https://www.statmodel.com/download/Rec.pdf

Asparouhov, T. & Muthén B. (2010). Bayesian analysis using Mplus: Technical implementation. http://statmodel.com/download/Bayes3.pdf

Asparouhov, T. & Muthén, B. O. (2011). Using Bayesian priors for more flexible latent class analysis. In *Joint Statistical Proceedings, Government Statistics Section* (pp. 4979–4993). Alexandria, VA: American Statistical Association.

Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*, 329–341. doi: 10.1080/10705511.2014.915181

Asparouhov, T., & Muthén, B. (2021). *Auxiliary variables in mixture modeling: Using the BCH method in Mplus to estimate a distal outcome model and an arbitrary second model.* Mplus Webnote No.21. Muthén & Muthén. https://www.statmodel.com/examples/webnotes/webnote21.pdf

Bakk, Z., & Kuha, J. (2018). T wo-step estimation of models between latent classes and external variables. *Psychometrika, 83*(4), 871–892. doi: 10.1007/s11336-017-9592-7

Bakk, Z., Oberski, D., & Vermunt, J. (2014). Relating latent class assignments to external variables: Standard errors for correct inference. *Political Analysis*, *22*, 520–540.

Bakk, Z., Tekle, F. B., & Vermunt, J. K. (2013). Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. *Sociological Methodology, 43*, 272–311. doi: 10.1177/0081175012470644

Bakk, Z. & Vermunt, J. K. (2016) Robustness of stepwise latent class modeling with continuous distal outcomes. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(1), 20-31. doi: 10.1080/10705511.2014.955104

Bandeen-Roche, K., Miglioretti, D. L., Zeger, S. L., Rathouz, P. J. (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association, 92*(440), 1375-1386. doi: 10.1080/01621459.1997.10473658

Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research, 42*, 757–786. doi: 10.1080/00273170701710338

Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis, 12*(1), 3-27. doi: 10.1093/pan/mph001

Brewsaugh, K., Masyn, K. E., & Salloum, A. (2018). Child welfare workers' sexism and beliefs about father involvement. *Children and Youth Services Review, 89*, 132-144. doi: 10.1016/j.childyouth.2018.04.029

Brooks, S. E. & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics, 7(4), 434-455. doi: 10.1080/10618600.1998.10474787

Celeux, G., Hurn, M., & Robert, C. P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, *95*(451), 957–970. doi: 10.1080/00031305.1992.10475878.

Chung, H., Loken, E., & Schafer, J. L. (2004). Difficulties in drawing inferences with finite-mixture models. *The American Statistician, 58*(2), 152-158. doi: 10.1198/0003130043286

Clogg, C. C., & Goodman, L. A. (1985). Simultaneous latent structure analysis in several groups. In N. B. Tuma (Ed.), *Sociological methodology 1985* (pp. 81–110). Jossey-Bass.

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and Health Sciences.* New York, NY: Wiley. doi : 10.1002/9780470567333

Collins, L. M., Schafer, J. L., & Kam, C. K. (2001). A comparison of inclusive and restrictive strategies in modern missing-data procedures. *Psychological Methods, 6*(4), 330-351. doi: 10.1037/1082-989X.6.4.330

Dayton, C. M. & Macready, G. B. (1988). Concomitant-variable latent class models. Journal of the American Statistical Association, 83, 173-178. doi: 10.2307/2288938

Depaoli, S. (2012). Measurement and structural model class separation in mixture CFA: ML/EM versus MCMC. *Structural Equation Modeling: A Multidisciplinary Journal, 19*(2), 178–203. doi: 10.1080/10705511.2012.659614

Depaoli, S. (2013). Mixture class recovery in GMM under varying degrees of class separation: Frequentist versus Bayesian estimation. *Psychological Methods, 18*(2), 186–219. doi: 10.1037/a0031609

Depaoli, S. (2014). The impact of inaccurate “informative” priors for growth parameters in Bayesian growth mixture modeling. *Structural Equation Modeling, 21*, 239-252. doi: 10.1080/10705511.2014.882686

Depaoli, S. (2021). *Bayesian Structural Equation Modeling.* The Guilford Press.

Depaoli, S., Rus, H., Clifton, J., van de Schoot, R., & Tiemensma, J. (2017). An introduction to Bayesian statistics in health psychology. *Health Psychology Review, 11,* 248-264.

Depaoli, S., Winter, S. D., & Visser, M. (2020). The Importance of Prior Sensitivity Analysis in Bayesian Statistics: Demonstrations Using an Interactive Shiny App. *Frontiers in Psychology*, *11*(November), 1–18. doi: 10.3389/fpsyg.2020.608045

Depaoli, S., Clifton, J. P., and Cobb, P. R. (2016). Just Another Gibbs Sampler (JAGS): A Flexible Software for MCMC Implementation. *Journal of Educational and Behavioral Statistics, 41,* 628-649.

Depaoli, S., Winter, S. D., & Visser, M. (2020). The Importance of Prior Sensitivity Analysis in Bayesian Statistics: Demonstrations Using an Interactive Shiny App. *Frontiers in Psychology*, *11*(November), 1–18. doi: 10.3389/fpsyg.2020.608045

Enders, C. K. (2010). Applied missing data analysis. New York: Guilford Press.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*(3), 430–457. doi: 10.1207/S15328007SEM0803_5

Enders, C. K., & Gottschall, A. C. (2011). The impact of missing data on the ethical quality of a research study. *In Handbook of Ethics in quantitative methodology*, Panter and Serba (Eds.), New York: Taylor and Francis Group, 357-381.

Enders, C. K., & Mansolf, M. (2018). Assessing the fit of structural equation models with

multiply imputed data. *Psychological Methods*, *23*(1), 76–93. doi: 10.1037/met0000102
Finch, H. (2005). Comparison of distance measures in cluster analysis with dichotomous data. *Journal of Data Science, 3*(1), 85-100. doi: 10.6339/JDS.2005.03(1).192
Fruhwirth-Schnatter, S. (2001) Markov chain Monte Carlo estimation of classical and Dynamic switching and mixture models. *Journal of the American Statistical Association, 96*(453), 194-209. doi: 10.1198/016214501750333063
Garrett, E.S. & Zeger, S. L. (2000). Latent Class Model Diagnosis. *Biometrics, 56*(4), 1055-1067. doi: 10.1111/j.0006-341x.2000.01055.x
Gelman, A., & Rubin, D. B. (1992a). Inference from iterative simulation using multiple sequences. *Statistical Science,* 7, 457–511. doi: 10.1214/ss/1177011136
Gelman, A., & Rubin, D. B. (1992b). "A single series from the Gibbs sampler provides a false sense of security," in *Bayesian Statistics 4*, eds J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Oxford: Oxford University Press), 625–631.
Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika, 61*, 215–231. doi: 10.1093/biomet/61.2.215
Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling, 10*(1), 80-100.
Haberman, S. J. (1979). *Analysis of qualitative data, Volume 2: New developments.* Academic Press.
Hagenaars, J. A. (1990). *Categorical Longitudinal Data: Log-Linear Panel, Trend, and Cohort Analysis*. Sage. doi: 10.2307/1165391
Hagenaars, J. A. (1993). *Loglinear models with latent variables*. Sage. doi: 10.4135/9781412984850
Hallquist, M. N. & Wiley, J. F. (2018). *MplusAutomation*: An R Package for Facilitating Large-Scale Latent Variable Analyses in M*plus*. *Structural Equations Modeling, 25*(4), 621-638. doi: 10.1080/10705511.2017.1402334
Henry, K. L. & Muthén, B. (2010). Multilevel latent class analysis: An application of adolescent smoking typologies with individual and contextual predictors. *Structural Equation Modeling, 17*(2), 193-215. doi: 10.1080/10705511003659342
Herman, K. C., Prewett, S. L., Eddy, C. L., Savala, A., Reinke, W. M. (2020). Profiles of middle school teacher stress and coping: Concurrent and prospective correlates. *Journal of School Psychology, 78*, 54-68. doi: 10.1016/j.jsp.2019.11.003
Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods, 11*(1), 36–53. doi: 10.1037/1082-989X.11.1.36
Hoijtink, H. & van de Schoot, R. (2018). Testing small variance priors using prior-posterior predictive p values. *Psychological Methods, 23*(5), 561-569. doi: 10.1037/met0000131
Huh, J., Riggs, N. R., Spruijt-Metz, D. Chou, C., Huang, Z., & Pentz, M. (2011). Identifying patterns of eating and physical activity in children: A latent class analysis of obesity risk. *Obesity (Silver Spring), 19*(3), 652-658. doi: 10.1038/oby.2010.228
Janssen, J. H. M., van Laar, S., de Rooij, M. J., Kuha, J., & Bakk, S. (2019). The Detection and Modeling of Direct Effects in Latent Class Analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 26*(2), 280-290. doi: 10.1080/10705511.2018.1541745
Jasra, A., Holmes, C. C., & Stephens, D. A. (2005). Markov Chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science, 20*(1), 50-67. doi: 10.1214/088342305000000016
Jorgensen T. D., Garnier-Villarreal M., Pornprasermanit S., Lee J. (2019). Small-variance priors can prevent detecting important misspecifications in Bayesian confirmatory factor analysis. In Wiberg M., Culpepper S., Janssen R., González J., Molenaar D.

(Eds.), *Quantitative Psychology: IMPS 2017. Springer Proceedings in Mathematics & Statistics* (pp. 255–263). Cham, Switzerland: Springer. doi:10.1007/978-3-030-01310-3_23
Kamakura, W. A., Wedel, M., & Agrawal, J. (1994). Concomitant variable latent class models for conjoint analysis. *International Journal of Research in Marketing, 11*(5), 451-464. doi: 10.1016/0167-8116(94)00004-2
Kim, S. (2014) Determining the number of latent classes in single- and multiphase growth mixture models. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(2), 263-279. doi: 10.1080/10705511.2014.882690
Kolb, R. R. & Dayton, C. M. (1996). Correcting for nonresponse in latent class analysis. *Multivariate Behavioral Research, 31*(1), 7-32. doi: 10.1207/s15327906mbr3101_2
Lanza, S. T. & Rhoades, B. L. (2013). Latent class analysis: An alternative perspective on subgroup analysis in prevention and treatment. *Prevention Science, 14*(2), 157–168. doi: 10.1007/s11121-011-0201-1
Lazarfeld, P. & Henry, N. (1968). *Latent Structure Analysis.* New York: Houghton Mifflin.
Lee, J., Jung, K., & Park, J. (2020). Detecting conditional dependence using flexible Bayesian latent class analysis. *Frontiers in Psychology, 11*, 1987. doi: 10.3389/fpsyg.2020.01987
Li, L., & Hser, Y. I. (2011). On inclusion of covariates for class enumeration of growth mixture models. *Multivariate Behavioral Research, 46*(2), 266–302. doi: 10.1080/00273171.2011.556549
Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*(404), 1198-1202.
Little, R. J. (1992). Regression with Missing X's: A Review. *Journal of the American Statistical Association, 87*(420), 1227-1237. doi: 10.1080/01621459.1992.10476282
Little, R. J. (2002). *Statistical analysis of missing data* (2^{nd} edition). Wiley.
Little, R. J. & Zhang, N. (2011). Subsample ignorable likelihood for regression analysis with missing data. *Journal of the Royal Statistical Society, Series C, Applied Statistics, 60*, 591-605. doi: 10.1111/j.1467-9876.2011.00763.x
Llyod J., Doll, H., Hawton, K., Dutton, W. H., Geddes, J. R., Goodwin, G. M., & Rogers, R. D. (2010). How psychological symptoms relate to different motivations for gambling: An online study of internet gamblers. *Biological Psychiatry, 68*(8), 733-740. doi: 10.1016/j.biopsych.2010.03.038
Lu, Z. L., Zhang, Z., & Lubke, G. (2011). Bayesian inference for growth mixture models with latent class dependent missing data. Multivariate Behavioral Research, 46(4), 567-597. doi: 10.1080/00273171.2011.589261
Lubke, G., & Muthén, B. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. *Structural Equation Modeling: A Multidisciplinary Journal, 14*, 26-47. doi: 10.1080/10705510709336735
Lubke, G. & Tueller, S. (2010). Latent class detection and class assignment: A comparison of the MAXEIG taxometric procedure and factor mixture modeling approaches. *Structural Equation Modeling, 17*(4), 605-628. doi: 10.1080/10705511.2010.510050
Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge, UK: University of Cambridge Press.
Magidson, J., and Vermunt, J.K. (2004). Latent class models. In: Kaplan., D (Ed.), *The Sage Handbook of Quantitative Methodology for the Social Sciences* (pp. 175-198). Sage.

Masyn, K. (2013). Latent class analysis and finite mixture modeling. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods in psychology* (Vol. 2, pp. 551–611). New York, NY: Oxford University Press.

Masyn, K. E. (2017) Measurement Invariance and Differential Item Functioning in Latent Class Analysis With Stepwise Multiple Indicator Multiple Cause Modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 24*(2), 180-197. doi: 10.1080/10705511.2016.1254049

McCutcheon, A. L. (1987). *Latent class analysis*. Sage. doi: 10.4135/9781412984713

Merkle, E. C. (2011). A comparison of imputation methods for Bayesian factor analysis models. *Journal of Educational and Behavioral Statistics, 36*(2), 257-276. doi: 10.3102/1076998610375833

Moss, H. B., Chen, C. M., & Yi, H. (2007). Subtypes of alcohol dependence in a nationally representative sample. *Drug Alcohol Dependence, 91*(2-3), 149-158. doi: 10.1016/j.drugalcdep.2007.05.016

Muthén, B. (1985). A method for studying the homogeneity of test items with respect to other relevant variables. *Journal of Educational Statistics, 10*(2), 121-132. doi: 10.3102/10769986010002121

Muthén, B. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika, 29*(1), 81–117. doi:10.2333/bhmk.29.81

Muthén, B. (2004). Latent Variable Analysis: Growth Mixture Modeling and Related Techniques for Longitudinal Data. In D. Kaplan (Ed.), *Handbook of Quantitative Methodology for the Social Sciences* (pp. 345-368). Sage. doi: 10.4135/9781412986311.n19

Muthén, B. O. & Asparouhov, T. (2012). Bayesian SEM: A more flexible representation of substantive theory. *Psychological Methods, 17*(3), 315-335. doi: 10.1037/a0026802

Muthén, B. O. & Asparouhov, T. (2013). BSEM Measurement Invariance Analysis. M*plus* Web Note No.17. Muthén & Muthén. https://www.statmodel.com/examples/webnotes/webnote17.pdf

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User's Guide. Eighth Edition*. Los Angeles, CA: Muthén & Muthén.

Nasiopoulou, P. Williams, P., Sheridan, S. & Hansen, K. Y. (2019). Exploring preschool teachers' professional profiles in Swedish preschool: A latent class analysis. *Early Child Development and Care, 189*(8), 1306-1324. doi: 10.1080/03004430.2017.1375482

Neelon, B., Swamy, G. K., Burgette, L. F. & Miranda, M. L. (2011). A Bayesian growth mixture model to examine maternal hypertension and birth outcomes. *Statistics in Medicine, 30*, 2721-2735. doi: 10.1002/sim.4291

Nylund, K. L., Asparouhov, T., & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14*(4), 535–569. doi: 10.1080/10705510701575396

Nylund K, Bellmore A, Nishina A, & Graham S. (2007). Subtypes, severity, and structural stability of peer victimization: what does latent class analysis say? *Child Development, 78*(6), 1705-1722. doi: 10.1111/j.1467-8624.2007.01097.x. PMID: 17988316

Nylund-Gibson, K., Grimm, R. P., & Masyn, K. E. (2019): Prediction from latent classes: A demonstration of different approaches to include distal outcomes in mixture models. *Structural Equation Modeling: A Multidisciplinary Journal, 26*(6), 967-985. doi: 10.1080/10705511.2019.1590146

Nylund-Gibson, K. & Masyn, K. E. (2016). Covariates and Mixture Modeling: Results from a Simulation Study Exploring the Impact of Misspecified Effects on Class Enumeration. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(6), 782-797. doi: 10.1080/10705511.2016.1221313

Oberski, D. L., van Kollenburg, G. H., & Vermunt, J. K. (2013). A Monte Carlo evaluation of three methods to detect local dependence in binary data latent class models. *Advances in Data Analysis and Classification, 7*, 267–279. doi: 10.1007/s11634-013-0146-2

Oberski, D. L., & Vermunt, J. K. (2014). *The expected parameter change (EPC) for local dependence assessment in binary data latent class models.* Submitted article. Retrieved from http://daob.nl/wp-content/ uploads/2014/07/lca-epc-revision2.pdf

Oberski, D. L. (2015). Beyond the number of classes: Separating substantive from non-substantive dependence in latent class analysis. *Advances in Data Analysis and Classification, 10*, 171-182. doi:10.1007/s11634-015-0211-0

Petras, H., & Masyn, K. (2010). General growth mixture analysis with ante- cedents and consequences of change. In A. R. Piquero & D. Weisburd (Eds.), *Handbooks of quantitative criminology* (pp. 69–100). Springer.

Quirk, M., Nylund-Gibson, K., & Furlong, M. (2013). Exploring patterns of Latino/a children's school readiness at kindergarten entry and their relations with grade 2 achievement. *Early Childhood Research Quarterly, 28*, 437– 449. doi: 10.1016/j.ecresq.2012.11.002

R Core Team. (2019). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*(3), 581-592. doi: 10.2307/2335739

Schafer, J. L. (1997). *Analysis of incomplete multivariate data.* CRC Press.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147.

Sedding, D. & Leitgöb, H. (2018). Exact and Bayesian approximate measurement invariance. In Davidov, E., Schmidt, P., Billiet, J., & Meuleman, B. (2nd edition), *Cross-cultural analysis methods and applications* (pp. 553-579). doi: 10.4324/9781315537078-20

Skrondal, A. & Rabe-Hesketh, S. (2014). Handling initial conditions and endogenous covariates in dynamic/transition models for binary data with unobserved heterogeneity. *Journal of the Royal Statistical Society, Series C, Applied Statistics, 63*(2), 211-237. doi: 10.1111/rssc.12023

Sterba, S. K. (2014). Handling Missing Covariates in Conditional mixture Models Under Missing at Random Assumptions. *Multivariate Behavioral Research, 49*, 614–632. doi: 10.1080/00273171.2014.950719

Sterba, S. K. (2017). Pattern mixture models for quantifying missing data uncertainty in longitudinal invariance testing. *Structural Equation Modeling, 24*(2), 283–300. doi: 10.1080/10705511.2016.1250635

Stromeyer, W. R., Miller, J. W., Sriramachandramurthy, R., & DeMartino, R. (2015). The prowess and pitfalls of Bayesian structural equation modeling: Important considerations for management research. *Journal of Management, 41*(2), 491-520. doi: 10.1177/014920631455196

Torrance-Rynard, V. L., & Walter, S. D. (1998). Effects of dependent errors in the assessment of diagnostic test performance. *Statistics in Medicine, 16*(19), 2157-2175. doi: 10.1002/(SICI)1097-0258(19971015)16:19<2157::AID-SIM653>3.0.CO;2-X

Tueller, S., Drotar, S., & Lubke, G. (2011). Addressing the problem of switched class labels in latent variable mixture model simulation studies. *Structural Equation Modeling*, *18*(1), 1 - 22. doi: 10.1080/10705511.2011.534695

Tueller, S., & Lubke, G. (2010). Evaluation of structural equation mixture models: Parameter estimates and correct class assignment. *Structural Equation Modeling, 17*(2)*,* 165–192. doi: 10.1080/10705511003659318

Vacek, P. M. (1985). The effect of conditional dependence on the evaluation of diagnostic tests. *Biometrics, 41*(4), 959-968. doi: 10.1111/j.1467-985X.2008.00544.x

van de Schoot, R., Winter, S., Zondervan-Zwijnenburg, M., Ryan, O., and Depaoli, S. (2017). A systematic review of Bayesian applications in psychology: The last 25 years. *Psychological Methods, 22,* 217-239. doi: 10.1037/met0000100.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P. C. (2020). Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC. *Bayesian Analysis, 16*(2), 667-718. doi: 10.1214/20-BA1221

Vera, J. D., & Enders, C. K. (2021). Is item imputation always better? An investigation of wave-missing data in growth models. *Structural Equation Modeling: A Multidisciplinary Journal, 28*(4), 506-517. doi: 10.1080/10705511.2020.1850289

Vermunt, J. K. (1997). *Log-linear models for event histories.* Sage.

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis, 18*(4), 450-469. doi: 10.1093/pan/mpq025

Vermunt, J. K., & Magidson, J. (2021). How to perform three-step latent class analysis in the presence of measurement non-invariance or differential item functioning. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(3), 356–364. doi: 10.1080/10705511.2020.1818084

Visser, M., & Depaoli, S. (2022). A Guide to Detecting and Modeling Local Dependence in Latent Class Analysis Models. *Structural Equation Modeling: A Multidisciplinary Journal.* doi: 10.1080/10705511.2022.2033622

Von Hippel, P. T. (2009). How to impute interactions, squares, and other transformed variables. *Sociological Methodology, 39*(1), 265–91. doi: 10.1111/j.1467-9531.2009.01215.x

Wang, W.-C., Shih, C.-L., & Yang, C.-C. (2009). The MIMIC method with scale purification for detecting differential item functioning. *Educational and Psychological Measurement, 69*(5), 713–731. doi: 10.1177/0013164409332228

White, I. R., P. Royston, and A. M. Wood (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30(4), 377–99. doi: 10.1002/sim.4067

Willse, J. T. & Goodman, J. T. (2008). Comparison of multiple-indicators, multiple-causes- and item response theory-based analyses of subgroup differences. Educational and Psychological Measurement, 68(4), 587-602, doi: 10.1177/0013164407312601

Winter, S. D., & Depaoli, S. (2020). An illustration of Bayesian approximate measurement invariance with longitudinal data and a small sample size. *International Journal of Behavioral Development, 44*(4), 371-382. doi: 10.1177/0165025419880610

Woods, C. M. (2009). Empirical selection of anchors for tests of differential item functioning. *Applied Psychological Measurement, 33*(1), 42–57. doi: 10.1177/0146621607314044

Woods, C. M. & Grimm, K. J. (2011). Testing for nonuniform differential item function with multiple indicator multiple cause models. *Applied Psychological Measurement, 35*(5), 339-361. doi: 10.1177/01466216114059

Wothke, W. (2000). Longitudinal and multigroup modeling with missing data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 219–240, 269–281). Lawrence Erlbaum Associates Publishers.

Xiao, Y., Liu, H., & Hau, K. T. (2019). A comparison of CFA, ESEM, and BSEM in test structure analysis. *Structural Equation Modeling, 26*(5), 665-677. doi: 10.1080/10705511.2018.1562928

Yamaguchi, K. (2000). Multinomial logit latent class regression models: An analysis of predictors of gender role attitudes among Japanese women. American Journal of Sociology, 105(6), 1702-1740. doi: 10.1086/210470

Appendix A: Additional Tables for Study 2

Appendix A contains the remaining tables and figures for Study 2. Figures 15-17 contain the relative bias in $\beta_4, \beta_5, \gamma_1$ from the P3 population model when using three levels of priors on the γ_1 and three levels of priors on the direct effect. Table 15 contains the relative bias and percentage of significant coefficients for P6 parameters when using three levels of priors on β_4 and β_5 and two levels of prior misspecifications on γ_1.

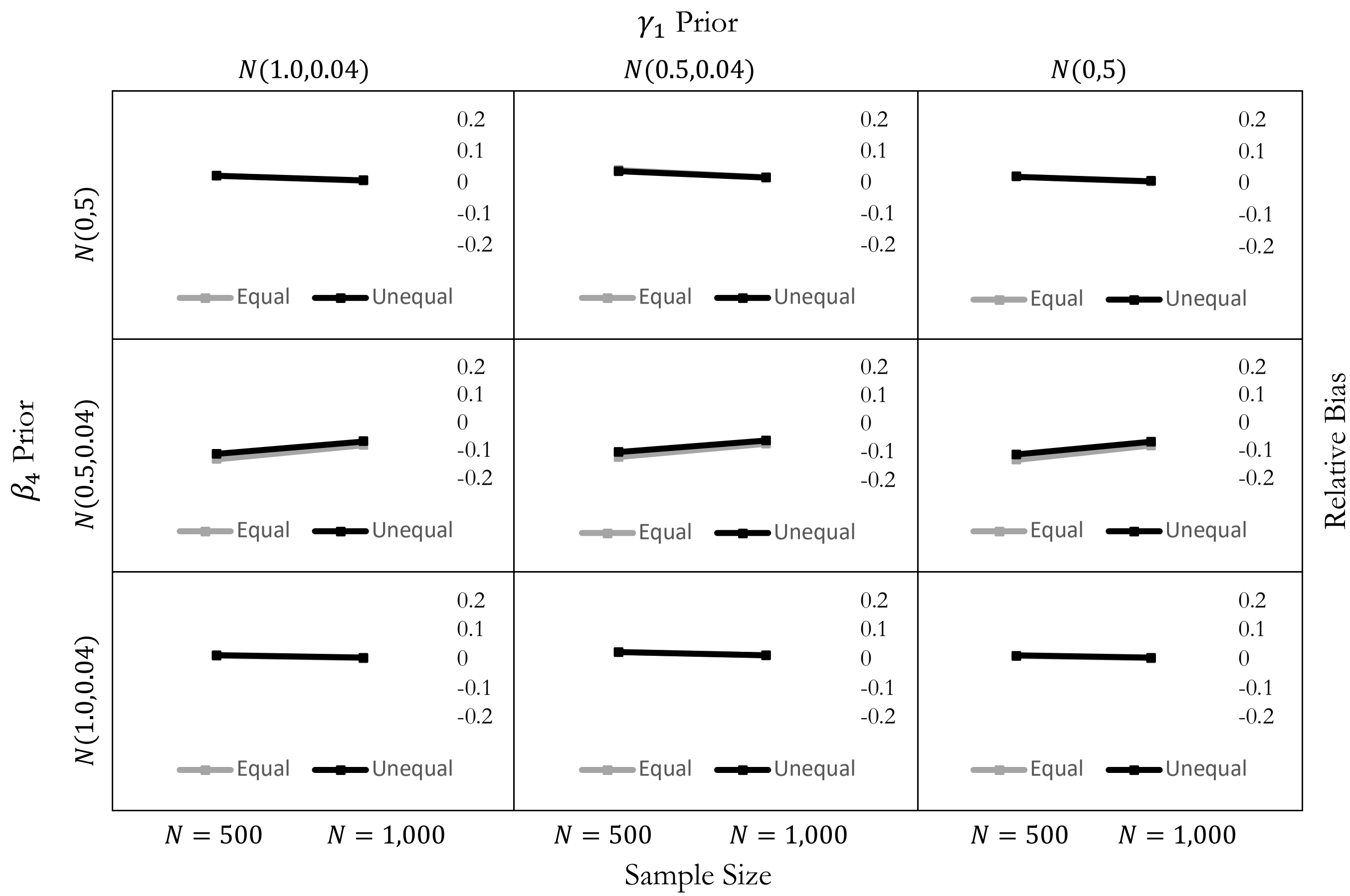


Figure 15. The relative bias in β_4 from P3 when using three levels of priors on γ_1 and three levels of priors on the direct effect.

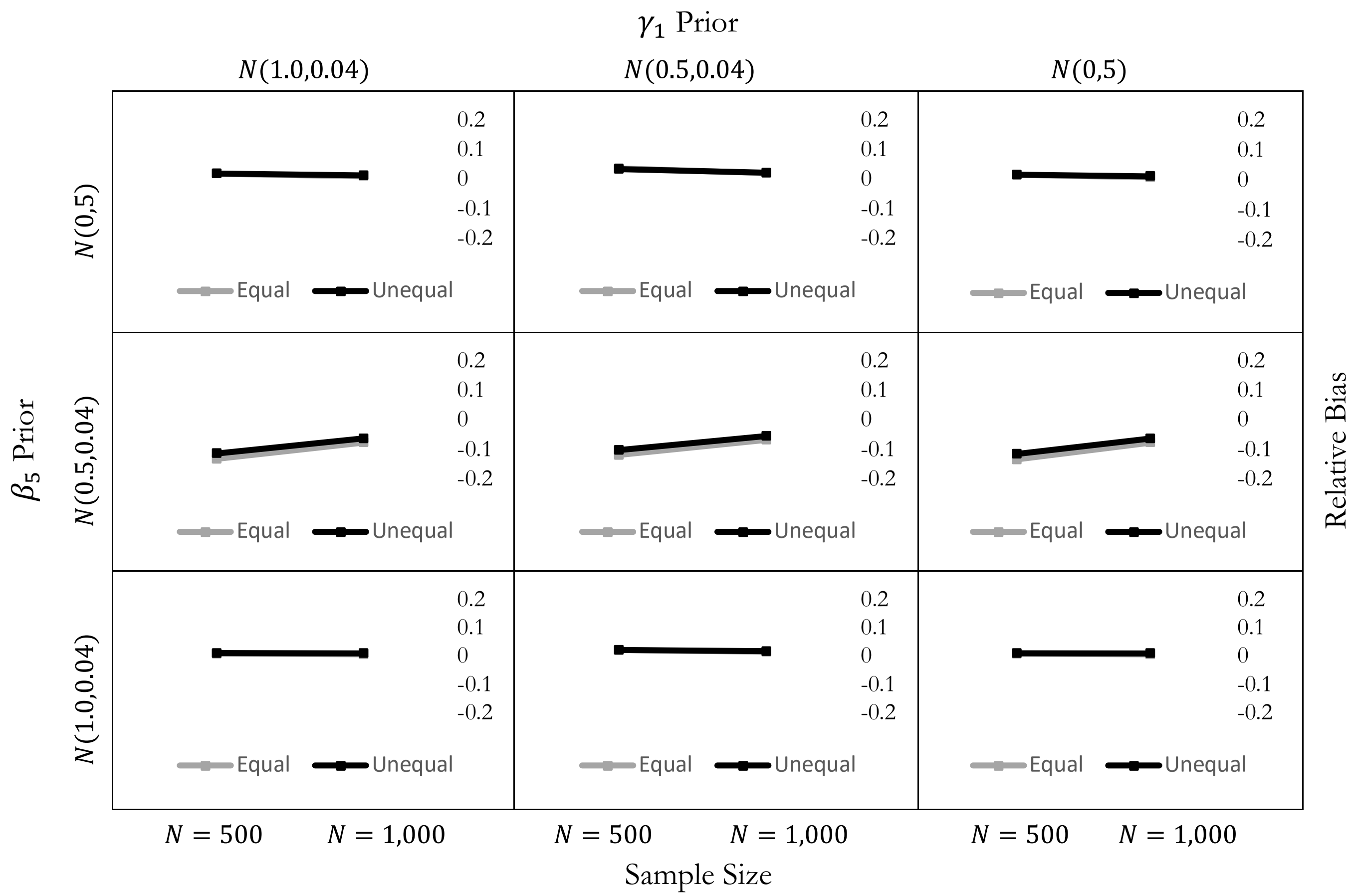


Figure 16. The relative bias in β_5 from P3 when using three levels of priors on γ_1 and three levels of priors on the direct effect.

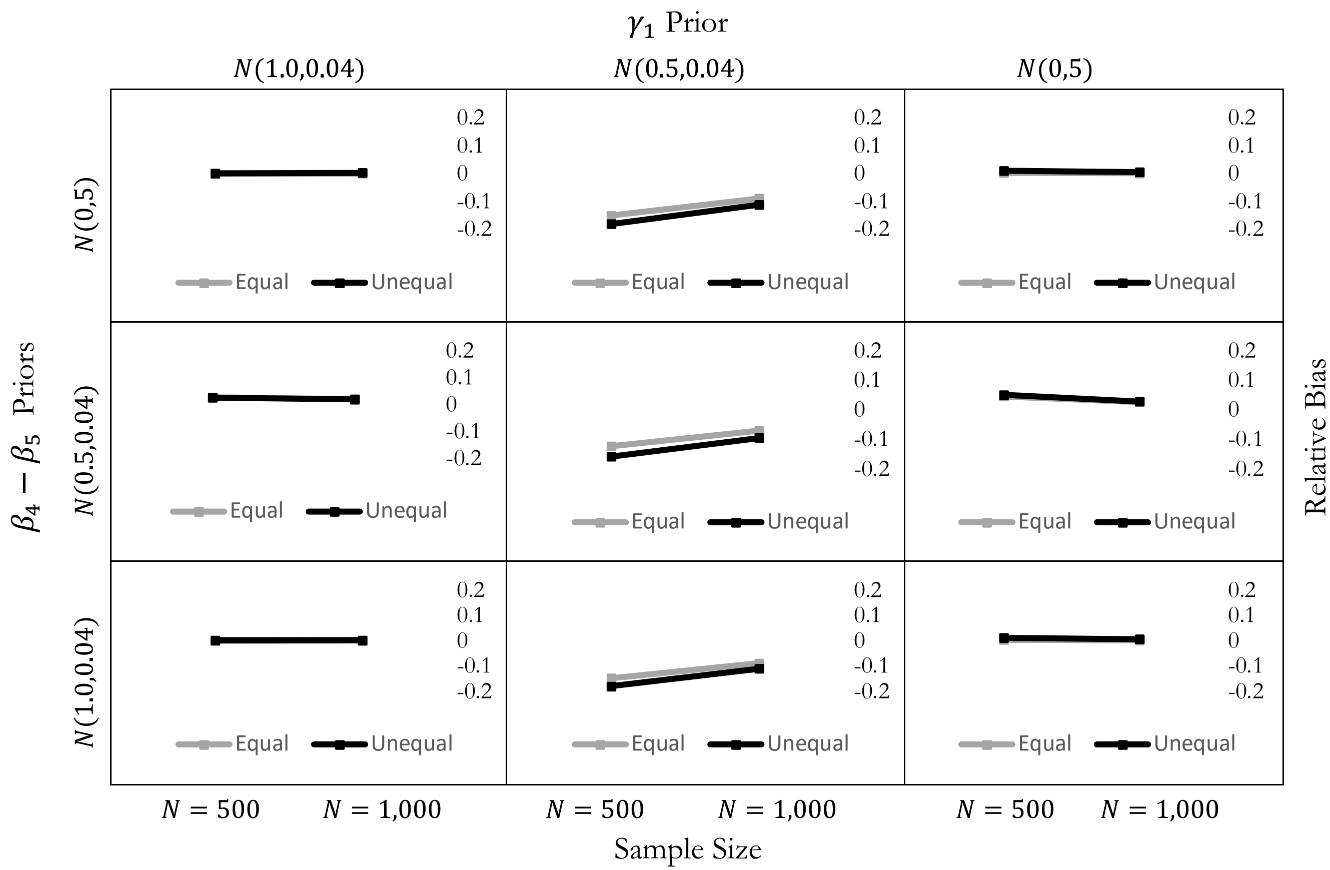


Figure 17. The relative bias in γ_1 from P3 when using three levels of priors on γ_1 and three levels of priors on the direct effect.

Table 15. The relative bias and % of significant coefficients for P6 parameters when using three levels of priors (diffuse, informative-correct, and informative-wrong) on β_4 and β_5 and two levels of prior misspecifications (informative misspecification, diffuse misspecification) on γ_1.

Sample Size		$N = 500$				$N = 1{,}000$				$N = 500$				$N = 1{,}000$			
Class Prop.		50%/50%		82%/18%		50%/50%		82%/18%		50%/50%		82%/18%		50%/50%		82%/18%	
Parameter	Pop. Value	Bias	% Sig	Bias	% Sig	Bias	% Sig	Bias	% Sig	Bias	% Sig	Bias	% Sig	Bias	% Sig	Bias	% Sig
		$\beta_4 - \beta_5$-~N(1,0.04) & γ_1~N(1,0.04)								$\beta_4 - \beta_5$-~N(1,0.04) & γ_1~N(0,5)							
β_4	1	.006	100	.004	100	.003	100	.001	100	.014	100	.011	100	.007	100	.006	100
β_5	1	.006	100	.005	100	.007	100	.006	100	.013	100	.012	100	.012	100	.011	100
γ_1	0	.097	16	.154	26	.054	12	.088	15	-.004	5	-.001	5	-.002	4	-.004	4
		$\beta_4 - \beta_5$-~N(0.5,0.04) & γ_1~N(1,0.04)								$\beta_4 - \beta_5$-~N(0.5,0.04) & γ_1~N(0,5)							
β_4	1	-.102	100	-.100	100	-.059	100	-.057	100	-.048	100	-.093	100	-.055	100	.052	100
β_5	1	-.102	100	-.099	100	-.054	100	-.053	100	-.051	100	-.092	100	-.055	100	.048	100
γ_1	0	.112	22	.170	30	.063	15	.098	17	.014	4	.021	5	.008	4	.008	4
		$\beta_4 - \beta_5$-~N(0,5) & γ_1~N(1,0.04)								$\beta_4 - \beta_5$-~N(0,5) & γ_1~N(0,5)							
β_4	1	.011	100	.007	100	.003	100	.001	100	.020	100	.016	100	.008	100	.007	100
β_5	1	.009	100	.009	100	.009	100	.007	100	.019	100	.018	100	.013	100	.013	100
γ_1	0	.097	17	.154	26	.054	12	.088	15	-.004	5	-.002	6	-.002	4	-.004	4