Published in final edited form as:

Struct Equ Modeling. 2017 ; 24(6): 911–925. doi:10.1080/10705511.2017.1340844.

Latent Class Analysis for Multiple Discrete Latent Variables: A Study on the Association Between Violent Behavior Drug-Using Behaviors

Saebom Jeon^a, Jungwun Lee^b, James C. Anthony^c, and Hwan Chung^{b,1}

^aDepartment of Marketing Information Consulting, Mokwon University

^bDepartment of Statistics, Korea University

^cDepartment of Epidemiology and Biostatistics, Michigan State University

Abstract

This paper proposes a new type of latent class analysis, joint latent class analysis (JLCA), which provides a set of principles for the systematic identification of the subsets of joint patterns for multiple discrete latent variables. Inferences about the parameters are obtained by a hybrid method of EM and Newton-Raphson algorithms. We apply JLCA in an investigation of adolescent violent behavior and drug-using behaviors. The data are from 4,957 male high-school students who participated in the Youth Risk Behavior Surveillance System 2015. The JLCA approach identifies the different joint patterns of four latent variables: violent behavior, alcohol consumption, tobacco cigarette smoking, and other drug use. The JLCA uncovers four common violent behaviors and three representative behavioral patterns for each of three other latent variables. In addition, the JLCA supports three common joint classes, representing the most probable simultaneous patterns for being violent and being a drug user among adolescent males.

Keywords

drug-using behavior; joint patterns of multiple latent variables; latent class analysis; violent behavior

Violent behavior (both victimization and aggression) and drug-using behavior are major social issue among adolescents in the United States (US). Each signal is an enhanced risk of adverse health and social outcomes (Lowry et al., 1999; Herrenkohl et al., 2000). During adolescence, the frequency of both violent behavior and drug-using behavior start to increase, with maladaptive consequences that can include incarceration, as well as both fatal and non-fatal injuries, often linked with co-occurrence of violence and drug use (Lowry et al., 1999; Herrenkohl et al., 2000; Center for Behavioral Health Statistics and Quality, 2015; Harford, Chen, & Grant, 2016).

¹ Correspondence should be sent to Hwan Chung, Professor, Department of Statistics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul 02841, Korea. hwanch@korea.ac.kr. Phone: +82-2-3290-2246.

The co-occurrence of violent behavior and drug use is not limited to cocaine or other internationally regulated drugs (IRD), for which associations are well established (Epstein-Ngo et al., 2013; White, Loeber, Stouthamer-Loeber, & Farrington, 1999; Walton et al., 2009; Harford et al., 2016). Many studies of adolescents have found that *alcohol use* is strongly associated with violent behavior (Lowry et al., 1999; Swahn, Simon, Hammig, & Guerrero, 2004; Swahn & Donovan, 2006; Rothman, Reyes, Johnson, & LaValley, 2012; Harford et al., 2016). In a national study of adolescents, Swahn et al. (2004) found that those who had drinking problems (e.g., frequent or binge drinking) were more likely to be involved in physical fighting than their peers. *Smoking tobacco cigarettes* is also found in association with youth violence (Bourque, 2016; Duncan et al., 2016).

The mechanisms that give rise to patterns of co-occurrence of adolescent violent behavior and adolescent drug-using behavior remain uncertain. Clearly, intoxication secondary to drug use can provoke and predict bouts of violent behavior that is not premeditated (DuRant et al., 2000). Nevertheless, some observers regard the overall association, in general, to be a spurious one, with other background characteristics responsible for observed associations that for the most part have been quantified using correlation coefficients, odds ratios, regression coefficients, or have been evaluated using test statistics such as the chi-square (White et al., 1999).

As latent variables, these behavioral constructs are measured via fallible sets of self-report items (e.g., from a questionnaire), and their association can be estimated by partitioning a true value from noise or error in measurement. In order to investigate adolescent violence and drug use along these lines, a latent class analysis (LCA) can be applied that identifies the underlying subgroups of the population, treating the response values for questionnaire items as mixtures of signal and noise. In this fashion, the LCA can be said to explain the association among items by assuming the unobserved group (i.e., the latent class), and adolescents can be classified into subgroups according to the nature or degree of their violent behavior or drug-using behavior.

In this respect, we are motivated to develop a generalized latent class type model in order to discover the joint behavioral patterns of violent behavior and the simultaneous use of multiple drug compounds, including alcohol, tobacco, and cannabis as the most frequently occurring examples. The proposed model seeks identification of the joint patterns by distinguishing between the questionnaire item(s) used to measure these behaviors versus the behaviors themselves, and considering these behaviors as latent variables, and by representing the simultaneous associations among these latent variables. The model is referred to here as joint latent class analysis (JLCA). Once elicited, the joint patterns of these several latent variables be studied in relation to background characteristics such as sex, age, and race or ethnicity. The proposed JLCA can incorporate these characteristics as potentially influential covariates via logistic regression.

Recently, several variations of LCA have been proposed, including multilevel LCA, the latent Markov model (LMM), latent class-profile analysis (LCPA), and LCA with causal inference (Chung, Park, & Lanza, 2005; Chung, Anthony, & Schafer, 2011; Kaplan & Keller, 2011; Lanza, Coffman, & Xu, 2013; Roesch, Villodas, & Villodas, 2010; Van Horn

et al., 2008; Vermunt, 2003). These variations have been designed for specific purposes. For example, LCPA can identify latent clusters or paths of latent classes using sets of repeatedly measured manifest items. Lanza et al. (2013) incorporated causal inference in LCA by considering college enrollment as a causal influence that exerts its causal effect on membership in different classes of drug-using behavior, which is considered a latent variable.

JLCA, as proposed here, encompasses several latent variables and is conceptually similar to the traditional structural equation model (SEM) in that both offer a framework for estimation of the association between unobserved latent variables based on various observed variables. However, JLCA can be distinguished from traditional latent variable approaches such as factor analysis, SEM, and random effects regression models in that its latent variables are categorical. We believe that the resulting categorical latent variables (i.e., latent class variable) in JLCA are easy to interpret, making the model more intuitive when the research aim is to uncover hidden subgroups within a population that is categorized similar simultaneous behavioral patterns of violent behavior and drug-using behavior. The JLCA results may have special appeal for program planners and policy makers because the resulting segmented subgroups may lead to more detailed insights into the problems of co-occurring violence and drug use. The JLCA findings also may be useful when the aim is to specify targets and to provide direction for intervention policy.

In the remainder of this paper, we provide methodological details on JLCA, and we then use the JLCA to investigate associations linking violent behavior and use of multiple drug compounds. This aim is achieved by identifying a small number of representative simultaneous behavioral patterns of violence and drug use. Whereas this paper is focused upon the JLCA as a method, we do attempt to provide some speculative interpretation of the observed results based on this model.

MODEL

Joint Latent Class Analysis

Traditional LCA infers unobserved latent class variables using multiple polytomous measurement variables. We propose a new model for multiple latent classes, calling it by the name of “joint latent class analysis” (JLCA) because it seeks to explain the joint patterns of two or more latent class variables. It does so by identifying subgroups of each latent class variable. In other words, JLCA simultaneously examines individuals’ set of class memberships for all latent variables and establishes two or more subgroups that consist of individuals who have a similar combination of class memberships. Here, we refer the latent class membership as a *class* variable and we denote a subgroup of those who have a similar combination of latent class membership in relation to values of a *joint class* variable. JLCA postulates that *joint class* variables are a mixture of *class* variables. JLCA specifies the association among multiple *class* variables (which are not directly observable) in an efficient yet parsimonious way.

Let C_j denote the j th latent class membership variable (i.e., *class* variable) with K_j nominal categories for $j = 1, \dots, J$. In other words, there are J *class* variables and the

number of classes of the j th *class* variable is K_j for $j = 1, \dots, J$. Furthermore, let a *joint class* variable U have S nominal categories describing the most common patterns of the combinations of *class* variables. If individuals' class memberships were known, a joint distribution of the multiple *class* variables, $\mathbf{C} = (C_1, \dots, C_J)$, could be expressed as frequencies in a contingency table with $\prod_{j=1}^J K_j$ cells, representing all possible combinations of latent class memberships. Inferences about the common patterns of the multiple *class* variables (i.e., *joint-class* variable U) should be based on this contingency table. However, as *latent class* variables, their values cannot be directly observed. Therefore, sets of multiple manifest items must be used to identify the *class* variables. Let $\mathbf{y}_{ij} = (y_{i1j}, \dots, y_{iM_jj})^T$ be the i th individual's discrete responses to a set of M_j manifest items measuring the j th *class* variable C_j , where each response y_{im_jj} can take values from 1 to r_{m_j} for $m_j = 1, \dots, M_j$ and $j = 1, \dots, J$. The probability that the i th individual then belongs to *joint class* s , the *classes* $\mathbf{c} = (c_1, \dots, c_J)$, and the responses $\mathbf{y}_i = (\mathbf{y}_{i1}, \dots, \mathbf{y}_{iJ})$ would be

$$L_i^* = P(U = s, \mathbf{C} = \mathbf{c}, \mathbf{Y} = \mathbf{y}_i) \quad (1)$$

$$= P(U = s)P(\mathbf{C} = \mathbf{c} | U = s)P(\mathbf{Y} = \mathbf{y}_i | U = s, \mathbf{C} = \mathbf{c})$$

$$= P(U = s) \prod_{j=1}^J \left\{ P(C_j = c_j | U = s) \prod_{m_j=1}^{M_j} P(Y_{m_jj} = y_{im_jj} | C_j = c_j) \right\}$$

$$= \gamma_s \prod_{j=1}^J \left\{ \eta_{c_j|s}^{(j)} \prod_{m_j=1}^{M_j} \prod_{k=1}^{r_{m_j}} \rho_{m_jk|c_j}^{I(y_{im_jj}=k)} \right\},$$

where $I(y_{im_jj} = k)$ is an indicator function which is equal to 1 when $y_{im_jj} = k$ and 0

otherwise. The complete-data likelihood given in (1) consists of three sets of parameters:

- (a) $\rho_{m_jk|c_j} = P(Y_{m_jj} = k | C_j = c_j)$ denotes the probability of the response k to the m_j th item measuring the j th *class* variable when the membership for the j th *class* variable is c_j (i.e., $C_j = c_j$);
- (b) $\eta_{c_j|s}^{(j)} = P(C_j = c_j | U = s)$ is the conditional probability of belonging to class c_j for the j th *class* variable given a *joint class* $U = s$; and
- (c) $\gamma_s = P(U = s)$ represents the probability of belonging to a *joint class* $U = s$.

The ρ -parameter, which is referred to as the *primary measurement parameter*, depicts to what extent individuals in each class to the m_j th item of the *class* variable C_j for $m_j = 1, \dots, M_j$ and $j = 1, \dots, J$. The η -parameter, which is called the *secondary measurement parameter*,

denotes the relationship between the j th *class* variable C_j and the *joint class* variable U . An identified *joint class* can be explained through a set of estimated η s as a particular individual's joint collection of class membership. The proposed JLCA is illustrated in Figure 1.

The marginal probability of item responses without the latent variables, which is referred to as the observed-data likelihood, is obtained by summing up all of the latent variables in (1):

$$\begin{aligned}
 L_i &= P(\mathbf{Y} = \mathbf{y}_i) \\
 &= \sum_{s=1}^S \sum_{C_1=1}^{K_1} \cdots \sum_{C_J=1}^{K_J} L_i^* \\
 &= \sum_{s=1}^S \sum_{C_1=1}^{K_1} \cdots \sum_{C_J=1}^{K_J} \gamma_s \prod_{j=1}^J \left\{ \eta_{c_j,ls}^{(j)} \prod_{m_j=1}^{M_j} \prod_{r=1}^{r_{m_j}} \rho_{m_j,kj|c_j}^{I(y_{im_j,j}=k)} \right\}.
 \end{aligned} \tag{2}$$

The formulation of JLCA assumes the following properties: (a) the *joint-class* variable is related to the manifest items only through J *class* variables; (b) the *class* variables are conditionally independent given a *joint-class* variable; and (c) manifest items for the j th *class* variable are conditionally independent given a class membership of C_j . Assumptions (b) and (c) correspond with the condition of 'local independence', which allows for inference of unobserved *class* and *joint class* variables (Lazarsfeld & Henry, 1968).

Joint Latent Class Analysis with Covariates

The prevalence of *joint class* membership can depend on the demographic characteristics or other suspected determinants, and these suspected determinants can be incorporated into the JLCA as illustrated in Figure 2. Consider a vector of subject-specific covariates $\mathbf{X}_i = (x_{i1}, \dots, x_{ip})^T$ which might influence the probability of belonging to *joint class* $U = s$, so that $P(U = s)$ is replaced by $P(U = s | \mathbf{x}_i)$. In other words, the estimated probability of belonging to a certain multinomial logistic regression.

The observed-data likelihood then is specified as

$$\begin{aligned}
 L_i &= P(\mathbf{Y} = \mathbf{y}_i | \mathbf{x}_i) \\
 &= \sum_{s=1}^S \gamma_s(\mathbf{x}_i) \prod_{j=1}^J \left\{ \sum_{c_j=1}^{K_j} \eta_{c_j,ls}^{(j)} \prod_{m_j=1}^{M_j} \prod_{k=1}^{r_{m_j}} \rho_{m_j,kj|c_j}^{I(y_{im_j,j}=k)} \right\},
 \end{aligned} \tag{3}$$

where $\gamma_s(\mathbf{x}_i) = P(U = s | \mathbf{x}_i) = \exp(\mathbf{x}_i \boldsymbol{\beta}_s) / \sum_{l=1}^S \exp(\mathbf{x}_i \boldsymbol{\beta}_l)$ and $\boldsymbol{\beta}_S = \mathbf{0}$ (i.e., *joint class S* is the of logistic regression coefficients, $\boldsymbol{\beta}_s = (\beta_{1s}, \dots, \beta_{ps})^T$ for $s = 1, \dots, S - 1$, can be interpreted regression slope estimate with respect to the log-odds that an individual belongs to *joint class joint class s*. versus the baseline *joint class S*.

ESTIMATION AND MODEL DIAGNOSIS

Expectation-Maximization Algorithm

The expectation-maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977) is a typical estimation method for LCA models, providing robust estimates located in parameter space regardless of the initial value. However, when covariates are included in the latent class structure, the expected complete-data likelihood does not have a closed-form, and requires an iterative procedure, such as the Newton-Raphson method, in the M-step (Van der Heijden, Dessens, & Bockenholt, 1996; Bandeen-Roche, Miglioretti, Zeger, & Rathouz, 1997). In this paper, we provide ML estimation methods including EM, Newton-Raphson, and hybrid EM algorithms for use in JLCA with covariates.

Parameter estimation in JLCA can be regarded as a special case of the missing-data problem because the values of the *joint class* and *class* variables are missing for individuals. The EM algorithm is one of the most popular iterative procedures for solving the missing-data problem when ML estimates cannot be obtained directly from the observed-data log-likelihood. EM consists of two steps, the E-step and M-step. The E-step computes the conditional probability that the i th individual belongs to the specific *joint class s* and the specific combination of *class* variables $\mathbf{c} = (c_1, \dots, c_J)$ given a set of item responses $\mathbf{y}_i$:

$$\theta_{i(s, c_1, \dots, c_J)} = P(U = s, C_1 = c_1, \dots, C_J = c_J | \mathbf{y}_i, \mathbf{x}_i) \quad (4)$$

$$= \frac{\gamma_s(\mathbf{x}_i) \prod_{j=1}^J \left\{ \eta_{c_j|s}^{(j)} \prod_{m_j=1}^{M_j} \prod_{k=1}^{r_{m_j}} \rho_{m_j k j | c_j}^{I(y_{im_j j} = k)} \right\}}{\sum_{s=1}^S \gamma_s(\mathbf{x}_i) \prod_{j=1}^J \left\{ \sum_{c_j=1}^{K_j} \eta_{c_j|s}^{(j)} \prod_{m_j=1}^{M_j} \prod_{k=1}^{r_{m_j}} \rho_{m_j k j | c_j}^{I(y_{im_j j} = k)} \right\}}.$$

In the E-step we implement the EM algorithm based on the complete-data log-likelihood to avoid the computational complexity when maximizing the observed-data log-likelihood using posterior probabilities given in (4). The expectation of the complete-data log-likelihood can be obtained by

$$E\left(\sum_{i=1}^n \log L_i^*\right) = \sum_{i=1}^n \sum_{s=1}^S \theta_{i(s)} \log \gamma_s(\mathbf{x}_i) + \sum_{i=1}^n \sum_{s=1}^S \sum_{j=1}^J \sum_{c_j=1}^{K_j} \theta_{i(s, c_j)}^{(j)} \log \eta_{c_j|s}^{(j)} \quad (5)$$

$$+ \sum_{i=1}^n \sum_{j=1}^J \sum_{c_j=1}^{K_j} \left\{ \theta_{i(c_j)}^{(j)} \sum_{m_j=1}^{M_j} \sum_{k=1}^{r_{m_j}} I(y_{im_j} = k) \log \rho_{m_j k | c_j} \right\},$$

where $\theta_{i(s, c_j)}^{(j)} = \sum_{c_1} \cdots \sum_{c_{j-1}} \sum_{c_{j+1}} \cdots \sum_{c_J} \theta_{i(s, c_1, \dots, c_J)}$, $\theta_{i(s)} = \sum_{c_j} \theta_{i(s, c_j)}^{(j)}$, and $\theta_{i(c_j)}^{(j)} = \sum_s \theta_{i(s, c_j)}^{(j)}$.

The M-step maximizes the expected complete-data log-likelihood given in (5) with respect to the model parameters. Thus, ML estimators for JLCA can be obtained using a Lagrange multiplier, which is a commonly used strategy for finding the local maxima and minima of a function subject to equality constraints. As a result, the parameters are updated with these resulting parameter estimators as follows:

$$\hat{\eta}_{c_j|s}^{(j)} = \frac{\sum_{i=1}^n \theta_{i(s, c_j)}^{(j)}}{\sum_{i=1}^n \theta_{i(s)}}, \quad \hat{\rho}_{m_j k | c_j} = \frac{\sum_{i=1}^n \theta_{i(c_j)}^{(j)} I(y_{im_j} = k)}{\sum_{i=1}^n \theta_{i(c_j)}^{(j)}}. \quad (6)$$

Since covariates are considered in the model, the prevalence parameters $\gamma_s(\mathbf{x}_i)$ for $s = 1, \dots, S$ in (5) are replaced with into the multinomial logistic regression coefficients, $\boldsymbol{\beta}_s = (\boldsymbol{\beta}_{1s}, \dots, \boldsymbol{\beta}_{ps})^T$ for $s = 1, \dots, S-1$. Conventional multinomial logistic regression models require polytomous response variables, but in JLCA the response is a multinomial latent variable identified by posterior probability given the manifest item responses. The estimated $\boldsymbol{\beta}_s$ express the associations between the *Joint class* variable and the covariates. Because the closed form of the ML estimates for $\boldsymbol{\beta}_s$ cannot be expressed, the Newton-Raphson algorithm can be used to estimate the regression coefficients in the M-step while the ρ and η -parameters are updated simultaneously in (6).

By performing the E-step and M-step iteratively, the parameter estimates converge to the ML solutions. The stopping rule is based on a numerical convergence threshold within the algorithm that depends upon either an increase in log-likelihood during the procedure, or the maximum absolute deviation (MAD) between the updated and the previous parameter estimates. As a result, the algorithm stops when the MAD becomes smaller than a pre-specified limit, or when the increase in the log-likelihood during the iteration becomes smaller than a pre-specified threshold.

Using the missingness at random (MAR) assumption (Rubin, 1976), we can easily extend the EM algorithm to a model with missing observations on measurement items. In the E-step, the conditional probability of $\theta_{i(s, c_1, \dots, c_J)}$ is calculated only with the observed

responses of y_j . We denote $\theta_{i(s, c_1, \dots, c_J)}^{obs}$ to distinguish it from the previous $\theta_{i(s, c_1, \dots, c_J)}$ given in (4). The primary measurement parameter is then obtained from the provisional estimates $\rho_{m_j k j | c_j}^*$ by

$$\hat{\rho}_{m_j k j | c_j} = \frac{\sum_{i \in obs_{m_j}^{(j)}} \theta_{i(c_j)}^{obs_j} I(y_{im_j j} = k) + \sum_{i \in mis_{m_j}^{(j)}} \theta_{i(c_j)}^{obs_j} \rho_{m_j k j | c_j}^*}{\sum_{i=1}^n \theta_{i(c_j)}^{obs_j}},$$

where $\theta_{i(c_j)}^{obs_j} = \sum_s \sum_{c_1} \dots \sum_{c_{j-1}} \sum_{c_{j+1}} \dots \sum_{c_J} \theta_{i(s, c_1, \dots, c_J)}^{obs}$, and $obs_{m_j}^{(j)}$ and $mis_{m_j}^{(j)}$ are the sets of individuals who respond and fail to respond to the m_j th item measuring the j th class variable, respectively.

Newton-Raphson Method

The Newton-Raphson (NR) method is also a widely-used iterative method used to calculate maximum likelihood estimates. It tends to require a much smaller number of iterations than does EM. The ML estimates can be obtained by solving the first derivative of the log-likelihood equation by updating the parameters as

$$\hat{\Theta}^{(t+1)} = \hat{\Theta}^{(t)} - H(\hat{\Theta}^{(t)})^{-1} S(\hat{\Theta}^{(t)}), \quad (7)$$

where $\hat{\Theta}^{(t)}$ is vector of all free parameter estimates at the t -th iteration, $H(\hat{\Theta}^{(t)})$ is the Hessian matrix, and $S(\hat{\Theta}^{(t)})$ is the score function of the log-likelihood evaluated at the t -th iteration. The details of the Hessian matrix and the score function are provided in the Appendix. Note that the estimated covariance matrix for parameter estimates is the inverse of the negative Hessian.

Hybrid Algorithm

Although the EM algorithm can yield stable estimates regardless of the boundary solutions, it requires a large number of iterations for convergence to occur. In contrast, the NR method requires a relatively small number of iterations and provides standard errors as a byproduct. However, if boundary solutions appear or the shape of the log-likelihood is not concave, NR does not work because the Hessian becomes (near) singular. We thus suggest a new hybrid method utilizing both NR and EM that harnesses their respective strengths and to compensate for their weaknesses. The hybrid method consists of three steps: EM, NR, and the Hessian-step.

The EM step.—The EM algorithm is repeated until the increase in the log-likelihood is small enough to be considered near the solution. This procedure stops if the increase in the log-likelihood is less than a pre-specified level, or the maximum absolute deviation (MAD)

of the successive estimates during the procedure is less than a pre-specified threshold. The stopping rule in the EM step is relatively weaker than that of a conventional EM algorithm, because the EM method tends to experience small increases in the log-likelihood near the solution. Any boundary solution appearing in the algorithm is constrained at zero or 1 under an appropriate threshold, and then the NR step occurs.

The NR step.—The NR algorithm is repeated until the increase in the log-likelihood or the MAD of the estimates becomes sufficiently small. In the NR step, the stopping criteria for the algorithm is stricter than that of the EM step, because NR requires a small number of iterations for convergence near the solution, so it can derive precise solutions faster than the EM algorithm can. If the Hessian matrix becomes singular due to the appearance of a new boundary solution, the process returns to the EM step, with this boundary solution fixed as constant, and thus not estimated.

The Hessian step.—Once the parameter estimates converge following the iterations in the above two steps, the asymptotic standard error of the estimates can be obtained with the inverted negative Hessian matrix evaluated at the converged estimates.

Model Diagnosis

As with other statistical models, it is important to assess model fit with JLCA using objective measures and a balanced judgment that takes into account the principle of parsimony as well as substantive knowledge. The first and most crucial step in JLCA is to choose appropriate structures for the *class* variables. JLCA consists of multiple *class* variables, and the number of latent subgroups for each *class* variable can differ. The Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are the standard criteria used to compare models with a different number of classes, because the asymptotic assumption for the likelihood ratio test statistic (G^2) based on a chi-square test generally does not hold. Note that JLCA models with a different number of classes are not nested. Models with smaller AIC (or BIC) are preferred but, in addition to these objective measures, interpretability should also be considered when determining the number of classes.

It is also important to examine the absolute model fit by comparing the predicted frequencies with the observed frequencies appearing in the data. Bootstrap samples of G^2 can be generated using the parameter estimates to determine the empirical distribution of G^2 as follows: (1) fit JLCA to the data set and obtain the observed G^2 based on the ML estimates; (2) generate a new data set with the ML estimates from the previous step, (3) fit JLCA to the generated data set from step (2) and compute G^2 . Bootstrap samples of the G^2 statistic are produced by repeating steps (2)-(3) many times, with the right tail probability of the observed G^2 from step (1) regarded as a bootstrap p -value.

Once the number of subgroups for each *class* variable is determined, the number of latent subgroups for *joint class* variable should be selected based on similar criteria such as a smaller AIC (or BIC) and a bootstrap p -value larger than 0.05. Trying large sets of different starting values helps one avoid local maxima, and the solution with the maximum log-likelihood should be selected. Finally, the model with covariates will be fitted with the selected number of classes and joint classes.

Simulation Study

Here, we performed simulation study in order to verify the validity of the proposed model and estimation methods in the previous section. When determining the number of joint classes, we need to fix the number of classes for each *class* variable because *joint class* variable is measured by representative joint patterns of each *class* variable. For this reason, we fit the traditional LCA model to data containing each set of items relevant to the specific *class* variable. Once the number of latent classes are determined, we find the most appropriate number of joint classes given the prefixed number of latent classes. To check the validity of this method, we generated one random dataset, and performed the model selection procedure to investigate whether the true latent structure of the JLCA was properly determined. A generated data has 500 observations and two sets of six binary manifest items measuring two *class* variables, respectively. The data are generated under the latent structure, where each of two *class* variables has three latent classes and these *class* variables form a three-class *joint class* variable. We considered two levels of the measurement relationship between items and classes: strong ($\rho = 0.1$ or 0.9) and mixed ($\rho = 0.1, 0.3, 0.7$, or 0.9). We began by fitting two LCA models to each set of manifest items, and chose the number of latent classes using AIC, BIC, and bootstrap p -value.

The results of model selection procedure for *class* variables are shown in Table 1. For JLCA with strong measurement, LCA models with three classes produce the smallest AIC (2921.52, 2928.03), BIC (3005.81, 3012.32) and an acceptable bootstrap p -value (0.45, 0.25) for both *class* variables. Therefore, we may conclude that the model selection procedure successfully discovered the number of latent classes. With the selected number of latent classes (i.e., three classes for both *class* variables), we fitted a series of JLCA models with differing numbers of joint classes. The goodness-of-fit measures for different JLCA models are also displayed in Table 1. The JLCA with three joint classes provides the smallest AIC (5732.07) and BIC (5942.80), indicating that it is the best model among the candidates, and it also has an acceptable bootstrap p -value (0.59). A similar result appeared with a mixed measurement JLCA model. LCA models for *Class 1* variable showed the best model fit with three latent classes, having the smallest AIC (3403.50), BIC (3487.79) and an acceptable bootstrap p -value (0.42). For *Class 2*, LCA model with three latent class also showed the smallest AIC (3464.94), BIC (3549.24) and an acceptable bootstrap p -value (0.68). With three classes for both *class* variables, the JLCA with three joint classes provides the smallest AIC (6779.82), BIC (6990.55) and an acceptable bootstrap p -value (0.75). Since the selected structure of the latent variables agreed with true value of the model, we may conclude that the model selection procedure successfully discovered the number of latent classes.

In this manner, we repeated independent simulation study by 50 times and counted how frequently the model decision is done properly. The repeated simulations were performed with diversified sample size ($n = 250$ and $n = 500$) and ρ -parameters (strong and mixed measurements). Tables 2 and 3 show the frequency of the selected model using different goodness-of-fit measures. For every independent trials, we selected the model that has the smallest AIC and BIC among three candidates (i.e., two-, three-, and four-class models). Note that the bootstrap p -value is the left tail probability of the observed G^2 on the empirical

distribution of G^2 under null hypothesis that the model is appropriate for the data. As the model becomes complicated, it becomes more similar with the saturated model and thus its p -value becomes larger. As a result, a model selected by the bootstrap p -value is the simplest one whose p -value is larger than 0.05. The number of success in Tables 2 and 3 was counted based on AIC, BIC, and bootstrap p -value: AIC and BIC might select different number of classes and thus we have at most two choices (i.e., one with the smallest AIC and the other with the smallest BIC). When AIC and BIC selected identical model in our simulation study, we simply checked whether the selected model had p -value larger than 0.05. If AIC and BIC selected different models, the simpler model was finally chosen under the principle of parsimony. This simpler model was then evaluated by the bootstrap p -value. When the sample size was small (i.e., $n = 250$), the model selection procedure is appropriate only for the model with strong measurement: The numbers of successes are 41 and 26 for strong mixed measurements, respectively. When the sample size was moderate (i.e., $n = 500$), however, the proposed procedure can be considered to be successful: the number of successes is 48 for strong measurement and 47 for mixed measurement out of 50 trials.

APPLICATION TO VIOLENT BEHAVIOR AND DRUG-USING BEHAVIORS

Data

The Youth Risk Behavior Surveillance System 2015 (YRBSS 2015) is a biennial survey of the health risk behaviors among school-attending adolescents in the US (Kann et al., 2016). The total number of survey participants in the YRBSS 2015 is 15,624 male and female students, but we only consider our application of the JLCA model is based on the 4,957 high-school male students aged 16 to 18 (grades 10 to 12) who have non-missing responses for YRBS questionnaire items on violence and drug use, defined to encompass the use of alcohol and tobacco cigarettes, as well as cannabis or other internationally regulated drugs, and who have answered survey questions on age at the time of assessment and race-ethnicity as classified for the US Bureau of Census.

In this study, we consider 18 self-report items on the topics of being a perpetrator or victim of violent behavior or being involved in drug-related behaviors. For *violent behavior*, five questions are used: (1) During the past 30 days, on how many days did you carry a weapon such as a gun, knife, or club? (2) During the past 30 days, on how many days did you not go to school because you felt you would be unsafe at school or on your way to or from school? (3) During the past 12 months, how many times has someone threatened or injured you with a weapon such as a gun, knife, or club on school property? (4) During the past 12 months, how many times were you in a physical fight? and (5) During the past 12 months, how many times were you in a physical fight in which you were injured and had to be treated by a doctor or nurse?

The four items for *alcohol drinking* behavior are as follows: (1) During your life, on how many days have you had at least one drink of alcohol? (2) How old were you when you had your first drink of alcohol other than a few sips? (3) During the past 30 days, on how many days did you have at least one drink of alcohol? and (4) During the past 30 days, on how many days did you have 5 or more drinks of alcohol in a row, that is, within a couple of hours?

For *tobacco cigarette smoking* behavior, four questions are selected: (1) Have you ever tried cigarettes, even one or two puffs? (2) How old were you when you smoked a whole cigarette for the first time? (3) During the past 30 days, on how many days did you smoke cigarettes? and (4) During the past 30 days, on the days you smoked, how many cigarettes did you smoke per day?

Finally, five items for *other drug use* behavior are: (1) During your life, how many times have you used marijuana? (2) How old were you when you tried marijuana for the first time? (3) During the past 30 days, how many times did you use marijuana? (4) Had you ever tried one of these illegal drugs: cocaine, glue, heroin, methamphetamines, or ecstasy? and (5) During the past 12 months, has anyone offered, sold, or given you an illegal drug on school property?

Table 4 shows the percentages of students responding with an implicative affirmative answer to these manifest items and the extent of missing data for the sample considered in this paper. When questions asked about quantitative matters such as the number of days (or times) and the age of first use, the responses were transformed into yes/no questions for Table 4. The items asking for the number of days (or times) are thus take the value of 1 if the answer is 1 day or more (or one time or more) and 0 otherwise. The items for the age of first use are re-framed to indicate precocious involvement, with a code of 1 for first use under the age of 13 years, and 0 otherwise.

By studying these questionnaire responses, along with covariate terms for age and race-ethnicity, we aimed to investigate the co-occurrence of violent behavior and drug-using behaviors among mid-adolescent male students in the US, and to answer the following questions: (i) What kinds of latent classes exist for each substance use and violence behavior? (ii) Are there common joint patterns of these behaviors? (iii) What kinds of joint latent classes exist to represent those common patterns identified in (ii)? and (iv) Does the prevalence of these common patterns of joint latent classes differ across subgroups specified in terms of age and in terms of race-ethnicity?

Model Selection

The first step in JLCA is to assess an the appropriate number of latent subgroups for each *class* variable. This can be determined before considering the *joint class* variable. In this step, it is not necessary to consider covariates when selecting the number of classes due to marginalization property (Bandeem-Roche et al., 1997). Therefore, we first performed four LCAs for the respective four *class* variables (i.e., violent behavior, alcohol drinking, tobacco cigarette smoking, and other drug use) formed in relation to the respective sets of manifest item responses. Our decision-making about the number of latent subgroups was based on AIC, BIC, and bootstrap p -values with a significance level of $\alpha = 0.05$, with preference for a smaller fit index and a bootstrap p -value larger than 0.05. A large number of sets of starting values were used to avoid local maxima, and the result with the highest likelihood was chosen to be the final ML solution. For the calculation of the bootstrap p -value, we performed a bootstrap procedure by generating 100 bootstrap samples of the G^2 statistic and calculated the proportion of the bootstrapped G^2 s that were larger than the observed G^2 .

Table 5 presents the AIC, BIC, and bootstrap p -values for LCA models with various numbers of latent subgroups for the *class* variables. The *class* variable “*violent behavior*” has the smallest BIC when the number of latent subgroups is three, but its bootstrap p -value is almost 0.00, indicating a lack of suitable model fit. Therefore, we chose a four-class model for “*violent behavior*”, which has the smallest AIC and an acceptable bootstrap p -value. For “*alcohol drinking*” and “*tobacco cigarette smoking*”, the maximum number of latent subgroups is three for the identifiable LC model, because a four-class model with four binary items requires $4 \times 4 + (4 - 1) = 19$ parameters, exceeding the $2^4 - 1 = 15$ available. The goodness-of-fit measures for *alcohol drinking*, *tobacco cigarette smoking*, and *other drug use* indicate the best model has three latent subgroups, producing the smallest AIC, the smallest BIC, and an acceptable bootstrap p -value (Table 5).

The next step in JLCA is to determine the number of subgroups for the *joint class* variables. With the number of latent subgroups for each *class* variable given in Table 5, we enlarged the JLCA model from two to six joint classes and calculated the AIC, BIC, and bootstrap p -value for each JLCA model. We used 100 sets of starting values randomly generated for the η and γ -parameters, but the starting values for the p -parameter were set to the p -parameter estimates from the previous LCA results. As in the first step, the result with the largest likelihood in the final solution was chosen as the final result for the JLCA model.

Table 6 displays goodness-of-fit measures for a series of JLCA models. The JLCA with three joint classes provided the smallest AIC and BIC, indicating that it is the best model among the candidates, and it also had an acceptable bootstrap p -value (> 0.05).

The JLCA with three joint classes included 88 parameters (59 primary measurement parameters, 27 secondary measurement parameters, and 2 prevalence parameters) to be estimated. However, the EM algorithm produced 25 estimates located close to the boundary (24 primary measurement parameters and one secondary measurement parameter), leading to a singular Hessian matrix; accordingly, we set those parameters to a fixed value of 0 or 1. Under these constraints, the number of parameters to be estimated was reduced to 63, and we performed the Newton-Raphson algorithm utilizing the EM estimates as a set of initial values.

Finally, in order to confirm the local identifiability of the selected JLCA model, the singularity of the Hessian matrix must be checked. The Hessian matrix from the ML solution using Newton-Raphson was non-singular under this condition. Its negative inverse matrix (i.e., the asymptotic covariance matrix of the parameter estimates) was positive definite, and the maximum value of the Hessian-based asymptotic variances was 0.0072. Therefore, we concluded that the JLCA is locally identifiable with decent goodness-of-fit when there were three joint classes identified by three *class* variables (i.e., *violent behavior* with four classes, *alcohol drinking* with three classes, *tobacco cigarette smoking* with three classes, and *other drug use* with three classes). Finally, we incorporated covariate terms for age (from 16 to 18) and race-ethnicity (white, black, Hispanic and others) to investigate variations in the relative prevalence of joint-class membership. Race is considered a multinomial variable with three categories, so two dummy variables are used with white as the baseline category. The selected JLCA model for YRBSS 2015 is illustrated in Figure 3.

Parameter Estimates

The estimated primary measurement parameters (i.e., the ρ -parameter) for violence and drug-using behaviors are presented in Tables 7 and 8, respectively. Table 7 represents the estimated probabilities of a ‘Yes’ or ‘Code 1’ response to the five binary items for each *violent behavior* class. The first latent class has low probabilities for all binary items, indicating a sub-group with little violence involvement, so this class is named the ‘little violence-involved’ group. The second latent class has relatively high probabilities for four items compared to the first sub-group, so this class is named the ‘slightly violence-involved’ group. The third latent class has high probabilities for the *carry weapon* and *physical fight* items, and is therefore named the ‘weapon carry and fight’ group. The last latent class, which has high probabilities for all the items, is referred to as the ‘seriously violence-involved’ group. The estimated prevalence of these four latent classes can be derived by calculating the marginal probability as follows:

$$\begin{aligned}\hat{P}(C_j = c_j) &= \frac{1}{n} \sum_{i=1}^n \sum_{s=1}^S \hat{P}(C_j = c_j | U = s) \hat{P}(U = s | \mathbf{x}_i) \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{s=1}^S \hat{\eta}_{c_j | s}^{(j)} \hat{s}_s(\mathbf{x}_i)\end{aligned}$$

for $c_j = 1, \dots, K_j$. Of the four latent classes for Violent behavior, the class ‘little violence-involved’ is the most prevalent (0.610). The second largest latent class is the ‘slightly violence-involved’ group whose estimated prevalence is 0.263. Estimated prevalence proportions for the ‘weapon carry and fight’ and ‘seriously violence-involved’ groups are 0.118 and 0.009, respectively.

Table 7 shows the identified classes for each of the drug involvement *class* variables with the estimates for the primary measurement parameters. The values indicate the estimated probabilities of a ‘Yes’ or ‘Code 1’ response to the items for a given latent class membership. For the *alcohol drinking* class, the first latent class has low probabilities for all items, indicating a ‘non-drinker’ group. The second class has a high probability for the *lifetime drinking* item only, so this class is named ‘lifetime drinker.’ The third class, with high probabilities for all items except for the *early alc onset* item can be referred to as the ‘current alcohol-involved’ group. Similarly, the three classes of *tobacco cigarette smoking* class can be named the ‘non-smoker’, ‘lifetime smoker’ and ‘current smoker with early cigarette onset’ classes respectively. For *other drug use*, the three classes can be named ‘non-user’, ‘current marijuana user’, and ‘early marijuana onset multiple drug user’ based on the manifest items. Table 7 also presents estimated prevalence proportions for the three latent classes. For *alcohol drinking*, the most prevalent class is ‘non-drinker’ with a rate of 0.377. For *Tobacco cigarette smoking*, ‘non-smoker’ is the largest class (0.613), while the prevalence estimate for ‘lifetime smoker’ is 0.30. Likewise, the prevalence of *other drug use* is presented in the last row of Table 7. Just over 50% of the US mid-adolescent male students can be labeled as a ‘non-user,’ roughly 32% qualified as recently active marijuana involvement, and about 14% qualified as involved in use of multiple drugs and with earlier marijuana onset.

Table 9 presents the representative joint patterns for the behavioral classes elicited by JLCA. All of the values except those in the last row in Table 9 are estimates of the secondary measurement parameters (i.e., η -parameters, the conditional probabilities of the class membership of each *Class* variable for a given subgroup of a *joint class* variable). For example, the values under the column ‘Joint class 1’ are the estimated probabilities of belonging to each class given first joint class membership. The majority of this joint class belongs to the non-involved or low level involvement classes for all *Class* variables. This implies little involvement with violent behavior as well as drug-related behavior or experiences in this set of US adolescent males. The JLCA model identified two joint classes for more involved individuals. Of the adolescents in the second joint class (see the column ‘Joint class 2’ in Table 9), about 40% belonged to the ‘slightly violence involved’ class, a higher probability than for those in the least involved class. For drug involvement, almost 60% of the adolescent males in the second joint class have had some past drinking and smoking experiences, and 72% of those qualified as recently active marijuana users. This classification might lead one to adopt a label such as ‘current marijuana user with slight violence involvement and with past drinking and tobacco smoking experiences’. The third joint class has higher probabilities for ‘slightly violence involved’ (0.462) and ‘weapon carry and fight’ (0.467) in the *Violence behavior* class than do the other joint classes. Moreover, the more prominent drug involvement classes are found with substantial frequency for each *class* variable in this joint class (estimated prevalence proportions of 0.910, 0.577, and 0.932, respectively, conditional on joint class membership). Therefore, one might label this class as ‘both violence-involved and drug-involved’ adolescents. The last row in Table 9 presents the prevalence of the three joint classes. Here, the prevalence estimates are the average of the ML estimates for the individual-specific joint class probabilities over the entire sample. The most common joint class lacks both violence involvement and drug involvement, and appears with an estimated prevalence proportion of 0.444, while the least common joint class includes the ‘both violence-involved and drug-involved’ adolescents, with an estimated prevalence of 0.145.

With age and race-ethnicity covariate terms added to the JLCA model, Table 10 provides the estimated odds ratios linking the covariates with membership in each joint class, along with their 95% confidence intervals. As explained in the previous section, the asymptotic standard errors of the estimates are derived from the Hessian matrix calculated during the hybrid algorithm. The regression slope estimate for ‘Age’ conveys the estimated likelihood of belonging to a particular joint class membership in the baseline class of adolescent males with little violent involvement and little drug involvement. As age increases across the range from 16 years to 18 years, the adolescent males are more likely to members of the joint classes with at least minimal violence and drug involvement, including the joint class that has more serious violence and drug involvement. In contrast, race-ethnicity is found to have little or no association with membership in these joint classes.

CONCLUSION

This paper proposes a new latent variable model that can be used to examine underlying joint patterns of multiple latent class variables, whereas the traditional LCA only identifies a small number of classes of a single discrete latent variable. By allowing one or more latent

variables, JLCA has advantages over LCA in not only dimension expansion but also dimension reduction. First, JLCA is a generalization of LCA allowing several latent variables simultaneously. Each latent variable can be considered as different attribute, and JLCA enables to consider various characteristics through relevant manifest items. Such a dimensional expansion also allows to explore the joint action or association of different latent variables, which cannot be explained by the traditional LCA. Second, LCA with a single discrete latent variable is not appropriate when the considered manifest items are not representing homogenous characteristic. In that case, due to the lack of communality among items, LCA approach can cause unnecessarily large number of latent classes, and the interpretability of subgroups become obscure and hard, implying no benefits of subgrouping or dimension reduction. In particular, it is difficult to disentangle the substantive problem regarding associations among latent class variables because the resulting classes from the standard LCA contain information on both of the *class* and *joint class* variables. In contrast, JLCA provides more intuitive insights of joint pattern of several discrete latent variables, with respect to different characteristics. As demonstrated in the example of research on adolescent violence involvement and drug involvement, the new JLCA model makes it possible to evoke a classification of multiple sets of classes for multiple discrete latent variables, and also provides estimates for the associations that link joint class membership with covariates and suspected determinants. Therefore, we believe that the JLCA should qualify as a valuable extension of traditional LCA.

Our empirical analysis of the YRBSS 2015 data produces a reasonable set of latent subgroups for the observed domains of violence involvement and drug involvement. Four latent subgroups are identified by the JLCA for *Violence behavior* across a range of violence involvement. An analogous gradient of involvement in drug-related behaviors and experiences can be seen in the joint classes observed on the basis of YRBS responses to questions about these topics. With these four sets of latent subgroups, there are $4 \times 3 \times 3 \times 3 = 108$ possible combinations of violence and drug-related behaviors and experiences. The JLCA has made it possible to summarize these 108 possible combinations in terms of three joint classes of the adolescent males under study. These three latent subgroups characterize the membership of the adolescent males in patterns of violence involvement and drug involvement that may have substantive importance in the planning of outreach or early intervention programs and in policy-making. The value of these types of joint classes may be enhanced via incorporation of predictors and hypothesized causal influences or covariates, as illustrated here with odds ratios estimated to show associations with age but not with race-ethnicity. That is, membership in the joint classes with greater violence involvement and drug involvement was more likely to be seen for the older adolescent males, but not for race-ethnicity subgroups.

We anticipate that the proposed JLCA with covariates will find application in relation to a fairly large set of problems faced in quantitative research in biomedical, anthropological, behavioral, psychological, and social domains, including quantitative genetics where latent class analysis already has been used in behavioral genetics investigations of Tourette's syndrome (Hirschtritt et al., 2016). We do note that JLCA is limited by the assumptions of local independence, in a similar way to the other models based on item responses such as latent class analysis and latent transition analysis. Therefore, the violations of local

independence may occur in many applications, and researchers should carefully draw inferences about the latent variables by considering the various sets of initial values and observing convergence to the global maximum.

In this research, JLCA postulates a single joint latent class variable which can be identified by the associations among several latent class variables. In this process, latent class variables are considered to be present in parallel and contribute equally to the identification of joint latent class variables. However, the associations among the latent class variables do not have to be equivalent. For example, the structure of associations among multiple latent variables may involve cause-effect linkages. If so, one might consider a specific latent class variable of interest as a response variable, and the joint class variable formulated by other latent class variables can be considered an explanatory variable. During future development of the JLCA model, we hope to incorporate, Bayesian inference via Markov chain Monte Carlo (MCMC) methods in order to overcome limitations sometimes faced in ML estimation.

ACKNOWLEDGMENTS

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2015R1D1A1A01056846 to Chung) and by a US National Institute of Health National Institute on Drug Abuse Senior Scientist and Mentorship Award (K05DA015799 to Anthony).

APPENDIX

Elements of the Score Function

Let Θ be a vector of all free parameters for the JLCA. The score function $S(\Theta)$ is obtained by the first-ordered derivatives of the log-likelihood of the JLCA given in (3) with respect to the model parameters Θ . Let β be the vectorized β parameters in the JLCA model. The elements of the first-derivative vector with respect to β (i.e., $\sum_{i=1}^n \partial \log L_i / \partial \beta$) are given by

$$\sum_{i=1}^n \frac{\partial \log L_i}{\partial \beta_{qs}} = \sum_{i=1}^n x_{iq} [\theta_{i(s)} - \gamma_s(x_i)]$$

for $q = 1, \dots, p$, $s = 1, \dots, S-1$. Likewise, let $\rho_{m_j j | c_j} = (\rho_{m_j 1 j | c_j}, \dots, \rho_{m_j J j | c_j})^T$ and

$\eta_s^{(j)} = (\eta_{1|s}^{(j)}, \dots, \eta_{K_j|s}^{(j)})^T$ be the vectorized ρ - and η -parameters, respectively for $m_j = 1, \dots,$

M_j , $c_j = 1, \dots, K_j$, $J = 1, \dots, J$, and $s = 1, \dots, S$. The elements of the first-derivative vector with respect to $\rho_{m_j j | c_j}$ and $\eta_s^{(j)}$ are obtained by

$$\sum_{i=1}^n \frac{\partial \log L_i}{\partial \rho_{m_j k j | c_j}} = \sum_{i=1}^n \frac{\theta_{i(c_j)} \zeta_{y_{im_j j}^k}}{\rho_{m_j k j | c_j}},$$

$$\sum_{i=1}^n \frac{\partial \log L_i}{\partial \eta_{c_j | s}^{(j)}} = \sum_{i=1}^n \frac{\theta_{i(s, c_j)}}{\eta_{c_j | s}^{(j)}}.$$

Here, $\zeta_{y_{im_j j}^k}$ is the indicator function which has the value of 1 if $y_{im_j j} = k$, otherwise 0. Note that there are $r_{m_j} - 1$ and $K_j - 1$ free parameters in $\rho_{m_j k j | c_j}$ and $\eta_s^{(j)}$, respectively. Therefore, the score function of the free parameters for $\rho_{m_j k j | c_j}$ and $\eta_s^{(j)}$ can be obtained by

$$\sum_{i=1}^n \frac{\partial \log L_i}{\partial \rho_{m_j k j | c_j}} \mathbf{A}_{r_{m_j}}^T \quad \text{and} \quad \sum_{i=1}^n \frac{\partial \log L_i}{\partial \eta_s^{(j)}} \mathbf{A}_{K_j}^T,$$

where $\mathbf{A}_k$ is a $(k-1) \times k$ matrix, composed of an identity matrix in the first $k-1$ columns and a column vector of -1 in the last column for $m_j = 1, \dots, M_j$, $c_j = 1, \dots, K_j$, $j = 1, \dots, J$, and $s = 1, \dots, S$.

Elements of the Hessian Matrix

The Hessian matrix is the negative inverse matrix of the second derivatives of the log-likelihood with respect to the model parameter Θ . The elements of the second derivative matrix with respect to β are given by

$$\sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \beta_{qs} \partial \beta_{q's'}} = \sum_{i=1}^n x_{iq} x_{iq'} [\theta_{is} (\zeta_{ss'} - \theta_{is'}) - \gamma_s(x_i) \zeta_{ss'} - \gamma_{s'}(x_i)]$$

for $q, q' = 1, \dots, p$, $s, s' = 1, \dots, S-1$, where $\zeta_{ss'} = 1$ if $s = s'$, otherwise 0 if $s \neq s'$. The elements of the second derivative matrices with respect to β and $\rho_{m_j k j | c_j}$ and $\eta_s^{(j)}$ can be obtained by

$$\sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \beta_{qs} \partial \rho_{m_j k j | c_j}} = \sum_{i=1}^n \frac{x_{iq} [\theta_{i(s, c_j)} - \theta_{i(s)} \theta_{i(c_j)}] \zeta_{y_{im_j j}^k}}{\rho_{m_j k j | c_j}},$$

$$\sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \beta_{qs} \partial \eta_{c_j | s}^{(j)}} = \sum_{i=1}^n \frac{x_{iq} [\theta_{i(s', c_j)} (\zeta_{ss'} - \theta_{i(s)})]}{\eta_{c_j | s}^{(j)}}$$

for $m_j = 1, \dots, M_j$, $k = 1, \dots, r_{m_j}$, $c_j = 1, \dots, K_j$, $j = 1, \dots, J$, and $s = 1, \dots, S$. Here, $\zeta_{y_{im_j j}^k}$ is an indicator function which has the value of 1 if $y_{im_j j} = k$, otherwise 0. Note that there are $r_{m_j} - 1$ and $K_j - 1$ free parameters in $\rho_{m_j j | c_j}$ and $\eta_s^{(j)}$ respectively. Therefore, the Hessian elements of the free parameters for $\rho_{m_j j | c_j}$ and $\eta_s^{(j)}$ can be obtained by

$$\sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \beta \partial \rho_{m_j j | c_j}} \mathbf{A}_{r_{m_j}}^T \quad \text{and} \quad \sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \beta \partial \eta_s^{(j)}} \mathbf{A}_{K_j}^T.$$

The elements of the second derivative matrices with respect to ρ and η are as follows:

$$\begin{aligned} \sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \rho_{m_j k j | c_j} \partial \rho_{m_j k' j' | c_j}} &= \sum_{i=1}^n \frac{\left[\theta_{i(c_j, c_j')} (1 - \zeta_{j' j}) + \theta_{i(c_j)} \left\{ \zeta_{j' j} [(1 - \zeta_{cc'}) + (1 - \zeta_{mm'})] - \theta_{i(c_j')} \right\} \right] \zeta_{y_{im_j j}^k} \zeta_{y_{im_j j'}^{k'}}}{\rho_{m_j k j | c_j} \rho_{m_j k' j' | c_j}} \\ \sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \eta_{c_j' | s}^{(j')} \partial \rho_{m_j k j | c_j}} &= \sum_{i=1}^n \frac{(1 - \zeta_{j' j}) \theta_{i(s, c_j, c_j')} + \theta_{i(s, c_j')} \left(\zeta_{c_j c_j'} - \theta_{i(c_j')} \right) \zeta_{y_{im_j j}^k}}{\eta_{c_j' | s}^{(j')} \rho_{m_j k j | c_j}} \end{aligned}$$

for $m_j = 1, \dots, M_j$, $k, k' = 1, \dots, r_{m_j}$, $c_j = 1, \dots, K_j$, $s, s' = 1, \dots, S$, and $j, j' = 1, \dots, J$.

Here, $\zeta_{y_{im_j j}^k}$ is an indicator function which has the value of 1 if $y_{im_j j} = k$, otherwise 0. Note that there are $r_{m_j} - 1$ and $K_j - 1$ free parameters in $\rho_{m_j j | c_j}$ and $\eta_s^{(j)}$, respectively. Therefore, the Hessian elements of the free parameters for $\rho_{m_j j | c_j}$ and $\eta_s^{(j)}$ can be obtained by

$$\mathbf{A}_{r_{m_j}} \sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \rho_{m_j j | c_j} \partial \rho_{m_j j' | c_j}} \mathbf{A}_{r_{m_j'}}^T \quad \text{and} \quad \mathbf{A}_{K_j} \sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \eta_s^{(j)} \partial \rho_{m_j j | c_j}} \mathbf{A}_{r_{m_j}}^T.$$

The elements of the second derivative vector with respect to η are as follows:

$$\sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \eta_{c_j | s}^{(j)} \partial \eta_{c_j' | s'}^{(j')}} = \sum_{i=1}^n \frac{\theta_{i(s, c_j, c_j')} \zeta_{ss'} (1 - \zeta_{j' j}) - \theta_{i(s, c_j)} \theta_{i(s', c_j')}}{\eta_{c_j | s}^{(j)} \eta_{c_j' | s'}^{(j')}}.$$

for $c_j = 1, \dots, K_j$, $j, j' = 1, \dots, J$, and $s, s' = 1, \dots, S$. There are $K_j - 1$ free parameters in $\eta_s^{(j)}$ and $K_{j'} - 1$ free parameters in $\eta_{s'}^{(j')}$, respectively. Therefore, the Hessian elements of the free parameters for $\eta_s^{(j)}$ and $\eta_{s'}^{(j')}$ can be obtained by

$$\mathbf{A}_{K_{j_i}} \sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \eta_s^{(j)} \partial \eta_{s'}^{(j')}} \mathbf{A}_{K_{j'}}^T$$

for $c_j = 1, \dots, K_j$, $j, j' = 1, \dots, J$, and $s, s' = 1, \dots, S$.

References

- Bandeem-Roche K, Miglioretti DL, Zeger SL, & Rathouz PJ (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association*, 92(440), 1375–1386.
- Bourque S (2016). Association between adolescent substance use and weapon carrying: Analysis of 2013 YRBS (Unpublished master's thesis). Georgia State University.
- Center for Behavioral Health Statistics and Quality. (2015). Behavioral health trends in the United States: Results from the 2014 national survey on drug use and health. (HHS Publication No. SMA 15-4927, NSDUH Series H-50) Retrieved from <http://www.samhsa.gov/data/>.
- Chung H, Anthony JC, & Schafer JL (2011). Latent class profile analysis: an application to stage sequential processes in early onset drinking behaviours. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(3), 689–712.
- Chung H, Park Y, & Lanza ST (2005). Latent transition analysis with covariates: Pubertal timing and substance use behaviours in adolescent females. *Statistics in Medicine*, 24(18), 2895–2910. [PubMed: 16134129]
- Dempster AP, Laird NM, & Rubin DB (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1–38.
- Duncan DT, Rienti MJ, Kulldorff M, Aldstadt J, Castro MC, Frounfelker R, . . . Williams DR (2016). Local spatial clustering in youths' use of tobacco, alcohol, and marijuana in boston. *The American Journal of Drug and Alcohol Abuse*, 42(4), 412–421. [PubMed: 27096932]
- DuRant RH, Altman D, Wolfson M, Barkin S, Kreiter S, & Krowchuk D (2000). Exposure to violence and victimization, depression, substance use, and the use of violence by young adolescents. *The Journal of Pediatrics*, 137(5), 707–713. [PubMed: 11060539]
- Epstein-Ngo QM, Cunningham RM, Whiteside LK, Chermack ST, Booth BM, Zimmerman MA, & Walton MA (2013). A daily calendar analysis of substance use and dating violence among high risk urban youth. *Drug and Alcohol Dependence*, 130(1), 194–200. [PubMed: 23219602]
- Harford TC, Chen CM, & Grant BF (2016). Other-and self-directed forms of violence and their relationship with number of substance use disorder criteria among youth ages 12–17: Results from the national survey on drug use and health. *Journal of studies on alcohol and drugs*, 77(2), 277–286. [PubMed: 26997186]
- Herrenkohl TI, Maguin E, Hill KG, Hawkins JD, Abbott RD, & Catalano RF (2000). Developmental risk factors for youth violence. *Journal of Adolescent Health*, 26(3), 176–186. [PubMed: 10706165]
- Hirschtritt ME, Darrow SM, Illmann C, Osiecki L, Grados M, Sandor P, . . . others (2016). Social disinhibition is a heritable subphenotype of tics in tourette syndrome. *Neurology*, 87(5), 497–504. [PubMed: 27371487]
- Kann L, McManus T, A HW, Shanklin SL, Flint KH, Hawkins B, J an Queen, . . . Zaza S (2016). Youth risk behavior surveillance - united states, 2015. *MMWR Surveillance Summaries*, 65(SS-6), 1–174.

- Kaplan D, & Keller B (2011). A note on cluster effects in latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 18(4), 525–536.
- Lanza ST, Coffman DL, & Xu S (2013). Causal inference in latent class analysis. *Structural Equation Modeling: a Multidisciplinary Journal*, 20(3), 361–383. [PubMed: 25419097]
- Lazarsfeld PF, & Henry NW (1968). *Latent structure analysis*. Houghton Mifflin Co.
- Lowry R, Cohen LR, Modzeleski W, Kann L, Collins JL, & Kolbe LJ (1999). School violence, substance use, and availability of illegal drugs on school property among us high school students. *Journal of School Health*, 69(9), 347–355. [PubMed: 10633319]
- Roesch SC, Villodas M, & Villodas F (2010). Latent class/profile analysis in maltreatment research: A commentary on Noonan et al., Pears et al., and looking beyond. *Child Abuse & Neglect*, 34(3), 155–160. [PubMed: 20207416]
- Rothman EF, Reyes LM, Johnson RM, & LaValley M (2012). Does the alcohol make them do it? Dating violence perpetration and drinking among youth. *Epidemiologic Reviews*, 34(1), 103–119. [PubMed: 22128086]
- Rubin DB (1976). Inference and missing data. *Biometrika*, 63(3), 581–592.
- Swahn MH, & Donovan JE (2006). Alcohol and violence: Comparison of the psychosocial correlates of adolescent involvement in alcohol-related physical fighting versus other physical fighting. *Addictive Behaviors*, 31(11), 2014–2029. [PubMed: 16571368]
- Swahn MH, Simon TR, Hammig BJ, & Guerrero JL (2004). Alcohol-consumption behaviors and risk for physical fighting and injuries among adolescent drinkers. *Addictive Behaviors*, 29(5), 959–963. [PubMed: 15219342]
- Van der Heijden PG, Dessens J, & Bockenholt U (1996). Estimating the concomitant-variable latent-class model with the em algorithm. *Journal of Educational and Behavioral Statistics*, 21(3), 215–229.
- Van Horn ML, Fagan AA, Jaki T, Brown EC, Hawkins JD, Arthur MW, . . . Catalano RF (2008). Using multilevel mixtures to evaluate intervention effects in group randomized trials. *Multivariate Behavioral Research*, 43(2), 289–326. [PubMed: 26765664]
- Vermunt JK (2003). Multilevel latent class models. *Sociological Methodology*, 33(1), 213–239.
- Walton MA, Cunningham RM, Goldstein AL, Chermack ST, Zimmerman MA, Bingham CR, . . . Blow FC (2009). Rates and correlates of violent behaviors among adolescents treated in an urban emergency department. *Journal of Adolescent Health*, 45(1), 77–83. [PubMed: 19541253]
- White HR, Loeber R, Stouthamer-Loeber M, & Farrington DP (1999). Developmental associations between substance use and violence. *Development and Psychopathology*, 11(04), 785–803. [PubMed: 10624726]

Figure 1:
The structure of a joint latent class analysis

Figure 2:
The structure of a joint latent class analysis with covariates

Figure 3:
The structure of joint latent class analysis for the YRBSS 2015 data

Table 1:

Goodness-of-fit measures for a series of LCA models with different numbers of classes for two different *class* variables in a simulated data[†].

Latent variable	Number of (joint) classes	Strong measurement			Mixed measurement		
		AIC	BIC	<i>p</i> -value	AIC	BIC	<i>p</i> -value
<i>Class 1</i>	2	3230.03	3284.82	0.00	3566.02	3620.81	0.00
	3	2921.52	3005.81	0.45	3403.50	3487.79	0.42
	4	2929.70	3043.49	0.79	3566.02	3522.21	0.61
<i>Class 2</i>	2	3229.32	3284.11	0.00	3581.58	3636.37	0.00
	3	2928.03	3012.32	0.25	3464.94	3549.24	0.68
	4	2936.51	3050.30	0.51	3476.27	3590.06	0.99
<i>Joint class</i>	2	5765.27	5954.93	0.08	6810.53	7000.19	0.49
	3	5732.07	5942.80	0.59	6779.82	6990.55	0.75
	4	5742.07	5973.87	0.77	6789.82	7021.62	0.99

[†]The true number of (joint) classes is three for all latent variables.

Table 2.

The frequency of the selected model via different goodness-of-fit measures over 50 repetitions with $n = 250$.

Latent variable	Number of (joint) classes	<u>Strong measurement</u>			<u>Mixed measurement</u>		
		AIC	BIC	<i>p</i> -value	AIC	BIC	<i>p</i> -value
<i>Class 1</i>	2	0	0	0	0	0	0
	3	46	50	44	45	50	49
	4	4	0	6	5	0	1
<i>Class 2</i>	2	0	0	0	0	1	0
	3	49	50	49	47	49	48
	4	1	0	1	3	0	2
<i>Joint class</i>	2	0	9	26	0	31	42
	3	50	41	24	50	19	7
	4	0	0	0	0	0	1

Table 3:

The frequency of the selected model via different goodness-of-fit measures over 50 repetitions with $n = 500$.

Latent variable	Number of (joint) classes	<u>Strong measurement</u>			<u>Mixed measurement</u>		
		AIC	BIC	<i>p</i> -value	AIC	BIC	<i>p</i> -value
<i>Class 1</i>	2	0	0	0	0	0	0
	3	44	50	47	50	44	49
	4	6	0	3	0	6	1
<i>Class 2</i>	2	0	0	0	1	0	0
	3	46	50	48	49	44	47
	4	4	0	2	0	6	3
<i>Joint class</i>	2	0	0	22	0	2	34
	3	50	50	26	50	48	15
	4	0	0	2	0	0	1

Table 4:

Percentages of students responding “Yes” to the manifest items for each latent variable and the frequency of missing responses[†].

<i>Class</i>		Yes	Missing
variable	Manifest item	(%)	(%)
<i>Violent behavior</i>	Carried a weapon during recent 30 days (<i>Carry Weapon</i>)	24.26	8.21
	Absent to school due to feeling unsafe recent 12 months (<i>Feeling Unsafe</i>)	5.14	0.32
	Threatened or injured by weapon on school recent 12 months (<i>Being Threatened</i>)	7.10	3.87
	Involved in a physical fight recent 12 months (<i>Physical Fight</i>)	22.47	13.71
	Seriously injured in a physical fight recent 12 months (<i>Seriously Injured</i>)	3.31	12.16
<i>Alcohol drinking</i>	Ever drank alcohol (<i>Lifetime Drinking</i>)	64.95	3.18
	drank alcohol before age 13 years for the first time (<i>Early ALC Onset</i>)	19.42	2.78
	drank alcohol during the recent 30 days (<i>Recent Drinking</i>)	33.10	16.76
	drank five or more drinks of alcohol in a row (<i>Heavy Episodic Drinking</i>)	21.44	5.26
<i>Tobacco cigarette smoking</i>	Ever tried tobacco cigarette smoking (<i>Lifetime Smoking</i>)	36.35	9.96
	Smoked tobacco before age 13 years for the first time (<i>Early CIG Onset</i>)	8.73	5.85
	Smoked tobacco cigarettes during the recent 30 days (<i>Recent Smoking</i>)	13.33	5.18
	Smoked more than 10 cigarettes per day (<i>Heavy Smoking</i>)	1.45	5.38
<i>Other drug use</i>	Ever tried marijuana (<i>Lifetime MJ Use</i>)	45.87	4.17
	Tried marijuana before age 13 years for the first time (<i>Early MJ Onset</i>)	10.65	2.54
	Used marijuana during the recent 30 days (<i>Recent MJ Use</i>)	26.62	10.53
	Ever tried other illegal drugs (<i>Lifetime Drug Use</i>)	15.29	5.41
	Ever was sold or offered an illegal drug on school property (<i>Drug Offered on School</i>)	25.47	3.77

[†] Note that some of these behaviors represent ‘involvement’ or ‘experience’ rather than the individual’s own behavior per se.

Table 5:

Goodness-of-fit measures for a series of LCA models with different numbers of classes for four different *class* variables.

Class variable	Number of classes	AIC	BIC	<i>p</i> -value
<i>Violent behavior</i>	2	14779.8	14851.4	0.00
	3	14687.0	14797.6	0.00
	4	14652.0	14801.7	0.34
	5	14663.7	14852.4	0.49
<i>Alcohol drinking</i>	2	16826.8	16885.4	0.00
	3	16607.9	16699.0	0.41
<i>Tobacco cigarette smoking</i>	2	11195.1	11253.6	0.00
	3	11058.8	11150.0	0.34
<i>Other drug use</i>	2	20381.3	20452.9	0.00
	3	20275.3	20386.0	0.28
	4	20279.7	20429.4	0.45
	5	20291.2	20480.0	0.68

Table 6:

Goodness-of-fit measures for a series of JLCA models with differing numbers of joint classes.

Number of joint classes	AIC	BIC	<i>p</i> -value
2	92849.5	93357.2	0.02
3	92062.3	92635.1	0.14
4	94555.2	95193.0	0.24
5	99348.1	100051.0	0.26
6	102272.9	103040.9	0.34

Table 7:

The estimated class prevalence and probabilities of responding “Yes” to the manifest items for each *Violent behavior* class (i.e., the ρ -parameters for *Violent behavior*)[†].

Manifest item	Class for <i>Violent behavior</i>			
	Little violence involved	Slightly violence involved	Weapon carry and fight	Seriously violence involved
<i>Carry Weapon</i>	0.113	0.444	0.642	0.791
<i>Feeling Unsafe</i>	0.010	0.066	0.165	1.000 [†]
<i>Being Threatened</i>	0.011	0.092	0.289	1.000 [†]
<i>Physical Fight</i>	0.064	0.394	1.000 [†]	1.000 [†]
<i>Seriously Injured</i>	0.000 [†]	0.000 [†]	0.271	1.000 [†]
Class prevalence estimate	0.610	0.263	0.118	0.009

[†]The estimated probabilities are constrained to be zero or one.

Table 8:

The estimated class prevalence and probabilities of responding ‘yes’ to the manifest items for each drug involvement class (i.e., ρ -parameters for *Alcohol drinking*, *Tobacco cigarette smoking*, and *Other drug use*)[†].

Class for <i>Alcohol drinking</i>			
Manifest item	Non-drinker	Lifetime drinker	Current alcohol-involved
<i>Lifetime Drinking</i>	0.159	1.000 [†]	1.000 [†]
<i>Early ALC Onset</i>	0.036	0.233	0.383
<i>Recent Drinking</i>	0.000 [†]	0.305	1.000 [†]
<i>Heavy Episodic Drinking</i>	0.000 [†]	0.0001	0.885
Class prevalence estimate	0.377	0.352	0.271
Class for <i>Tobacco cigarette smoking</i>			
Manifest item	Non-smoker	Lifetime smoker	Early onset current smoker
<i>Lifetime Smoking</i>	0.046	1.000 [†]	1.000 [†]
<i>Early CIG Onset</i>	0.000 [†]	0.158	0.539
<i>Recent Smoking</i>	0.000 [†]	0.218	1.000 [†]
<i>Heavy Smoking</i>	0.000 [†]	0.000 [†]	0.200
Class prevalence estimate	0.613	0.300	0.087
Class for <i>Other drug use</i>			
Manifest item	Non-user	Current MJ user	Early MJ onset multiple drug user
<i>Lifetime MJ Use</i>	0.041	1.000 [†]	1.000 [†]
<i>Early MJ Onset</i>	0.000 [†]	0.127	0.503
<i>Recent MJ Use</i>	0.000 [†]	0.503	0.836
<i>Lifetime Drug Use</i>	0.032	0.202	0.814
<i>Drug Offered on School</i>	0.165	0.309	0.579
Class prevalence estimate	0.539	0.318	0.143

[†]The probabilities are constrained to be zero or one.

Table 9:

The estimated probabilities of belonging to a class for each joint class (η -parameter) and the estimated prevalence of joint classes (γ -parameter) [†].

Class	Joint class		
	Joint class 1	Joint class 2	Joint class 3
<i>Violent behavior</i>			
Little Violence involved	0.895	0.513	0.015
Slightly violence involved	0.069	0.401	0.462
Weapon carry and fight	0.034	0.084	0.467
Seriously violence involved	0.002	0.002	0.056
<i>Alcohol drinking</i>			
Non-drinker	0.760	0.087	0.034
Lifetime drinker	0.220	0.598	0.056
Current heavier drinker	0.020	0.315	0.910
<i>Tobacco cigarette smoking</i>			
Non-smoker	0.989	0.399	0.076
Lifetime smoker	0.005	0.599	0.347
Early onset current smoker	0.006	0.002	0.577
<i>Other drug use</i>			
Non-user	0.963	0.260	0.045
Current MJ user	0.037	0.722	0.023
Early MJ onset multiple drug user	0.000 [†]	0.018	0.932
Joint class prevalence estimate	0.444	0.411	0.145

[†]The estimated probability is constrained to be zero.

Table 10:

The estimated odds ratios for joint classes and their 95% confidence intervals, with baseline (reference) odds set for membership in the joint class with the lowest levels of involvement in violence and drug-related behaviors and experiences.

Covariate	Joint class		<i>d.f.</i>	Chi-square	<i>p</i> -value
	Joint class 2	Joint class 3			
Age (from 16 to 18 years)	1.413 [1.282, 1.558]	1.467 [1.234, 1.743]	2	102.24	< 0.001
Race-ethnicity (White as reference)					
Black	1.161 [0.892, 1.510]	0.863 [0.618, 1.611]	4	4.78	0.311
Others	1.009 [0.858, 1.187]	1.157 [0.867, 1.544]			