

New Approaches for Examining Associations With Latent Categorical Variables: Applications to Substance Abuse and Aggression

AQ: au

Alan Feingold, Stacey S. Tiberio, and Deborah M. Capaldi
Oregon Social Learning Center

Assessments of substance use behaviors often include categorical variables that are frequently related to other measures using logistic regression or chi-square analysis. When the categorical variable is latent (e.g., extracted from a latent class analysis [LCA]), classification of observations is often used to create an observed nominal variable from the latent one for use in a subsequent analysis. However, recent simulation studies have found that this *classical 3-step analysis* championed by the pioneers of LCA produces underestimates of the associations of latent classes with other variables. Two preferable but underused alternatives for examining such linkages—each of which is most appropriate under certain conditions—are (a) *3-step analysis*, which corrects the underestimation bias of the classical approach, and (b) *1-step analysis*. The purpose of this article is to dissuade researchers from conducting classical 3-step analysis and to promote the use of the 2 newer approaches that are described and compared. In addition, the applications of these newer models—for use when the independent, the dependent, or both categorical variables are latent—are illustrated through substantive analyses relating classes of substance abusers to classes of intimate partner aggressors.

Keywords: categorical analysis, intimate partner violence, latent class analysis, structural equation modeling, substance abuse

Categorical variables are frequently used in research in the behavioral sciences, especially in the addictions field. The nature of categorical data is that observations fall into discrete groups and the analysis examines group membership (Agresti, 2002), including proportion of respondents in a given category (e.g., prevalence of a substance use disorder) and probability of different patients attaining clinical goals (e.g., completing treatment). These variables may be composed using psychological indicators that define a *behavioral taxonomy* (e.g., Moffitt, 1993; sometimes called a *typology*; see Jackson, Sher, & Wood, 2000) developed from clinical observations or psychological theory. An alcohol use disorder (AUD), for example, is a clinically derived taxonomy that categorizes respondents into three mutually exclusive and exhaustive groups—for alcohol dependence, alcohol abuse, and no AUD—according to diagnostic criteria in the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.; DSM-

IV-TR; American Psychiatric Association, 2000). Examples of theory-driven taxonomies include *batterer typology* models of intimate partner violence (IPV; e.g., Holtzworth-Munroe, 2000; Holtzworth-Munroe, Meehan, Stuart, Herron, & Rehman, 2000; see also Capaldi & Kim, 2007, for a critique of those models).

Randomized clinical trials often use categorical outcomes (especially binary variables, such as alcohol consumption vs. abstinence) to determine treatment efficacy. Treatment completion is also an example of a dichotomous outcome measure often considered in program evaluations.

Although the distributions of categorical variables are of primary interest to epidemiologists studying the prevalence rates of disorders, psychologists generally are more concerned with examining associations between independent and dependent variables to test theories and hypotheses about human behavior. When only the independent variable is categorical (as is generally the case with experiments, including randomized clinical trials), dummy variable or other coding schemes can be used to capture the variable. The multiple regression analysis then proceeds exactly as if the variable was continuous, and is mathematically equivalent to analysis of variance (Cohen, Cohen, West, & Aiken, 2003).

When the dependent variable is nominal, however, *categorical analysis* methods, including chi-square and logistic regression analyses, are needed. The latter examines logit-transformed probabilities of falling into different outcome categories conditional on the independent variables (Hosmer & Lemeshow, 2000). When the outcome is dichotomous, binary logistic regression analysis may be used to compare the probabilities of being in one group (e.g., the treatment completion group) across the levels of the independent variable (e.g., treatment A vs. treatment B). A commonly used effect size for this difference in probabilities is the odds ratio (OR;

Alan Feingold, Stacey S. Tiberio, and Deborah M. Capaldi, Oregon Social Learning Center, Eugene, Oregon.

This project was supported by awards from National Institutes of Health (NIH) Grant RC1DA028344 from the National Institute of Drug Abuse (NIDA), Grant R01AA018669 from the National Institute of Alcoholism and Alcohol Abuse (NIAAA), and Grant R01HD46364 from the National Institute of Child Health and Human Development (NICHD). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH, NIDA, NIAAA, or NICHD. We thank Isaac Washburn for his input on the Mplus syntax.

Correspondence concerning this article should be addressed to Alan Feingold, Oregon Social Learning Center, 10 Shelton McMurfhey Boulevard, Eugene, OR 97401-4928. E-mail: alanf@oslc.org

AQ: 1

AQ:2,Fn Fleiss & Berlin, 2009; Haddock, Rindskopf, & Shadish, 1998),¹ which is the categorical analogue of Cohen's *d* for group differences in continuous outcomes (Feingold, 2009).

When the outcome measure consists of more than two unordered categories, multinomial logistic regression analysis (MLRA) can be used. In MLRA, one group is designated the *reference category*, and the analysis compares the probability of an observation belonging to each nonreference group relative with that of the reference group across levels of the independent variable. (The reference category in this case is very similar to the role of the reference class in dummy variable coding of categorical predictors in linear regression.)

Consider, for example, a prevention study examining whether a hypothesized childhood risk factor (e.g., family history of mental illness) predicts diagnosis of an AUD in adulthood. Given an adult AUD outcome with three categories (abuse, dependence, or neither), MLRA could be used to determine whether a family history of mental illness (a binary variable) increases children's risk of succumbing to an AUD in adulthood. Adults who do not have a diagnosis of an AUD would serve as the reference category; one OR would convey the difference in the probability of the respondent being in the alcohol dependence category versus the AUD-free category (as a function of family history status); a second OR would be obtained for the corresponding difference in the probability of being in the alcohol abuse category versus the AUD-free category.

The categorical outcomes described above are examples of observed variables. There has been an increasing use of *latent categorical variables*, which capture unobserved heterogeneity within a sample, to study addictive behaviors (e.g., B. O. Muthén, 2006). Latent class analysis (LCA) is a popular technique that can be used to identify homogeneous subsamples of respondents from item endorsement patterns (Collins & Lanza, 2010; B. O. Muthén, 2008).² When groups (called *classes*) are derived from a statistical analysis of item response patterns instead of measured directly, the categorical variable is latent rather than observed.

B. O. Muthén (2006), for example, conducted LCA of symptoms of alcohol dependence and alcohol abuse among current drinkers and identified four classes of people. The identification of subpopulations from such item response patterns should be distinguished from the classification approach taken by the psychiatric committees that developed diagnostic criteria for addictive disorders based on clinical observations. LCA has also recently been used to identify distinct subgroups of perpetrators of IPV (Ansara & Hindin, 2010; Carbone-Lopez, Kruttschnitt, & Macmillan, 2006; Klostermann, Mignone, & Chen, 2009).

A common method traditionally used to examine associations involving a latent categorical dependent variable is *classical three-step analysis*, which uses the posterior probabilities from an LCA to classify observations (e.g., people) on the basis of their likely class membership to create an observed variable that can serve as an outcome in a subsequent logistic regression or chi-square analysis (e.g., Agrawal, Lynskey, Madden, Bucholz, & Heath, 2007; Bornovalova, Levy, Gratz, & Lejuez, 2010). However, simulation studies have consistently found that classical three-step analysis underestimates the strength of associations between latent classes and observed covariates (Bakk, 2011; Bolck, Croon, & Hagenars, 2004; Vermunt, 2010).

Alternatively, one-step analysis extracts latent classes and examines the association between latent categorical and observed variables simultaneously via a general latent variable modeling framework (e.g., B. O. Muthén & Shedden, 1999). As shown in Table 1, different one-step models are used when the independent, the dependent, or both variables are latent: (a) *LCA with covariates* (LCA-C)—also known as *latent class regression analysis* (e.g., Bandeen-Roche, Miglioretti, Zeger, & Rathouz, 1997)—predicts a latent categorical dependent variable (classes) from an observed variable (Goodman, 1974; Hagenars, 1993); (b) *LCA with distal outcomes* (LCA-D) predicts an observed variable from a latent one (Asparouhov & Muthén, 2013); and (c) *structural equation modeling with categorical variables* (categorical SEM; Skrondal & Rabe-Hesketh, 2005) is used to predict one latent categorical variable from another. In these models (which can be thought of as *logistic regression with latent variables*), one or more ORs are obtained that capture the associations between the two categorical variables.

Unfortunately, one-step analysis is sometimes problematic, such as when large numbers of covariates are used in an exploratory analysis (Vermunt, 2010) or an observed covariate has a bimodal distribution (Asparouhov & Muthén, 2013). Bolck et al. (2004) were among the pioneers of corrected three-step analysis (referred to as *three-step analysis* in the contemporary literature) to relate a latent class variable to an observed covariate when one-step analysis is not ideal (or the investigator wants to ensure that the formation of the latent classes is not influenced by the observed variables to which the classes are to be related).

Although the improved three-step methods developed by Bolck et al. (2004) were not as efficient as one-step analysis, enhancements to them were introduced shortly thereafter (Clark & Muthén, 2009; Mplus Technical Appendixes, 2010; Wang, Brown, & Bandeen-Roche, 2005). Further refinements have resulted in methods of three-step analysis that are almost as efficient as one-step analysis in many cases (Asparouhov & Muthén, 2013; Vermunt, 2010).

These newest three-step procedures calculate the LCA first; the most likely class memberships are then obtained from the posterior probabilities of the LCA along with the classification uncertainty rate (i.e., measurement error); and the most likely class member-

¹ In probability theory, the *odds* of an event occurring is the probability that the event occurs divided by the probability that the event does not occur (Agresti, 2002; Feingold, 2012). The odds of an event (e.g., perpetration of aggression) can be calculated separately for each of two groups (e.g., men with and without a substance use disorder) using the observed proportions, and the ratio of the odds between the two groups is the OR. For example, if 50% of men without a drug problem and 80% of men with such a problem commit an act of violence, the odds of a man without a drug problem aggressing are $0.50/0.50 = 1.00$; for a man with a drug problem, the odds are $0.80/0.20 = 4.00$. The OR for the prediction of aggression from drug problem status would then be 4.00.

² Although LCA is increasing in popularity, cluster analysis is an older method that can also identify homogeneous subsamples based on item responses and has been widely used in addictions research over the past 2 decades. Cluster-analytic studies have typically uncovered two broad classes of addicted patients: a less severely impaired class and a more severely impaired class (Ehlers, Gilder, Gizer, & Wilhelmsen, 2009; Feingold, Ball, Kranzler, & Rounsaville, 1996; Kranzler et al., 2008; Zucker, Ellis, Fitzgerald, Bingham, & Sanford, 1996). The use of LCA in the same studies would likely have identified the same subpopulations of substance abusers.

Table 1
Overview of One- and Three-Step Models for Examining Associations With Categorical Variables

Independent variable	Dependent variable	
	Observed	Latent
Observed	Logistic regression	LCA with covariates
Latent	LCA with distal outcomes	Categorical SEM

Note. LCA = latent class analysis; SEM = structural equation modeling.

ship variables are then analyzed together with covariates or distal outcomes accounting for the measurement error in classification. However, Asparouhov and Muthén's (2013) simulation studies identified conditions in which one-step analysis remains more efficient than three-step analysis. If, for example, low separation exists between the classes (i.e., entropy < .60) in one-step analysis, the covariates can influence class assignment, resulting in more separation between the classes and a more efficient estimation procedure compared with the three-step analysis. Thus, both one-step and three-step analysis methods need to be in the armamentarium of investigators working with categorical data, whereas the classical three-step approach was an interim solution that has largely outlived its usefulness.

The main purpose of this article is to promote one-step and three-step categorical modeling of categorical data as preferable alternatives to a classical three-step analysis that uses posterior probabilities from a latent variable model to create observed variables that can be included in a logistic regression or chi-square analysis. In addition to the explications of these different procedures for linking latent categorical variables to other categorical or continuous variables, worked examples of these models that relate classes of substance abusers to classes of perpetrators of IPV are provided.

Method

Participants

The participants were drawn from a sample of 206 men who were enrolled in a long-term longitudinal study, the Oregon Youth Study (OYS)/Couples Study (Capaldi & Clark, 1998; Feingold, Kerr, & Capaldi, 2008), when they were in Grade 4 from schools with a higher incidence of delinquency in the neighborhood of a medium-sized city in the Pacific Northwest.³ The men were evaluated annually or biannually over 2 decades. In addition, data were obtained from the men's romantic partners in several biannual waves that began once the OYS men entered their late teens. The present study analyzed data collected from the men when they were in their mid-20s ($M = 25.9$ years, $SD = 0.7$) and assessed for (a) lifetime *DSM-IV-TR* symptoms of substance dependence and abuse for different substances and (b) a history of perpetration of aggression toward partners who had participated with them in the Couples Study. The majority of participants (89.6%) were White. At time of last data collection, 30.7% were married and 22.8% were cohabiting.

Measurements and Procedure

Substance use symptoms. Substance abuse/dependence symptoms were examined in 199 men who were administered a structured psychiatric interview—Version 2.0 of the Composite International Diagnostic Interview (CIDI; World Health Organization, 1997)—that assesses lifetime *DSM-IV-TR* symptoms of dependence and abuse for each of nine psychoactive substances (nicotine, alcohol, cannabis, cocaine, opiates, hallucinogens, sedatives, amphetamines, and “other drugs”—although abuse of nicotine is not measured). The CIDI was used to determine whether participants had ever experienced at least one symptom of dependence and at least one symptom of abuse for each substance.

Five dichotomous variables were first created by assigning participants scores of 1 if they had ever had experienced one or more symptoms of dependence for nicotine, alcohol, and marijuana and one or more symptoms of abuse for alcohol and marijuana (else = 0). Second, two items for *other drugs* were created. Participants were assigned a value of 1 for *other drug dependence* if they had ever experienced a symptom of dependence on one or more of the CIDI-examined substances other than nicotine, alcohol, and marijuana (else = 0), and a value of 1 for *other drug abuse* if they had reported one or more symptoms of abuse for any substance other than alcohol and marijuana (else = 0). Thus, seven dichotomous items tapping dependence on nicotine, alcohol, marijuana, and other drugs and abuse of alcohol, marijuana, and other drugs were created from the CIDI.

Aggression. Commencing with the second wave of data in which men participated with romantic partners, the men's partners at each time reported men's aggression via four physical and five psychological aggression items on the Conflict Tactics Scales (CTS; Straus, 1979).⁴ There were up to three waves of CTS data collected from each man's partner or partners by the time he had reached his mid-20s, which were examined once with the CIDI. For each of the 185 men who had participated with a romantic partner at one or more of these waves, partners' CTS reports of victimization were used to determine whether he had perpetrated aggression during the year prior to each assessment period.

Because the CTS uses Likert scaling, *never* was scored 0 and other categories were recoded to 1. Next, for each CTS item, a score of 1 was assigned if the man was reported to have ever committed that act of aggression during at least one of the years before he was examined with a partner (else = 0).

Design and Analysis

The analyses began by extracting two independent sets of two-, three-, four-, and five-class solutions—one from the LCA of the substance dependence/abuse items and the other from the LCA of the aggression items. Two statistics obtained with LCA, the Bayesian information criterion (BIC) and the bootstrap likelihood ratio test (BLRT), have been found to be the most effective at identi-

³ Although the sample size for the illustrative analyses is small, we feel it is adequate for analyses that are primarily pedagogical than substantive, especially given that our models were found to have high entropy values and no clusters with few cases assigned to them.

⁴ Two CTS items that concerned knives and guns were never endorsed in our sample and thus could not be included in the LCA of aggressive behaviors.

fyng the number of latent classes that should be extracted from the indicator variables (Nylund, Asparouhov, & Muthén, 2007), and both criteria were considered in selecting from among the four different LCA solutions for each of the two constructs. With the BIC, the solution with the smallest value is identified as the optimal model, whereas the BLRT tests the statistical significance of the improvement in the model when an additional class is extracted.

Next, separate observed categorical variables were created for substance use and IPV from the posterior probabilities associated with the chosen solutions from the respective LCAs. To illustrate the use of classical three-step analysis, we first determined the associations between the two LCA-derived variables with an MLRA. The now-observed substance problems groups (the predictor variable) were dummy coded in the familiar manner used with nominal independent variables in ordinary least squares multiple regression (Cohen et al., 2003). One aggressor group (non-aggressors) was designated as the reference category, as a reference category is required for MLRA (Hosmer & Lemeshow, 2000). The MLRA thus yielded ORs for predicting the probability of being in an aggressor class (vs. the nonaggressor reference class) as a function of being in a substance abuse class (vs. a reference class of nonabusers).

Given that the two categorical variables were both latent, one- or three-step categorical SEM should be used to establish associations between them. However, to illustrate use of LCA-C and LCA-D, we used the posterior probabilities from one of the two variables to create an observed variable that was assumed for didactic purposes to be a naturally occurring observed categorical variable (e.g., gender). In each of the models that associated an "observed" categorical variable with a latent one (one- and three-step LCA-C and one-step LCA-D), we included one of the two LCA-derived variables created for use in the prior MLRA in the analysis as observed and treated the other variable as latent. Thus, the LCA-Cs were LCAs of the aggression items that included as a covariate the groups derived from the LCA of the substance-related symptoms (and assumed for pedagogical purpose to be an observed nominal variable). The one-step LCA-D was an LCA of the substance use items, with the aggressor groups derived from the LCA of the aggression items included as the observed categorical distal outcome. Finally, one- and three-step categorical SEM analyses, which examined both substance use symptoms and aggression as latent categorical variables, were used to establish the linkages between the two constructs by regressing the latter on the former.

Results

All analyses were conducted using Mplus (Aspourhov & Muthén, 2013; L. K. Muthén & B. O. Muthén, 2012). Appendix A contains the Mplus input statements used to conduct the one-step analyses (LCA-C, LCA-D, and categorical SEM) and Appendix B contains commands for conducting the three-step analyses (LCA-C and categorical SEM).

LCA of Substance Use Symptoms

In the LCA of the substance abuse/dependence items, the BIC was smallest for the three-class solution and the BLRT was sta-

tistically significant for the three-class solution but not for the four-class solution (see Table 2), both of which indicated that three classes—with about an equal number of men in each class (see Table 3)—should be extracted. The entropy for the three-class solution was .85, indicating that the men could adequately be assigned to substance abuse classes. The top panel in Table 3 reports the structure matrix for this solution.

The men in the first class (the *drug problem-free* class) generally never had a *DSM-IV-TR* symptom of substance dependence or substance abuse, although a significant minority (nearly one third) had experienced at least one symptom of nicotine dependence during their lifetimes. All the men in the second class (the *soft drug-abusing* class) had a history of symptoms related to alcohol problems but a notable minority also reported symptoms for marijuana dependence or abuse and more than one half had been dependent on nicotine. However, almost none of the men in this class had experienced problems with hard drugs. The men in the remaining class (the *polysubstance-abusing* class) generally had a history of symptoms associated with abuse of and dependence on nicotine, alcohol, marijuana, and hard drugs.

LCA of Aggression Items

The BIC for the three-class LCA solution for the aggression items was smaller than the BIC found for either of the other two solutions, indicating three classes should be extracted (see Table 2). The BLRT, however, was statistically significant for the four- but not the five-class solution, indicating that four rather than three classes should be extracted. However, the four-class solution extracted two similar classes of men with a history of perpetration of both physical and psychological aggression (called *global aggressors*), suggesting that the three-class solution should be accepted. The entropy for the three-class solution was .83, indicating that participants could adequately be assigned to IPV classes. The bottom panel in Table 3 reports the structure matrix for the three-class LCA solution for the aggression items.

The first IPV class consisted of nonaggressive men, although a small majority of the members in that class had engaged in the mildest form of psychological aggression (sulking). The second, and largest class (which included almost one half of the sample),

Table 2
Bayesian Information Criterion Values and Bootstrap Likelihood Ratio Tests for Latent Class Analysis of Substance Abuse Symptoms and Conflict Tactics Scales (Aggression) Items

LCA	Number of classes extracted		
	2	3	4
Substance abuse symptoms			
BIC	1509.18	1508.11	1535.79
BLRT	329.70*	43.40*	14.67
Conflict Tactics Scales (aggression)			
BIC	1429.12	1382.72	1410.18
BLRT	386.41*	98.60*	24.74*

Note. LCA = latent class analysis; BIC = Bayesian information criterion; BLRT = bootstrap likelihood ratio test; $df = 8$ for BLRT for LCA of substance abuse/dependence symptoms; $df = 10$ for BLRT for LCA of aggression items.

* $p < .001$.

Table 3
Item Endorsement Probabilities From Three-Class Solutions From Latent Class Analysis of Substance Use Problems (n = 199) and the Conflict Tactics Scales (n = 185)

	Classes of substance abusers		
	None	Soft drugs	Polysubstance
<i>n</i>	59	77	63
%	30	39	32
Substance dependence/abuse items			
Nicotine dependence	0.30	0.56	0.90
Alcohol abuse	0.10	0.54	0.69
Alcohol dependence	0.08	0.94	0.83
Cannabis abuse	0.00	0.17	0.78
Cannabis dependence	0.04	0.24	0.87
Other drug abuse	0.00	0.00	0.67
Other drug dependence	0.02	0.07	0.85
Classes of intimate partner aggressors			
	None	Psych	Global
<i>n</i>	60	85	40
%	32	46	22
Conflict Tactics Scales items			
Yelled/insulted	0.29	0.96	1.00
Sulked/refused to talk	0.55	0.93	0.90
Stomped out of the room	0.29	0.96	0.98
Threw/smashed	0.00	0.39	1.00
Threatened to hit/throw	0.00	0.11	0.90
Threw something	0.00	0.10	0.65
Pushed/grabbed/shoved	0.00	0.28	0.98
Hit (or attempt), not with object	0.00	0.02	0.64
Hit (or attempt), hard object	0.02	0.00	0.30

Note. LCA = latent class analysis; Soft drugs = alcohol and/or cannabis problem; Polysubstance = polysubstance abuse problem; Psych = psychological aggression; Global = global aggression (psychological + physical aggression).

was composed of men who had perpetrated psychologically aggressive behaviors but had not physically harmed (or threatened to harm) at least one of their partners. The final, and smallest class (accounting for about one fifth of the sample), was composed of global aggressors.

Associating Latent Classes of Substance Abusers and Aggressors

MLRA. The classical three-step analysis (MLRA following LCA) was used to determine whether the probabilities of being in each of the two aggressor groups versus the nonaggressive group differed as a function of being in either of the two substance-abusing groups versus being in the drug problem-free group. Thus, two logistic functions were obtained for predicting (a) the probability of being a psychological aggressor versus a nonaggressor and (b) the probability of being a global aggressor versus a nonaggressor from the LCA-derived substance use categories. Because the substance use variable was composed of three groups, two dummy-coded variables were used to capture it. To form the first variable, we assigned a score of 1 to each participant if he was in the soft drug-abusing group and 0 otherwise. For the second

variable, a score of 1 was assigned if he was in the polysubstance-abusing group and 0 otherwise. With two logistic functions and two dummy-coded variables as predictors, the analyses yielded four independent ORs for the probability of being in either of the two aggressor groups versus the nonaggressive group as a function of being in either of the two drug problem groups versus the drug problem-free group.

As shown in the top row of Table 4, the ORs for the differences in predicted probabilities obtained using MLRA indicated that male polysubstance abusers were more likely than men with no history of drug problems to be assigned to each of the two classes of aggressors than to the nonaggressor class. Men with a history of problems limited to abuse of soft drugs, however, were not more likely than men without any history of symptoms of substance abuse to be assigned to either aggressor class than to the nonaggressor class.

LCA-C. In both the one- and three-step LCA-Cs, the dummy-coded groups derived from the LCA of substance use items that were previously used in the MLRA were now treated as observed covariates used to predict the latent classes of aggressors from groups of substance abusers. The structure matrix of the three-class solution of the aggression items from the one-step LCA-C was compared with the one obtained previously from the LCA conducted without the substance abuse groups included as a covariate (see Table 3 for the latter). None of the coefficients (item probabilities) for the aggression items from the LCA-C differed by more than .01 from respective coefficients obtained with the LCA that had not included a substance abuse variable as a covariate. Because LCA-C is conceptually equivalent to an MLRA with a latent covariate, the analyses produced ORs that reflected the same hypotheses of associations tested in the previous (MLRA) analysis. As predicted from the simulation studies that showed a downward bias in associations obtained with classical three-step analysis, the respective significant ORs were larger when obtained with LCA-C than with MLRA (see Table 4).

LCA-D. In the one-step LCA-D, the classes from LCA of substance use-related items were used to predict groups of aggressors. The structure matrix of the three-class solution of the substance abuse/dependence items from this analyses was compared with the one obtained previously from the LCA of the substance use items conducted without the aggression variable included as a distal outcome (see Table 3 for the latter). None of the coefficients (item probabilities) for the substance use symptoms from the LCA-D differed by more than .01 from respective coefficients obtained with the LCA that had not included aggression as a distal outcome. The two ORs obtained using the one-step LCA-D indicated the same associations found in the previous analyses but were expectedly larger than corresponding values obtained with MLRA that had treated both variables as observed.

Categorical SEM. Finally, one- and three-step categorical SEM analyses that specified three classes for each variable (with the numbers of classes predetermined from the LCAs that did not include auxiliary variables) were used to link the substance abuse and aggression classes. The item probabilities were also compared across classes for each of the two variables obtained from the one-step categorical SEM with the corresponding coefficients found in the separate LCAs of the same variables. In the analyses of the aggression items, the coefficients (item probabilities) were virtually identical (i.e., never differing by more than .01) in the

Table 4
Regression of Aggressor Classes on Substance Abuser Classes by Analytic Method

Predictor: Class predicted:	Odds ratios for associations between latent variable classes				Entropy
	Soft drug class		Polysubstance class		
	Psych	Global	Psych	Global	
<i>Analytic Method</i>					
MLRA	1.90	1.47	2.94*	2.86*	NA
95% CI	[0.84, 4.29]	[0.54, 3.97]	[1.20, 7.16]	[1.01, 8.07]	
1-step LCA-C	1.70	1.69	4.44*	4.43*	.85
95% CI	[0.51, 5.60]	[0.49, 5.77]	[1.41, 14.01]	[1.27, 15.52]	
3-step LCA-C	2.05	1.52	3.40*	3.17*	.85
95% CI	[0.80, 5.24]	[0.53, 4.37]	[1.18, 9.75]	[1.03, 9.78]	
1-step LCA-D	2.07	1.78	3.35*	3.81*	.80
95% CI	[0.80, 5.32]	[0.55, 5.72]	[1.32, 8.52]	[1.27, 11.42]	
1-step categorical SEM	1.96	2.71	4.76**	6.03**	.79
95% CI	[0.52, 7.34]	[0.37, 19.77]	[1.49, 15.18]	[1.56, 23.34]	
3-step categorical SEM	2.11	1.56	3.80*	3.40*	.73
95% CI	[0.72, 6.16]	[0.47, 5.21]	[1.23, 11.73]	[1.03, 11.25]	

Note. All three-step analyses are corrected using the modal maximum likelihood method. Three-step LCA-D with a nominal distal outcome cannot be conducted in Mplus 7.1. Odds ratios greater than 1.00 indicate higher probabilities of perpetrating intimate partner aggression for men in a substance use problems class compared with men in the drug problem-free class. MLRA = multinomial logistic regression analysis; LCA = latent class analysis; LCA-C = LCA with covariates; LCA-D = LCA with distal outcome; Psych = psychologically aggressive versus nonaggressive (reference) class; Global = globally aggressive vs. nonaggressive class; SEM = structural equation modeling; NA = not applicable; CI = confidence interval.

* $p < .05$. ** $p < .01$.

LCA and categorical SEM, but that was not always the case with the substance use symptoms items. Comparisons between the two sets of coefficients indicated differences not exceeding .01 for 19 of the 21 item probabilities and a negligible .03 (.54 vs. .57) for another item. However, the probability of having an alcohol dependence symptom for those in the drug problem-free class was notably higher (.14 vs. .08) in the one-step categorical SEM than the LCA of substance use symptoms. The significance tests of the ORs from the one- and three-step categorical analyses yielded the same results obtained previously in the MLRA, LCA-C, and LCA-D analyses (see Table 4). The respective significant ORs, however, were found to be the largest in the one-step categorical SEM analysis, which were the only associations to attain significance at the .01 level.

Discussion

Classical three-step analysis is the traditional approach used to examine the relationship between two variables when at least one of them is latent and categorical: The posterior probabilities from an LCA are used to classify individuals, thus forming groups that can be treated in a subsequent categorical analysis (e.g., MLRA) as an observed nominal variable. To illustrate this approach, LCA was first used to generate three classes of substance abusers and three classes of aggressors from two independent sets of dichotomous response (indicator) variables. Next, MLRA was used to determine the association between the two now-observed categorical variables. The analysis of this 3×3 contingency table found that male polysubstance abusers were significantly more likely to have committed IPV than men with no history of drug problems, but men who had abused alcohol and/or marijuana but not hard drugs were not significantly more likely to have perpetrated such violence than men who had no history of substance abuse.

However, recent simulation studies have documented that the classical three-step approach using classification and MLRA considerably underestimates the true associations between variables (e.g., Bolck et al., 2004), a weakness not found when latent categorical variables are duly included in a one-step analysis. Because the categorical substance abuse and aggression variables used in the present study are both latent, the appropriate one-step analysis for these data is categorical SEM, in which the latent aggression construct is regressed on the latent substance abuse construct. Categorical SEM is thus a categorical counterpart to ordinary SEM, with the latter extracting dimensions (i.e., factors) from indicators and regressing a continuous outcome factor on a predictor (exogenous) factor.

As expected from simulation studies, the results from the categorical SEM and the MLRA were similar, but the ORs for statistically significant associations were consistently larger in the one-step analysis. Moreover, the two significant ORs were each significant at the .01 level in the one-step analysis but only at the .05 level in the MLRA.

To illustrate the application of two additional (and more common) types of one-step analyses of categorical data (LCA-C and LCA-D), we pretended that the LCA-derived observed group of either substance abusers or aggressors was a true nominally scaled variable and then conducted an LCA-C and an LCA-D, with substance abuse treated as a covariate in the former and aggression as a distal outcome in the latter. Both analyses yielded the same two significant associations between substance abuse and aggression classes found in the prior analyses. As expected, the magnitude of respective ORs obtained with LCA-C/LCA-D was somewhere in between values generated from MLRA and categorical SEM because one rather than neither or both of the variables was appropriately modeled as latent in LCA-C and LCA-D.

Modern methods of three-step analysis that are refinements of classical three step-analysis have recently been introduced and in some cases are preferable to the more established one-step analysis when calculating associations with latent categorical variables (Asparouhov & Muthén, 2013; Vermunt, 2010). These new three-step analyses are free from the strong downward bias in estimates characteristic of the prevalent classical three-step approach and have an advantage over one-step analysis in that the formation of the latent classes is not influenced by the observed covariate or the distal outcome (Asparouhov & Muthén, 2013; Vermunt, 2010). Accordingly, we also ran three-step LCA-C to afford comparisons with MLRA and one-step LCA-C. (Unfortunately, current statistical programs do not conduct a three-step LCA-D with an observed categorical distal outcome, precluding its illustration with our data.)

As expected, ORs for significant associations from the three-step LCA-C were larger than ORs obtained with classical three-step analysis (i.e., using MLRA) but smaller than ORs found with one-step LCA-C. Note that the findings from simulation studies of a slightly downward bias in the three-step analysis and a slight upward bias in the one-step approach (Vermunt, 2010) would mean the latter would be expected to yield larger estimates than would the former, which was indeed the case in our illustrative analyses of the two types of LCA-C.

Comparison of results from MLRA with one- and three-step categorical SEM found the same pattern of estimates observed in the previous analyses: The respective ORs were the largest in one-step analysis and smallest in MLRA.

Limitations

Because of this article's focus on LCA with auxiliary variables, familiarity with the fundamentals of LCA (but not with methods for relating latent categorical variables to auxiliary variables) was assumed. Thus, many important issues regarding LCA were evaded, precluding use of this article as a comprehensive primer on it. These issues pertain to item selection, number of items to be used (and its relation to the maximum number of latent classes that can be observed), base rates of endorsement of the items, choosing from competing solutions when different criteria suggest different numbers of classes should be extracted, and requisite sample size for different kinds of analyses. However, there are numerous introductions to LCA (e.g., Bartholomew, Knott, & Moustaki, 2011; Collins & Lanza, 2010; Hagenaaers & McCutcheon, 2002) that address these basic issues and that can be used in conjunction with this article by investigators new to categorical data analysis who want to apply these new methods to test their hypotheses.

Applications to Longitudinal Analysis

Although we have discussed modeling of cross-sectional categorical data, latent classes may also be derived from longitudinal analyses, such as growth mixture modeling (GMM) of alcohol use (e.g., Capaldi, Feingold, Kim, Yoerger, & Washburn, 2013; Sher, Jackson, & Steinley, 2011). Whereas LCA derives classes from multiple responses collected at a single time, GMM extracts classes from the same typically continuous outcome measured over time and defined by differences in trajectories of outcomes (L. K. Muthén & B. O. Muthén, 2012).

Both one-step and three-step approaches to LCA can also be used to examine associations involving classes obtained with GMM (Asparouhov & Muthén, 2013; L. K. Muthén & B. O. Muthén, 2012). Thus, categorical modeling can be used to link GMM classes to an observed variable through either GMM with covariates or GMM with distal outcomes, which correspond to LCA-C and LCA-D, respectively. Categorical SEM also can be used to examine associations (a) between independent and dependent latent categorical variables when both sets of classes have been extracted using outcomes that have been measured repeatedly and (b) between classes extracted from cross-sectional analysis and trajectory classes. For example, analysis of a continuous outcome measured over time could be used to extract latent trajectory classes to be predicted from latent classes formed from baseline item responses tapping, say, risk factors. The results would then address whether unobserved heterogeneity in risk characteristics at study onset predicts future growth in problematic behaviors. Finally, covariates and distal outcomes in LCA and GMM with auxiliary variables are not limited to categorical variables but may be observed or latent continuous variables (e.g., Guo & Wall, 2006).

Continuous Indicator Variables

Although LCA is not appropriate for examination of indicator variables that are continuous, *latent profile analysis* (LPA; Lazarsfeld & Henry, 1968) is an LCA analogue for use with continuous indicators. Mplus can be used to conduct one-step LPA-C and LPA-D that correspond to LCA-C and LCA-D, respectively. Three-step approaches for handling latent classes derived from continuous indicators have recently been formulated (Gudicha & Vermunt, in press).

Conclusions

Although we are not contending that classes derived from latent class analysis are necessarily more meaningful than groups formed from observed variables (e.g., *DSM-IV-TR* diagnoses), one-step and three-step analyses are generally preferable to the prevailing classical three-step analysis because they yield less biased ORs. The type of LCA models that should be used depends on the scaling of the variables, as some categorical variables (e.g., gender, diagnosis) are inherently observed rather than latent. Thus, LCA-C, LCA-D, and categorical SEM (with one-step or three-step analysis) are all suitable for particular categorical analyses, but the widely used classical three-step analysis is to be avoided because it underestimates the true associations between latent classes and auxiliary variables.

References

- Agrawal, A., Lynskey, M. T., Madden, P. A. F., Bucholz, K. K., & Heath, A. C. (2007). A latent class analysis of illicit drug abuse/dependence: Results from the National Epidemiological Survey on Alcohol and Related Conditions. *Addiction, 102*, 94–104. doi:10.1111/j.1360-0443.2006.01630.x
- Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York, NY: Wiley. doi:10.1002/0471249688
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- Ansara, D. L., & Hindin, M. J. (2010). Exploring gender differences in the patterns of intimate partner violence in Canada: A latent class approach.

- Journal of Epidemiology and Community Health*, 64, 849–854. doi:10.1136/jech.2009.095208
- Asparouhov, T., & Muthén, B. O. (2013). *Auxiliary variables in mixture modeling: A 3-step approach using Mplus* (Mplus Web Notes: No. 15, Version 6). Retrieved from http://statmodel.com/examples/webnotes/AuxMixture_submitted_corrected_webnote
- Bakk, Z. (2011). *Bias correction methods for three-step latent class modeling with covariates* (Unpublished master's thesis). Tilberg University, Tilberg, Netherlands.
- Bandein-Roche, K., Miglioretti, D. L., Zeger, S. L., & Rathouz, P. J. (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association*, 92, 1375–1386. doi:10.1080/01621459.1997.10473658
- Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). *Latent variable models and factor analysis: A unified approach* (3rd ed.). West Sussex, UK: Wiley. doi:10.1002/9781119970583
- Bolck, A., Croon, M. A., & Hagenaars, J. A. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis*, 12, 3–27. doi:10.1093/pan/mp001
- Bornovalova, M. A., Levy, R., Gratz, K. L., & Lejuez, C. W. (2010). Understanding the heterogeneity of BPD symptoms through latent class analysis: Initial results and clinical correlates among inner-city substance users. *Psychological Assessment*, 22, 233–245. doi:10.1037/a0018493
- Capaldi, D. M., & Clark, S. (1998). Prospective family predictors of aggression toward female partners for at-risk young men. *Developmental Psychology*, 34, 1175–1188. doi:10.1037/0012-1649.34.6.1175
- Capaldi, D. M., Feingold, A., Kim, H. K., Yoerger, K., & Washburn, I. J. (2013). Heterogeneity in growth and desistance of alcohol use for men in their 20s: Prediction from early risk factors and association with treatment. *Alcoholism: Clinical and Experimental Research*, 37, Issue Supplement s1, E347–E355. doi:10.1111/j.1530-0277.2012.01876.x
- Capaldi, D. M., & Kim, H. K. (2007). Typological approaches to violence in couples: A critique and alternative conceptual approach. *Clinical Psychology Review*, 27, 253–265. doi:10.1016/j.cpr.2006.09.001
- Carbone-Lopez, K., Kruttschnitt, C., & Macmillan, R. (2006). Patterns of intimate partner violence and their associations with physical health, psychological distress, and substance use. *Public Health Reports*, 121, 382–392.
- Clark, C. L., & Muthén, B. (2009). *Relating latent class analysis results to variables not included in the analysis*. Manuscript submitted for publication. Retrieved from <http://www.statmodel.com/download/relatinglca.pdf>
- Cohen, P., Cohen, J., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral analysis* (3rd ed.). Mahwah, NJ: Erlbaum.
- Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis*. Hoboken, NJ: Wiley.
- Ehlers, C. L., Gilder, D. A., Gizer, I. R., & Wilhelmsen, K. C. (2009). Heritability and a genome-wide linkage analysis of a Type II/B cluster construct for cannabis dependence in an American Indian community. *Addiction Biology*, 14, 338–348. doi:10.1111/j.1369-1600.2009.00160.x
- Feingold, A. (2009). Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological Methods*, 14, 43–53. doi:10.1037/a0014699
- Feingold, A. (2012). A regression framework for effect size assessments in longitudinal modeling of group differences. *Review of General Psychology*. Advance online publication. doi:10.1037/a0030048
- Feingold, A., Ball, S. A., Kranzler, H. R., & Rounsaville, B. J. (1996). Generalizability of the Type A/Type B distinction across different psychoactive substances. *The American Journal of Drug and Alcohol Abuse*, 22, 449–462. doi:10.3109/00952999609001671
- Feingold, A., Kerr, D. C. R., & Capaldi, D. M. (2008). Associations of substance use problems with intimate partner violence for at-risk men in long-term relationships. *Journal of Family Psychology*, 22, 429–438. doi:10.1037/0893-3200.22.3.429
- Fleiss, J. L., & Berlin, J. A. (2009). Effect sizes for dichotomous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds., 2nd ed), *The handbook of research synthesis* (pp. 237–253). New York: Russell Sage.
- Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61, 215–231. doi:10.1093/biomet/61.2.215
- Gudicha, D. W., & Vermunt, J. K. (in press). Mixture model clustering with covariates using adjusted three-step approaches. In B. Lausen, D. van den Poel, & A. Ultsch (Eds.), *Algorithms from and for nature and life: Studies in classification, data analysis, and knowledge organization*. Heidelberg, Germany: Springer-Verlag GmbH.
- Guo, J., & Wall, E. (2006). Latent class regression on latent factors. *Biostatistics*, 7, 145–163.
- Haddock, C. K., Rindskopf, D., & Shadish, W. R. (1998). Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological Methods*, 3, 339–353. doi:10.1037/1082-989X.3.3.339
- Hagenaars, J. A. (1993). *Loglinear models with latent variables*. Newbury Park, CA: Sage.
- Hagenaars, J. A., & McCutcheon, A. L. (Eds.). (2002). *Applied latent class analysis*. New York, NY: Cambridge University Press. doi:10.1017/CBO9780511499531
- Holtzworth-Munroe, A. (2000). A typology of men who are violent toward their female partners: Making sense of the heterogeneity in husband violence. *Current Directions in Psychological Science*, 9, 140–143. doi:10.1111/1467-8721.00079
- Holtzworth-Munroe, A., Meehan, J. C., Stuart, G. L., Herron, K., & Rehman, U. (2000). Testing the Holtzworth-Munroe and Stuart (1994) batterer typology. *Journal of Consulting and Clinical Psychology*, 68, 1000–1019. doi:10.1037/0022-006X.68.6.1000
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York, NY: Wiley. doi:10.1002/0471722146
- Jackson, K. M., Sher, K. J., & Wood, P. K. (2000). Trajectories of concurrent substance use disorders: A developmental, typological approach to comorbidity. *Alcoholism: Clinical and Experimental Research*, 24, 902–913. doi:10.1111/j.1530-0277.2000.tb02072.x
- Klostermann, K., Mignone, T., & Chen, R. (2009). Subtypes of alcohol and intimate partner violence: A latent class analysis. *Violence and Victims*, 24, 563–576. doi:10.1891/0886-6708.24.5.563
- Kranzler, H. R., Wilcox, M., Weiss, R. D., Brady, K., Hesselbrock, V., Rounsaville, B., . . . Gelernter, J. (2008). The validity of cocaine dependence subtypes. *Addictive Behaviors*, 33, 41–53. doi:10.1016/j.addbeh.2007.05.011
- Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston, MA: Houghton Mifflin.
- Moffitt, T. E. (1993). Adolescence-limited and life-course antisocial behavior: A developmental taxonomy. *Psychological Review*, 100, 674–701. doi:10.1037/0033-295X.100.4.674
- Mplus Technical Appendices. (2010). *Wald test of mean equality for potential latent class predictors*. Retrieved from <http://www.statmodel.com/download/meantest2.pdf>
- Muthén, B. O. (2006). Should substance use disorders be considered as categorical or dimensional? *Addiction*, 101, 6–16. doi:10.1111/j.1360-0443.2006.01583.x
- Muthén, B. O. (2008). Latent variable hybrids: Overview of old and new models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 1–24). Charlotte, NC: Information Age Publishing.
- Muthén, B. O., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, 55, 463–469. doi:10.1111/j.0006-341X.1999.00463.x

- Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.
- Nylund, K., Asparouhov, T., & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14*, 535–569. doi:10.1080/10705510701575396
- Sher, K. J., Jackson, K. M., & Steinley, D. (2011). Alcohol use trajectories and the ubiquitous cat's cradle: Cause for concern? *Journal of Abnormal Psychology, 120*, 322–335. doi:10.1037/a0021813
- Skrondal, A., & Rabe-Hesketh, S. (2005). Structural equation modeling: Categorical variables. In B. S. Everitt & D. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (Vol. 4, pp. 1905–1910). Hoboken, NJ: Wiley. doi:10.1002/0470013192.bsa596
- Straus, M. A. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics (CT) Scales. *Journal of Marriage and Family, 41*, 75–88. doi:10.2307/351733
- Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis, 18*, 450–469. doi:10.1093/pan/mpq025
- Wang, C.-P., Brown, C. H., & Bandeen-Roche, K. (2005). 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, 100*, 1054–1076. doi:10.1198/016214505000000501
- World Health Organization. (1997). *Composite International Diagnostic Interview (CIDI): Core version 2.1*. Geneva, Switzerland: Author.
- Zucker, R. A., Ellis, D. A., Fitzgerald, H. E., Bingham, C. R., & Sanford, K. (1996). Other evidence for at least two alcoholisms: II. Life course variation in antisociality and heterogeneity of alcoholic outcome. *Development and Psychopathology, 8*, 831–848. doi:10.1017/S0954579400007458

AQ: 3

Appendix A

Mplus Input Statements for One-Step Categorical Analysis With Auxiliary Variables

Latent Class Analysis With Covariates (LCA-C)

TITLE: LCA-C: LCA of IPV indicators with
SUD group as covariate

DATA: FILE IS F:\Data.dat; ! file containing IPV indicators and
SUD group

VARIABLE:
NAMES ARE Family IPV1-IPV9 SU1-SU7 IPVClass Poly
Soft;
USEVARIABLES ARE IPV1-IPV9 Poly Soft;
CATEGORICAL = IPV1-IPV9; ! defining indicators as categorical
variables
MISSING = ALL(-9999); ! defining missing data value
CLASSES = cIPV (3); ! defining latent class name and number
of classes

ANALYSIS:
TYPE = MIXTURE; ! requesting a mixture analysis
STARTS = 100 10; ! requesting 100 initial random starting
values and 10 final stage optimizations

MODEL: ! model specification follows
%OVERALL% ! defining effects are for all classes
cIPV ON Soft; ! regressing IPV indicators on soft drug
SUD group
cIPV ON Poly; ! regressing IPV indicators on poly substance
SUD group

OUTPUT:
CINTERVAL; ! requesting confidence intervals

Latent Class Analysis With a Distal Outcome (LCA-D)

TITLE: LCA-D: LCA of substance use indicators with
IPV group as distal outcome

DATA: FILE IS F:\Data.dat; ! file containing
SUD indicators and IPV group

VARIABLE:
NAMES ARE Family IPV1-IPV9 SU1-SU7 IPVClass Poly
Soft;
USEVARIABLES ARE SU1-SU7 IPVClass;
CATEGORICAL = SU1-SU7; ! defining indicators as categorical
variables
NOMINAL = IPVClass; ! defining IPV class as a nominal
variable
MISSING = ALL(-9999); ! defining missing data value
CLASSES = cSU (3); ! defining latent class name and number
of classes

ANALYSIS:
TYPE = MIXTURE; ! requesting a mixture analysis
STARTS = 100 10; ! requesting 100 initial random starting
values and 10 final stage optimizations

MODEL: ! model specification follows
%cSU#1% ! defining below effects are for first SUD class only
[SU1\$1 - SU7\$1]; ! SUD indicators
[IPVClass#1] (IPV1SUD1m); ! class mean for IPV group 1,
SUD class 1
[IPVClass#2] (IPV2SUD1m); ! class mean for IPV group 2,
SUD class 1
%cSU#2% ! defining below effects are for second SUD class
only
[SU1\$1 - SU7\$1]; ! SUD indicators
[IPVClass#1] (IPV1SUD2m); ! class mean for IPV group 1,
SUD class 2
[IPVClass#2] (IPV2SUD2m); ! class mean for IPV group 2,
SUD class 2
%cSU#3% ! defining below effects are for third SUD class only
[SU1\$1 - SU7\$1]; ! SUD indicators

(Appendices continue)

[IPVClass#1] (IPV1SUD3m); ! class mean for IPV group 1,
SUD class 3
[IPVClass#2] (IPV2SUD3m); ! class mean for IPV group 2,
SUD class 3
MODEL CONSTRAINT: ! calculating log odds ratios and signif-
icance tests using difference between class means
NEW LogGLP LogPAP LogGLS LogPAS;
LogGLP = IPV1SUD1m - IPV1SUD3m;
LogPAP = IPV2SUD1m - IPV2SUD3m;
LogGLS = IPV1SUD2m - IPV1SUD3m;
LogPAS = IPV2SUD2m - IPV2SUD3m;
OUTPUT:
CINTERVAL; ! requesting confidence intervals

Structural Equation Modeling With Categorical Variables (Categorical SEM)

TITLE: C-SEM: Categorical SEM where SUD class predicts IPV
class
DATA: FILE IS F:\Data.dat; ! file containing IPV and SUD
indicators
VARIABLE:
NAMES ARE Family IPV1-IPV9 SU1-SU7 IPVClass Poly
Soft;
USEVARIABLES ARE IPV1-IPV9 SU1-SU7;
CATEGORICAL = IPV1-IPV9 SU1-SU7; ! defining indicators
as categorical variables

MISSING = ALL(999); ! defining missing data value
CLASSES = cSU (3) cIPV (3); ! defining latent class names
and number of classes
ANALYSIS:
TYPE = MIXTURE; ! requesting a mixture analysis
STARTS = 100 10; ! requesting 100 initial random starting
values and 10 final stage optimizations
MODEL: ! model specification follows
%OVERALL% ! defining effects are for all IPV and SUD
classes
cIPV ON cSU; ! regressing IPV class on SUD class
MODEL cIPV: ! defining LCA for IPV
%cIPV#1%
%cIPV#2%
%cIPV#3%
[IPV1\$1-IPV9\$1]; ! IPV indicators for all 3 IPV classes
MODEL cSU: ! defining LCA for SUD
%cSU#1%
%cSU#2%
%cSU#3%
[SU1\$1-SU7\$1]; ! SUD indicators for all 3 SUD classes
OUTPUT:
CINTERVAL; ! requesting confidence intervals

Note. SUD = substance use disorder symptoms; IPV = intimate partner
violence (aggression).

(Appendices continue)

Appendix B

Mplus Input Statements for Three-Step Categorical Analyses With Auxiliary Variables

Latent Class Analysis With Covariates (LCA-C)

TITLE: Three-Step LCA-C: LCA of IPV indicators with SUD group as covariate
 DATA: FILE IS F:\Data.dat; ! file containing IPV indicators and SUD group
 VARIABLE:
 NAMES ARE Family IPV1-IPV9 SU1-SU7 IPVClass Poly Soft;
 USEVARIABLES ARE IPV1-IPV9;
 CATEGORICAL = IPV1-IPV9; ! defining indicators as categorical variables
 MISSING = ALL(-9999); ! defining missing data value
 CLASSES = cIPV (3); ! defining latent class name and number of classes
 AUXILIARY = Poly (r3step) Soft (r3step); ! defining covariates using corrected 3-step method
 ANALYSIS:
 TYPE = MIXTURE; ! requesting a mixture analysis
 STARTS = 100 10; ! requesting 100 initial random starting values and 10 final stage optimizations
 MODEL:
 %OVERALL% ! defining effects are for all classes
 [IPV1\$1-IPV9\$1]; ! IPV indicators
 OUTPUT:
 CINTERVAL; ! requesting confidence intervals

Structural Equation Modeling With Categorical Variables (Categorical SEM)

TITLE: Three-Step Categorical SEM. See Mplus Web Notes: No. 15, Version 6 (Asparouhov & Muthén, 2013) for method to calculate the classification uncertainty rates used in the MODEL command
 DATA: FILE IS F:\Data.dat; ! file containing predicted IPV and SUD classes
 VARIABLE:
 NAMES ARE Family SUDClass IPVClass;
 USEVARIABLES = SUDClass IPVClass;
 NOMINAL = SUDClass IPVClass; ! defining indicators as nominal
 MISSING = ALL(-9999); ! defining missing data value
 CLASSES = cSUD(3) cIPV(3); ! defining latent class names and number of classes
 ANALYSIS:
 TYPE = MIXTURE; ! requesting a mixture analysis
 MODEL: ! model specification follows

%OVERALL% ! defining effects are for all classes
 cIPV ON cSUD; ! regression IPV class on SUD class
 MODEL cSUD: ! defining LCA for SUD variables
 %cSUD#1% ! defining below effects are for first SUD class only
 [SUDClass#1 @8.279]; ! fixing class means to account for classification uncertainty rate
 [SUDClass#2 @5.180]; ! fixing class means to account for classification uncertainty rate
 %cSUD#2% ! defining below effects are for second SUD class only
 [SUDClass#1 @0.209]; ! fixing class means to account for classification uncertainty rate
 [SUDClass#2 @3.407]; ! fixing class means to account for classification uncertainty rate
 %cSUD#3% ! defining below effects are for third SUD class only
 [SUDClass#1 @-4.989]; ! fixing class means to account for classification uncertainty rate
 [SUDClass#2 @-2.439]; ! fixing class means to account for classification uncertainty rate
 MODEL cIPV: ! defining LCA for IPV variables
 %cIPV#1% ! defining below effects are for first IPV class only
 [IPVClass#1 @13.766]; ! fixing class means to account for classification uncertainty rate
 [IPVClass#2 @10.789]; ! fixing class means to account for classification uncertainty rate
 %cIPV#2% ! defining below effects are for second IPV class only
 [IPVClass#1 @-1.242]; ! fixing class means to account for classification uncertainty rate
 [IPVClass#2 @2.697]; ! fixing class means to account for classification uncertainty rate
 %cIPV#3% ! defining below effects are for third IPV class only
 [IPVClass#1 @-13.761]; ! fixing class means to account for classification uncertainty rate
 [IPVClass#2 @-2.881]; ! fixing class means to account for classification uncertainty rate
 OUTPUT:
 CINTERVAL; ! requesting confidence intervals

Note. For commands to conduct LCA-C and LCA-D with continuous auxiliary variables in Mplus, see [Asparouhov and Muthén \(2013\)](#).

Received July 30, 2012

Revision received December 3, 2012

Accepted December 4, 2012 ■