0145-6008/00/2406-0882$03.00/0
ALCOHOLISM: CLINICAL AND EXPERIMENTAL RESEARCH

Vol. 24, No. 6
June 2000

Integrating Person-Centered and Variable-Centered Analyses: Growth Mixture Modeling With Latent Trajectory Classes

Bengt Muthén and Linda K. Muthén

Background: Many alcohol research questions require methods that take a person-centered approach because the interest is in finding heterogeneous groups of individuals, such as those who are susceptible to alcohol dependence and those who are not. A person-centered focus also is useful with longitudinal data to represent heterogeneity in developmental trajectories. In alcohol, drug, and mental health research the recognition of heterogeneity has led to theories of multiple developmental pathways.

Methods: This paper gives a brief overview of new methods that integrate variable- and person-centered analyses. Methods discussed include latent class analysis, latent transition analysis, latent class growth analysis, growth mixture modeling, and general growth mixture modeling. These methods are presented in a general latent variable modeling framework that expands traditional latent variable modeling by including not only continuous latent variables but also categorical latent variables.

Results: Four examples that use the National Longitudinal Survey of Youth (NLSY) data are presented to illustrate latent class analysis, latent class growth analysis, growth mixture modeling, and general growth mixture modeling. Latent class analysis of antisocial behavior found four classes. Four heavy drinking trajectory classes were found. The relationship between the latent classes and background variables and consequences was studied.

Conclusions: Person-centered and variable-centered analyses typically have been seen as different activities that use different types of models and software. This paper gives a brief overview of new methods that integrate variable- and person-centered analyses. The general framework makes it possible to combine these models and to study new models serving as a stimulus for asking research questions that have both person- and variable-centered aspects.

Key Words: Latent Variables, Mixtures, Latent Trajectory Classes, Unobserved Heterogeneity, Developmental Pathways.

COMMONLY USED STATISTICAL methods such as regression analysis, factor analysis, and structural equation modeling take a variable-centered approach to data analysis. In these methods, the focus is on relationships among variables. The goal is to predict outcomes, study how constructs influence their indicators, and relate independent and dependent variables in structural equations. Many alcohol research questions, however, require methods that also take a person-centered approach. Examples of such methods include cluster analysis, finite mixture analysis, latent class analysis, and latent transition analysis. In these methods, the focus is on relationships among individuals. The goal is to group individuals into categories, each one of which contains individuals who are similar to each other and different from individuals in other categories.

A person-centered focus is useful in alcohol research, where data often include heterogeneous groups of individuals such as those who are susceptible to alcohol dependence and those who are not. A person-centered focus also is useful with longitudinal data to represent heterogeneity in developmental trajectories. In alcohol, drug, and mental health research the recognition of heterogeneity has led to theories of multiple developmental pathways. For example, Schulenberg et al. (1996) studied binge drinking patterns for young adults in the Monitoring the Future Study by following high-school seniors after graduation during four waves of observations through age 26. Converging or diverging trajectories that end up or start at the same level

From the Graduate School of Education and Information Studies (B.M), University of California–Los Angeles, Los Angeles, California, and Muthén & Muthén (L.K.M.), Los Angeles, California.

Received for publication November 5, 1999; accepted February 15, 2000.

Presented at the Annual Meeting of the Research Society on Alcoholism, June 26–July 1, 1999, Santa Barbara, California.

Supported by Grant K02 AA 00230 from NIAAA, by Grant R21 AA10948 from NIAAA, by Grant 016636 from ABMRF, and by SBIR Contract No. N43AA42008 from NIAAA.

The statistical program (Mplus) used for the analysis of data in this paper was developed by the authors and is commercially available through a private company, Muthén & Muthén, Los Angeles, CA, which they own.

Reprint requests: Bengt O. Muthén, Graduate School of Education and Information Studies, UCLA, Moore Hall, Box 951521, Los Angeles CA 90095-1521; Fax: 310-397-8621; E-mail: bmuthen@ucla.edu

but have different starting or end points were related to individual background characteristics. Multiple pathways also were discussed in terms of subtypes of alcoholism (Zucker, 1994), stages of progression in drug involvement from adolescence to adulthood (Kandel et al., 1992), adolescent-limited versus life-course-persistent antisocial behavior (Moffitt, 1993), and normal versus pathogenic prostate-specific antigen development (Pearson et al., 1994). Seminal work on the methodolgy of developmental pathways has been carried out by Nagin (Nagin and Land, 1993; Nagin, 1999) using applications in criminology.

This paper gives a brief overview of new methods that integrate variable- and person-centered analyses. Methods to be discussed include latent class analysis, latent transition analysis, latent class growth analysis, growth mixture modeling, and general growth mixture modeling. These methods are presented in a general latent variable modeling framework that expands traditional latent variable modeling by including not only continuous latent variables but also categorical latent variables. Categorical latent variables correspond to the person-centered component in that the categories describe groups of individuals who are homogeneous within a given category and are heterogeneous across categories. This type of modeling can be carried out using new methodology that is part of the Mplus program (Muthén and Muthén, 1998). Four examples relevant to alcohol research are presented using data from the National Longitudinal Survey of Youth (NLSY). The NLSY is an annual national survey of young people in the United States and includes alcohol-related measures collected over ages 18–37.

OVERVIEW OF METHODS

Alcohol data often come from a heterogeneous population. Often population heterogeneity is observed and can be represented by variables in a model. Even when population heterogeneity is not observed, however, it can be taken into account by using latent classes. Methods that take unobserved heterogeneity into account often are referred to as person-centered. A categorical latent variable can be used to represent the latent classes. The general idea is that each latent class corresponds to a subpopulation that has its own set of parameter values. This approach to heterogeneity can be used in conjunction with the estimation of any model from a univariate mean to a complex growth model. A set of increasingly more complex methods that address the problem of unobserved population heterogeneity is discussed.

Latent Class Analysis

Latent class analysis (LCA) describes how the probabilities of a set of observed categorical variables or indicators vary across groups of individuals where group membership is not observed. For example, the observed categorical variables may correspond to a set of dichotomous diagnostic criteria or symptom items, and the latent classes may describe the presence or absence of an alcohol use disorder such as alcohol dependence. LCA refers to the unobserved groups of individuals as latent classes. The object of LCA is to find the smallest number of latent classes that can describe the associations among a set of observed categorical variables. The analysis adds classes stepwise until the model fits the data well. The parameters of the model are the probabilities of being in each class and the probabilities of fulfilling each criterion given class membership. In addition, the latent class model provides estimates of class probabilities for each individual. These values are called posterior probabilities. LCA may relate the probability of class membership to a set of background variables.

Several applications of LCA can be found in the alcohol and mental health literature. For example, Bucholz et al. (1996) attempted to identify distinctive subtypes of alcoholics. They applied latent class analysis to a set of lifetime symptoms of alcohol dependence using data from the Collaborative Study of the Genetics of Alcoholism and related the four-class solution to a set of background variables. Nestadt et al. (1994) studied diagnostic criteria for schizophrenia. Latent class analysis has been proposed for use in medical diagnosis by Rindskopf and Rindskopf (1986), Uebersax and Grove (1990), and Clogg (1995), among others. The papers by Jackson, Sher, and Wood and by Bucholz in this issue provide examples of LCA in the areas of alcohol and tobacco use. For an overview of LCA methods and applications, see Clogg (1995).

LCA can be compared to factor analysis. The primary objective of LCA is to find groups of individuals who are similar using a categorical latent variable. The primary objective of factor analysis is to find the smallest number of factors or dimensions that can explain the relationships among a set of observed variables using continuous latent variables. The choice of factor analysis or LCA is a matter of which model is most useful in practice. It cannot be determined statistically, because data that have been generated by an m-dimensional factor analysis model can be fit perfectly by a latent class model with m+1 classes (see Bartholomew, 1987, pp. 36–38). The categorical conceptualization is particularly suitable when diagnosis is the primary focus and diagnostic criteria or symptom items are analyzed. In LCA the choice of cutpoints on underlying dimensions is avoided, and the classification is provided directly by the model.

Latent Transition Analysis

LCA studies class membership in cross-sectional data, whereas latent transition analysis (LTA) studies change in class membership using longitudinal data. LTA uses multiple indicators at each time point to define a latent class variable for each time point. The main objective of the analysis is to study the probability of a transition from a class at one time point to a class at the next time point.

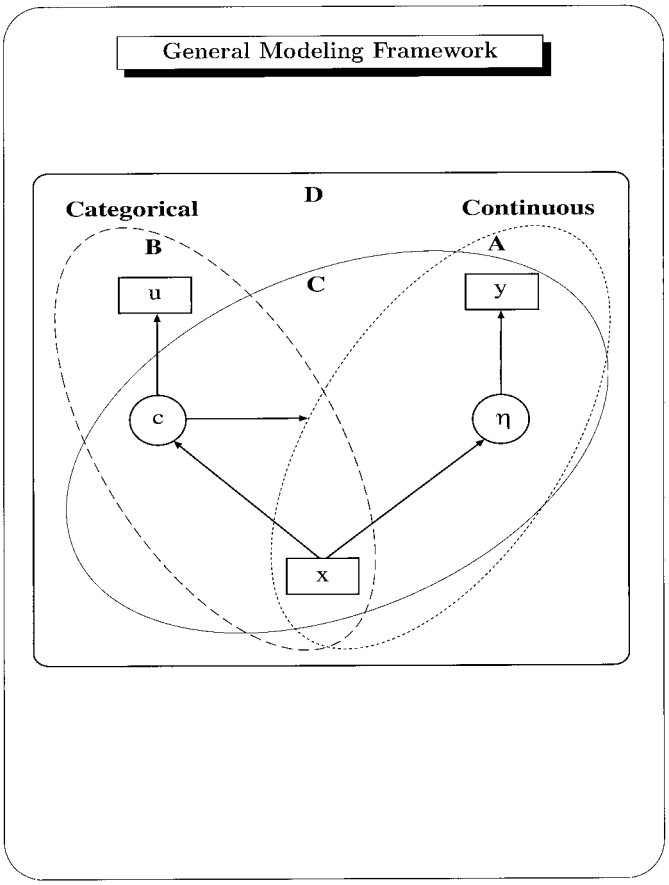

Fig. 1. General modeling framework.

LTA has been proposed for public health research by Graham et al. (1991), Collins and Wugalter (1992), Bandeen-Roche et al. (1997), and Reboussin et al. (1998). Graham et al. (1991) used LTA to study adolescent substance use onset, contrasting models with starting classes defined by alcohol only versus alcohol and tobacco, and evaluating the effects of a prevention program in grades seven and eight. Reboussin et al. (1998) studied transitions between health-risk states related to weapons-carrying behavior in a school sample measured annually for 5 years. Bandeen-Roche et al. (1997) studied physical disability classes related to age and arthritis. The paper by Bucholz in this issue provides an example of LTA in the area of alcohol use. An overview of LTA is given in Collins et al. (1997).

Latent Class Growth Analysis

Latent class growth analysis (LCGA) uses a single outcome variable measured at multiple time points to define a latent class model in which the latent classes correspond to different growth curve shapes for the outcome variable. For example, in a two-class model, one class may have linear growth on the outcome variable and the other may have quadratic growth. The object of the analysis is to estimate the different growth curve shapes and the class probabilities. In addition, posterior probabilities of class membership for all individuals can be computed. The model may include background variables influencing the class probabilities.

Nagin and Land (1993), Nagin and Tremblay (1999), Jones et al. (1998), and Nagin (1999) used LCGA to study the developmental classes of adolescent-limited and chronic criminal involvement in a sample of males from ages 8 to 32 years. The article by Hill, White, Chung, and Hawkins in this issue provides an example of LCGA in the area of binge drinking, and Nagin (1999) gives an overview with applications.

Growth Mixture Modeling

Growth mixture modeling (GMM) is based on conventional growth modeling. Conventional growth modeling is

Table 1. LCA and EFA for Antisocial Behavior (n = 7326)

	LCA Solution: Categorical Factors				EFA Solution: Continuous Factors		
	Class 1	Class 2	Class 3	Class 4	Factor 1	Factor 2	Factor 3
Property	**0.78**	0.18	0.28	0.02	**0.65**	0.19	−0.04
Fighting	0.74	0.20	**0.53**	0.09	0.19	**0.60**	−0.13
Shoplifting	**0.82**	0.42	0.31	0.07	**0.61**	−0.03	0.18
Stole <$50	**0.70**	0.27	0.21	0.05	**0.85**	−0.21	0.05
Stole > $50	**0.37**	0.03	0.03	0.00	**0.81**	0.00	0.01
Use of force	0.26	0.02	0.08	0.01	0.34	0.37	−0.01
Seriously threatening	0.82	0.38	**0.68**	0.10	−0.11	**0.89**	0.03
Intent to injure	0.40	0.08	0.19	0.00	−0.11	**0.83**	0.08
Use marijuana	0.90	**0.96**	0.36	**0.22**	−0.02	0.00	**0.88**
Use other drugs	0.60	**0.58**	0.03	0.01	0.01	−0.02	**0.88**
Sold marijuana	0.51	0.26	0.01	0.00	0.15	0.07	**0.74**
Sold hard drugs	0.14	0.04	0.00	0.00	0.19	0.09	**0.59**
"Con" someone	0.64	0.20	**0.38**	0.07	**0.43**	0.25	−0.07
Take auto	0.36	0.08	0.11	0.01	**0.45**	0.15	0.07
Broke into building	**0.43**	0.03	0.04	0.00	**0.80**	0.03	0.01
Held stolen goods	**0.67**	0.10	0.12	0.00	**0.69**	0.11	0.06
Gambling operation	0.15	0.02	0.03	0.00	0.28	0.36	0.08
Class probability	0.09	0.18	0.25	0.47			

EFA, exploratory factor analysis; LCA, latent class analysis. Bold numbers, see text for explanation.

used to analyze longitudinal data by relating an observed outcome variable to time or to a time-related variable such as age. Individual variation in growth is captured by the fact that coefficients in the growth model are random—that is, they vary across individuals. In growth modeling in a structural equation modeling framework, the random coefficients are continuous latent variables or growth factors. Growth modeling estimates the variation of the growth factors and the influence of background variables on this variation. For an introductory overview of latent variable growth modeling and applications of this methodology see Muthén and Khoo (1998).

GMM combines features of conventional growth modeling and LCGA. Conventional growth modeling estimates a mean growth curve under the assumption that all individuals in the sample come from a single population. Individual variation around the mean growth curve is captured by the estimation of the growth factor variances. LCGA estimates a mean growth curve for each class. No individual variation around the mean growth curves is allowed. As a result, the variation in the growth factors within each class is assumed to be zero. GMM both estimates mean growth curves for each class and captures individual variation around these growth curves by the estimation of growth factor variances for each class.

GMM draws on finite mixture models for growth (see Verbeke and Lesaffre, 1996, and Muthén and Shedden, 1999). Muthén and Shedden (1999) studied normative and non-normative development of heavy drinking using data from the National Longitudinal Study of Youth. For an overview of methods and applications, see Muthén (1998).

General Growth Mixture Modeling

GMM can be incorporated into a more general latent variable framework that allows combinations of the models previously discussed. This is referred to as general growth mixture modeling (GGMM). It is the statistical framework used in Mplus (Muthén and Muthén, 1998) that is summarized in graphical form in Figure 1. Here, u represents categorical indicators of a categorical latent variable c, y represents continuous indicators of a continuous latent variable η, and x represents background variables. Factor analysis and conventional growth modeling can be handled by the model part marked by the ellipse A. LCA, LTA, and LCGA can be handled by the model part marked by the ellipse B. GMM can be handled by the model part marked C. And, combinations of such models can be handled in the most general framework D. For an overview of methods and applications related to longitudinal analysis, see Muthén et al. (1998), Muthén (1998), and Muthén and Muthén (1998).

METHODS

Subjects

The four examples in this paper use data from the National Longitudinal Survey of Youth (Ohio State University, 1994). The NLSY was initiated in 1979 and included a representative sample of 12,686 men and women born between 1957 and 1964. The NLSY data were collected as a multistage probability sample with an oversampling of black, Hispanic, and economically disadvantaged nonblack and non-Hispanic youth and a cross-sectional sample designed to represent the military population. The NLSY is composed of eight birth-year cohorts. As of 1989, the overall retention rate of the sample was 93%. The National Institute on Alcohol Abuse and Alcoholism supported an alcohol supplement in 1982, 1983, 1984, 1985, 1988, 1989, and 1994. Example 1 looks at 7326 respondents. These respondents are the original sample of 10,893, excluding the 1793 respondents in the military subsample that was dropped in 1985, who provided complete information on the variables considered in Example 1. Examples 2, 3, and 4 use the respondents from Example 1, who were born in 1964 and have complete data for the variables used in each example. Sample sizes for examples 2, 3, and 4 are 924, 922, and 1225, respectively.

Measures

Outcome Variables. Antisocial behavior, excluding the use of alcohol, is measured by 17 items administered in 1980. These items assessed the frequency of various behaviors, excluding alcohol use, during the past year (Table 1). With the exception of the use of marijuana, there were few responses for frequencies in excess of 2 or more times. As a result, the items were dichotomized, to 1 or more times in the past year compared to not at all in the past year. Heavy drinking is measured by the question: "How often have you had 6 or more drinks on one occasion during the last 30 days?" The responses are recorded as: never (0); once (1); 2 or 3 times (2); 4 or 5 times (3); 6 or 7 times (4); 8 or 9 times (5); and 10 or more times (6). Data on this variable are used for 1982, 1983, 1984, 1988, 1989, and 1994, whereas survey year 1985 is excluded because a question format

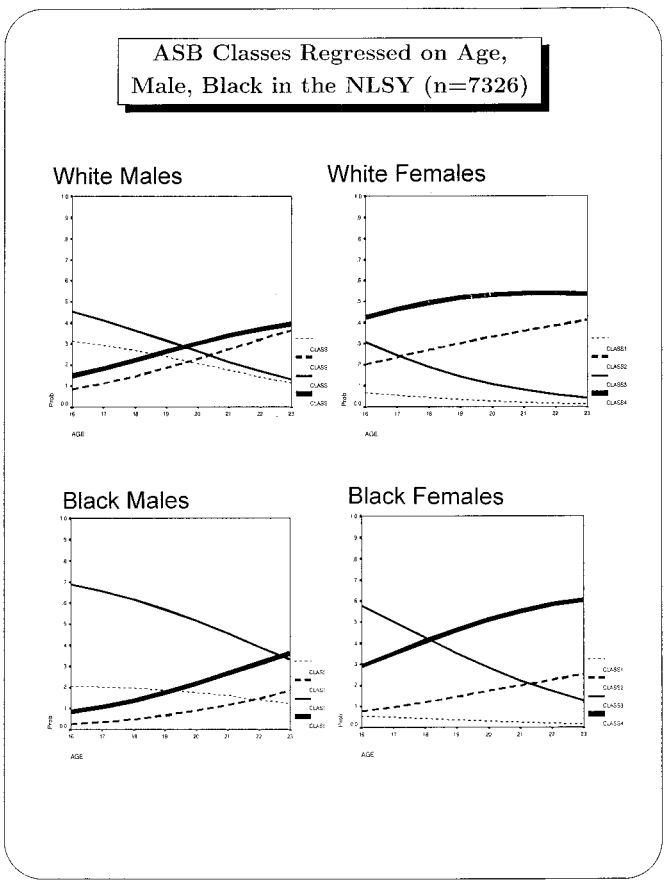

Fig. 2. Antisocial behavior classes related to age, gender, and ethnicity (n = 7326).

change was made that year that may have made across-time comparisons invalid (Harford, 1994). Alcohol dependence is a dichotomous variable based on 22 symptom items measured in 1989 and 1994. The symptom items are designed to approximate the DSM-IV diagnoses of alcohol dependence and abuse. Each symptom item is scored 1 if the event occurred at least once in the last year and 0 otherwise. A factor analysis of these items found two factors. The 17 items that measured the most severe factor were summed and then dichotomized, with 1 representing a sum score of 2 or more and 0 representing a sum score of less than 2. This dichotomous variable is referred to as alcohol dependence.

Background Variables. Background variables are gender, ethnicity, family history of alcohol problems, early onset of drinking, age, high school dropout status, and college education. Gender (Male) is represented by a 0/1 dummy variable, with 1 representing male. Ethnicity (Black, Hisp) is captured by two 0/1 dummy variables, one for black and one for Hispanic, with 1 representing minority status. Family history of alcoholism (FHA) is measured by the 1988 question: "Have any of your relatives listed on this card been alcoholics or problem drinkers at any time in their lives?" Three dummy variables are considered for FHA: first-degree relatives (FH1) only; second- or third-degree relatives (FH23) only; first- or second- and third-degree relatives (FH123). Early onset of drinking (ES) is measured by the 1982 question: "How old were you when you first started drinking?" Early onset is scored as a dummy variable, with early onset defined as starting to drink at age 14 or younger. Dropping out of high school (HSDRP) is measured as a dummy variable defined as not having completed high school by age 22. College education (Coll) is measured as having some college education by age 22 and is scored as a 0/1 dummy variable, with 1 representing some college education by age 22.

EXAMPLES

Four examples that use the NLSY data to illustrate LCA, LCGA, GMM, and GGMM are presented in this section. These examples are presented as methodological illustrations and should not be used for substantive inferences. All analyses were carried out using the Mplus program (Muthén and Muthén, 1998). Mplus input specifications and data for these examples are available at the Mplus web site, http://www.statmodel.com/.

Example 1: LCA and EFA of Antisocial Behavior

In the first example, latent class analysis (LCA) and exploratory factor analysis (EFA) are used to examine 17 dichotomous antisocial behavior items from the NLSY. The person-centered results of LCA are compared to the variable-centered results of EFA. In addition, the probabilities of latent class membership from the LCA are related to the background variables of age, gender, and ethnicity in an LCA with covariates.

Table 1 contains results from an LCA and EFA of the 17 dichotomous antisocial behavior items from the NLSY. In the LCA, the entries are the probabilities of individuals in a class endorsing an item. In the LCA, Class 4 is the most prevalent class (47%). Individuals in Class 4 do not endorse any item with a high probability. The highest probability is for using marijuana, with a value of 0.22. Class 4 can be considered to be the normative class. Class 3 (25%), in contrast to Class 4, endorses the items of fighting, threatening, and conning with much higher probabilities of 0.53, 0.68, and 0.38, respectively. This class can be considered a person offense class. Although individuals in Class 3 endorse smoking marijuana, with a probability of 0.36, this is not considerably different from Class 4. Class 2 (18%), in contrast to Classes 3 and 4, strongly endorses the items of using marijuana and using drugs. This class can be considered a substance involvement class. Class 1 (9%), in contrast to Classes 2, 3, and 4, endorses the items of damaging property, shoplifting, stealing less than $50, stealing more than $50, entering a building, and stealing goods. This class can be considered a property offense class. Probabilities are bolded in the table for items that differentiate the classes.

The four LCA classes are not ordered—that is, the probabilities of all items do not decrease from Class 1 to Class 4. For example, fighting and seriously threatening have lower probabilities for Class 2 than for Class 3. This suggests that there is not a single dimension of antisocial behavior with the four classes representing decreasing levels of severity on this dimension; instead, the classes represent different kinds of antisocial behavior, corresponding to different class profiles of high and low item probabilities.

In the EFA, the entries are factor loadings that show how strongly each factor influences each item. Because the items are dichotomous, the factor loadings are coefficients from probit regressions of the items on the factors. The bolded factor loadings represent the values that are considerably higher on one factor than on any other factor. In the EFA, the first factor relates to property offense, the second factor relates to person offense, and the third factor relates to substance use. A further discussion of the factor analysis of these items is offered in Harford and Muthén (1999).

Three EFA factors and four LCA classes were found to best fit the data. The choice of which method to use depends on the research question being asked. LCA groups individuals and provides information on class profiles. EFA groups items and provides information on which dimensions they measure. Consequently, the pattern of high/low EFA loadings and high/low LCA probabilities is not always the same.

Although the focuses of LCA and EFA differ, there is a relationship between the EFA dimensions and the LCA classes. Individuals in Class 1 are likely to score high on factors 1, 2, and 3, because they endorse many of the items in these factors with a high probability. Individuals in Class 2 are likely to rank high on factor 3, because they endorse many of the factor 3 items with a high probability. They are likely to be lower on the other factors, because they do not endorse the items in these factors with a high probability. Individuals in Class 3 are likely to be high on factor 2, because they endorse many of the factor 2 items with a high probability. They are likely to be lower on the other factors. Individuals in Class 4 are likely to be low on all factors, because their endorsement of all items is low. The patterns of loadings in EFA and probabilities in LCA are not always the same. For example, the second EFA factor is defined by fighting, serious threatening, and intent to injure. Although fighting and seriously threatening also characterize the LCA Class 3, intent to injure has a relatively low probability in Class 3. The probability may be low because individuals who have high probabilities for intent to injure are in Class 1.

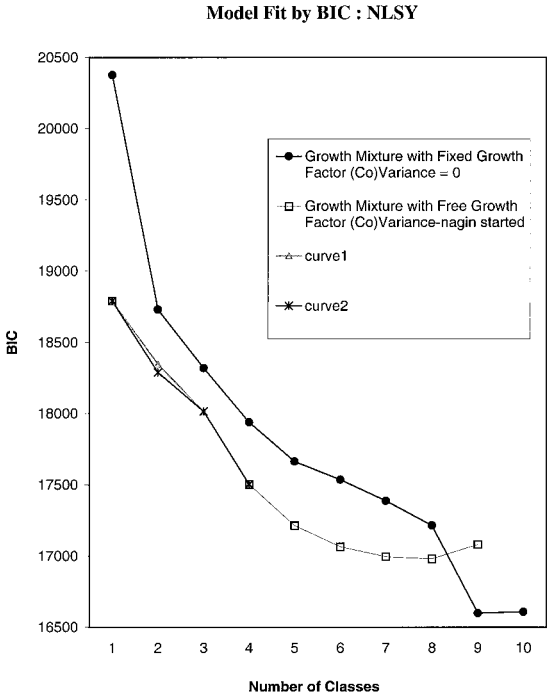

Fig. 3. Model fit evaluated by BIC.

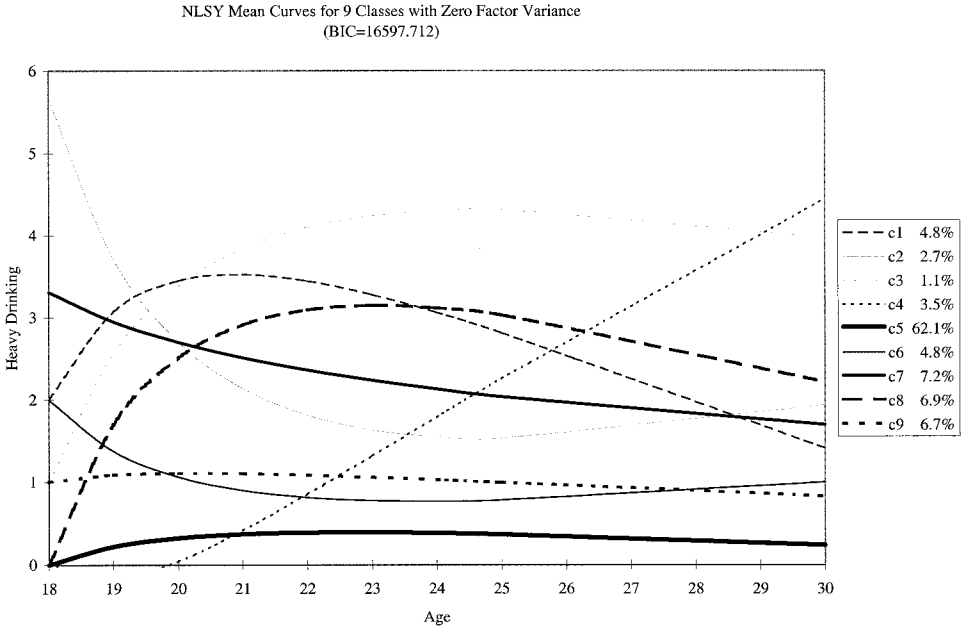

Fig. 4. Estimated heavy drinking curves for the LCGA nine-class solution.

Figure 2 shows the relationships between the probability of membership in each of the four antisocial behavior classes and the background variables of age, gender, and ethnicity. These results are based on LCA with covariates. The probabilities are expressed in terms of multinomial logit regression coefficients (see Agresti, 1990, p. 313). It is seen that Class 3 behavior (person offense) is most prevalent for younger individuals, except for white females. The age at which Class 4 behavior (normative) becomes more prevalent varies greatly across the four groups defined by gender and ethnicity. Also, whereas Class 1 behavior (property offense) declines with age for all four groups, Class 2 behavior (substance involvement) increases with age for all four groups, although it is lower overall for black males and black females.

Example 2: LCGA and GMM of Heavy Drinking

In the second example, latent class growth analysis (LCGA) and growth mixture modeling (GMM) are used to analyze the variable of heavy drinking for the NLSY cohort born in 1964, measured across ages 18–30. The goal is to find different trajectory classes corresponding to individuals following normative and non-normative developmental pathways. LCGA is useful as a first step in determining major types of trajectories, because LCGA is a special case of GMM where the growth factor variances within each class are 0. The GMM analyses build on previous growth studies of these data carried out in Muthén and Muthén (2000) and Muthén and Shedden (1999).

In this example, three different considerations are used to decide on the number of latent classes. The first is the Bayesian information criterion (BIC) statistic that balances two components, maximizing the likelihood and keeping the model parsimonious. A low BIC value indicates a well-fitting model. The second is the classification quality that can be determined by examining the posterior probabilities. The average posterior probability for each class for individuals whose highest probability is for that class should be considerably higher than the average posterior probabilities for the other classes for those individuals. The third consideration is the usefulness of the latent classes in practice. This can be determined by examining the trajectory shapes for similarity, the number of individuals in each class, the number of estimated parameters, and the differences in predictions of consequences.

Figure 3 gives the LCGA BIC values for one through ten classes and the GMM BIC values for one through nine classes (a tenth class could not be found). BIC is a measure that considers both the likelihood value of a model and the number of parameters estimated. A good model, according to BIC, has a high likelihood value without using many parameters. This combination results in a low BIC value.

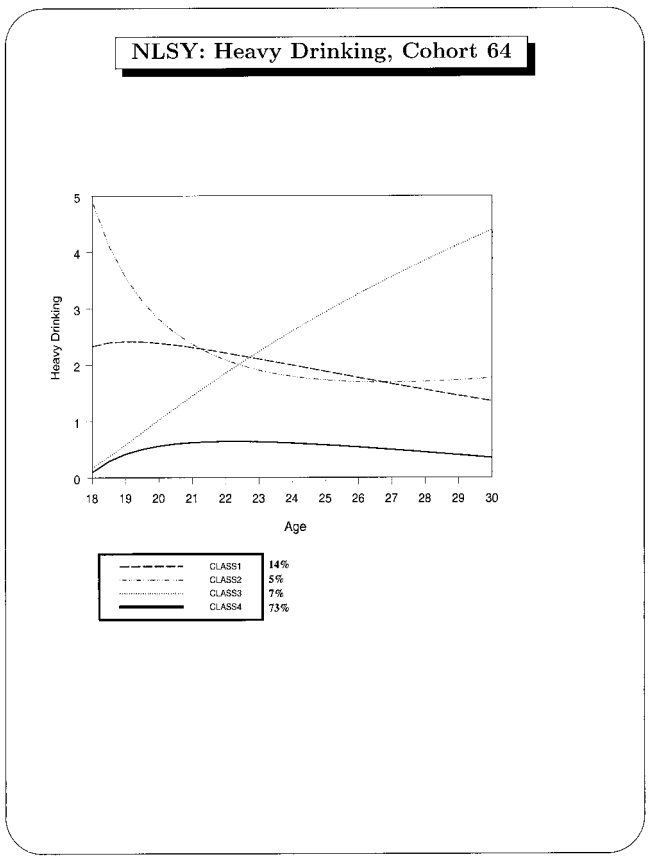

Fig. 5. Estimated heavy drinking curves for the GMM four-class solution (n = 64). Classes 1 and 2 represent early heavy drinking that decreases over time, Class 3 represents increasing heavy drinking, and Class 4 represents the normative class with heavy drinking escalating in the early 20s and later declining.

Table 2. Average Posterior Probabilities for Four- and Nine-Class HD Solutions (n = 924)

	Nine-Class HD Solutions									
Class	1	2	3	4	5	6	7	8	9	*n*
1	**0.93**	0.00	0.00	0.00	0.00	**0.07**	0.00	0.00	0.00	45
2	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	25
3	0.00	0.00	**0.99**	0.00	0.00	0.00	0.00	0.00	**0.01**	10
4	0.00	0.00	0.00	**0.95**	**0.02**	0.00	0.00	**0.02**	0.00	26
5	0.00	0.00	0.00	**0.01**	**0.99**	0.00	0.00	**0.01**	0.00	577
6	**0.05**	0.00	0.00	0.00	0.00	**0.95**	0.00	0.00	0.00	44
7	0.00	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	67
8	0.00	0.00	0.00	**0.06**	**0.07**	0.00	0.00	**0.87**	0.00	67
9	0.00	0.00	**0.01**	0.00	0.00	0.00	0.00	0.00	**0.99**	63

	Four-Class HD Solutions				
Class	1	2	3	4	*n*
1	**1.00**	0.00	0.00	0.00	135
2	0.00	**1.00**	0.00	0.00	46
3	0.00	0.00	**0.95**	**0.05**	60
4	0.00	0.00	**0.01**	**0.99**	683

HD, heavy drinking. Non-zero entries are bolded. Each row contains information for individuals who were most likely to be in the class represented by that row. See text for further information.

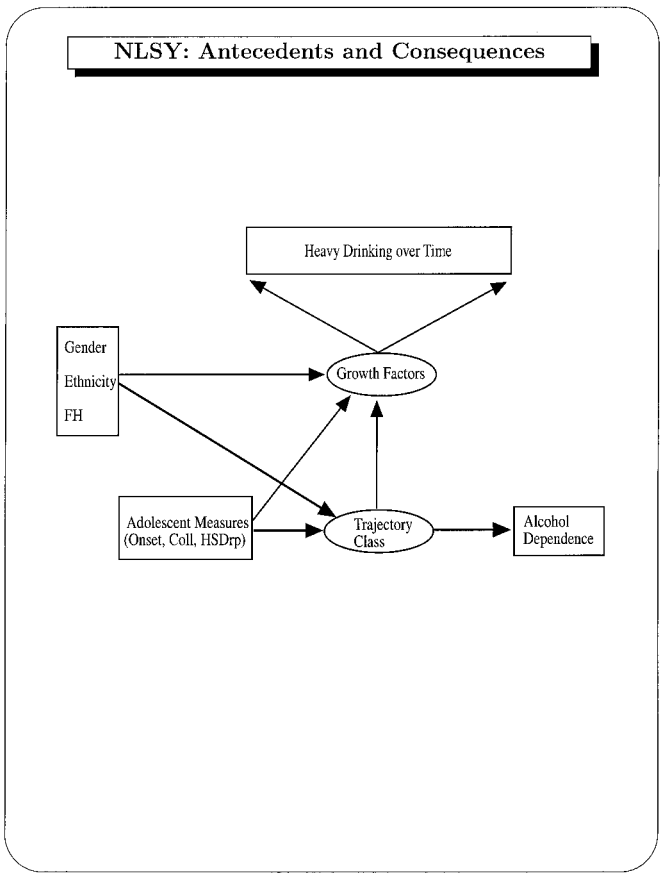

Fig. 6. Heavy drinking related to antecedents and consequences (general growth mixture modeling [GGMM]).

The LCGA Class 1 solution represents a null model where the covariances among the observed variables are 0. The GMM Class 1 solution is a conventional growth model. The BIC curves in Figure 3 show that both LCGA and GMM with two or more classes fit the data better than the null model or conventional growth model. In this example, the best fitting model based on BIC is not obtained until a high number of classes is included, eight for GMM and nine for LCGA. The overall best model as judged by BIC is the nine-class LCGA.

Table 3. Alcohol Dependence as a Function of Heavy Drinking Class (n = 922)

Class	Probability	Odds Ratio
	Nine-Class Solution	
1	0.28	11.61
2	0.24	9.58
3	0.57	40.82
4	0.64	53.39
5	0.03	1.00
6	0.09	2.97
7	0.21	8.00
8	0.22	8.39
9	0.18	6.82
	Four-Class Solution	
1	0.16	3.92
2	0.26	7.06
3	0.60	30.00
4	0.05	1.00

Figure 4 shows the estimated heavy drinking curves for the LCGA nine-class solution and the percentages of individuals in the nine classes.

There appear to be four major types of curves represented by the nine classes. Class 5 represents normative growth. Classes 1, 3, and 8 represent similar classes where heavy drinking has a considerably higher peak around age 21. However, for Class 3 there is not a substantial decrease after age 21. Classes 2, 6, and 7 represent similar classes where heavy drinking already was high at age 18 but declines thereafter. Class 4 represents a class with increasing heavy drinking throughout the age range. Class 9 appears to be an intermediate class between the normative Class 5 and Classes 1, 3, and 8. The eight-class GMM solution also shows these four major types of curves.

Because four major types of curves with some variation are seen, the four-class GMM solution is studied (see Fig. 4). Figure 5 shows the estimated heavy drinking curves for the GMM four-class solution and the percentages of individuals in the four classes. Class 4 represents the normative class with heavy drinking escalating in the early 20s and later declining. Classes 1 and 2 represent early heavy drinking that decreases over time. Class 3 represents increasing heavy drinking.

Table 2 contains information about classification quality for the nine-class LCGA and the four-class GMM solutions. All non-zero entries are bolded. Each row contains information for individuals who were most likely to be in the class represented by that row. For example, row 5 contains information about individuals whose posterior probability for Class 5 was higher than their posterior probabilities for the other eight classes. The averages of their posterior probabilities for being in each class

Table 4. Heavy Drinking Related to Covariates (n = 922)

	Heavy Drinking Classes					
	1 (Down)		2 (High 18)		3 (Up)	
	Estimated	t	Estimated	t	Estimated	t
Male	**1.21**	5.52	**1.25**	3.48	**1.45**	4.73
Black	**−0.89**	−3.43	**−3.14**	−2.86	−0.06	−0.17
Hispanic	**−0.65**	−2.22	−0.35	−0.86	−0.01	−0.03
ES	**1.24**	4.79	**2.05**	5.72	0.71	1.78
FH1	0.03	0.09	−0.21	−0.41	−0.08	−0.16
FH23	0.04	0.15	0.25	0.56	0.08	0.23
FH123	−0.23	−0.58	**1.18**	2.59	**1.00**	2.60
HSDRP	0.57	1.98	0.32	0.76	**0.91**	2.93
College	−0.07	−0.31	**−1.31**	−2.85	**−1.08**	−2.59

ES, early onset of drinking; FH1, first-degree relative; FH23, second- or third-degree relatives only; FH123, first- or second- and third-degree relatives; HSDRP, high school dropout. Bold numbers, see text for explanation. t, ratio of estimate to standard error.

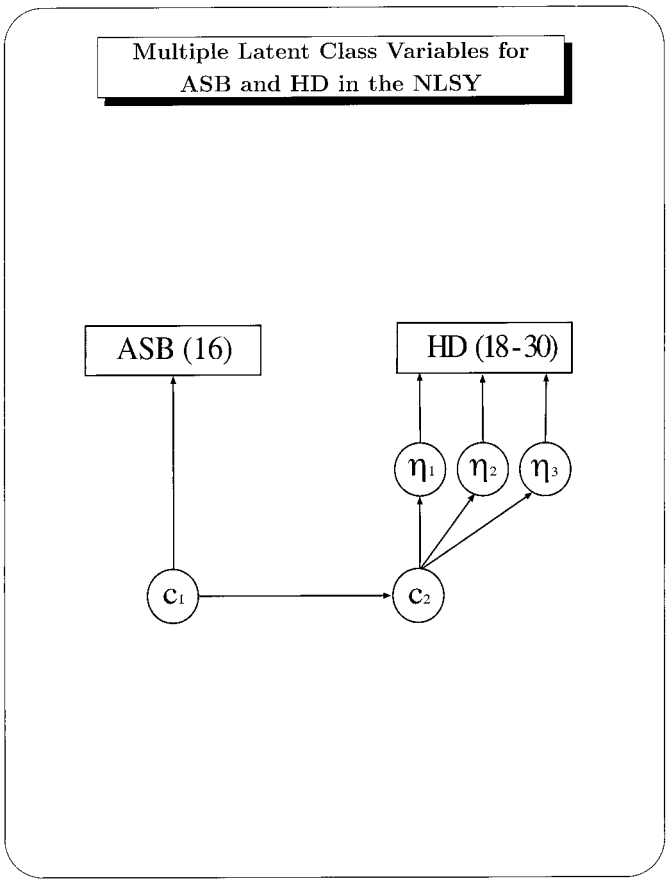

Fig. 7. Heavy drinking classes related to antisocial behavior classes. To the left is the four-class LCA model for 9 of the 17 antisocial items and to the right is the four-class heavy drinking GMM.

are shown in row 5. Good classification quality is obtained when diagonal elements are high and off-diagonal elements are low. Classification quality is high for both solutions, but somewhat better for the four-class solution.

The third consideration for determining the number of latent classes is the usefulness of the latent classes in practice. This can be determined by examining the trajectory shapes for similarity, the number of individuals in each class, the number of estimated parameters, and the differences in predictions of consequences based on different numbers of classes. The first three classes are discussed here, and the fourth is discussed in Example 3.

As previously discussed, the heavy drinking curves for the nine-class LCGA solution shown in Figure 4 are represented by four major curve types, indicating that the LCGA solution may use more classes than necessary to describe the data. The four LCGA curve types are similar to the four estimated heavy drinking curves for the four-class GMM solution shown in Figure 5. The four-class GMM solution apparently incorporates the nine classes of the LCGA by allowing for within-class variation. The numbers of individuals in each class are shown in Table 2. The nine-class LCGA solution has five classes with class size less than 50. Only one GMM class has class size less than 50. It is doubtful that classes with so few individuals allow a trustworthy generalization. The number of estimated parameters for the LCGA solution is 41; the number of estimated parameters for the GMM solution is 27. The smaller number of parameters in the GMM makes it more likely that this model will cross-validate well in a new sample.

Example 3: GGMM of Heavy Drinking

In the third example, the nine LCGA heavy drinking classes and the four GMM heavy drinking classes from Example 2 are related to alcohol dependence at age 30. This analysis also is the final step in determining the optimal number of latent classes for heavy drinking. In addition, the four GMM heavy drinking curve classes are related to the background variables of gender, ethnicity, early onset of drinking, family history of alcohol problems, high school dropout, and attending college.

Figure 6 depicts general growth mixture modeling (GGMM) for development of heavy drinking. It shows how the model for heavy drinking development is incorporated into a larger model, relating heavy drinking to the antecedents of gender, ethnicity, family history of alcohol problems, early onset of drinking, high school dropout status, and college education and the consequence of alcohol dependence.

Table 3 shows the relationship between heavy drinking trajectory class membership and alcohol dependence at age 30 for the nine- and four-class solutions. Probabilities of alcohol dependence and the odds ratios of alcohol dependence relative to the normative class are shown.

For the nine-class solution, Class 5, the normative class, has a probability of 0.03 of being alcohol dependent at age 30. The normative class for the four-class solution is Class 4, with a probability of 0.05 of being alcohol-dependent at age 30. Classes 3 and 4 from the nine-class solution are the two classes with escalating heavy drinking over time and the highest heavy drinking at age 30. They have probabilities of 0.57 and 0.64, respectively, of being alcohol-dependent at age 30. Class 3 of the four-class solution has a probability of 0.60 of being alcohol-dependent at age 30. The other six classes from the nine-class solution have probabilities ranging from 0.09 to 0.28 of being alcohol-dependent at age 30. The other two classes from the four-class solution have probabilities of 0.16 and 0.26 of being alcohol-dependent at age 30. Based on the prediction of alcohol dependence, the four-class solution seems to give sufficient discrimination between the classes.

Considering the information presented in Example 2 and the results presented in the preceding paragraphs, it appears that the four-class solution is most useful in practice. Although, the BIC value was better for

Table 5. Antisocial Behavior and Heavy Drinking (n = 1225)

ASB Class*	ASB Class Probability	Conditional Probabilities Given ASB Class: HD Class† 1	2	3	4
1	0.15	0.38	0.15	0.09	0.38
2	0.14	0.24	0.14	0.05	0.57
3	0.40	0.13	0.04	0.09	0.75
4	0.32	0.05	0.01	0.03	0.93
HD Probability		0.16	0.06	0.06	0.73

* ASB classes (based on 9 selected items): 1 = Property offense; 2 = substance involvement; 3 = person offense; 4 = Normative.
† HD classes (based on ages 18–30): 1 = Down; 2 = High 18; 3 = Up; 4 = Normative.
ASB, antisocial behavior; HD, heavy drinking.

the nine-class solution, the nine-class solution does not appear to have nine distinct curve shapes. Four curve shapes similar to the four-class solution appeared to describe the data. In addition, many classifications were slightly worse in the nine-class solution, and there were several small class sizes. Because the four-class solution is more parsimonious and predicts dependence adequately, it seems the better choice. It may be that three classes are sufficient, given that Class 1 and Class 2 of the four-class solution have similar probabilities for dependence.

Table 4 shows how membership in the four trajectory classes relates to the background variables. The results are multinomial logistic regression coefficients (see Agresti, 1990, p. 313). Class 4 represents the normative class, with heavy drinking escalating in the early 20s and later declining. The values for Class 4 are standardized to 0, which means that the Table 4 coefficients are log odds of being in a given class relative to Class 4. Class 1 (Down) represents a class with heavy drinking that decreases over time. Class 2 (High18) represents a class with very high heavy drinking that decreases over time. Class 3 (Up) represent a class with increasing heavy drinking over time. Bolded entries are statistically significant at the 0.05 level.

Table 4 shows that Class 1 membership is influenced by being male, nonminority, and having early onset of drinking. Class 2 membership is influenced by being male, nonblack, having early onset of drinking, having a positive family history of alcohol problems, and not attending college. Class 3 membership is influenced by being male, having a positive family history of alcohol problems, being a high school dropout, and not attending college.

Example 4: GGMM Relating Heavy Drinking Class to Antisocial Class

In the fourth example, the four latent classes for antisocial behavior at age 16 are related to the four trajectory classes for heavy drinking at ages 18–30. This is an example of a GGMM.

Figure 7 shows the combination of two model parts. The NLSY data for this illustration are from birth cohort 1964 so that the antisocial behavior items are measured at age 16 and the heavy drinking measures are from ages 18 to 30. As shown in Figure 7, this GGMM is a generalized form of LTA. As in LTA, a latent class variable at two different time points is considered. It is of interest to study the conditional probabilities (transition probabilities) of heavy drinking class membership as a function of antisocial behavior class membership. As opposed to LTA, two different kinds of latent class variables are considered and the multiple indicators for heavy drinking represent repeated measures.

Table 5 contains the estimated marginal and conditional probabilities for the antisocial behavior and heavy drinking classes. The marginal probabilities for the antisocial classes are somewhat different than from the analysis in Example 1, because only the youngest cohort and 9 antisocial behavior items were used. The conditional probabilities show the relationship between antisocial behavior and heavy drinking class membership. Individuals in ASB Class 4, the normative antisocial behavior class, and ASB Class 3, person offense, have high odds of being in heavy drinking (HD) Class 4, the normative heavy drinking class, relative to the other HD classes. Individuals in ASB Class 2 (substance involvement) and individuals in ASB Class 1 (property offense) have much lower odds of being in HD Class 4. Individuals in these classes have a nontrivial probability of being in HD Class 1. ASB class membership does not seem to explain HD Class 3.

DISCUSSION

Person-centered and variable-centered analyses typically have been seen as different activities that use different types of models and software. This paper gives a brief overview of new methods that integrate variable- and person-centered analyses. These methods include latent class analysis (LCA), latent transition analysis (LTA), latent class growth analysis (LCGA), growth mixture modeling (GMM), and general growth mixture modeling (GGMM). LCA describes how the probabilities of a set of observed categorical variables vary across groups of individuals where group membership is unobserved. The goal of LCA is the find the smallest number of latent classes that describe the association among a set of observed categorical variables. LCA studies class membership in cross-sectional data. In contrast, LTA uses multiple indicators at each time point to define a latent class variable for each time point. The main objective of the analysis is to study the probability of a transition from a class at one time point to a class at the next time point. LCGA uses a single outcome variable at multiple time points to define a latent class model where the latent classes correspond to different growth curve shapes for the outcome variable. GMM combines features of conventional growth modeling with LCGA. GMM both estimates a mean growth curve for each class and captures individual variation around these growth curves by the estimation of growth factor variances for each class. GGMM takes GMM further by incorporating the GMM model into a more complex model with, for example, a distal outcome predicted by class membership.

These methods are presented in a general latent variable modeling framework that expands traditional latent variable modeling by including not only continuous latent variables but also categorical latent variables. The examples presented in the paper can be seen as special cases of this general framework. The general framework makes it possible to combine these models and to study new models

serving as a stimulus for asking research questions that have both person- and variable-centered aspects.

REFERENCES

Agresti A (1990) *Categorical Data Analysis*. John Wiley & Sons, New York.

Bandeen-Roche K, Miglioretti DL, Zeger SL, Rathouz PJ (1997) Latent variable regression for multiple discrete outcomes. J Am Statistical Assn 92:1375–1386.

Bartholomew DJ (1987) *Latent Variable Models and Factor Analysis*. Oxford University Press, New York.

Bucholz KK, Heath AC, Reich T, Hesselbrock VM, Kramer JR, Nurnberger JI, Schuckit MA (1996) Can we subtype alcoholism? A latent class analysis of data from relatives of alcoholics in a multi-center family study of alcoholism. Alcohol Clin Exp Res 20:1462–1471.

Clogg CC (1995) Latent class models, in *Handbook of Statistical Modeling for the Social and Behavioral Sciences* (Arminger G, Clogg CC, Sobel ME eds), pp. 311–359, Plenum Press, New York.

Collins LM, Graham JW, Rousculp SS, Hansen WB (1997) Heavy caffeine use and the beginning of the substance use onset process: An illustration of latent transition analysis, in *The Science of Prevention: Methodological Advances from Alcohol and Substance Use Research* (Bryant K, Windle KM, West S eds), pp. 79–99, American Psychological Association, Washington, DC.

Collins LM, Wugalter SE (1992) Latent class models for stage-sequential dynamic latent variables. Multivariate Behavioral Research 27:131–157.

Graham JW, Collins LM, Wugalter SE, Chung NK, Hansen WB (1991) Modeling transitions in latent stage- sequential processes: A substance use prevention example. J Consult Clin Psychol 59:48–57.

Harford T (1994) The effects of order of questions on reported alcohol consumption. Addiction 89:421–424.

Harford T, Muthén B (1999) Adolescent and young adult antisocial behavior and adult alcohol use disorders: A 14-year prospective follow-up in a national survey. Technical report, University of California, Los Angeles.

Jones BL, Nagin DS, Roeder K (1998) *A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories,* Working Paper No. 684, Carnegie Mellon University, Department of Statistics, Pittsburgh.

Kandel DB, Yamaguchi K, Chen K (1992) Stages of progression in drug involvement from adolescence to adulthood: Further evidence for the gateway theory. J Stud Alcohol 53:447–457.

Moffitt TE (1993) Adolescence-limited and life-course persistent antisocial behavior. Psychol Rev 100:674–701.

Muthén B (1998) Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling, in *New Methods for the Analysis of Change* (Sayer, A, Collins L eds), APA, Washington, DC, in press.

Muthén B, Khoo ST (1998) Longitudinal studies of achievement growth using latent variable modeling. Learning and Individual Differences 10:73–101.

Muthén B, Muthén L (2000) The development of heavy drinking from age 18 to 37 in a U.S. national sample. J Stud Alcohol 61:290–300.

Muthén B, Shedden K (1999) Finite mixture modeling with mixture outcomes using the EM algorithm. Biometrics 55:463–469.

Muthén B, Brown CH, Khoo S, Yang C, Jo B (1998) General growth mixture modeling of latent trajectory classes: Perspectives and prospects. Paper presented at the meeting of the Prevention Science and Methodology Group, Tempe, AZ.

Muthén LK, Muthén B (1998) *Mplus User's Gui*de. Muthén & Muthén, Los Angeles.

Nagin DS (1999) Analyzing developmental trajectories: A semiparametric, group-based approach. Psychological Methods 4:139–157.

Nagin DS, Land KC (1993) Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed Poisson model. Criminology 31:327–362.

Nagin DS, Tremblay RE (1999) Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and non violent juvenile delinquency. Child Dev 70:1181–1196.

Nestadt G, Hanfelt J, Liang KY, Lamacz M, Wolyniec P, Pulver AE (1994) An evaluation of the structure of schizophrenia spectrum personality disorders. J Personal Disord 8:288–298.

The Ohio State University National Longitudinal Surveys. Columbus, (1994) Ohio: Center for Human Resource Research.

Pearson JD, Morrell CH, Landis PK, Carter HB, Brant LJ (1994) Mixed-effect regression models for studying the natural history of prostate disease. Stat Med 13:587–601.

Reboussin BA, Reboussin DM, Liang KY, Anthony JC (1998) Latent transition modeling of progression of health-risk behavior. Multivariate Behavioral Research 33:457–478.

Rindskopf D, Rindskopf W (1986) The value of latent class analysis in medical diagnosis. Stat Med 5:21–27.

Schulenberg J, O'Malley PM, Bachman JG, Wadsworth KN, Johnston LD (1996) Getting drunk and growing up: Trajectories of frequent binge drinking during the transition to young adulthood. J Stud Alcohol 57:289–304.

Uebersax JS, Grove WM (1990) Latent class analysis of diagnostic agreement. Stat Med 9:559–572.

Verbeke G, Lesaffre E (1996) A linear mixed-effects model with heterogeneity in the random-effects population. J Am Stat Assn 91:217–221.

Zucker RA (1994) Pathways to alcohol problems and alcoholism: A developmental account of the evidence for multiple alcoholisms and for contextual contributions to risk, in *The Development of Alcohol Problems: Exploring the Biopsychosocial Matrix of Risk* (Zucker RA, Howard J, Boyd GM eds), pp. 255–289, NIAAA Research Monograph No. 26, U.S. Department of Health and Human Services, Rockville, MD.