

Latent class analysis with multiple latent group variables

Jung Wun Lee^a, Hwan Chung^{1, a}

^aDepartment of Statistics, Korea University, Korea

Abstract

This study develops a new type of latent class analysis (LCA) in order to explain the associations between one latent variable and several other categorical latent variables. Our model postulates that the prevalence of the latent variable of interest is affected by another latent variable composed of other several latent variables. For the parameter estimation, we propose deterministic annealing EM (DAEM) to deal with local maxima problem in the proposed model. We perform simulation study to demonstrate how DAEM can find the set of parameter estimates at the global maximum of the likelihood over the repeated samples. We apply the proposed LCA model in an investigation of the effect of and joint patterns for drug-using behavior to violent behavior among US high school male students using data from the Youth Risk Behavior Surveillance System 2015. Considering the age of male adolescents as a covariate influencing violent behavior, we identified three classes of violent behavior and three classes of drug-using behavior. We also discovered that the prevalence of violent behavior is affected by the type of drug used for drug-using behavior.

Keywords: deterministic annealing EM, drug-using behavior, joint latent class analysis, multiple modalities, violent behavior

1. Introduction

Violent and drug-using behavior are a major social issue among US adolescents that contribute to premature death, disability, and other social problems (Van Horn *et al.*, 2014). Miller *et al.* (2007) revealed that young injection drug users may be exposed to the risk of premature mortality due to health risks. These delinquent behavior were studied through previous research that showed strong associations between drug-using and violent behavior. White *et al.* (1999) showed that there are concurrent associations between frequency of drug use and violence among adolescence teenagers. Lundholm (2013) revealed that acute substance intake and abuse increase the risks of both interpersonal and self-directed violence. Some studies have also found that students who drank alcohol on school property were more likely to carry weapons at school than those who did not drink alcohol (Lowry *et al.*, 1999) and that the severity of adolescent violent behavior significantly correlated to the frequency of cigarette smoking, alcohol use, and multiple substance use (DuRant *et al.*, 2000). Dawkins (1997) found that the use of alcohol and marijuana is more important in terms of their effects on some violent offenses than other drug use. Many of these types of studies used several questionnaires to measure drug-using behavior and violent behavior. Standard statistical methods were used to find associations among variables of interests that transformed the original manifest items to scale or categorical variables; calculated standard statistics such as chi-square, Spearman ranked correlation, Pearson's

¹ Corresponding author: Department of Statistics, Korea University, 145, Anam-ro, Seongbuk-gu, Seoul 02841, Korea.
E-mail: hwanch@korea.ac.kr

correlation, Kruskal-Wallis, and odds ratio; and conducted multiple linear regression or logistic regression to verify the effect of drug-using behavior on some violent offenses. However, variables like drug-using behavior or violent behavior cannot be measured directly with a single item and should be considered as latent variables. The latent variable may be identified through multiple manifest items that measure some hypothetical concepts. The latent class analysis (LCA) can be very helpful in this line of research. LCA identifies a small number of underlying subgroups of the population, using several manifest items. These subgroups are called latent classes. Manifest items measuring a latent variable are strongly correlated and LCA can identify individuals whose response patterns for the manifest items are similar and classify them into the same latent class.

In this paper, a new type of LCA with multiple latent groups (LCA-MLG) has been proposed so that we can investigate the effect of drug-using behavior on violent behavior. In our model, the subgroups of drug-using behavior are identified by the joint latent class analysis (JLCA) using the framework of LCA with multiple groups. We adopted deterministic annealing EM (DAEM) as a parameter estimation strategy to overcome the local maxima problem. We then use the proposed model to the data from the Youth Risk Behavior Surveillance System 2015 (YRBSS 2015) in an investigation of the effect of the joint patterns of drug-using behavior to violent behavior among the US high school male students (Centers for Disease Control and Prevention, 2015).

The remainder of this paper presents the description of the proposed model LCA-MLG and estimation methods for the model parameters in Sections 2 and 3, respectively. In Section 4, we evaluate the performance of DAEM over repeated sampling. We then apply the proposed model to the real dataset from the YRBSS 2015 in Section 6.

2. Model

2.1. Latent class analysis

The LCA is a finite mixture model for dividing population into several subgroups based on individuals' responses to the manifest items. LCA assumes that the population is composed of several unobservable subgroups (i.e., latent classes) which can be measured by multiple manifest items indirectly, implying that associations among the manifest items are totally explained by latent class variable. Suppose there are P categorical manifest items Z_1, \dots, Z_P . The responses of each manifest item for the i^{th} individual are obtained as a P -dimensional vector $\mathbf{z}_i = [z_{i1}, \dots, z_{iP}]^T$, where z_{ip} can take any value from $1, \dots, r_p$ for $p = 1, \dots, P$. Let the latent class variable W has D categories, then the observed-data likelihood of LCA can be specified as

$$\begin{aligned}
 P(\mathbf{Z} = \mathbf{z}_i) &= \sum_{w=1}^D P(W = w, \mathbf{Z} = \mathbf{z}_i) = \sum_{w=1}^D P(W = w)P(\mathbf{Z} = \mathbf{z}_i | W = w) \\
 &= \sum_{w=1}^D P(W = w) \prod_{p=1}^P P(Z_p = z_{ip} | W = w) \\
 &= \sum_{w=1}^D \delta_w \prod_{p=1}^P \prod_{h=1}^{r_p} \phi_{ph|w}^{I(z_{ip}=h)}, \tag{2.1}
 \end{aligned}$$

where $I(z_{ip} = h)$ is the indicator function which is 1 when $z_{ip} = h$ and 0 otherwise. The likelihood of LCA given in (2.1) is constructed under the local independence assumptions, implying that the manifest items are conditionally independent given a latent class membership. Here, $\phi_{ph|w} = P(Z_p =$

$h \mid W = w$), referred as the primary measurement parameter, explains the relationship between the latent class and the p^{th} manifest item, and $\delta_w = P(W = w)$ represents the prevalence of latent class membership w ; therefore, parameters in (2.1) are conditional probabilities, the sum-to-one and non-negative constraints are explicit (i.e., $\sum_{w=1}^D \delta_w = 1$ and $\sum_{h=1}^{r_p} \phi_{ph|w} = 1$ for $p = 1, \dots, P$, $w = 1, \dots, D$).

2.2. Joint latent class analysis (JLCA)

The JLCA is an extended version of the LCA model to deal with multiple latent class variables. Suppose there are J latent class variables $\mathbf{C} = [C_1, \dots, C_J]^T$, which can be identified by the J sets of manifest items $\mathbf{Y}_1, \dots, \mathbf{Y}_J$, respectively. Let the j^{th} latent class variable C_j have K_j nominal categories for $j = 1, \dots, J$. Here, $\mathbf{Y}_j = [Y_{1j}, \dots, Y_{M_jj}]^T$ is the vector of the j^{th} set of manifest items measuring the latent class C_j for $j = 1, \dots, J$. Further, let $\mathbf{y}_{ij} = [y_{i1j}, \dots, y_{iM_jj}]^T$ be the i^{th} individual's responses to a set of M_j manifest items measuring the j^{th} latent class C_j , where each item response y_{im_jj} can take any value from 1 to r_{m_j} for $m_j = 1, \dots, M_j$ and $j = 1, \dots, J$. In JLCA, the association among J latent class variables $\mathbf{C} = [C_1, \dots, C_J]^T$ can be explained by other latent subgroups (i.e., joint latent classes). The population can be divided into several subgroups that have similar joint patterns of the latent class variables $\mathbf{C} = [C_1, \dots, C_J]^T$.

Let the joint latent class variable U have S nominal categories describing the most common joint patterns of the latent classes. Then, the i^{th} individual's contribution to the likelihood function for JLCA can be specified as

$$\begin{aligned}
 P(\mathbf{Y} = \mathbf{y}_i) &= \sum_{u=1}^S \sum_{c_1=1}^{K_1} \cdots \sum_{c_J=1}^{K_J} P(U = u, \mathbf{C} = \mathbf{c}, \mathbf{Y} = \mathbf{y}_i) \\
 &= \sum_{u=1}^S \sum_{c_1=1}^{K_1} \cdots \sum_{c_J=1}^{K_J} P(U = u) P(\mathbf{C} = \mathbf{c} \mid U = u) P(\mathbf{Y} = \mathbf{y}_i \mid \mathbf{C} = \mathbf{c}) \\
 &= \sum_{u=1}^S \sum_{c_1=1}^{K_1} \cdots \sum_{c_J=1}^{K_J} P(U = u) \prod_{j=1}^J \left\{ P(C_j = c_j \mid U = u) \prod_{m_j=1}^{M_j} P(Y_{m_jj} = y_{im_jj} \mid C_j = c_j) \right\} \\
 &= \sum_{u=1}^S \sum_{c_1=1}^{K_1} \cdots \sum_{c_J=1}^{K_J} \gamma_u \prod_{j=1}^J \left\{ \eta_{c_j|u}^{(j)} \prod_{m_j=1}^{M_j} \prod_{k=1}^{r_{m_j}} \rho_{m_jk|c_j}^{I(y_{im_jj}=k)} \right\}, \tag{2.2}
 \end{aligned}$$

where $I(y_{im_jj} = k)$ is the indicator function which is 1 when $y_{im_jj} = k$ and 0 otherwise. The primary measurement parameter, $\rho_{m_jk|c_j} = P(Y_{m_jj} = k \mid C_j = c_j)$, is the probability of responding k for the m_j^{th} item measuring the j^{th} latent class variable, given that the j^{th} latent class membership is c_j . The secondary measurement parameter $\eta_{c_j|u}^{(j)} = P(C_j = c_j \mid U = u)$ represents the probability of the j^{th} latent class variable having membership c_j given that the joint latent class membership is u .

The prevalence parameter, $\gamma_u = P(U = u)$, represents the probability of having u^{th} joint latent class membership. As LCA, all parameters in (2.2) are conditional probabilities, the sum-to-one and non-negative constraints are explicit (i.e., $\sum_{k=1}^{r_{m_j}} \rho_{m_jk|c_j} = 1$ for $m_j = 1, \dots, M_j$, $c_j = 1, \dots, K_j$, and $j = 1, \dots, J$; $\sum_{c_j=1}^{K_j} \eta_{c_j|u}^{(j)} = 1$ for $j = 1, \dots, J$ and $u = 1, \dots, S$; and $\sum_{u=1}^S \gamma_u = 1$). The likelihood function of JLCA given in (2.2) is based on the following three assumptions: (a) The joint class is related to manifest items only through the latent class variables; (b) the manifest items are conditionally independent given the latent class membership; and (c) the latent class membership is unrelated each other given joint latent class u .

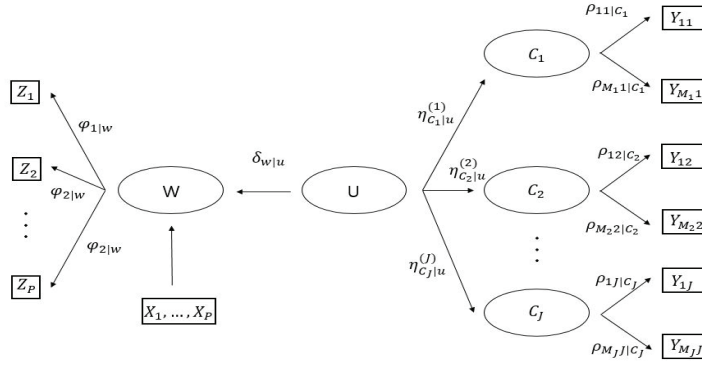


Figure 1: A diagram of latent class analysis with multiple latent group variables.

2.3. Latent class analysis with multiple latent group variables (LCA-MLG)

The LCA-MLG postulates that latent class variable may be affected by joint latent class variable which can be identified by the JLCA model. We therefore consider joint latent class as a latent group variable in the traditional LCA. We propose the LCA-MLG and illustrate the model in Figure 1. The right side of Figure 1 is a JLCA with joint latent class variable U uncovered by the vector of latent class variable $\mathbf{C} = [C_1, \dots, C_J]^T$, and each latent variable C_j is identified by the j^{th} set of manifest items $\mathbf{Y}_j = [Y_{1j}, \dots, Y_{M_j j}]^T$. The left side of Figure 1 is an ordinary LCA with latent variable W which can be identified through the manifest items $\mathbf{Z} = [Z_1, \dots, Z_P]^T$. Consequently, the outcome latent class variable W is affected by the joint latent class variable U , and the distribution of the outcome latent class variable W is varied as the value of latent group variable U changes. In this manner, the association between latent group variable and the outcome latent class variable can be investigated. Using the notation given in (2.1) and (2.2), the complete-data likelihood of the LCA-MLG for the i^{th} observation is obtained by

$$\begin{aligned}
 L_i^* &= P(W = w, \mathbf{Z} = \mathbf{z}_i, U = u, \mathbf{C} = \mathbf{c}, \mathbf{Y} = \mathbf{y}_i) \\
 &= P(W = w, \mathbf{Z} = \mathbf{z}_i \mid U = u)P(U = u, \mathbf{C} = \mathbf{c}, \mathbf{Y} = \mathbf{y}_i) \\
 &= P(W = w \mid U = u)P(\mathbf{Z} = \mathbf{z}_i \mid W = w) \times P(U = u)P(\mathbf{C} = \mathbf{c} \mid U = u)P(\mathbf{Y} = \mathbf{y}_i \mid \mathbf{C} = \mathbf{c}) \\
 &= \delta_{w|u} \prod_{p=1}^P \prod_{h=1}^{r_p} \phi_{ph|d}^{I(z_{ip}=h)} \gamma_u \prod_{j=1}^J \left\{ \eta_{c_j|u}^{(j)} \prod_{m_j=1}^{M_j} \prod_{k=1}^{r_{m_j}} \rho_{m_j k|c_j}^{I(y_{im_j j}=k)} \right\}. \tag{2.3}
 \end{aligned}$$

The likelihood of the manifest items (i.e., the observed-data likelihood) can be derived by the marginal summation of (2.3) with respect to all considered latent variables:

$$L_i = P(\mathbf{Z}_i = \mathbf{z}_i, \mathbf{Y}_i = \mathbf{y}_i) = \sum_{u=1}^S \sum_{c_1=1}^{K_1} \cdots \sum_{c_J=1}^{K_J} \sum_{w=1}^D L_i^*. \tag{2.4}$$

The prevalence of the outcome latent class variable may be affected by the demographic characteristics or other individual information, and these characteristics can be considered as covariates in the proposed model (Figure 1). Suppose we have a vector of covariates $\mathbf{x}_i = [x_{i1}, \dots, x_{ip}]^T$ for the i^{th} observation which may influence the prevalence of the outcome latent class $P(W = w \mid U = u)$. Then the prevalence of latent class can be modeled with the multinomial logistic regression by substituting

$\delta_{w|u}$ given in (2.3) into $\delta_{w|u}(\mathbf{x}_i) = \exp(\mathbf{x}_i^T \boldsymbol{\beta}_{w|u}) / \sum_{l=1}^D \exp(\mathbf{x}_i^T \boldsymbol{\beta}_{l|u})$, where the vector of logistic regression coefficients $\boldsymbol{\beta}_{w|u} = [\beta_{1w|u}, \dots, \beta_{pw|u}]^T$ is interpreted as the log-odds ratio that an individual belongs to a specific latent class w versus to a baseline latent class, given the latent group membership u .

3. Parameter estimation and model diagnosis

3.1. Deterministic annealing EM

The LCA with latent group variable is composed of an unobservable latent structure; therefore, the parameter estimation may be regarded as a missing-data problem. The expectation-maximization (EM) is the standard method to estimate the model parameters for LCA and JLCA. However, the EM algorithm is highly influenced by its initial value. Once the inappropriate initial values are given, the final solution of the EM may be deviated from the global maximum and one of the local maxima will be provided. To overcome this problem, a large number of sets of starting values should be tried for the standard EM algorithm, and we may choose the estimates with the highest log-likelihood. This may help to resolve the local maxima problem; however, the time and computational cost would be very expensive. There is also no guarantee that the result with the highest log-likelihood among the candidates is actually the global maximum. To overcome the difficulty in local maxima for the proposed model, we adopt DAEM method (Ueda and Nakano, 1998).

The DAEM is proposed to overcome the local maxima problem by using the principle of maximum entropy. In the DAEM process, a modified posterior is introduced by additional factor ω and the iteration loop for the ω is located after the E-step and M-step in the conventional EM method. The E-step and M-step are repeated until the estimates converged given ω , then the ω increases to apply another run for the EM algorithm with the updated ω . It is repeated until the ω becomes 1, and we may take the final estimates from the iteration with $\omega = 1$ as the global maximum.

- E-step: The DAEM maximizes the modified observed-data log-likelihood which can be defined as

$$F(\omega) = \sum_{i=1}^n F_i(\omega) = \sum_{i=1}^n \frac{1}{\omega} \log \sum_{u=1}^S \sum_{c_1=1}^{K_1} \cdots \sum_{c_J=1}^{K_J} \sum_{w=1}^D (L_i^*)^\omega, \quad (3.1)$$

where L_i^* is the complete-data likelihood derived in (2.3). Note that $F(\omega)$ is identical to the observed-data log-likelihood when ω is equal to 1. To maximize (3.1), we adopt a density function $q(U = u, \mathbf{C} = \mathbf{c}, W = w | \mathbf{y}_i, \mathbf{z}_i)$ to express (3.1) as an expectation form. Then by Jensen's inequality, the lower bound of $F_i(\omega)$ is calculated as

$$\begin{aligned} F_i(\omega) &= \frac{1}{\omega} \log \sum_{u=1}^S \sum_{c_1=1}^{K_1} \cdots \sum_{c_J=1}^{K_J} \sum_{w=1}^D (L_i^*)^\omega \frac{q(U = u, \mathbf{C} = \mathbf{c}, W = w | \mathbf{y}_i, \mathbf{z}_i)}{q(U = u, \mathbf{C} = \mathbf{c}, W = w | \mathbf{y}_i, \mathbf{z}_i)} \\ &\geq \frac{1}{\omega} \sum_{u=1}^S \sum_{c_1=1}^{K_1} \cdots \sum_{c_J=1}^{K_J} \sum_{w=1}^D \log \left\{ \frac{(L_i^*)^\omega}{q(U = u, \mathbf{C} = \mathbf{c}, W = w | \mathbf{y}_i, \mathbf{z}_i)} \right\} q(U = u, \mathbf{C} = \mathbf{c}, W = w | \mathbf{y}_i, \mathbf{z}_i) \\ &= \sum_{u=1}^S \sum_{c_1=1}^{K_1} \cdots \sum_{c_J=1}^{K_J} \sum_{w=1}^D q(U = u, \mathbf{C} = \mathbf{c}, W = w | \mathbf{y}_i, \mathbf{z}_i) \log (L_i^*)^\omega \\ &\quad - \frac{1}{\omega} \sum_{u=1}^S \sum_{c_1=1}^{K_1} \cdots \sum_{c_J=1}^{K_J} \sum_{w=1}^D q(U = u, \mathbf{C} = \mathbf{c}, W = w | \mathbf{y}_i, \mathbf{z}_i) \log q(U = u, \mathbf{C} = \mathbf{c}, W = w | \mathbf{y}_i, \mathbf{z}_i). \end{aligned}$$

To determine the optimal choice of q , we take functional derivatives with respect to q and set it zero under the constraint $\sum_{u=1}^S \sum_{c_1=1}^{K_1} \cdots \sum_{c_J=1}^{K_J} \sum_{w=1}^D q(U = u, C = c, W = w \mid y_i, z_i) = 1$ (Chang and Chung, 2013). The optimal choice for q is then:

$$q(U = u, C = c, W = w \mid y_i, z_i) = \frac{(L_i^*)^\omega}{\sum_{s=1}^S \sum_{c_1=1}^{K_1} \cdots \sum_{c_J=1}^{K_J} \sum_{d=1}^D (L_i^*)^\omega} = \theta_{i(u, c_1, \dots, c_J, w)}(\omega).$$

The optimal choice q may be considered as modified posterior and we can calculate the expectation of the modified complete-data log-likelihood which can be driven from (3.1) as

$$\begin{aligned} E \left\{ \sum_{i=1}^n \frac{1}{\omega} \log (L_i^*)^\omega \right\} &= \sum_{i=1}^n \sum_{u=1}^S \sum_{w=1}^D \theta_{i(u, w)}(\omega) \log (\delta_{w|s}) + \sum_{i=1}^n \sum_{w=1}^D \theta_{i(w)}(\omega) \sum_{p=1}^P \sum_{h=1}^{r_p} I(Z_{ip} = h) \log (\phi_{ph|w}) \\ &+ \sum_{i=1}^n \sum_{u=1}^S \theta_{i(u)}(\omega) \log (\gamma_u) + \sum_{i=1}^n \sum_{j=1}^J \sum_{c_j=1}^{K_j} \sum_{u=1}^S \theta_{i(u, c_j)}(\omega) \log (\eta_{c_j|u}^{(j)}) \\ &+ \sum_{i=1}^n \sum_{j=1}^J \left\{ \sum_{c_j=1}^{K_j} \theta_{i(c_j)}(\omega) \sum_{m_j=1}^{M_j} \sum_{k=1}^{r_{m_j}} I(y_{imj} = k) \log (\rho_{mkj|c_j}) \right\}, \end{aligned} \quad (3.2)$$

where the modified posterior probabilities for the specific dimensions are defined as $\theta_{i(u, w)}(\omega) = \prod_j \sum_{c_j} \theta_{i(u, c_1, \dots, c_J, w)}(\omega)$, $\theta_{i(w)}(\omega) = \sum_u \theta_{i(u, w)}(\omega)$, $\theta_{i(u)}(\omega) = \sum_w \theta_{i(u, w)}(\omega)$, $\theta_{i(u, c_j)}(\omega) = \sum_w \prod_{k \neq j} \sum_{c_k} \theta_{i(u, c_1, \dots, c_J, w)}(\omega)$, and $\theta_{i(c_j)}(\omega) = \sum_{u=1}^S \theta_{i(u, c_j)}(\omega)$.

- M-step: We may obtain the estimators that maximize the expectation given in (3.2) by using the Lagrange multiplier. The updated parameter estimates can be calculated as

$$\begin{aligned} \hat{\gamma}_u &= \frac{\sum_{i=1}^n \theta_{i(u)}(\omega)}{n}, \quad \hat{\delta}_{w|u} = \frac{\sum_{i=1}^n \theta_{i(u, w)}(\omega)}{\sum_{i=1}^n \theta_{i(u)}(\omega)}, \quad \hat{\eta}_{c_j|u}^{(j)} = \frac{\sum_{i=1}^n \theta_{i(u, c_j)}(\omega)}{\sum_{i=1}^n \theta_{i(u)}(\omega)} \\ \hat{\rho}_{mkj|c_j} &= \frac{\sum_{i=1}^n \theta_{i(c_j)}(\omega) I(y_{imj} = k)}{\sum_{i=1}^n \theta_{i(c_j)}(\omega)}, \quad \hat{\phi}_{ph|w} = \frac{\sum_{i=1}^n \theta_{i(w)}(\omega) I(z_{ip} = h)}{\sum_{i=1}^n \theta_{i(w)}(\omega)}. \end{aligned}$$

Starting with the fixed value of ω as 0, parameter estimates are converged through the DAEM process. Next, we increase ω , and apply the same procedure using the converged parameter estimates at the previous step as the initial value. DAEM iterates this procedure until we have $\omega = 1$. In this manner, ω -loop intervenes the conventional E- and M-step, and we may obtain the parameter estimates at the global maximum without significant consideration for the initial value. If the model contains any covariates, the δ part in the M-step is substituted with the β parameters as a coefficients of multinomial logistic regression. In this case, the standard Newton-Raphson method may be applied at the phase of $\omega = 1$. The details of the Hessian matrix and the score function for β elements are provided in Appendices A and B.

3.2. Model diagnosis and selection

It is important for LCA to assess model fit with a balanced judgement that considers objective measures as well as substantive knowledge in order to understand distinctive features and underlying

structure of the data in a simple manner. The chi-square asymptotic assumption for the likelihood ratio test statistic (G^2) generally does not hold since latent class models with different number of classes are not nested each other (Collins and Lanza, 2013). Instead, Akaike information criterion (AIC) and Bayesian information criterion (BIC) are popular criteria to assess the relative model fit among candidate models with different number of classes. The model with smaller AIC (or BIC) is preferred, but interpretability of the latent structure also should be considered when choosing the appropriate number of classes.

It is also important to examine the absolute model fit by calculating the difference between expected and the observed frequencies. Jeong and Lee (2009) suggested the parametric bootstrap testing procedure to obtain the asymptotic distribution of test statistics proposed as a goodness-of-fit statistic for the cumulative logit model. Chung *et al.* (2011) used parametric bootstrap p -value that may be obtained from the empirical distribution of G^2 . This empirical distribution is generated by the parametric bootstrap method, and the bootstrap p -value can be computed from the proportion of the empirical G^2 s larger than the observed G^2 . Using the maximum-likelihood (ML) parameter estimates from the data, the empirical distribution of G^2 can be constructed as: (a) fit the LCA with latent group variables to the data set and calculate the observed G^2 ; (b) generate a bootstrap data set with the ML estimates; (c) fit the LCA with latent group variables to the bootstrap data set generated at step (b), and compute the G^2 . By independently repeating the procedures (b) and (c), the bootstrap samples of G^2 are produced, and the right tail probability of the observed G^2 from step (a) may be regarded as a bootstrap p -value.

Once the number of latent classes for each latent variable is determined, the number of group latent classes (i.e., joint latent classes) may be selected in similar criteria such as smaller AIC (or BIC) and bootstrap p -value. The covariates, then, can be considered with the selected number of latent classes.

4. Simulation study

In this section, we performed two sets of simulation studies to check how fairly the DAEM method works. The first study confirms that the DAEM method is superior to the EM in finding ML solutions at the global maximum. The second study evaluates how properly the DAEM estimates model parameters in LCA with the latent group variable with covariates. We construct confidence intervals based on asymptotic standard errors from Hessian matrix given in Appendix B.

In the first study, we generated one target dataset whose number of observation is 500, and we randomly generated 30 sets of initial values. With these starting values, we independently performed parameter estimation using (a) the conventional EM method and (b) the DAEM method with $\omega = [0.01, 0.1, 0.2, 0.4, 0.61, 0.64, 0.69, 0.71, 0.83, 0.91]$, and compare the values of log-likelihood from each of 30 sets of initial values. The left plot in Figure 2 is the histogram of log-likelihood values from the results by the EM algorithm using 30 sets of initial values. We can notice that three different values appeared, representing that the log-likelihood has two local maxima. Among the 30 trials, only 11 starting values succeeded in converging to the global maximum, and the other 19 trials are trapped in the local maxima. In DAEM, however, all 30 trials successfully provided the global maxima whose log-likelihood was -4162.916 , implying the DAEM is not affected by the initial value choice when finding the global maximum.

The second study examined if DAEM properly operates to provide ML estimates of the proposed LCA model. We generated 200 data sets with a sample size of 500 and calculated ML estimates using the DAEM. The calculated parameter estimates and the standard errors from the Hessian matrix for the one sample were then used to construct a 95% confidence interval that checked if it covered the true value of the parameter or not. These procedures were independently repeated for 200 generated

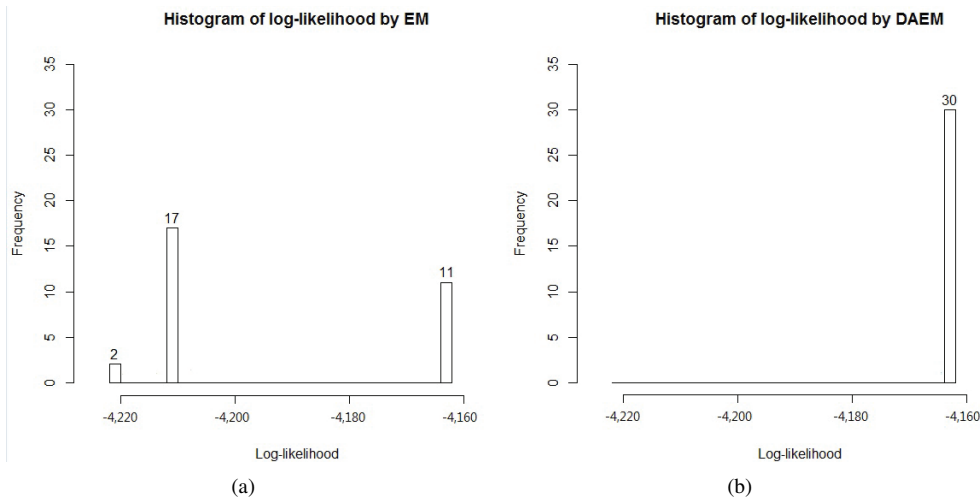


Figure 2: A histogram of the log-likelihoods by EM and DAEM. EM = expectation-maximization; DAEM = deterministic annealing expectation-maximization.

data sets and the coverage of the confidence intervals were subsequently calculated.

The structure of the generated data set is as follows. There are two latent variables which have two latent classes measured by four manifest items, respectively. These two latent variables form a group latent variable whose number of joint class is two. There is also one outcome latent variable which has two latent classes measured by four manifest items. For the measurement parameters, both of strong measurement (Table 1) and mixed measurement (Table 2) were considered. The average estimates from the DAEM are considerably similar with the true values, and the coverage probabilities of the 95% confidence intervals are fairly near the 0.95 both in strong and mixed measurements. This implies that the parameter estimation and model identification work properly.

5. Application to youth survey data

5.1. Data description

The Youth Risk Behavior Surveillance System 2015 (YRBSS 2015) is a biennial survey research about the health risk behavior and drug-using behavior among US adolescents. Among the 15,624 survey participants in the data, we focus on 4,957 high-school male students 16 to 18 years of age.

In this paper, we have 18 self-report items to measure violent and drug-using behavior. Five items were used to measure violent behavior: (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? (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? Cigarette smoking was measured by four manifest items: (1) Have you ever tried cigarette smoking, 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? (4) During the past 30 days, on the days you smoked, how many

Table 1: Average EST, MSE, and CP of 95% confidence intervals for parameter estimates (strong measurement)

Parameter	True	EST	MSE	CP	Parameter	True	EST	MSE	CP
$\rho_{111 1}$	0.10	0.101	0.0003	0.93	γ_1	0.50	0.495	0.0020	0.93
$\rho_{211 1}$	0.10	0.098	0.0003	0.95	$\eta_{11 1}^{(1)}$	0.90	0.903	0.0016	0.96
$\rho_{311 1}$	0.10	0.101	0.0003	0.95	$\eta_{11 2}^{(1)}$	0.10	0.101	0.0018	0.94
$\rho_{411 1}$	0.10	0.100	0.0003	0.96	$\eta_{11 1}^{(2)}$	0.10	0.098	0.0017	0.93
$\rho_{112 1}$	0.90	0.900	0.0003	0.93	$\eta_{11 2}^{(2)}$	0.90	0.901	0.0018	0.93
$\rho_{212 1}$	0.90	0.901	0.0004	0.93	$\beta_{01 1}$	-1.00	-1.055	0.0823	0.96
$\rho_{312 1}$	0.90	0.900	0.0004	0.94	$\beta_{11 1}$	1.00	1.038	0.0511	0.93
$\rho_{412 1}$	0.90	0.900	0.0004	0.95	$\beta_{01 2}$	1.00	1.023	0.0672	0.98
$\rho_{111 2}$	0.90	0.902	0.0004	0.95	$\beta_{11 2}$	-1.00	-1.032	0.0488	0.96
$\rho_{211 2}$	0.90	0.900	0.0003	0.93	$\phi_{11 1}$	0.10	0.099	0.0004	0.97
$\rho_{311 2}$	0.90	0.900	0.0003	0.93	$\phi_{21 1}$	0.10	0.101	0.0004	0.95
$\rho_{411 2}$	0.90	0.899	0.0004	0.94	$\phi_{31 1}$	0.10	0.099	0.0004	0.93
$\rho_{112 2}$	0.10	0.100	0.0003	0.97	$\phi_{41 1}$	0.10	0.101	0.0004	0.93
$\rho_{212 2}$	0.10	0.098	0.0003	0.94	$\phi_{11 2}$	0.90	0.898	0.0003	0.91
$\rho_{312 2}$	0.10	0.099	0.0004	0.95	$\phi_{21 2}$	0.90	0.899	0.0004	0.95
$\rho_{412 2}$	0.10	0.101	0.0003	0.95	$\phi_{31 2}$	0.90	0.899	0.0004	0.94
					$\phi_{41 2}$	0.90	0.902	0.0003	0.94

EST = estimates; MSE = mean square error; CP = coverage probability.

Table 2: Average EST, MSE, and CP of 95% confidence intervals for parameter estimates (mixed measurement)

Parameter	True	EST	MSE	CP	Parameter	True	EST	MSE	CP
$\rho_{111 1}$	0.10	0.097	0.0011	0.91	γ_1	0.50	0.513	0.0133	0.97
$\rho_{211 1}$	0.10	0.098	0.0011	0.95	$\eta_{11 1}^{(1)}$	0.70	0.700	0.0074	0.96
$\rho_{311 1}$	0.30	0.299	0.0009	0.98	$\eta_{11 2}^{(1)}$	0.20	0.187	0.0066	0.97
$\rho_{411 1}$	0.30	0.298	0.0011	0.95	$\eta_{11 1}^{(2)}$	0.30	0.298	0.0076	0.98
$\rho_{112 1}$	0.90	0.901	0.0008	0.96	$\eta_{11 2}^{(2)}$	0.80	0.813	0.0074	0.96
$\rho_{212 1}$	0.90	0.902	0.0006	0.97	$\beta_{01 1}$	-1.00	-1.242	0.7771	0.98
$\rho_{312 1}$	0.70	0.700	0.0008	0.98	$\beta_{11 1}$	1.00	1.212	0.7228	0.98
$\rho_{412 1}$	0.70	0.693	0.0008	0.95	$\beta_{01 2}$	1.00	1.331	0.6397	0.98
$\rho_{111 2}$	0.90	0.899	0.0005	0.98	$\beta_{11 2}$	-1.00	-1.349	0.6534	0.99
$\rho_{211 2}$	0.90	0.898	0.0006	0.97	$\phi_{11 1}$	0.10	0.098	0.0015	0.92
$\rho_{311 2}$	0.70	0.709	0.0006	0.98	$\phi_{21 1}$	0.10	0.087	0.0016	0.95
$\rho_{411 2}$	0.70	0.702	0.0009	0.93	$\phi_{31 1}$	0.30	0.310	0.0007	0.97
$\rho_{112 2}$	0.10	0.103	0.0007	0.95	$\phi_{41 1}$	0.30	0.302	0.0009	0.95
$\rho_{212 2}$	0.10	0.096	0.0008	0.98	$\phi_{11 2}$	0.70	0.691	0.0017	0.93
$\rho_{312 2}$	0.30	0.299	0.0009	0.98	$\phi_{21 2}$	0.70	0.701	0.0007	0.93
$\rho_{412 2}$	0.30	0.300	0.0015	0.96	$\phi_{31 2}$	0.10	0.103	0.0009	0.95
					$\phi_{41 2}$	0.10	0.102	0.0017	0.96

EST = estimates; MSE = mean square error; CP = coverage probability.

Table 3: Percentages of responding ‘yes’ to the manifest items for each latent variable and their missing rates

Latent variable	Manifest item Questionnaires on “Have you...?”	Yes (%)	Missing (%)
<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 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>Cigarette smoking</i>	Ever tried cigarette smoking (<i>Lifetime Smoking</i>)	36.35	9.96
	Smoked before age 13 years for the first time (<i>Early CIG Onset</i>)	8.73	5.85
	Smoked 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>Alcohol consumption</i>	Ever drunken alcohol (<i>Lifetime Drinking</i>)	64.95	3.18
	Drunken alcohol before age 13 years for the first time (<i>Early ALC Onset</i>)	19.42	2.78
	Drunken alcohol during the recent 30 days (<i>Recent Drinking</i>)	33.10	16.76
	Drunken five or more drinks of alcohol in a row (<i>Binge Drinking</i>)	21.44	5.26
<i>Other illegal drug use</i>	Ever tried MJ (<i>Lifetime MJ Use</i>)	45.87	4.17
	Tried MJ before age 13 years for the first time (<i>Early MJ Onset</i>)	10.65	2.54
	Used MJ during the recent 30 days (<i>Recent MJ Use</i>)	26.62	10.53
	Ever tried other illegal drugs (<i>Lifetime Illegal Drug Use</i>)	15.29	5.41
	Ever sold or offered illegal drug on school property (<i>Sell Drug on School</i>)	25.47	3.77

MJ = marijuana.

during the past 12 months, did you ever try to quit smoking cigarettes did you smoke per day? Four items on the alcohol consumption 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? (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? Finally, five items on the other illegal drug-using behavior were: (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) Have you ever tried one of these illegal drugs: cocaine, sniffed solvents, heroin, methamphetamines, or ecstasy? (5) During the past 12 months, has anyone offered, sold, or given you an illegal drug on school property?

Among the original 18 manifest items, the quantitative variables are changed into binomial items: the variables about number of days (or times) are categorized as 1 if one day or more (or one time or more), and 0 otherwise, changing the responses into binary patterns (i.e., whether an individual has experience in something or not). The items on the age of first use are categorized as 1 under 13 years old, and 0 otherwise, indicating whether or not early exposure. Table 3 shows the proportions of responding ‘yes’ to the manifest items and the missing rates.

Using these 18 items along with their age as a covariate, we inspect the effect of drug-using behavior towards violent behavior, by answering following questions: (a) What kinds of latent classes may be found for each drug use and violent behavior? (b) What kinds of common joint patterns can be identified for cigarette, alcohol, and other illegal drug use behavior? (c) How does the prevalence of violent behavior change as the joint latent class membership of drug use varies?

5.2. Model selection

To construct the LCA-MLG model we need to determine the number of latent classes for each latent variable. Firstly, we perform four LCAs to select the number of classes for the respective latent

Table 4: Goodness-of-fit measures for a series of latent class analysis models with the different number of classes for four different class variables

Latent variable	Number of classes	AIC	BIC	Bootstrap p -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>Cigarette smoking</i>	2	11195.1	11253.6	0.00
	3	11058.8	11150.0	0.34
<i>Alcohol consumption</i>	2	16826.8	16885.4	0.00
	3	16607.9	16699.0	0.41
<i>Other illegal 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

AIC = Akaike information criterion; BIC = Bayesian information criterion.

variables (i.e., violent behavior, cigarette smoking, alcohol consumption, and other illegal drug use) based on each set of manifest items. In this step, covariates are not necessary to be included due to the marginalization property (Bandein-Roche *et al.*, 1997). We adopt AIC, BIC, and bootstrap p -value with significance level $\alpha = 0.05$ for the goodness-of-fit statistic: the model with smaller AIC (or BIC) and bootstrap p -value larger than 0.05 were preferred. We also assess the interpretability of identified latent classes. For the bootstrap p -value, we generate 100 bootstrap samples of G^2 -statistic, and compute the proportion of the bootstrap samples of G^2 that are larger than the observed G^2 .

Table 4 shows the goodness-of-fit statistics with the different number of classes for each latent variable. Note that only 2- and 3-class models are allowed to be fitted for *Cigarette smoking* and *Alcohol consumption* because 4-class model has $4 \times 4 + (4 - 1) = 19$ parameters to be estimated with four binary items having $2^4 - 1 = 15$ degrees of freedom only. Table 4 shows that all drug-using behavior (i.e., *Cigarette smoking*, *Alcohol consumption*, and *Other illegal drug use*) can be adequately summarized by the 3-class model. For *Violent behavior*, the 3-class model shows the best fit in terms of BIC; however, the 4-class model shows the best in terms of AIC. In the 4-class model, however, the meaning of the identified classes are not distinct: the ρ -parameter estimates of the two classes are very similar, implying that these two classes should have a similar label. Both 3-class and 4-class models should therefore be considered as candidates for constructing the LCA-MLG model.

Given the number of latent classes for each of drug-using behavior, we determine the number of joint latent classes in the LCA-MLG model. Table 5 lists a series of LCA-MLG models fitted with various number of joint latent classes for drug-using behavior and their goodness of fit measures. The 3-class model and the 4-class model were both considered since the number of latent classes for *Violent behavior* were inconclusive in the previous step. Table 5 shows that the LCA-MLG model with four outcome classes and three or four joint classes can be considered as the model with the best fit. However, we decide to select the model with three outcome classes because the interpretational difference between the two outcome classes are still obscure. Among the models with three outcome classes, the LCA-MLG with three joint classes shows the smallest BIC and the model with four joint classes shows the smallest AIC. We select the LCA-MLG with three outcome classes and three joint classes as a final model because both show an acceptable bootstrap p -value. During these procedures, parameter estimation was conducted using the DAEM method, and the Hessian matrix of the model was investigated to confirm that negative Hessian is positive definite. Any boundary solution with smaller than 0.001 that makes the Hessian singular was constrained as 0 (or 1 if it is larger than

Table 5: Goodness-of-fit measures for a series of LCA-MLG models with the different number of joint classes of drug-using behavior (i.e., *Cigarette smoking*, *Alcohol consumption*, and *Other illegal drug use*) and outcome latent classes of *Violent behavior*

Number of classes for <i>Violent behavior</i>	Number of joint classes for drug-using behavior	AIC	BIC	Bootstrap p -value
3	2	78725.8	79187.9	0.01
	3	78590.1	79110.8	0.06
	4	78555.8	79135.0	0.24
	5	78547.7	79185.5	0.22
	6	78611.0	79307.4	0.33
4	2	78661.5	79169.2	0.04
	3	77987.4	78560.1	0.16
	4	77911.2	78614.1	0.16
	5	77979.7	78617.5	0.32
	6	163880.7	164648.7	0.34

LCA-MLG = latent class analysis with multiple latent groups; AIC = Akaike information criterion; BIC = Bayesian information criterion.

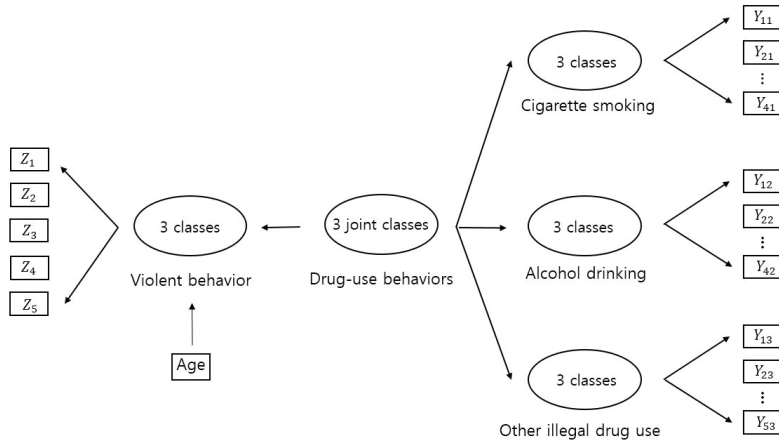


Figure 3: A structure of the selected latent class analysis with the multiple latent groups model.

0.999). Finally, age was considered as a covariate with the selected LCA-MLG illustrated in Figure 3.

5.3. Parameter estimates for multiple latent group variables

The LCA-MLG model considered drug-using behavior as a group variable that may affect the outcome latent variable of *Violent behavior*. There were three latent variables that measured drug-using behavior: *Cigarette smoking*, *Alcohol consumption*, and *Other illegal drug use*.

Under the selected model structure, the primary measurement parameter estimates for drug-using behavior (i.e., ρ -parameters); the prevalences of the latent classes are given in Table 6. For *Cigarette smoking*, class 1 indicates ‘non-smoker’ group, having all ρ -parameters with very small value. Class 2 shows high probability for *Lifetime Smoking* only, so it is labeled as ‘lifetime smoker’ group. Class 3 shows the high probability for *Lifetime Smoking*, *Early CIG Onset*, and *Recent Smoking*, implying that this group can be named as ‘early onset current smoker.’ For *Alcohol consumption*, class 1 shows small values for all items, so it can be named as ‘non-drinker’ group. Class 2 shows high tendency of responding ‘yes’ to the *Lifetime Drinking* item only, so it may be called ‘lifetime drinker’ group.

Table 6: The estimated probabilities of responding ‘yes’ to the manifest items and class prevalences for each of the latent variables

Manifest item	Latent class for <i>Cigarette smoking</i>		
	Non-smoker	Lifetime smoker	Early onset current smoker
<i>Lifetime Smoking</i>	0.047	1.000*	1.000*
<i>Early CIG Onset</i>	0.000*	0.158	0.535
<i>Recent Smoking</i>	0.000*	0.219	1.000*
<i>Heavy Smoking</i>	0.000*	0.000*	0.200
Class prevalence	0.614	0.298	0.088
Manifest item	Latent class for <i>Alcohol consumption</i>		
	Non-drinker	Lifetime drinker	Binge drinker
<i>Lifetime Drinking</i>	0.160	1.000*	1.000*
<i>Early ALC Onset</i>	0.036	0.233	0.382
<i>Recent Drinking</i>	0.000*	0.306	1.000*
<i>Binge Drinking</i>	0.000*	0.000*	0.886
Class prevalence	0.378	0.352	0.270
Manifest item	Latent class for <i>Other illegal drug use</i>		
	Non-user	Current marijuana user	Early and multiple drug user
<i>Lifetime MJ Use</i>	0.038	1.000*	1.000*
<i>Early MJ Onset</i>	0.000*	0.124	0.497
<i>Recent MJ Use</i>	0.000*	0.497	0.835
<i>Lifetime Illegal Drug Use</i>	0.032	0.199	0.803
<i>Drugs Offered at School</i>	0.165	0.306	0.577
Class prevalence	0.539	0.314	0.147

MJ = marijuana. * The probabilities are constrained to be zero or one.

High probabilities for all items except for *Early ALC Onset* in class 3 implies that this group can be represented as ‘binge drinker’ group. In the similar manner, three classes of *Other illegal drug use* can be named as ‘non-user,’ ‘current marijuana user,’ and ‘early and multiple drug user,’ respectively. The ρ -parameters in the first class are all close to zero, and second latent class has high probability for *Lifetime MJ Use* and *Recent MJ Use*. The third latent class has high ρ -parameters for *Lifetime MJ Use*, *Recent MJ Use*, *Lifetime Illegal Drug Use* and *Drugs Offered at School*, showing an early onset behavior on marijuana, and lifetime-in school exposure of illegal drugs.

The estimated class prevalence for drug-using behavior may be estimated as

$$\begin{aligned}
 \hat{P}(C_j = c_j) &= \sum_{u=1}^S \hat{P}(C_j = c_j | U = u) \hat{P}(U = u) \\
 &= \sum_{u=1}^S \hat{\eta}_{c_j|u}^{(j)} \hat{\gamma}_u,
 \end{aligned}$$

for $c_j = 1, \dots, K_j$ and $j = 1, \dots, J$. For example, Table 6 shows that the largest class in Alcohol drinking is non-drinker with prevalence 0.378. Three latent variables (i.e., *Cigarette smoking*, *Alcohol drinking* and *Other illegal drug use*), identified by their respective manifest items, form a joint latent variable which then becomes a group variable.

Table 7 shows the secondary measurement parameters (i.e., η -parameters) for measuring common patterns of three latent variables associated with drug-using behavior. As introduced in Section 2, η -parameter represents the conditional probability of being in the specific class for each of the latent variables given a joint class membership. In the first joint class, individuals tend to be ‘non-smoker’

Table 7: The estimated probabilities of belonging to a latent class for a given joint class membership and the prevalence of joint classes

Latent class		Joint latent class for drug-using behavior		
		Non drug user	Current drug user	Multiple drug user
<i>Cigarette smoking</i>	Non-smoker	0.988	0.401	0.079
	Lifetime smoker	0.005	0.598	0.341
	Early onset current smoker	0.007	0.001	0.580
<i>Alcohol consumption</i>	Non-drinker	0.760	0.089	0.034
	Lifetime drinker	0.220	0.597	0.059
	Binge drinker	0.020	0.314	0.907
<i>Other illegal drug use</i>	Non-user	0.964	0.259	0.038
	Current marijuana user	0.036	0.720	0.009
	Early and multiple drug user	0.000*	0.021	0.953
Joint class prevalence		0.443	0.412	0.145

* The probabilities are constrained to be zero or one.

for *Cigarette smoking*, ‘non-drinker’ for *Alcohol consumption*, and ‘non-user’ for *Other illegal drug use*. This implies that the first joint class represents the subgroup that does not have tendency to use any drug and can therefore be named as ‘non drug user.’ In the second joint class, individuals shows 59.8% probability of ‘lifetime smoker,’ 59.7% probability of ‘lifetime drinker,’ and 72% probability of ‘current marijuana user.’

As a result, the second joint latent class may be interpreted as ‘marijuana user with cigarette and alcohol onset’ group. For the third joint latent class, individuals show probabilities of 0.580 for ‘early onset current smoker,’ 0.907 for ‘binge drinker,’ and 0.953 for ‘early and multiple drug user.’ Therefore, the third joint latent class can be labeled as ‘multiple drug user’ group. The last row in Table 7 indicates γ estimates, representing the prevalence of the joint latent classes. The most common joint class is ‘not drug user’ group with a prevalence rate of 0.443, while ‘multiple drug user’ has the smallest prevalence rate 0.145.

5.4. Parameter estimates for outcome latent variable

Under the selected LCA-MLG model, the outcome latent variable is *Violent behavior*. The primary measurement parameter estimates for the outcome latent variable (i.e., ϕ -parameters) and the prevalence of the identified latent class is given in Table 8. We can see that class 1 has small ϕ -estimates for all five binary items, implying ‘not violent’ group. Class 2 has high probabilities for *Carry Weapon* and *Physical Fight* items, so it may be named as ‘weapon carry and fight’ group. Class 3, which has high probabilities for *Carry Weapon*, *Being Threatened* and *Physical Fight*, may be labelled as ‘seriously violent’ group. The estimated prevalences of latent classes for *Violent behavior* are obtained by

$$\begin{aligned}
 \hat{P}(W = w) &= \frac{1}{n} \sum_{i=1}^n \sum_{u=1}^S \hat{P}(W = w \mid U = u, \mathbf{x}_i) \hat{P}(U = u) \\
 &= \frac{1}{n} \sum_{i=1}^n \sum_{u=1}^S \hat{\delta}_{w|u}(\mathbf{x}_i) \hat{\gamma}_u,
 \end{aligned}$$

for $w = 1, \dots, D$. In Table 8, we can see that ‘not violent’ group is the most prevalent with 0.762. The second largest class is ‘weapon carry and fight’ group, whose class prevalence is 0.185. The prevalence of the ‘seriously violent’ group is 0.053.

Table 8: The estimated probabilities of responding ‘yes’ to the manifest items and class prevalences for each of latent variables

Manifest item	Latent class for <i>Violent behavior</i>		
	Not violent	Weapon carry and fight	Seriously violent
<i>Carry Weapon</i>	0.177	0.501	0.797
<i>Feeling Unsafe</i>	0.018	0.066	0.494
<i>Being Threatened</i>	0.021	0.113	0.724
<i>Physical Fight</i>	0.057	1.000*	0.827
<i>Seriously Injured</i>	0.000*	0.117	0.364
Class prevalence	0.762	0.185	0.053

* The probabilities are constrained to be zero or one.

Table 9: The estimated odds ratios of age for *Violent behavior* for a given joint class membership and 95% confidence intervals (‘not-violent’ group is the baseline)

Joint latent class	Latent class for <i>Violent behavior</i>	
	Weapon carry and fight	Seriously violent
Non drug user	0.723 [0.472, 1.106]	NA*
Current drug user	0.690 [0.558, 0.852]	0.815 [0.325, 2.041]
Multiple drug user	0.892 [0.640, 1.241]	0.857 [0.526, 1.396]

* The probability of belonging to ‘seriously violent’ for ‘non drug user’ group is constrained to be zero.

Table 10: The estimated prevalence of *Violent behavior* given a latent membership of group variable (i.e., drug-using behavior)

Group variable	Latent class for <i>Violent behavior</i>		
	Not violent	Weapon carry and fight	Seriously violent
Non drug user	0.943	0.057	0.000*
Current drug user	0.742	0.231	0.027
Multiple drug user	0.267	0.450	0.283

* The probability of belonging to ‘seriously violent’ for ‘non drug user’ group is constrained to be zero.

We consider age as a covariate to investigate its influence on *Violent behavior* with an effect of drug-using behavior. Table 9 illustrates the estimated odds ratios of age for the identified classes of *Violent behavior* and their 95% confidence intervals. As introduced in Section 3, the asymptotic standard errors of the estimates is obtained as the square root of the diagonal terms of the negative Hessian matrix evaluated by the ML estimates. The effects of age on the prevalence of *Violent behavior* are significant only for ‘current drug user’ group.

The estimated prevalences of *Violent behavior* given a latent membership of group variable (i.e., drug-using behavior) are given in Table 10. Note that the values in Table 10 can be obtained by

$$\begin{aligned}
 \hat{P}(W = w \mid U = u) &= \frac{1}{n} \sum_{i=1}^n \hat{\delta}_{w|u}(\mathbf{x}_i) \\
 &= \frac{1}{n} \sum_{i=1}^n \frac{\exp(\mathbf{x}_i' \hat{\boldsymbol{\beta}}_{w|u})}{\sum_{d=1}^D \exp(\mathbf{x}_i' \hat{\boldsymbol{\beta}}_{d|u})}.
 \end{aligned}$$

Table 10 shows that the prevalences of *Violent behavior* are moving from ‘not violent’ to ‘seriously violent’ as drug-using behavior from ‘non drug user’ to ‘multiple drug user.’ We may conclude that as drug-using behavior become more serious, a risk of exposure to violent behavior increases.

6. Conclusion

This paper proposes a new latent variable model to examine the relationship between violent behavior and multiple drug-using behavior among high-school male students. The conventional LCA may be able to deal with a single latent variable. The newly proposed LCA-MLG with covariates can investigate joint effects of several latent variables on the prevalence of outcome latent variable. EM algorithm is widely adopted for the parameter estimation of the incomplete data, but it has several problems of local maxima in the likelihood. We adopted the DAEM algorithm (Ueda and Nakano, 1998) to overcome this problem and to provide with precise ML estimates of latent variable model where the latent structure is quite complex. In addition, the Hessian matrix of the model was calculated to provide asymptotic standard error of the estimates. The analysis of YRBSS 2015 indicates three representative subgroups that show similar patterns in drug-using behavior including cigarette, alcohol and other illegal drugs. These common patterns form a joint latent variable whose joint latent classes can be referred as ‘non drug user,’ ‘current drug user,’ and ‘multiple drug user’ group depending on the extent of experiences or behavior towards various drugs. Similarly, three common subgroups were discovered for the violent behavior of high-school students, as measured by five binary items. The LCA-MLG model enables us to examine how the prevalence of *Violent behavior* are affected with respect to group latent membership and covariate. We found that the probability of belonging to ‘weapon carry and fight’ and ‘seriously violent’ compared to the ‘not violent’ increases as the degree of drug-using behavior becomes serious. For individuals who are in ‘current drug user,’ the probabilities of belonging to the ‘weapon carry and fight’ or ‘seriously violent’ compared to the ‘not violent’ decrease as age increases.

The structure of LCA-MLG tries to explain the association between several latent group variables and one latent outcome variable. However, the associations covered with LCA-MLG model are not exact causalities. Consequently, examining causal inference between several latent variable may be a valuable further research topic. Consequently, we have made a DAEM routine for LCA-MLG written in R language (version 3.3.1) which is available on request.

Appendix A: Elements of the score function

Let Θ be a vector of all free parameters for LCA-MLG. The score function $S(\Theta)$ is obtained by the first-ordered derivatives of the log-likelihood of the LCA-MLG given in (2.4) with respect to the model parameters Θ . Let β be the vectorized β parameters in the LCA-MLG model. The elements of the first-derivative vector with respect to β are given by

$$\sum_{i=1}^n \frac{\partial \log L_i}{\partial \beta_{qw|u}} = \sum_{i=1}^n x_{iq} [\theta_{i(u,w)} - \delta_{w|u}(x_i) \theta_{i(u)}],$$

for $q = 1, \dots, p$, $w = 1, \dots, D-1$, $\beta_{D|u} = 0$, $u = 1, \dots, S$. Likewise, let $\rho_{m_j|c_j} = [\rho_{m_j 1|c_j}, \dots, \rho_{m_j r_{m_j}|c_j}]^T$, $\eta_u^{(j)} = [\eta_{1|u}^{(j)}, \dots, \eta_{K_j|u}^{(j)}]^T$, $\phi_{p|w} = [\phi_{p1|w}, \dots, \phi_{ph|w}]^T$, and $\gamma = [\gamma_1, \dots, \gamma_S]^T$ be the vectorized ρ -, η -, ϕ , and γ -parameters, respectively. The elements of the first-derivative vector with respect to $\rho_{m_j|c_j}$, $\eta_u^{(j)}$, $\phi_{p|w}$, and γ are obtained by

$$\begin{aligned} \sum_{i=1}^n \frac{\partial \log L_i}{\partial \rho_{m_j k|c_j}} &= \sum_{i=1}^n \frac{\theta_{i(c_j)} \zeta_{y_{im_j k}}}{\rho_{m_j k|c_j}}, & \sum_{i=1}^n \frac{\partial \log L_i}{\partial \eta_{c_j|u}^{(j)}} &= \sum_{i=1}^n \frac{\theta_{i(u, c_j)}}{\eta_{c_j|u}^{(j)}}, \\ \sum_{i=1}^n \frac{\partial \log L_i}{\partial \phi_{ph|w}} &= \sum_{i=1}^n \frac{\theta_{i(w)} \zeta_{y_{ip h}}}{\phi_{ph|w}}, & \text{and } \sum_{i=1}^n \frac{\partial \log L_i}{\partial \gamma_u} &= \sum_{i=1}^n \frac{\theta_{i(u)}}{\gamma_u}, \end{aligned}$$

for $m_j = 1, \dots, M_j$, $c_j = 1, \dots, K_j$, $j = 1, \dots, J$, $p = 1, \dots, P$, $d = 1, \dots, D$, and $s = 1, \dots, S$. 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 $\rho_{m_j j | c_j}$ and $\eta_u^{(j)}$, respectively. Therefore, the score function of the free parameters for $\rho_{m_j j | c_j}$ and $\eta_u^{(j)}$ can be obtained by multiplying a constraint matrix A_C on the first-derivative vectors, where A_C is a $(C - 1) \times C$ matrix, composed of an identity matrix in the first $C - 1$ columns and a column vector of -1 in the last column.

Appendix B: 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 parameters Θ . The elements of the second derivative matrix with respect to β are given by.

$$\sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \beta_{q w | u} \partial \beta_{q' w' | u'}} = \sum_{i=1}^n x_{iq} x_{iq'} \{ \zeta_{uu'} [d_{i(u, w)} (\zeta_{w w'} - \delta_{w | u}(x_i)) - d_{i(u, w')} \delta_{w | u}(x_i)] - d_{i(u', w')} d_{i(u, w)} \},$$

where $d_{i(u, w)} = \theta_{i(u, w)} - \delta_{w | u}(x_i) \theta_{i(u)}$ for $q, q' = 1, \dots, l$, $w = 1, \dots, D - 1$, $\alpha_{D | u} = 0$, $u, u' = 1, \dots, S$, and $\zeta_{uu'} = 1$ if $u = u'$, 0 otherwise. The elements of the second derivative matrices with respect to β and $\rho_{m_j j | c_j}$, $\eta_u^{(j)}$, γ_u and $\phi_{p | w}$ are obtained by

$$\begin{aligned} \sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \rho_{m_j k | c_j} \partial \beta_{q w | u}} &= \sum_{i=1}^n \frac{x_{iq} \{ \theta_{i(u, c_j, w)} - \theta_{i(u, w)} \theta_{i(c_j)} - \delta_{w | u}(x_i) [\theta_{i(u, c_j)} - \theta_{i(u)} \theta_{i(c_j)}] \} \zeta_{y_{im_j j} k}}{\rho_{m_j k | c_j}}, \\ \sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \eta_{c_j | u}^{(j)} \partial \beta_{q w | u'}} &= \sum_{i=1}^n \frac{x_{iq} \{ \zeta_{uu'} \theta_{i(u, w, c_j)} - \theta_{i(u', w)} \theta_{i(u, c_j)} - \delta_{w | u'}(x_i) (\zeta_{uu'} - \theta_{i(u')}) \theta_{i(u, c_j)} \}}{\partial \eta_{c_j | u}^{(j)}}, \\ \sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \gamma_u \partial \beta_{q w | u'}} &= \sum_{i=1}^n \frac{x_{iq} \{ (\zeta_{uu'} - \theta_{i(u)}) \theta_{(u', w)} - \delta_{w | u'}(x_i) (\zeta_{uu'} - \theta_{i(u')}) \theta_{i(u)} \}}{\partial \gamma_u}, \\ \sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \phi_{p h | w} \partial \beta_{q w' | u}} &= \sum_{i=1}^n \frac{x_{iq} \{ \theta_{i(u, w')} (\zeta_{w w'} - \theta_{i(w)}) - \delta_{w' | u}(x_i) [\theta_{i(w, w)} - \theta_{i(u)} \theta_{i(w)}] \} \zeta_{z_{p h}}}{\phi_{p h | w}}, \end{aligned}$$

for $m_j = 1, \dots, M_j$, $k = 1, \dots, r_{m_j}$, $c_j = 1, \dots, K_j$, $w = 1, \dots, D - 1$, $p = 1, \dots, r_p$, $j = 1, \dots, J$, and $s = 1, \dots, S$. Here, $\zeta_{y_{im_j j} k}$ is an indicator function which has the value of 1 if $y_{im_j j} = k$, 0 otherwise. Note that there are $S - 1$, $r_p - 1$, $r_{m_j} - 1$, and $K_j - 1$ free parameters in γ , $\phi_{p | w}$, $\rho_{m_j j | c_j}$, and $\eta_u^{(j)}$, respectively. Therefore, the Hessian elements of the free parameters for γ , $\phi_{p | w}$, $\rho_{m_j j | c_j}$, and $\eta_u^{(j)}$ can be obtained by multiplying constraint matrix A_C on the second-derivative matrices. The dimension of the constraint matrix A_C is determined by the number of free parameters for each $\rho_{m_j j | c_j}$, $\eta_u^{(j)}$, γ_u and $\phi_{p | w}$, respectively.

The elements of the second derivative matrices with respect to ρ are:

$$\begin{aligned} \sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \rho_{m_j k | c_j} \partial \rho_{m'_j k' | c'_j}} &= \sum_{i=1}^n \frac{[\theta_{i(c_j, c'_j)} (1 - \zeta_{j' j}) + \theta_{i(c_j)} \{ \zeta_{j' j} [(1 - \zeta_{c c'}) + (1 - \zeta_{m m'})] - \theta_{i(c'_j)} \} \zeta_{y_{im_j j} k} \zeta_{y_{im'_j j'} k'}]}{\rho_{m_j k | c_j} \rho_{m'_j k' | c'_j}}, \\ \sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \eta_{c_j | u}^{(j')} \partial \rho_{m_j k | c'_j}} &= \sum_{i=1}^n \frac{(1 - \zeta_{j' j}) \theta_{i(u, c'_j, c'_j)} + \theta_{i(u, c'_j)} (\zeta_{c'_j c'_j} - \theta_{i(c'_j)}) \zeta_{y_{im_j j} k}}{\eta_{c_j | u}^{(j')} \rho_{m_j k | c'_j}}, \end{aligned}$$

$$\sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \gamma_u \partial \rho_{m_j k j | c_j}} = \sum_{i=1}^n \frac{(\theta_{i(u, c_j)} - \theta_{i(c_j)} \theta_{i(w)}) \zeta_{y_{im_j j} k}}{\gamma_u \rho_{m_j k j | c_j}},$$

$$\sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \phi_{ph|w} \partial \rho_{m_j k j | c_j}} = \sum_{i=1}^n \frac{(\theta_{i(c_j, w)} - \theta_{i(c_j)} \theta_{i(w)}) \zeta_{z_{ip} h} \zeta_{y_{im_j j} k}}{\phi_{ph|w} \rho_{m_j k j | c_j}},$$

for $w = 1, \dots, D$, $p = 1, \dots, P$, $h = 1, \dots, r_p$, $m_j = 1, \dots, M_j$, $k, k' = 1, \dots, r_{m_j}$, $c_j = 1, \dots, K_j$, $u, u' = 1, \dots, S$, and $j, j' = 1, \dots, J$. Here, $\zeta_{y_{im_j j} k}$ is an indicator function which has the value of 1 if $y_{im_j j} = k$, 0 otherwise. Note that there are $r_{m_j} - 1$ and $K_j - 1$ free parameters in $\rho_{m_j j | c_j}$ and $\eta_u^{(j')}$, respectively. Therefore, the Hessian elements of the free parameters for $\rho_{m_j j | c_j}$ and $\eta_u^{(j')}$ can be obtained by multiplying constraint matrices A_C and A_C^T on the second-derivative matrices. As discussed before, the dimension of the constraint matrices are determined by the number of free parameters for each $\rho_{m_j j | c_j}$, $\eta_u^{(j)}$, γ_u and $\phi_{p|w}$, respectively.

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

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

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

$$\sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \phi_{ph|w} \partial \eta_{c_j | u}^{(j)}} = \sum_{i=1}^n \frac{(\theta_{i(u, c_j, w)} - \theta_{i(u, c_j)} \theta_{i(w)}) \zeta_{z_{ip} h}}{\phi_{ph|w} \eta_{c_j | u}^{(j)}},$$

for $c_j = 1, \dots, K_j$, $j = 1, \dots, J$, $w = 1, \dots, D$ and $u, u' = 1, \dots, S$. There are $S - 1$ free parameters in γ , $r_p - 1$ free parameters in $\phi_{p|w}$ and $K_j - 1$ free parameters in $\eta_u^{(j')}$, respectively. Therefore, the Hessian elements of the free parameters for $\eta_u^{(j)}$, $\eta_u^{(j')}$, $\phi_{p|w}$ and γ can be obtained by multiplying constraint matrices A_C and A_C^T on the second-derivative matrices.

The elements of the second derivative vector with respect to γ are:

$$\sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \gamma_u \partial \gamma_{u'}} = - \sum_{i=1}^n \frac{\theta_{i(u)} \theta_{i(u')}}{\gamma_u \gamma_{u'}},$$

$$\sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \phi_{ph|w} \partial \gamma_u} = \sum_{i=1}^n \frac{(\theta_{i(u, w)} - \theta_{i(w)} \theta_{i(w')}) \zeta_{z_{ip} h}}{\phi_{ph|w} \gamma_u},$$

for $j, j' = 1, \dots, J$ and $u, u' = 1, \dots, S$. There are $S - 1$ free parameters in γ and $r_p - 1$ free parameters in $\phi_{p|w}$, respectively. Therefore, the Hessian elements of the free parameters for γ , $\phi_{p|w}$ and γ can be obtained by multiplying constraint matrices A_C and A_C^T on the second-derivative matrices.

The elements of the second derivative matrices with respect to ϕ are:

$$\sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \phi_{ph|w} \partial \phi_{p' h' | w'}} = \sum_{i=1}^n \frac{\theta_{i(w)} \{ (1 - \zeta_{ww'}) + (1 - \zeta_{pp'}) - \theta_{i(w')} \} \zeta_{z_{ip} h} \zeta_{z_{ip'} h'}}{\phi_{ph|w} \rho_{p' h' | w'}},$$

for $p = 1, \dots, P$, $k = 1, \dots, r_p$, $c_j = 1, \dots, K_j$. Here, $\zeta_{z_{ip} h}$ is an indicator function which has the value of 1 if $z_{ip} = h$, otherwise 0. Note that there are $r_p - 1$ and $r_{p'} - 1$ free parameters in $\phi_{p|w}$ and

$\phi_{p'|w'}$, respectively. Therefore, the Hessian elements of the free parameters for $\phi_{ph|w}$ and $\phi_{p'h'|w'}$ can be obtained by multiplying constraint matrices A_C and A_C^T on the second-derivative matrices.

Acknowledgement

This work was supported by a Korea University Grant (K1509141 to Hwan Chung) and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2015R1D1A1A01056846 to Hwan Chung).

References

- Bandeen-Roche K, Miglioretti DL, Zeger SL, and Rathouz PJ (1997). Latent variable regression for multiple discrete outcomes, *Journal of the American Statistical Association*, **92**, 1375–1386.
- Centers for Disease Control and Prevention (2015). Youth Risk Behavior Surveillance System, Available at <http://www.cdc.gov/YRBSS>
- Chang HC and Chung H (2013). Dealing with multiple local modalities in latent class profile analysis, *Computational Statistics and Data Analysis*, **68**, 296–310.
- Chung H, Anthony JC, and 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**, 689–712.
- Collins LM and Lanza ST (2013). *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*, John Wiley & Sons, Hoboken, NY.
- Dawkins MP (1997). Drug use and violent crime among adolescents, *Adolescence*, **32**, 395–405.
- DuRant RH, Altman D, Wolfson M, Barkin S, Kreiter S, and Krowchuk D (2000). Exposure to violence and victimization, depression, substance use, and the use of violence by young adolescents, *The Journal of Pediatrics*, **137**, 707–713.
- Jeong KM and Lee HY (2009). Goodness-of-fit tests for the ordinal response models with misspecified links, *Communications for Statistical Applications and Methods*, **16**, 697–705.
- Lowry R, Cohen LR, Modzeleski W, Kann L, Collins JL, and 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**, 347–355.
- Lundholm L (2013). Substance use and violence: influence of alcohol, illicit drugs and anabolic androgenic steroids on violent crime and self-directed violence (Doctoral dissertation), Acta Universitatis Upsaliensis, Uppsala.
- Miller CL, Kerr T, Strathdee SA, Li K, and Wood E (2007). Factors associated with premature mortality among young injection drug users in Vancouver, *Harm Reduction Journal*, **4**, 1–7. <https://doi.org/10.1186/1477-7517-4-1>
- Ueda N and Nakano R (1998). Deterministic annealing EM algorithm, *Neural Networks*, **11**, 271–282.
- Van Horn ML, Fagan AA, Hawkins JD, and Oesterle S (2014). Effects of the communities that care system on cross-sectional profiles of adolescent substance use and delinquency, *American Journal of Preventive Medicine*, **47**, 188–197.
- White HR, Loeber R, Stouthamer-Loeber M, and Farrington DP (1999). Developmental associations between substance use and violence, *Development and Psychopathology*, **11**, 785–803.