Published in final edited form as:
Struct Equ Modeling. 2013 January ; 20(1): 1–26. doi:10.1080/10705511.2013.742377.

Latent Class Analysis With Distal Outcomes: A Flexible Model-Based Approach

Stephanie T. Lanza[1,2], **Xianming Tan**[1], and **Bethany C. Bray**[3]

[1]The Methodology Center, The Pennsylvania State University

[2]College of Health and Human Development, The Pennsylvania State University

[3]Department of Psychology, Virginia Polytechnic Institute and State University

Abstract

Although prediction of class membership from observed variables in latent class analysis is well understood, predicting an observed distal outcome from latent class membership is more complicated. A flexible model-based approach is proposed to empirically derive and summarize the class-dependent density functions of distal outcomes with categorical, continuous, or count distributions. A Monte Carlo simulation study is conducted to compare the performance of the new technique to two commonly used classify-analyze techniques: maximum-probability assignment and multiple pseudo-class draws. Simulation results show that the model-based approach produces substantially less biased estimates of the effect compared to either classify-analyze technique, particularly when the association between the latent class variable and the distal outcome is strong. In addition, we show that only the model-based approach is consistent. The approach is demonstrated empirically: latent classes of adolescent depression are used to predict smoking, grades, and delinquency. SAS syntax for implementing this approach using PROC LCA and a corresponding macro are provided.

Keywords

latent class analysis; distal outcome; finite mixture model; pseudo-class draws

Finite mixture modeling (McLachlan & Peel, 2000), particularly latent class analysis (LCA; Collins & Lanza, 2010), has become a statistical tool that social and behavioral scientists turn to with increasing frequency. Scientific questions that can be addressed with this set of methods are different from, and often complementary to, those that are addressed with more traditional methods such as multiple regression and analysis of variance. Mixture models posit that there are two or more underlying subgroups in a population, and subgroup membership must be inferred from responses to multiple items. In other words, population heterogeneity is explained by the identification of latent classes that are unique from one another, but each class is comprised of individuals who are similar on a set of observed variables.

Correspondence should be addressed to, Stephanie T. Lanza, The Methodology Center, The Pennsylvania State University, 204 E. Calder Way, Suite 400, State College, PA 16801. SLanza@psu.edu; Office: 814-865-7095.

In many empirical studies, interest lies in understanding which characteristics predict latent class membership. For example, does poor parenting predict membership in trajectory classes of criminal offending (Roeder, Lynch, & Nagin, 1999)? Do adolescents' friendship goals predict substance use patterns (Lanza, Patrick, & Maggs, 2010)? When the predictor is observed and the outcome is latent (i.e., predicting latent class membership from an observed covariate), the mathematical model is well understood. LCA with covariates has been described in detail in the literature (see Collins & Lanza, 2010; Lanza, Collins, Lemmon, & Schafer, 2007) and is summarized below. However, scientists are often interested in an effect in the opposite direction, in which the predictor is latent and the outcome is manifest (i.e., predicting a distal outcome from latent class membership). To be more precise, we are interested in the conditional distribution of a distal outcome, *Z*, given a latent class variable, *C*. In this case, the problem is more difficult because the predictor (true subgroup membership) is unknown (see Figure 1; Lanza, Collins, Schafer, & Flaherty, 2005).

Being able to predict a distal outcome from latent class membership will provide etiological information about how the confluence of characteristics and/or behaviors at an initial time point predicts an outcome of interest; potential application abounds. Within public health, examples include predicting alcohol dependence from early substance use behavior and predicting contraction of a sexually transmitted infection from early sexual risk behavior. To date, researchers have typically used classify-analyze strategies in an attempt to approximate the effect of *C* on *Z*. These strategies assign individuals to a latent class in a first analysis step; then class membership is treated as observed and used to predict the distal outcome in a second analysis step (e.g., Clogg, 1995). Examples in the literature include predicting pain outcomes from latent classes defined by barriers to cancer pain management (Roberts & Ward, 2011) and predicting depression from peer victimization latent classes (Nylund, Bellmore, Nishina, & Graham, 2007).

The two most common approaches to LCA with a distal outcome are the maximum-probability assignment rule (Nagin, 2005) and the multiple pseudo-class draws approach (Bandeen-Roche, Miglioretti, Zeger, & Rathouz, 1997; Wang, Brown, & Bandeen-Roche, 2005). Because these two classify-analyze approaches involve assigning (i.e., imputing) latent class membership and conducting the outcome analysis in separate steps, conclusions drawn about the effect of *C* on *Z* may be incorrect for several reasons. First, there is uncertainty related to class membership, which is not taken into account in the maximum-probability assignment rule. Second, and more importantly, all standard classify-analyze approaches impute the latent variable under a model that is not sufficiently general; this may result in attenuated estimates of the relation between *C* and *Z*.

We propose a new model-based approach to LCA with distal outcomes that is flexible in terms of the metric of *Z* and straightforward to implement. After a brief introduction to the latent class model, we describe current classify-analyze approaches to estimating the effect of *C* on *Z*; we then introduce a model-based approach to LCA with a distal outcome, and perform a simulation study to demonstrate its performance relative to classify-analyze approaches; finally, we present an empirical demonstration of the model-based approach to LCA with a distal outcome. This paper has two goals: (1) We will present a new model-

based approach to LCA with distal outcomes; and (2) We will present SAS syntax for implementing this model-based approach, which relies on an add-on procedure and macro that are freely available.

A Brief Review of the Latent Class Model

The latent class model, which is described in detail by Collins and Lanza (2010) and Lanza et al. (2007), can be summarized as follows. Suppose that there are K latent subgroups that must be inferred from $j = 1, \ldots, J$ observed variables, and that variable j has $r_j = 1, \ldots, R_j$ response categories. Let $x = (r_1, \ldots, r_J)$ represent the vector of a particular subject's responses to the J variables. Let C represent the latent variable with latent classes $c = 1, \ldots, K$. Finally, $I(x_j = r_j)$ is an indicator function that equals 1 when the response to variable $j = r_j$, and equals 0 otherwise. The probability of observing a particular response pattern is

$$\Pr\{X=x\}=\sum_{c=1}^{K}\gamma_c\prod_{j=1}^{J}\prod_{r_j=1}^{R_j}\rho_{j,r_j|c}^{I(x_j=r_j)}, \quad (1)$$

where γ_c represents the probability of membership in latent class c and $\rho_{j,r_j|c}^{I(x_j=r_j)}$ represents the probability of response r_j to item j given membership in latent class c.

This model can be extended to include covariates (i.e., predictors of latent class membership) using a logistic regression model in which the outcome is a categorical latent variable (see (Bandeen-Roche, Miglioretti, Zeger, & Rathouz, 1997; Collins & Lanza, 2010; Dayton & Macready, 1988)). Suppose that a covariate U is used to predict latent class membership. Then the latent class model can be expressed as

$$\Pr\{X=x|U=u\}=\sum_{c=1}^{K}\gamma_c(u)\prod_{j=1}^{J}\prod_{r_j=1}^{R_j}\rho_{j,r_j|c}^{I(x_j=r_j)}, \quad (2)$$

where $\gamma_c(u) = \Pr\{C = c|U = u\}$ is a standard baseline-category multinomial logistic model (e.g., Agresti, 2002).

With a single covariate U, $\gamma_c(u)$ can be expressed as

$$\gamma_c(u)=\Pr\{C=c|U=u\}=\frac{e^{\beta_{0c}+\beta_{1c}u}}{1+\sum_{c'=1}^{K-1}e^{\beta_{0c'}+\beta_{1c'}u}} \quad (3)$$

for $c' = 1, \ldots, K-1$ and reference class K.

Individuals' posterior probabilities of membership in each latent class can be obtained from the resultant LCA parameters by applying Bayes' theorem (e.g., Gelman, Carlin, Stern, & Rubin, 2003; Lanza et al., 2007):

$$\Pr\{C=c|X=x\}=\frac{\Pr\{C=c\}\Pr\{X=x|C=c\}}{\Pr\{X=x\}}. \quad (4)$$

A model with a particular number of latent classes can be selected using a bootstrap likelihood-ratio test (McLachlan & Peel, 2000; McLachlan, 1987), as well as information criteria such as AIC (Akaike, 1974), BIC (Schwartz, 1978), CAIC (Bozdogan, 1987), and a-BIC (Sclove, 1987). Multiple sets of random starting values should be used to assess the degree of certainty that the global maximum (as opposed to a local maximum) in the likelihood function has been identified. In addition, the ability to interpret the latent classes in a solution can help guide model selection.

Effect sizes in LCA

It is possible to calculate an effect size (Cohen, 1992) indicating the strength of association between a latent class variable *C* and a distal outcome *Z*. The effect size is calculated as follows:

- For a categorical outcome *Z* with *m* categories,

$$\omega = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{K}\frac{(P_{ij} - P_{0ij})^2}{P_{0ij}}},$$

  where $P_{ij} = \Pr\{Z = i, C = j\} = \pi_j \Pr\{Z = i|C = j\}$, $P_{0ij} = \Pr\{Z = i\}$. We note that $\omega = 0$ if and only if $P_{ij} = \Pr\{Z = i, C = j\} = \Pr\{Z = i\}$. That is, $\omega = 0$ if and only if *C* and *Z* are independent.

- For a continuous or count outcome,

$$\omega = \sqrt{\sum_{c=1}^{K}\pi_c(\mu_c - \overline{\mu})^2},$$

  where $\pi_c = \Pr\{C = c\}$, $\mu_c = E(Z|C = c)$, and

  $\overline{\mu} = E(Z) = \sum_{c=1}^{K}\pi_c E(Z|C = c) = \sum_{c=1}^{K}\pi_c\mu_c$.

The actual effect size will vary depending on whether a model-based approach, maximum-probability assignment, or a multiple pseudo-class draws approach is used to estimate the effect. In addition, for a continuous distal outcome, the effect size will depend on whether the mean or mode is used to represent the distribution of the outcome given the latent class.

Common Approaches to Predicting a Distal Outcome From Latent Class Membership

There are two common classify-analyze approaches to estimating the effect of *C* on *Z*. The most straightforward approach is to assign individuals to latent classes based on their maximum posterior probability (see, e.g., Nagin, 2005). Specifically, a latent class model that only includes manifest indicators (i.e., an unconditional latent class model) is fit to the observed *X* variables and each individual's vector of posterior probabilities of membership is retained. Individuals are then assigned to the latent class that corresponds to their

maximum posterior probability. Then, class membership is treated as known (i.e., any uncertainty in each individual's true class membership is ignored) and a subsequent outcome analysis is conducted. All modern LCA programs provide the option to save posterior probabilities; any data management software could then be used to perform the outcome analysis. A significant drawback to this approach is that inference in the outcome analysis may be biased to the extent that there is uncertainty in latent class membership.

The second approach, multiple pseudo-class draws (Bandeen-Roche et al., 1997; Wang et al., 2005), mimics the maximum-probability assignment approach, but accounts for the uncertainty in class membership. As with maximum-probability assignment, the pseudo-class approach requires a first stage of analysis in which an unconditional latent class model is fit to the data, and posterior probabilities are retained. In contrast to maximum-probability assignment, however, with the pseudo-class approach latent class membership is assigned randomly according to each individual's posterior distribution. In a second stage of analysis, the distal outcome is, for example, regressed on latent class membership. This procedure is repeated multiple (typically 20) times in order to account for the uncertainty in latent class membership. Results are then combined across the multiple draws using rules derived for multiple imputation of missing data (Rubin, 1987).

The pseudo-class draws approach is also fairly straightforward to implement, and modern software packages for LCA such as M*plus* (L. K. Muthén & Muthén, 1998–2007) include this technique as an option. Equality of means on the distal outcome across latent classes can be tested (Asparouhov, 2010). Numerous applications of this approach appear in recent literature, including Fried et al. (2009) and Petras and Masyn (2010).

An important limitation of both of these classify-analyze approaches is that any association between C and Z is ignored in the classification/imputation stage. Thus, the effect of C on Z estimated in the second stage will be attenuated. The missing data literature provides relevant insight into the impact of imputing data under a model that is more restrictive than the analysis model (e.g., Collins, Schafer, & Kam, 2001; Schafer, 1997).

A Model-Based Approach to Predict a Distal Outcome from Latent Class Membership

Although there is considerable recent interest focusing on inferring the relationship between a latent class variable, C, and distal outcome, Z (Clark & Muthén, 2009; Petras & Masyn, 2010), we are not aware of literature that presents a precise statement of the underlying assumptions needed for such inference. In this section, we show why an additional assumption is needed to resolve a non-identifiability issue, and describe the proposed model-based approach to estimate $Z|C$.

Let us first restate the problem more precisely. We have multiple observed indicators X, a distal outcome Z, and a latent class variable C. We assume that (X, C) follows an LCA model with a fixed number of classes. Although C is not observable, we wish to estimate the conditional distribution of the distal outcome for each latent class ($Z|C$). However, without certain assumptions regarding the joint distribution of (X, Z, C), the estimation of $Z|C$ is not

possible. In general, the joint distribution of random variables is not identifiable from their marginal distributions alone (Casella & Berger, 1990). This is demonstrated in the following example.

In the contingency table shown below, choosing different $\varepsilon \in (-1/4, 1/4)$ leads to different joint distributions of (X, Z), but identical marginal distributions for X and Z. This implies that there exists an infinite number of joint distributions of (X, Z) which have the same marginal distribution of X and Z.

	$X = 1$	$X = 2$	Z margin
$Z = 1$	$1/4 + \varepsilon$	$1/4 - \varepsilon$	1/2
$Z = 2$	$1/4 - \varepsilon$	$1/4 + \varepsilon$	1/2
X margin	1/2	1/2	

Although the problem we consider in this article, which involves X, C, and Z, is more complicated than this example, the key idea still applies: marginal distributions cannot determine the joint distribution without additional information. That is, we cannot determine the joint distribution of (X, C, Z) given the marginal distributions of (X, C) and (X, Z), without additional assumptions. Hence, we cannot infer $Z|C$ solely based on the distributions of (X, Z) and (X, C).

An Important Assumption: Conditional Independence Between X and Z Given C

In order to be able to estimate the conditional distribution of Z given C, $f(Z|C)$, we propose making the assumption of conditional independence between X and Z given the latent class variable C. That is, we assume that $f(X, Z|C) = f(X|C)f(Z|C)$. Although there might be alternative assumptions which can also resolve the non-identifiability issue, we prefer this conditional independence assumption for its similarity to the local independence assumption underlying most LCA models (Collins & Lanza, 2010).

For completeness, the assumptions underlying the proposed model-based approach to LCA with distal outcomes can be explicitly listed as follows. First, we assume that in addition to the observed response indicator variables X and distal outcome Z, there exists a latent class variable C, and the marginal distribution of the latent class variable C is $\Pr\{C = c\} = \pi_c$ (c = 1, 2, …, K), with $0 < \pi_c < 1$ (c = 1, 2, …, K) and $\sum_{c=1}^{K} \pi_c = 1$. Second, we assume that the conditional distribution of X given C is implied by the fundamental LCA model, defined above. Third, we assume that the conditional distribution of C given Z can be summarized by a logistic regression model:

$$\Pr\{C=c|Z=z\}=\frac{e^{\beta_{0c}+\beta_{1c}z}}{1+\sum_{c'=1}^{K-1}e^{\beta_{0c'}+\beta_{1c'}z}}.$$

Assuming a logistic regression model for predicting *C* from a covariate is quite reasonable, and is standard practice in the LCA literature (e.g., Vermunt & Magidson, 2005).

Modeling the Latent Class Variable and the Effect of C on Z Simultaneously

In LCA with a distal outcome, interest lies in the density $f\{Z = z|C = c\}$. We can determine the desired distribution of $Z|C$ by applying Bayes' Theorem:

$$f\{z=z|C=c\}=\frac{f\{Z=z\}\times f\{C=c|Z=z\}}{f\{C=c\}}$$

Given the assumptions above, $f\{C = c\}$ is determined by the LCA model and $f\{C = c|Z = z\}$ is determined by the LCA model with *Z* included as a covariate. The final piece of necessary information, $f\{Z = z\}$, depends on the distribution of *Z*. We first present the case in which *Z* is a binary distal outcome, and discuss extending this approach to categorical outcomes with more than two categories and to count outcomes; we then present an approach for estimating the conditional distribution of a continuous *Z*. No assumption about the particular distributional form of *Z*, such as Gaussian, is required.

Prediction of a binary/categorical/count distal outcome—When *Z* is binary, including *Z* as an additional indicator in the LCA model, including *Z* as a grouping variable in the LCA model, and incorporating *Z* into the LCA model as a covariate are mathematically equivalent. All of these approaches require the assumption of conditional independence between *X* and *Z* given *C* (Roeder et al., 1999). We recommend the third approach of incorporating *Z* as a covariate because it can be readily extended to other types of distal outcomes without requiring distributional assumptions of *Z*. Then, the density of concern, $f\{Z = z|C = c\}$, can be expressed as

$$\Pr\{Z=z|C=c\}=\frac{\Pr\{Z=z\}e^{\beta_{0c}+\beta_{1c}z}}{\Pr\{C=c\}(1+\sum_{c'=1}^{K-1}e^{\beta_{0c'}+\beta_{1c'}z})}.$$

Using this approach, $\Pr\{Z = z\}$ is estimated from the empirical distribution of *Z* (i.e., from the proportions in the observed data); the estimates for $\{\beta_{0c}, \beta_{1c}; c = 1, 2, \ldots, K - 1\}$ are provided by the LCA with covariates model; and the marginal distribution $\Pr\{C = c\}$ can be obtained by multiplying $\Pr\{C = c \mid Z = z\}$ by the marginal distribution $\Pr\{Z = z\}$. Thus, we can estimate $\Pr\{Z = z|C = c\}$ given these estimates for $\Pr\{Z = z\}$, $\Pr\{C = c\}$, and $\{\beta_{0c}, \beta_{1c}; c = 1, 2, \ldots, K - 1\}$.

The above arguments can be extended to a categorical outcome with more than two categories (i.e., $Z \in \{1, 2, 3\ldots, m\}$ and $m \geq 2$), if we assume that

$$\Pr\{C=c|Z=i\}=\frac{e^{\beta_{0c}+\beta_{1ci}}}{1+\sum_{c'=1}^{K-1}e^{\beta_{0c'}+\beta_{1c'i}}}, \text{for} i=2,3,\ldots,m.$$

An Excel calculator has recently been published online (Lanza & Rhoades, 2011b) so that analysts can implement this approach to LCA with a binary distal outcome in their work. The calculator uses as inputs the logistic regression coefficients (β_{1c}) and the known marginal probabilities of the binary distal outcome; the calculator then provides the probabilities of Z given C. This approach is demonstrated in the corresponding article by Lanza and Rhoades (2011a).

The model for LCA with a binary distal outcome can also be extended to a count type outcome with more than two categories (i.e., $Z \in \{0, 1, 2, 3, \ldots\}$), if we assume that

$$\Pr\{Z=z|C=c\}=\frac{\Pr\{Z=z\}e^{\beta_{0c}+\beta_{1c}z}}{\Pr\{C=c\}(1+\sum_{c'=1}^{K-1}e^{\beta_{0c'}+\beta_{1c'}z})}, z=0,1,2,\ldots.$$

In this approach, $\Pr\{Z = z\}$ is also estimated from the empirical distribution of Z, instead of assuming a certain conditional distribution for $Z|C$, such as a conditional Poisson distribution $Z|(C = c) \sim \text{Poisson}(\lambda_c)$.

Prediction of a continuous distal outcome—Obtaining the distribution of a continuous distal outcome given C is a more complicated case than that of a categorical Z. We propose extending the approach described above for a binary/count distal outcome to continuous outcomes. That is, below we explain how to include the continuous distal outcome, Z, as a covariate in the latent class model. Similar to the binary/count case, using this approach we are able to obtain estimates for $\{\beta_{0c}, \beta_{1c}; c = 1, 2, \ldots, K - 1\}$ from the LCA with covariates model. Then, to estimate $f\{Z = z|C = c\}$ we need to estimate $f\{Z = z\}$, and the marginal distribution $\Pr\{C = c\}$ can be obtained by multiplying $\Pr\{C = c \mid Z = z\}$ by the marginal distribution $f\{Z = z\}$. We estimate the density of Z using kernel density estimates (Silverman, 1986) for continuous variables, which can be readily implemented using SAS PROC KDE (SAS Institute Inc., 2002–2004). The default bandwidth selection method in PROC KDE is based on the plug-in formula of Sheather and Jones, as suggested in Jones, Marron, and Sheather (1996). In sum, we propose a flexible, semi-parametric approach for modeling the effect of C on a continuous Z, in which we empirically estimate the distribution of Z. Using the conditional and marginal distributions we can obtain the mean (or mode) of Z for each latent class. Again, this approach does not require a specification of the conditional distribution of Z given C, such as a conditional normal distribution $Z|(C = c) \sim N(\mu_c, \sigma^2)$; instead, it uses the empirical distribution of Z.

Software—LCA, as well as the proposed model-based approach to LCA with a distal outcome, can be conducted in SAS. Syntax for conducting LCA with a distal outcome is included in the Appendix. The SAS procedure for conducting latent class analysis, PROC LCA (Lanza, Dziak, Huang, Xu, & Collins, 2011), and the new %LCA_distal macro, are available for download at methodology.psu.edu.

Summary—Using Bayes' theorem as the foundation, the proposed model-based approach is a general procedure for estimating the conditional distribution of Z given C, regardless of whether the distal outcome is a categorical, count, or continuous variable. The approach

empirically derives the class-specific distribution of *Z* using observed proportions or kernel density estimation, along with information provided by the latent class model. This approach also yields the information needed to test for the significance of the association between *C* and *Z*. Because this approach includes *Z* as a covariate, the difference in the log-likelihood between the latent class models with and without *Z* can be used to construct this test. This log-likelihood difference times −2 can be compared to a chi-square table with degrees of freedom equal to the number of latent classes minus one.

In order to examine the properties of this model-based approach to LCA with a distal outcome, we now move to a simulation study. The impact of four factors on performance of this technique is examined for binary, count, and continuous outcomes. Performance of the proposed model-based approach is compared to that of maximum-probability assignment and multiple pseudo-class draws.

A Comparison of Three Estimation Methods for LCA with a Distal Outcome

Design

In this simulation study, we examined the effect of four factors on the performance of the model-based approach, as well as the two classify-analyze approaches, to LCA with a distal outcome. The factors were the conditional distribution of the distal outcome, *Z*; the strength of the association between the latent class variable and the distal outcome (i.e., effect size); the quality of the LCA measurement model (i.e., the degree of association between the observed and latent variables, which in this case corresponds to the degree of separation between latent classes); and the sample size. Specifically, the levels of the factors considered were as follows.

Type of Z—Three types of the distal outcome were considered: binary, continuous, and count. In our simulation, we let $Z|C = c \sim \text{Binom}(p_c)$ for binary *Z*; $Z|C = c \sim N(\mu_c, 1)$ for continuous *Z*; and $Z|C = c \sim \text{Poisson}(\lambda_c)$ for count *Z*. We hypothesized that any attenuation observed when a model-based approach is not used would be present regardless of the distribution of *Z*.

Strength of the effect of C on Z—For each *Z* distribution listed above, four strengths of association between the latent class variable and the distal outcome were considered. These corresponded to no effect, weak effect, medium effect, and strong effect as defined by Cohen (1992). The corresponding population values of p_c (for binary *Z*), μ_c (for continuous *Z*), and λ_c (for count *Z*), are listed in the top, middle and lower panel of Table 1, respectively. We hypothesized that attenuation of the effect of *C* on *Z* would increase as the effect size increases, and that this attenuation would be much smaller for the model-based approach as compared to the two classify-analyze approaches.

LCA measurement model—Using the empirical example of latent classes of adolescent depression described in Lanza, Flaherty, & Collins (2003) as a basis, latent class models with eight binary indicators and five latent classes were considered. We specified latent class prevalences and measurement models that had a structure similar to that in the empirical study. For all models in this simulation study, the proportion of individuals in

Classes 1 through 5 were specified to be 40%, 20%, 20%, 10%, and 10%, respectively. Two levels of measurement quality were considered: moderate, characterized by item-response probabilities equal to .8 or .2, and high, characterized by item-response probabilities equal to .9 or .1. Table 2 shows the set of item-response probabilities specified to achieve these two levels of measurement. We hypothesized that high measurement quality would reduce bias under any other combination of factors, regardless of estimation method.

Sample size—We considered sample sizes of 500 and 1000. Assessing performance for very small sample sizes was not a goal of this study; rather, we were interested in examining whether any benefits are achieved by increasing *n* from a moderate size to a large size. We hypothesized that there would be little difference between these sample sizes.

The fully crossed factorial design consisted of 48 conditions. For each condition, we implemented three approaches for estimating the effect of *C* on *Z*: the proposed model-based approach, the maximum-probability assignment approach, and the multiple (in this case, 20) pseudo-class draws approach. For each condition, we replicated the analysis 1000 times and summarized the simulation outputs to assess how each factor affected performance of the three approaches.

Procedure

The following Monte Carlo procedure was used in each of the 48 simulation design cells.

Step 1: Generation of LCA data—Given the specified LCA model (i.e., latent class prevalences and item-response probabilities) and the specified strength of association between *C* and *Z*, to generate one random observation, we first generated a latent class variable *C* from a multinomial distribution specified by the latent class prevalences (i.e., mixing proportions); we then generated item responses based on the item-response probabilities (i.e., ρ parameters) for that cell, and then generated the distal outcome *Z* based on the *C*-*Z* model for that cell.

Step 2: LCA model fitting—For each replicate data set, two different LCA models were fit. The first model included no distal outcome *Z* (for the maximum-probability assignment and pseudo-class draws approaches), and the second model included the distal outcome *Z* as a covariate (for the model-based approach). We used 100 sets of random starting values for the LCA model that did not include *Z* in order to avoid local maxima and for an examination of model identification. The parameter estimates from the model that did not include *Z* were used as starting values for the LCA model with *Z* as a covariate.

Step 3: Calculation of Z given C for each approach—Given the LCA results derived in Step 2, along with the random sample, the estimation of the effect of *C* on *Z* was conducted for each approach. For the model-based approach we employed the procedure described above, which relies on the β, γ and ρ parameters from the LCA model with *Z* included as a covariate. For maximum-probability assignment and multiple pseudo-class draws, we first inferred the latent classes *C* for each observation using the corresponding approaches (described above), and then in a subsequent model we estimated the effect of *C*

on *Z*. For the pseudo-class draws approach, this final step was repeated 20 times and results were combined across draws.

Step 4: Summary of results—The goal of this step was to summarize results across the 1000 replicate data sets in order to draw comparisons between the three methods of estimation. For each approach, we first compared the estimated effect of *C* on *Z* to the true effect, shown in Table 1, and then summarized the results across replications to obtain the bias and root mean squared error (RMSE) for each parameter estimate. This step required that we address the issue that the ordering of the latent classes is random across the 1000 replicates. To impose a standard order on the latent classes, we wrote a SAS macro to take the LCA estimates and true LCA model parameters as inputs, then reordered the latent classes based on distance calculations comparing the estimated LCA parameters and the true LCA model parameters.

Results

Tables 3, 4, and 5 show simulation results for the binary, continuous, and count outcomes, respectively. Within each table, we present results for $n = 500$ in the top panel and for $n = 1000$ in the bottom panel. Moderate measurement quality is shown on the left side, and high measurement quality on the right side. For each effect size (zero, small, medium, large), we present results based on the three analytic approaches: the proposed model-based method (Model), maximum-probability assignment (Assign), and multiple pseudo-class draws (P-C). Each cell reflects the bias (i.e., mean estimated value minus true value) in the estimate of *Z* given *C*. For example, Table 3 shows that for moderate measurement quality, $n = 1000$, and large effect size, the bias in the estimated proportion of individuals in each latent class with a 1 on the binary outcome was 0.003, −0.002, −0.016, −0.061, and −0.010 for Latent Classes 1, 2, 3, 4, and 5, respectively. Recall from Table 1 that the true proportions for this cell were 0.006, 0.153, 0.300, 0.447, and 0.594. Negative values of bias indicate that the class-specific prevalence of the outcome is underestimated. For the same set of conditions, the bias was from 2 to 10 times larger for the maximum-probability assignment (0.035, 0.022, −0.105, −0.115, −0.091) and the multiple pseudo-class draws (0.042, 0.023, −0.112, −0.130, −0.120) approaches.

Several general patterns emerged across results for the binary, continuous, and count distal outcomes. First, as expected, when the effect size was set to zero, all three methods performed equally well, in that bias was less than 0.01 for each latent class regardless of sample size, measurement quality, or method. Second, because the prevalence of Latent Class 1 was considerably larger than that of other latent classes (0.4; see Table 2), bias was consistently smaller for this latent class. This was expected because, all other factors held constant, there is more information available related to larger latent classes, making estimation more accurate. Similarly, the bias was consistently larger for the smaller latent classes (Latent Classes 4 and 5) because there was less information available for estimation. Third, as expected, as the strength of the association between the latent class variable and the distal outcome strengthened, the potential for bias increased, and – importantly – the benefits of using a model-based approach became more significant. Fourth, when the methods performed differentially, the model-based approach consistently performed better

than the two classify-analyze approaches. In every case, the impact of using either maximum-probability assignment or multiple pseudo-class draws was manifested by an attenuation of the effect of *C* on *Z*. That is, the more negative biases seen in the two classify-analyze approaches confirmed our hypothesis that these methods would result in underestimation of the distal outcome for the latent classes that are furthest from the mean on *Z*.

A somewhat surprising finding was that maximum-probability assignment worked at least as well as the multiple pseudo-class draws technique in terms of bias/attenuation of the effect of *C* on *Z*. This suggests that, in the long run, this simple classify-analyze approach is preferable to the pseudo-class draws approach. However, the variability in the estimates across the 1000 replicates for the maximum-probability assignment approach was higher than that for the multiple pseudo-class draws approach (not shown). Therefore, in empirical studies the pseudo-class draws approach may be more reliable than maximum-probability assignment. Regardless of this fact, however, the model-based approach introduced here performed substantially better than either of the standard classify-analyze techniques.

One final important finding is that, in addition to the model-based approach being less biased in the long run, this new method was shown to be consistent. That is, as *n* increased, bias was reduced. However, sample size had essentially no effect on performance of the maximum-probability assignment or pseudo-class draws methods; neither classify-analyze strategy appeared to be consistent.

In sum, improving measurement quality (i.e., moving from item-response probabilities of .2 and .8 to probabilities of .1 and .9) had a substantial impact for all methods, such that bias was reduced consistently by more than half for all methods. As discussed above, as the effect size between *C* and *Z* increased, the potential for bias increased. With larger effect sizes, attenuation increased much more in the two classify-analyze approaches than it did in the model-based approach. All of these patterns emerged consistently for all types (binary, continuous, and count) of distal outcome. Thus, the model-based approach proposed here outperformed maximum-probability assignment and multiple pseudo-class draws under every condition.

We next move to an empirical demonstration of the model-based approach to LCA with a distal outcome. The motivating example involves latent classes of depression in adolescence. Three distal outcomes are included for demonstration purposes: a binary outcome (regular smoking), a continuous outcome (grades), and a count outcome (delinquency).

Empirical Example: Adolescent Depression Classes Predicting Later Outcomes

Method

Participants—The classify-analyze and model-based approaches to estimating LCA with a distal outcome were compared in the context of a latent class model for adolescent depression using data from The National Longitudinal Study of Adolescent Health (Add

Health; Harris, 2009; Harris et al., 2009). This latent class model was first demonstrated in a chapter by Lanza et al. (2003). Add Health was mandated by Congress to collect data for the purpose of measuring the effect of social context on the health and well-being of adolescents in the United States. The first wave of the sample included 11,796 students in 7th through 12th grades, surveyed between April and December, 1995; the second wave included the same individuals surveyed again between April and August, 1996. The sample in this demonstration draws from the 1,044 adolescents (48.4% female) in the public-use dataset who were in 11th grade at Wave I and 12th grade at Wave II. So that all methodological comparisons could be made based on the same data set, we chose to include in the sample only participants who provided data on at least one depression item at Time 1 (grade 11), as well as data on all three distal outcomes at Time 2 (grade 12). The final sample included n = 762 (49.3% female) adolescents.

Measures of the latent class variable—Indicators of Time 1 depression latent classes included eight observed variables assessing the frequency of experiencing various depression symptoms in the past week. Four of these were indicators of sadness (Could not shake blues, Felt depressed, Felt lonely, Felt sad), two were indicators of feeling disliked by others (People unfriendly to you, People dislike you), and two were indicators of feelings related to failing at life (Life been a failure, Life not worth living). The original six-level variables were recoded so that 1 represented never or rarely experiencing the symptom, and 2 represented experiencing the symptom sometimes, a lot, most of the time, or all of the time. See Lanza et al. (2003) for details on the measurement of depression in this latent class model.

Measures for distal outcome variables—At Time 2, 254 participants (28.4%) reported Yes to the question: “Since your last interview, have you smoked cigarettes regularly; that is, at least one cigarette every day for 30 days?”

Grades at Time 2 was measured by taking the mean of four variables assessing the grade received on academic subjects during the most recent grading period. Grades were reported on a four-point scale corresponding to A, B, C, or D/F. The four academic subjects included in this measure were English or Language Arts, Mathematics, History or Social Studies, and Science. The average score, coded such that higher scores corresponded to better grades, was transformed by taking the square root (to reduce skewness) and then standardized to facilitate interpretation of the results.

A count variable was created indicating the number of delinquent acts reported at Time 2. First, each of the 14 four-category delinquency items was recoded to 0 for individuals who reported never engaging in that act during the past 12 months, and 1 for individuals who reported engaging in that act at least once. Then, for each individual, the scores on the 14 items were summed, resulting in a count variable with a range of zero to 14. The mean score on the delinquency scale was 1.7.

Analytic Procedure—Before considering the distal outcomes, we examined several models of depression with different numbers of latent classes in order to confirm that the five-class model reported by Lanza et al. (2003) was optimal for this particular sample.

Once the latent class model was selected, adolescent depression latent class membership was used to predict three outcomes using separate models. All latent class models were fit using PROC LCA (Lanza et al., 2011). The two classify-analyze approaches were carried out in SAS, and the new model-based approach was implemented using the SAS macro, %LCA_distal.

Maximum-probability assignment was conducted as follows. First, each individual's posterior probability of membership in each latent class was retained from the latent class model. Second, individuals were assigned to the latent class corresponding to their most likely membership (i.e., their maximum posterior probability). Third, the class-specific proportions or means for the distal outcome were calculated.

To implement the multiple pseudo-class draws approach, we used the posterior probabilities derived for the maximum-probability assignment approach. For each individual, 20 independent random numbers were drawn from the multinomial distribution defined by that individual's posterior probabilities. These random numbers were used to assign individuals to latent classes with probabilities proportionate to their posterior probabilities. This resulted in 20 data sets with class assignment included in each set. This procedure is analogous to multiple imputation of missing data (Schafer, 1997). Next, the class-specific proportions or means for the distal outcome were calculated. Finally, results were averaged across the 20 data sets.

The proposed *model-based approach* required that we estimate the latent class model of depression with the distal outcome included as a covariate. From this model, we retained the multinomial logistic regression coefficients reflecting the association between the latent class variable and the distal outcome. We then constructed empirically derived distributions of the distal outcome given latent class, under the conditional independence assumption. From that conditional distribution we reported the conditional probability (in the case of a binary distal outcome), conditional mean count (in the case of a count outcome), or conditional mean (in the case of a continuous outcome) for each latent class. The SAS syntax used for implementing the model-based approach to estimate the probability of regular cigarette use conditional on latent class appears in the Appendix.

Results

Consistent with previous literature (Lanza et al., 2003), the model with five latent classes of depression was selected on the basis of fit statistics, information criteria, and interpretability of the model. Table 6 shows the five latent classes of depression. The item-response probabilities represent the conditional probability of endorsing a past-week depression symptom given latent class membership; that is, these values are column-conditional. Estimates near zero indicate that individuals in that latent class were unlikely to have experienced that symptom, whereas estimates near one indicate that those individuals were likely to have experienced it. Together, these item-response probabilities formed the basis for labeling the latent classes. The most prevalent latent class was Non-Depressed (43.8%), followed by Sad (27.0%), Sad+Disliked (14.9%), Disliked (8.8%), and Depressed (5.5%).

The maximum-probability assignment and multiple pseudo-class draws approaches relied on posterior probabilities, which varied considerably across participants. The average posterior probability for those in the Non-Depressed latent class was 0.93 (N = 396, *MIN* = 0.44, *MAX* = 0.99); for the Sad latent class, was 0.88 (N = 232, *MIN* = 0.35, *MAX* = 1.00); for the Disliked latent class, was 0.77 (N = 73, *MIN* = 0.50, *MAX* = 0.95); for the Sad+Disliked latent class, was 0.79 (N = 150, *MIN* = 0.48, *MAX* = 0.96); and for the Depressed latent class, was 0.90 (N = 45, *MIN* = 0.59, *MAX* = 0.99).

Table 7 presents the estimated proportion of each latent class engaging in regular cigarette use, the mean grade score given latent class, and the mean delinquency count given latent class using each of the three statistical approaches. We calculated the effect sizes for the various outcomes, indicating the overall strength of association between C and Z. For both regular cigarette use and grades, effect sizes fell in the weak to moderate range (Cohen, 1992). For regular cigarette use, the effect size was estimated to be 0.16 for the model-based approach, 0.14 for the maximum-probability assignment approach, and 0.13 for the pseudo-class draws approach. For grades, effect sizes based on the mean and mode were estimated to be 0.16 and 0.22, respectively, for the model-based approach; the effect size was 0.19 for maximum-probability assignment and 0.21 for pseudo-class draws. The effect size for the association between C and the delinquency count variable was 0.47 for the model-based approach, 0.43 for maximum probability class assignment, and 0.46 for the pseudo-class approach; this association fell in the moderate to strong range (Cohen, 1992).

Based on Table 7, it appears that the effects may have been attenuated, particularly for cigarette use and grades, when relying on maximum-probability assignment or the pseudo-class approach. This is detectable by focusing on the column for the Depressed latent class, where, regardless of approach used, rates of regular cigarette use were highest. Although we did not know the true proportion in this latent class that reported regular cigarette use (because latent class membership was not known with certainty), the estimated rate of regular cigarette use was highest based on the model-based approach, and was somewhat lower for the two classify-analyze approaches.

Mean standardized grades for the Depressed latent class were lowest based on the model-based approach (*mean* = −0.457), followed by the pseudo-class approach (*mean* = −0.401) and the maximum-probability assignment approach (*mean* = −0.346). Figure 2 shows the estimated distributions of grades conditional on latent class.

The association between depression class membership and the count variable had the largest effect size; thus, we expected to see the greatest attenuation of effects with this distal outcome. Because Z was not included in the class assignment step for either classify-analyze approach, we anticipated seeing more substantial differences across methods for larger effect sizes. Interestingly, for this count outcome there was virtually no difference across methods in the estimates of Z given C.

Discussion

A pervasive issue in behavioral and social sciences is that computational methods do not exist to estimate the effect of membership in latent classes on a distal outcome. Methods to do this would enable scientists to understand the level of risk associated with membership in a particular latent class. For example, this would enable estimation of the risk of developing nicotine dependence given a particular set of early smoking behavior experiences, or the risk of contracting a sexually transmitted infection given a particular profile of sexual risk behavior. Although the statistical model for predicting latent class membership from an observed covariate is well understood, it is the opposite direction of effect that is at the heart of such research questions.

By applying Bayes' theorem, we can capture information from a model that is well-understood (LCA with covariates) and transform it into information that addresses this exact research question. This is the foundation for the flexible model-based approach proposed here. The critical pieces of information come from two sources. First, a latent class model is specified with the distal outcome as a covariate in order to obtain the logistic regression coefficients reflecting their association. Second, the class-conditional marginal density of Z is estimated, for example using a kernel density estimation approach. The SAS macro LCA_distal, introduced here for estimating LCA with distal outcomes that are categorical, continuous, or count variables, automates this approach.

In comparison to current approaches commonly employed for LCA with a distal outcome, which all rely on some sort of classification step (using the posterior probabilities) followed by a step for the outcome analysis, the proposed approach directly models the association. It is this very association that is meant to be approximated using classify-analyze approaches.

Performance of the model-based approach

A Monte Carlo simulation study was conducted to compare the performance of this new approach to two classify-analyze approaches: maximum-probability assignment and multiple pseudo-class draws. Simulation results show that the model-based approach produces substantially less biased estimates of the effect compared to either classify-analyze technique, particularly when the association between the latent class variable and the distal outcome is strong. Although the RMSE was larger for the model-based approach in the case of no or small effect size, as the strength of the effect of C on Z increased the relative performance reversed, such that the model-based approach had smaller RMSE. Taken together, when a moderate to strong relation exists between the latent class variable and the distal outcome, we recommend the model-based approach because of its lower bias and lower RMSE. In addition, we show that only the model-based approach exhibits the property of consistency (i.e., its performance improves as n increases).

In addition, we made several hypotheses regarding the factors examined in the simulation study. We expected the performance of the model-based approach to be superior to that of both classify-analyze approaches, regardless of the metric of the distal outcome (categorical, continuous, and count). This was consistently supported in the simulation study. Our hypothesis that the attenuation of effects would increase as the effect size increased was

confirmed. In addition, improving measurement quality resulted in better performance (i.e., less bias) for the model-based approach and for both classify-analyze approaches. As expected, we observed no improvement in the performance of either classify-analyze approach as sample size increased. For the model-based approach, however, performance did improve as sample size increased, suggesting that this method is statistically consistent.

Maximum-probability assignment versus pseudo-class draws

The current state-of-the-art for conducting LCA with a distal outcome is to employ one of two classify-analyze approaches. The first approach, maximum-probability assignment, is simple to do, although caution against its use is well recognized. In particular, because classification uncertainty is not accounted for in this approach, inference in the subsequent outcome analysis is known to be potentially biased. The second approach, multiple pseudo-class draws, was originally proposed as a method to account for that classification uncertainty in order to conduct model diagnostics of a particular LCA model (Wang et al., 2005). For example, within each of the 20 or so data sets, violations of the assumption of conditional independence between the latent class indicators given *C* can be examined. This approach was readily adopted, and assumed to outperform maximum-probability assignment in other settings, despite its performance not being compared systematically in those new settings. In particular, this technique was adopted widely for conducting LCA with distal outcomes (e.g., Fried et al., 2009; Petras & Masyn, 2010).

The Monte Carlo study described here suggests that the benefits of multiple pseudo-class draws are not readily apparent. In fact, the maximum-probability assignment approach consistently resulted in less bias in terms of the point estimates. We also calculated the variability across estimates from the 1000 replicates, and found that the multiple pseudo-class draws approach was less variable. However, given the larger bias of this approach, it may not be a preferable technique despite the lower variability. We then made comparisons based on the RMSE, and found that when effect sizes were medium to large, maximum-probability assignment actually had lower RMSE, thus outperforming multiple pseudo-class draws for conducting LCA with a distal outcome.

Including Z as an additional LCA indicator

Theoretically, to conduct LCA with a distal outcome, it is possible to treat *Z* as an additional indicator in the LCA model (along with *X*). Indeed, if the distal outcome is binary, then incorporating it as an additional indicator in the latent class model (along with *X*), as a covariate, and as a grouping variable are mathematically equivalent for a given number of latent classes. All of these approaches assume conditional independence between *X* and *Z* given *C*. Each of these approaches provides the necessary information to obtain estimates of *Z* given *C*.

A noted drawback to including the distal outcome as another indicator is that this approach can alter the meaning of *C* (Petras & Masyn, 2010). This observation deserves further consideration. To the extent that *Z* provides information in the latent class model that is unique from the latent class indicators *X*, there are important implications for both model interpretation and model selection. That is, if one adds unique information via the additional

indicator in a latent class model, the model then is being used to summarize a different set of information than just what *X* provides. Based on the more extensive information, more latent classes than those identified only by information in *X* may be identified when using both *X* and *Z*. Another limitation of this approach is that including *Z* as an additional indicator does not lend itself to an overall test of a hypothesized effect of *C* on *Z*.

If the distal outcome is count or continuous, inclusion of *Z* as a discrete grouping variable is not possible. Further, the approach of including *Z* as an additional indicator requires restrictive assumptions regarding the distribution of *Z* given *C* (for example, that *Z* follows a mixture of normal distributions, i.e. $f\{Z = z|C = c\} \sim N(\mu_c, \sigma^2)$). In contrast, with the model-based approach, we do not need to specify the conditional distribution of *Z* given *C* if *Z* is treated as a covariate, but only need to assume that $\Pr\{C = c|Z\}$ follows a logistic regression model. This implies that the second approach (i.e., treating *Z* as a covariate) requires fewer assumptions, and hence is more flexible than the first approach (i.e., treating *Z* as an indicator). Specifically, this approach is less vulnerable to potential model mis-specification of the marginal distribution of *Z*, as the distribution of *Z* is estimated directly from the observed data.

The model-based approach described here only requires the fitting of an LCA model with *Z* as a covariate, which can be implemented with several publicly-available and proprietary software packages (e.g., PROC LCA, the R package poLCA, Mplus, Latent GOLD; Lanza et al., 2011; Linzer & Lewis, in press; L. K. Muthén & Muthén, 1998–2007; Vermunt & Magidson, 2005). The nonparametric estimation of the class-conditional marginal density of *Z* can be readily implemented with common statistical software programs, including SAS (via PROC KDE) and R.

Limitations

The model-based approach performed consistently better than the classify-analyze approaches in the Monte Carlo study. However, one somewhat inconsistent finding emerged in the empirical example involving prediction of a delinquency count variable from depression latent classes, in that very little attenuation could be detected in the classify-analyze approaches compared to the model-based approach despite the large effect size. This merits further study, for example into the sensitivity of the performance of each method to violations of assumptions being made by the latent class and/or logistic models.

All of the approaches compared here require that there be no missing data on the distal outcome. Future research is merited on handling missing data, in particular applying this model-based approach to LCA with distal outcomes when multiple imputation is employed.

Conclusions and future work

We proposed a conceptually straightforward, computationally simple approach to estimating the effect of latent class membership on a distal outcome. This early work sets the stage for comparisons between this approach and others, such as multiply imputing the latent class variable under a fully-Bayesian model for a potentially more rigorous classify-analyze approximation.

This approach to LCA with distal outcomes was shown to outperform current classify-analyze practices under a variety of conditions, regardless of whether individuals are assigned based on their maximum posterior probability or if a pseudo-class draws approach is taken. This new solution to predicting distal outcomes from latent class membership has broad applicability. In particular, this strategy is relevant not just for traditional latent class analysis, but for the broad set of mixture models that are relevant in the social and behavioral sciences. These include latent profile analysis (where the indicators are continuous variables), growth mixture modeling (B. O. Muthén & Sheddon, 1999; Nagin, 2005), factor mixture modeling (Lubke & Muthen, 2007; McLachlan & Peel, 2000), and mixture regression modeling (Kaplan, 2005).

Important work remains in applying this model-based approach to studies involving research questions that are more complex than just the effect of *C* on *Z*, such as moderation of effects. This would be relevant in various situations, such as allowing the effect to vary across race/ethnicity groups or controlling for baseline levels on the distal outcome. It is important to note that in latent class models, if a distal outcome is included as a covariate (per the model-based approach described here), including a grouping variable such as race/ethnicity implicity allows the interaction between the covariate and the grouping variable. That is, the effect of *C* on *Z* is estimated within each group. Grouping variables can be included in the downloadable Excel calculator (mentioned above, available for download at methodology.psu.edu) that implements the model-based approach for a categorical distal outcome (Lanza & Rhoades, 2011b). Extending the more general LCA_distal macro to handle more complex LCA models such as this is an important future direction.

Acknowledgments

This project was supported by Award Number P50-DA010075 from the National Institute on Drug Abuse. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Drug Abuse or the National Institutes of Health. The authors wish to thank John J. Dziak for his helpful feedback on an early version of this manuscript. This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

Appendix

SAS Syntax

This appendix provides the SAS syntax for implementing the model-based approach to estimating the probability of regular smoking in Grade 12 (*Z*) conditional on the depression latent class variable at Grade 11 (*C*).

```
*Estimate latent class model with binary distal outcome Z included as
covariate;
proc lca data=outcomes start=data.baseline_start
```

```
outparam=data.estimates_cigZ outpost=data.posterior_cigZ;
id newaid rcig_c2;
nclass 5;
items w1fs3 w1fs6 w1fs13 w1fs16 w1fs9 w1fs19 w1fs14 w1fs17;
categories 2 2 2 2 2 2 2 2;
covariates rcig_c2;
reference 4;
run;
*Execute macro to obtain distribution of Z given C;
%LCA_distal(input_data = sasf.Variables, /*input random sample*/
param = _beta_param, /*beta parameter part*/
post = sasf.Posterior_bin_y, /*posterior membership probabilities*/
id = newaid, /*ID variable*/
distal = rcig_c2, /*distal outcome variable*/
yc = 1, /*1=discrete, 2=continuous, 3=count*/
y_cat = 2, /*number of categories of Z, given yc=1*/
method = 1, /*model-based, max assignment or pseudo-class*/
output_dataset_name= res11 /*output results*/
);
```

References

Agresti, A. Categorical data analysis. Hoboken, NJ: John Wiley & Sons, Inc.; 2002.

Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control. 1974; 19:716–723.

Asparouhov, T. Wald test of mean equality for potential latent class predictors in mixture modeling [technical appendix]. 2010. Available for download at http://www.statmodel.com/papers.shtml.

Bandeen-Roche K, Miglioretti DL, Zeger SL, Rathouz PJ. Latent variable regression for multiple discrete outcomes. Journal of the American Statistical Association. 1997; 92(440):1375–1386.

Bozdogan H. Model selection and Akaike Information Criterion (AIC): The general theory and its analytical extension. Psychometrika. 1987; 52:345–370.

Casella, G.; Berger, RL. Statistical inference. Belmont, CA: Duxbury; 1990.

Clark, SL.; Muthén, BO. Relating latent class anlaysis results to variables not included in the analysis. 2009. Manuscript submitted for publication. Available for download at http://www.statmodel.com/papers.shtml.

Clogg, CC. Latent class models: Recent developments and prospects for the future. In: Arminger, G.; Clogg, CC.; Sobel, ME., editors. Handbook of statistical modeling for the social and behavioral sciences. New York, NY: Plenum Press; 1995. p. 311-359.

Cohen J. A power primer. Psychological Bulletin. 1992; 112:155–159. [PubMed: 19565683]

Collins, LM.; Lanza, ST. Latent class and latent transition analysis: With applications in the social, behavioral and health sciences. Hoboken, NJ: John Wiley & Sons, Inc.; 2010.

Collins LM, Schafer JL, Kam C-M. A comparison of inclusive and restrictive strategies in modern missing data procedures. Psychological Methods. 2001; 6:330–351. [PubMed: 11778676]

Dayton CM, Macready GB. Concomitant-variable latent-class models. Journal of the American Statistical Association. 1988; 83(401):173–178.

Fried LP, Xue Q-L, Cappola AR, Ferrucci L, Chassin L, Varadham R, Bandeen-Roche K. Nonlinear multisystem physiological dysregulation associated with frailty in older women: Implications for etiology and treatment. Journal of Gerontology. 2009; 64A:1049–1057.

Gelman, A.; Carlin, JB.; Stern, HS.; Rubin, DB. Bayesian data analysis. New York, NY: Taylor & Francis; 2003.

Harris, KM. The National Longitudinal Study of Adolescent Health (Add Health), Waves I & II, 1994–1996; Wave III, 2001–2002; Wave IV, 2007–2009 [machine-readable data file and documentation]. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill; 2009.

Harris, KM.; Halpern, CT.; Whitsel, E.; Hussey, J.; Tabor, J.; Entzel, P.; Udry, JR. The National Longitudinal Study of Adolescent Health: Research design [WWW document]. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill; 2009. URL: http://www.cpc.unc.edu/projects/addhealth/design.

Jones MC, Marron JS, Sheather SJ. A brief survey of bandwidth selection for density estimation. Journal of the American Statistical Association. 1996; 91:401–407.

Kaplan D. Finite mixture dynamic regression modeling of panel data with implications for response analysis. Journal of Educational and Behavioral Statistics. 2005; 30:169–187.

Lanza ST, Collins LM, Lemmon DR, Schafer JL. PROC LCA: A SAS procedure for latent class analysis. Structural Equation Modeling. 2007; 14(4):671–694. [PubMed: 19953201]

Lanza ST, Collins LM, Schafer JL, Flaherty BP. Using data augmentation to obtain standard errors and conduct hypothesis tests in latent class and latent transition analysis. Psychological Methods. 2005; 10:84–100. [PubMed: 15810870]

Lanza, ST.; Dziak, JJ.; Huang, L.; Xu, S.; Collins, LM. Proc LCA & Proc LTA users' guide (Version 1.2.6). University Park, PA: The Methodology Center, Penn State; 2011. Retrieved from http://methodology.psu.edu.

Lanza, ST.; Flaherty, BP.; Collins, LM. Latent class and latent transition analysis. In: Schinka, JA.; Velicer, WF., editors. Handbook of psychology: Vol. 2, research methods in psychology. Hoboken, NJ: Wiley; 2003. p. 663-685.

Lanza ST, Patrick ME, Maggs JL. Latent transition analysis: Benefits of a latent variable approach to modeling transitions in substance use. Journal of Drug Issues. 2010; 40(1):93–120. [PubMed: 20672019]

Lanza ST, Rhoades BL. Latent class analysis: An alternative perspective on subgroup analysis in prevention and treatment. Prevention Science. 2011a Advanced online publication.

Lanza, ST.; Rhoades, BL. LCA outcome probability calculator (Version 1.0). University Park, PA: The Methodology Center, Penn State; 2011b. Retrieved from The Methodology Center: http://methodology.psu.edu/ra/lcalta/calculator.

Linzer D, Lewis J. poLCA: An R package for polytomous variable latent class analysis. Journal of Statistical Software. (in press).

Lubke G, Muthen B. Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. Structural Equation Modeling. 2007; 14:26–47.

McLachlan GJ. On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. Journal of the Royal Statistical Society, Series C (Applied Statistics). 1987; 36(3): 318–324.

McLachlan, GJ.; Peel, D. Finite mixture models. New York, NY: John Wiley and Sons, Inc.; 2000.

Muthén BO, Sheddon K. Finite mixture modeling with mitxure outcomes using the EM algorith. Biometrics. 1999; 55:463–469. [PubMed: 11318201]

Muthén, LK.; Muthén, BO. Mplus user's guide. Fifth ed.. Los Angeles, CA: Muthén & Muthén; 1998–2007.

Nagin, DS. Group-based modeling of development. Cambridge, MA: Harvard University Press; 2005.

Nylund K, Bellmore A, Nishina A, Graham S. Subtypes, severity, and structural stability of peer victimization: What does latent class analysis say? Child Development. 2007; 78(6):1706–1722. [PubMed: 17988316]

Petras, H.; Masyn, K. General growth mixture analysis with antecedents and consequences of change. In: Piquero, AR.; Weisburd, D., editors. Handbook of quantitative criminology. New York, NY: Springer; 2010. p. 69-100.

Roberts TJ, Ward SE. Using latent transition analysis in nursing research to explore change over time. Nursing Research. 2011; 60(1):73–79. [PubMed: 21127448]

Roeder K, Lynch K, Nagin DS. Modeling uncertainty in latent class membership: A case study in criminology. Journal of the American Statistical Association. 1999; 94:766–776.

Rubin, DB. Multiple imputation for nonresponse in survey research. New York, NY: Wiley; 1987.

SAS Institute Inc.. SAS 9.1.3 help and documentation. Cary, NC: SAS Institute Inc.; 2002–2004.

Schafer, JL. Analysis of incomplete multivariate data. London: Chapman & Hall; 1997.

Schwartz G. Estimating the dimension of a model. The Annals of Statistics. 1978; 6:461–464.

Sclove SL. Application of model-selection criteria to some problems in multivariate analysis. Psychometrika. 1987; 52:333–343.

Silverman, BW. Density estimation. New York, NY: Chapman & Hall; 1986.

Vermunt, JK.; Magidson, J. Latent GOLD 4.0 users' guide. Belmont, MA: Statistical Innovations; 2005.

Wang C, Brown CH, Bandeen-Roche K. Residual diagnostics for growth mixture models: Examining the impact of a preventive intervention on multiple trajectories of aggressive behavior. Journal of the American Statistical Association. 2005; 100(471):1054–1076.

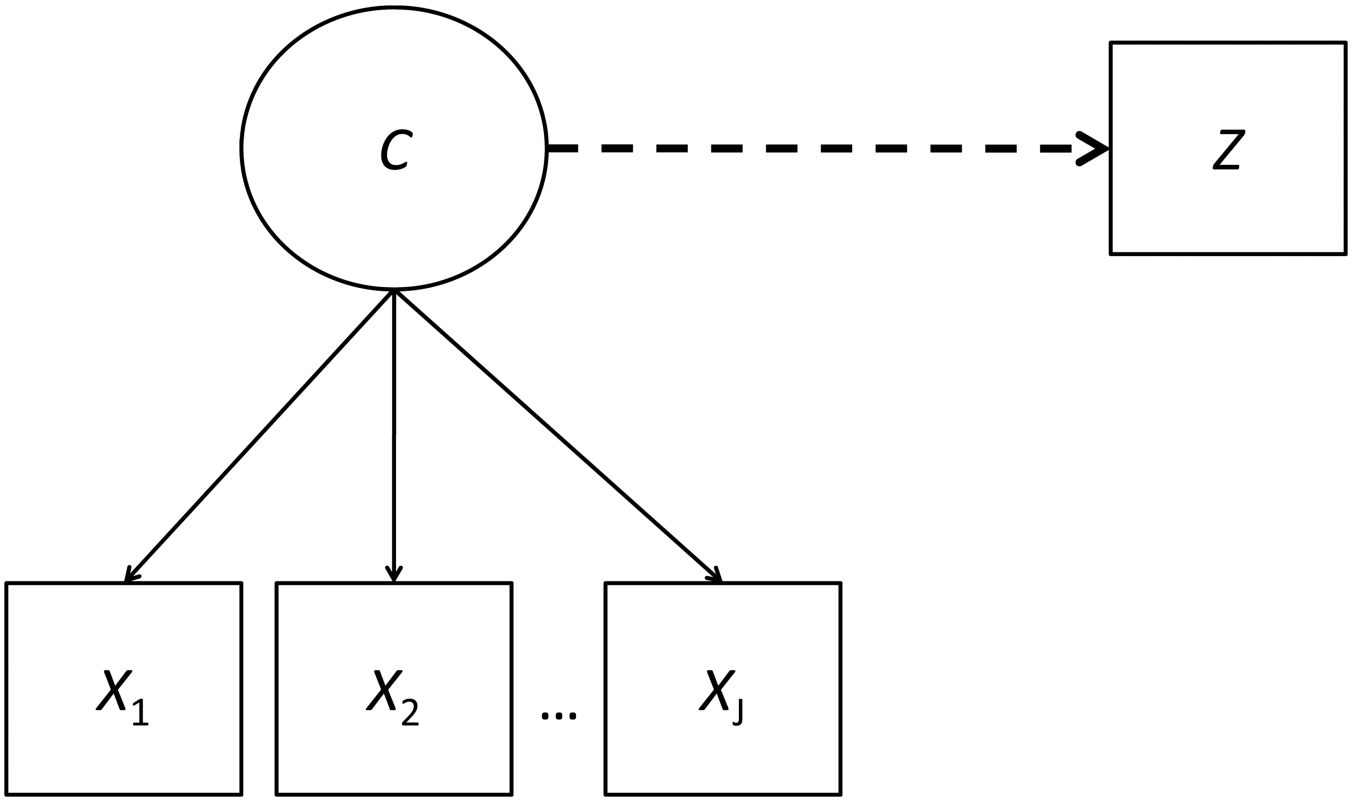

Figure 1.

Graphical representation of the latent class model with a distal outcome. *C* refers to the latent class variable, X_1, X_2, …X_J refer to manifest indicators of *C*, and *Z* refers to the distal outcome.

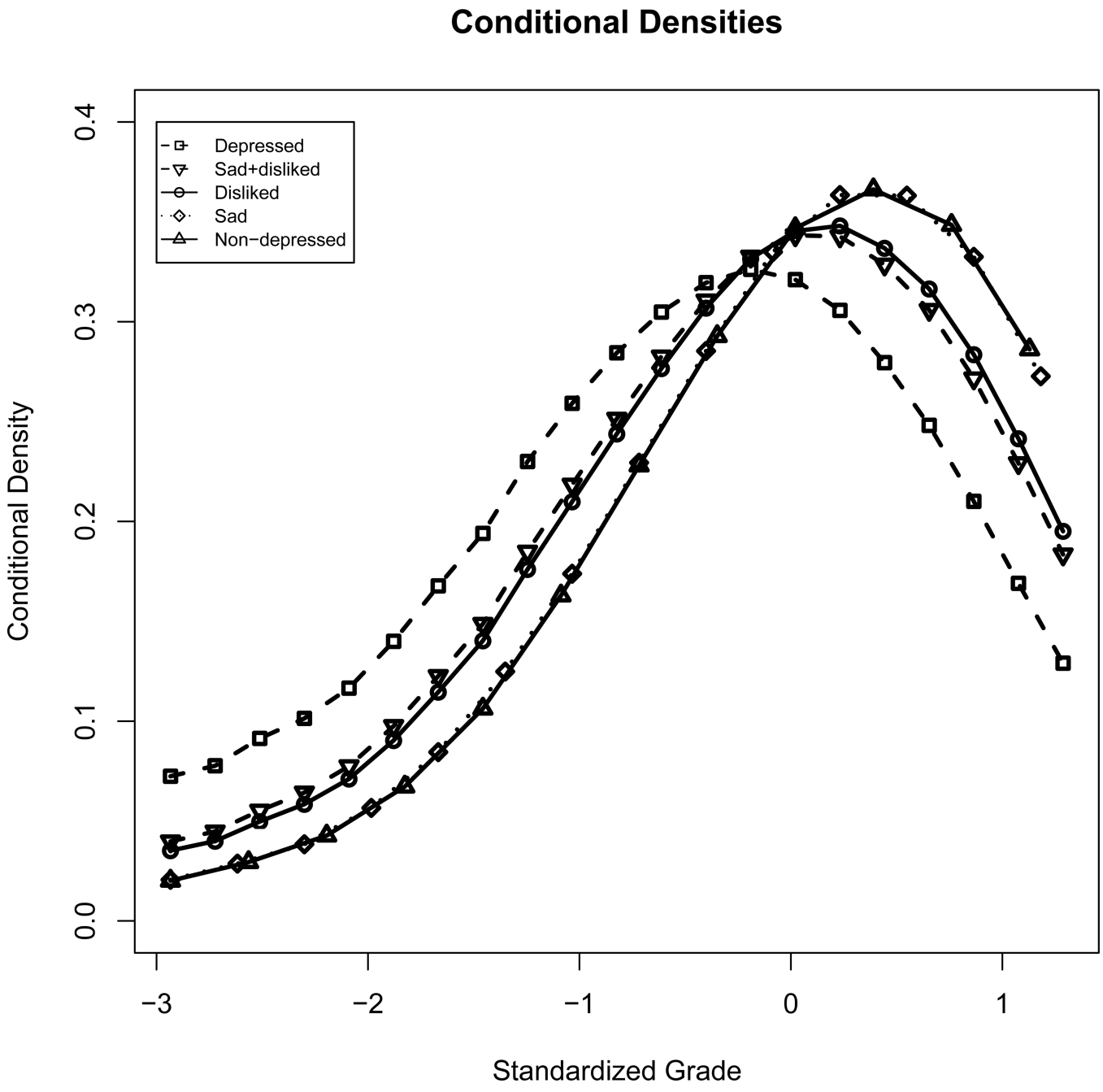

Figure 2.
Estimated density functions for grades at Time 2 conditional on depression latent class membership at Time 1.

Table 1

Patterns of Z|C: Specified true values for the distal outcome given latent class membership in the simulation study

	Latent Class					
	1	**2**	**3**	**4**	**5**	**Effect Size**
			Binary Z			
Pr{*Z*|*C*}	0.300	0.300	0.300	0.300	0.300	= 0.0
Pr{*Z*|*C*}	0.234	0.267	0.300	0.333	0.366	≈0.1
Pr{*Z*|*C*}	0.110	0.205	0.300	0.395	0.490	≈0.3
Pr{*Z*|*C*}	0.006	0.153	0.300	0.447	0.594	≈0.5
			Continuous Z (Conditional Normal)			
E{*Z*|*C*}	0.00	0.00	0.00	0.00	0.00	= 0.0
E{*Z*|*C*}	−0.14	−0.07	0.00	0.07	0.14	≈ 0.1
E{*Z*|*C*}	−0.38	−0.19	0.00	0.19	0.38	≈ 0.3
E{*Z*|*C*}	−0.64	−0.32	0.00	0.32	0.64	≈ 0.5
			Count Z (Conditional Poisson)			
E{*Z*|*C*}	0.80	0.80	0.80	0.80	0.80	= 0.0
E{*Z*|*C*}	0.66	0.73	0.80	0.87	0.94	≈ 0.1
E{*Z*|*C*}	0.42	0.61	0.80	0.99	1.18	≈ 0.3
E{*Z*|*C*}	0.16	0.48	0.80	1.12	1.44	≈ 0.5

Table 2

Patterns of item-response probabilities: Two conditions for item-response probabilities specified in the simulation study

	Latent Class				
	1	**2**	**3**	**4**	**5**
LC Membership Probabilities	0.4	0.2	0.2	0.1	0.1
Moderate Measurement Quality					
Could note shake blues	0.2	0.8	0.2	0.8	0.8
Felt depressed	0.2	0.8	0.2	0.8	0.8
Felt lonely	0.2	0.8	0.2	0.8	0.8
Felt sad	0.2	0.8	0.2	0.2	0.8
People unfriendly	0.2	0.2	0.8	0.2	0.8
Disliked by people	0.2	0.2	0.8	0.8	0.8
Life was failure	0.2	0.2	0.2	0.2	0.8
Life not worth living	0.2	0.2	0.2	0.2	0.8
High Measurement Quality					
Could not shake blues	0.1	0.9	0.1	0.9	0.9
Felt depressed	0.1	0.9	0.1	0.9	0.9
Felt lonely	0.1	0.9	0.1	0.9	0.9
Felt sad	0.1	0.9	0.1	0.9	0.9
People unfriendly	0.1	0.1	0.9	0.9	0.9
Disliked by people	0.1	0.1	0.9	0.9	0.9
Life was failure	0.1	0.1	0.1	0.1	0.9
Life not worth living	0.1	0:1	0.1	0.1	0.9

Struct Equ Modeling. Author manuscript; available in PMC 2014 November 21.

Table 3

Simulation results for LCA with a binary distal outcome: Bias in proportion with Z = 1 given latent class membership

		Moderate Measurement					Strong Measurement				
		Latent Class					Latent Class				
Method	**ES**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
						n = 500					
Model	Zero	−0.003	−0.000	0.009	0.005	0.006	−0.000	0.002	0.002	0.002	−0.002
Assign	Zero	−0.002	0.001	0.002	0.002	0.004	−0.000	0.002	0.003	0.002	−0.002
P-C	Zero	−0.002	0.000	0.003	0.002	0.004	−0.000	0.002	0.002	0.002	−0.001
Model	Sm	−0.000	0.001	−0.017	−0.017	−0.009	−0.000	−0.001	−0.012	−0.009	−0.003
Assign	Sm	0.008	0.007	−0.025	−0.031	−0.025	0.003	0.002	−0.017	−0.014	−0.007
P-C	Sm	0.010	0.008	−0.027	−0.034	−0.030	0.003	0.002	−0.018	−0.015	−0.009
Model	Med	0.001	0.005	−0.033	−0.072	−0.023	0.000	0.001	−0.030	−0.028	−0.005
Assign	Med	0.023	0.020	−0.066	−0.097	−0.078	0.009	0.006	−0.050	−0.041	−0.018
P-C	Med	0.028	0.020	−0.071	−0.104	−0.091	0.010	0.006	−0.053	−0.044	−0.022
Model	Lg	0.010	0.003	−0.037	−0.115	−0.029	0.001	−0.001	−0.034	−0.027	−0.008
Assign	Lg	0.039	0.028	−0.103	−0.150	−0.113	0.012	0.008	−0.074	−0.056	−0.026
P-C	Lg	0.046	0.030	−0.111	−0.159	−0.135	0.013	0.009	−0.080	−0.061	−0.032
						n = 1000					
Model	Zero	−0.000	−0.001	0.003	−0.002	0.001	0.001	−0.000	0.001	0.003	0.001
Assign	Zero	0.000	−0.000	0.002	−0.003	0.001	0.002	−0.000	0.000	0.002	0.001
P-C	Zero	0.000	−0.000	0.002	−0.002	0.000	0.001	0.000	0.001	0.002	0.001
Model	Sm	−0.003	0.001	−0.013	−0.011	−0.004	0.000	−0.002	−0.014	−0.009	0.000
Assign	Sm	0.008	0.007	−0.023	−0.023	−0.021	0.003	0.001	−0.019	−0.014	−0.004
P-C	Sm	0.009	0.006	−0.025	−0.027	−0.028	0.003	0.001	−0.020	−0.015	−0.006
Model	Med	−0.001	−0.001	−0.030	−0.040	−0.008	0.001	−0.002	−0.020	−0.019	−0.003
Assign	Med	0.023	0.015	−0.070	−0.077	−0.061	0.009	0.006	−0.048	−0.037	−0.016
P-C	Med	0.028	0.015	−0.075	−0.086	−0.079	0.010	0.006	−0.051	−0.041	−0.021
Model	Lg	0.003	−0.002	−0.016	−0.061	−0.010	0.000	0.001	−0.018	−0.022	−0.005

		Moderate Measurement					Strong Measurement				
		Latent Class					Latent Class				
Method	**ES**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
Assign	Lg	0.035	0.022	−0.105	−0.115	−0.091	0.012	0.011	−0.075	−0.056	−0.024
P-C	Lg	0.042	0.023	−0.112	−0.130	−0.120	0.013	0.010	−0.081	−0.062	−0.032

Note: Model = model-based approach; Assign = maximum-probability assignment rule; P-C = multiple pseudo-class draws.

Table 4

Simulation results for LCA with a continuous distal outcome: Bias in mean given latent class membership

		Moderate Measurement					Strong Measurement				
		Latent Class					Latent Class				
Method	**ES**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
						n = 500					
Model	Zero	0.004	−0.002	0.002	−0.001	0.002	−0.000	−0.001	−0.001	−0.007	−0.006
Assign	Zero	0.003	−0.001	0.001	−0.000	0.002	0.000	−0.002	−0.002	−0.006	−0.005
P-C	Zero	0.002	−0.000	0.001	−0.002	−0.001	−0.000	−0.001	−0.002	−0.006	−0.005
Model	Sm	−0.005	0.005	−0.034	−0.052	−0.010	−0.004	−0.001	−0.022	−0.018	0.010
Assign	Sm	0.018	0.019	−0.046	−0.073	−0.057	0.006	0.004	−0.035	−0.031	−0.009
P-C	Sm	0.022	0.018	−0.052	−0.078	−0.065	0.006	0.004	−0.037	−0.034	−0.011
Model	Med	−0.016	0.005	−0.056	−0.145	−0.018	−0.015	−0.004	−0.059	−0.042	0.012
Assign	Med	0.044	0.038	−0.131	−0.199	−0.147	0.012	0.009	−0.098	−0.082	−0.040
P-C	Med	0.054	0.040	−0.138	−0.212	−0.176	0.014	0.009	−0.105	−0.087	−0.047
Model	Lg	−0.015	0.012	−0.111	−0.218	−0.043	−0.014	0.001	−0.067	−0.045	0.029
Assign	Lg	0.082	0.061	−0.224	−0.314	−0.251	0.028	0.024	−0.165	−0.127	−0.053
P-C	Lg	0.097	0.065	−0.240	−0.334	−0.298	0.032	0.023	−0.177	−0.140	−0.066
						n = 1000					
Model	Zero	0.000	0.002	0.001	0.003	−0.006	−0.001	0.005	−0.001	−0.000	0.004
Assign	Zero	−0.000	0.001	0.001	0.001	−0.003	−0.001	0.004	0.000	−0.000	0.003
P-C	Zero	−0.000	0.001	−0.000	0.003	−0.003	−0.001	0.004	−0.000	−0.000	0.004
Model	Sm	−0.006	−0.001	−0.028	−0.040	0.002	−0.008	−0.003	−0.019	−0.013	0.005
Assign	Sm	0.018	0.011	−0.048	−0.061	−0.043	0.003	0.004	−0.032	−0.028	−0.013
P-C	Sm	0.022	0.012	−0.052	−0.066	−0.058	0.003	0.003	−0.035	−0.031	−0.017
Model	Med	−0.019	−0.001	−0.065	−0.076	0.003	−0.014	−0.004	−0.037	−0.021	0.017
Assign	Med	0.044	0.034	−0.135	−0.155	−0.117	0.015	0.013	−0.096	−0.072	−0.034
P-C	Med	0.054	0.035	−0.146	−0.172	−0.154	0.016	0.012	−0.104	−0.080	−0.043
Model	Lg	−0.027	−0.002	−0.067	−0.136	0.017	−0.016	0.002	−0.033	−0.031	0.022

		Moderate Measurement					Strong Measurement				
		Latent Class					Latent Class				
Method	**ES**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
Assign	Lg	0.074	0.053	−0.225	−0.266	−0.190	0.027	0.027	−0.156	−0.128	−0.057
P-C	Lg	0.090	0.053	−0.243	−0.292	−0.254	0.029	0.025	−0.170	−0.138	−0.075

Note: Model = model-based approach; Assign = maximum-probability assignment rule; P-C = multiple pseudo-class draws.

Table 5

Simulation results for LCA with a count distal outcome: Bias in mean count given latent class membership

		Moderate Measurement					Strong Measurement				
		Latent Class					Latent Class				
Method	**ES**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
						n = 500					
Model	Zero	−0.005	0.003	0.015	0.004	0.014	−0.000	−0.000	0.002	−0.004	−0.003
Assign	Zero	−0.002	0.003	0.007	0.001	0.007	0.000	−0.001	0.001	−0.003	−0.003
P-C	Zero	−0.001	0.002	0.006	0.001	0.008	−0.000	−0.001	0.001	−0.002	−0.003
Model	Sm	0.003	−0.001	−0.035	−0.062	−0.025	−0.002	0.001	−0.029	−0.020	−0.008
Assign	Sm	0.019	0.012	−0.051	−0.074	−0.064	0.005	0.007	−0.039	−0.027	−0.018
P-C	Sm	0.022	0.011	−0.055	−0.076	−0.073	0.006	0.006	−0.042	−0.029	−0.021
Model	Med	0.001	−0.001	−0.065	−0.133	−0.051	0.001	−0.000	−0.052	−0.060	−0.002
Assign	Med	0.047	0.032	−0.132	−0.189	−0.155	0.018	0.012	−0.093	−0.086	−0.030
P-C	Med	0.056	0.035	−0.142	−0.201	−0.182	0.020	0.012	−0.099	−0.093	−0.036
Model	Lg	0.004	0.010	−0.072	−0.217	−0.060	0.000	−0.001	−0.060	−0.064	−0.002
Assign	Lg	0.083	0.062	−0.222	−0.323	−0.241	0.027	0.019	−0.167	−0.124	−0.046
P-C	Lg	0.099	0.065	−0.235	−0.345	−0.288	0.029	0.019	−0.180	−0.135	−0.059
						n = 1000					
Model	Zero	0.005	−0.001	−0.004	0.000	−0.003	−0.001	−0.002	0.001	0.002	−0.001
Assign	Zero	0.003	−0.001	−0.003	−0.002	−0.002	−0.000	−0.001	0.000	0.002	0.000
P-C	Zero	0.002	−0.000	−0.002	−0.001	−0.001	−0.000	−0.002	0.001	0.002	−0.000
Model	Sm	−0.005	0.001	−0.030	−0.039	0.003	−0.002	0.001	−0.019	−0.023	0.003
Assign	Sm	0.015	0.012	−0.051	−0.056	−0.039	0.005	0.007	−0.033	−0.031	−0.008
P-C	Sm	0.019	0.013	−0.055	−0.062	−0.052	0.006	0.006	−0.036	−0.034	−0.012
Model	Med	−0.007	−0.001	−0.048	−0.080	−0.011	−0.002	−0.003	−0.038	−0.038	−0.002
Assign	Med	0.044	0.030	−0.131	−0.145	−0.118	0.016	0.012	−0.093	−0.077	−0.029
P-C	Med	0.054	0.032	−0.141	−0.166	−0.153	0.018	0.011	−0.102	−0.084	−0.040
Model	Lg	−0.001	0.003	−0.030	−0.133	−0.029	−0.001	−0.002	−0.029	−0.044	−0.007

		Moderate Measurement					Strong Measurement				
		Latent Class					Latent Class				
Method	**ES**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
Assign	Lg	0.075	0.049	−0.219	−0.252	−0.201	0.026	0.020	−0.158	−0.128	−0.050
P-C	Lg	0.091	0.050	−0.237	−0.285	−0.264	0.028	0.018	−0.171	−0.139	−0.068

Note: Model = model-based approach; Assign = maximum-probability assignment rule; P-C = multiple pseudo-class draws.

Table 6

Parameter estimates for five-class model of adolescent depression

		Latent Class				
		Non-Depressed	**Sad**	**Disliked**	**Sad + Disliked**	**Depressed**
Latent Class Membership Probabilities		0.438	0.270	0.088	0.149	0.055
Indicator	*Overall Proportion*	*Item−Response Probabilities*				
Could not shake blues	0.326	0.038	0.529	0.237	0.634	0.931
Felt depressed	0.434	0.069	0.744	0.288	0.823	1.000
Felt lonely	0.396	0.092	0.594	0.248	0.825	0.923
Felt sad	0.494	0.154	0.810	0.267	0.885	0.958
People unfriendly to you	0.342	0.160	0.203	0.697	0.784	0.704
People disliked you	0.331	0.057	0.098	1.000	1.000	0.781
Life been a failure	0.138	0.018	0.106	0.096	0.283	0.941
Life not worth living	0.097	0.007	0.055	0.114	0.141	0.880

Note: $n = 896$.

Table 7

Empirical results showing outcomes at Time 2 conditional on depression latent class membership at Time 1

	Latent Class				
Method	**Non-Depressed**	**Sad**	**Disliked**	**Sad + Disliked**	**Depressed**
Proportion Reporting Past-year Regular Cigarette Use (*ES* = 0.13 to 0.16)					
Model-Based Approach	0.194	0.307	0.275	0.261	0.474
Max Probability Assignment	0.207	0.296	0.250	0.301	0.450
Multiple Pseudo-Class Draws	0.208	0.289	0.270	0.275	0.448
Mean Standardized Grade Score (*ES* = 0.16 to 0.22)					
Model-Based Approach	−0.081	−0.070	−0.223	−0.268	−0.457
Max Probability Assignment	0.077	0.868	−0.113	−0.165	−0.346
Multiple Pseudo-Class Draws	0.080	0.096	−0.105	−0.146	−0.401
Mean Delinquency Count (*ES* = 0.43 to 0.47)					
Model-Based Approach	1.384	1.869	1.662	2.412	2.482
Max Probability Assignment	1.399	1.857	1.696	2.323	2.425
Multiple Pesudo-Class Draws	1.394	1.868	1.661	2.371	2.485

Note: *n* = 896; *ES* = effect size.

Struct Equ Modeling. Author manuscript; available in PMC 2014 November 21.