New Approaches to Studying Problem Behaviors: A Comparison of Methods for Modeling Longitudinal, Categorical Adolescent Drinking Data

Betsy J. Feldman
University of California, Berkeley

Katherine E. Masyn and Rand D. Conger
University of California, Davis

Analyzing problem-behavior trajectories can be difficult. The data are generally categorical and often quite skewed, violating distributional assumptions of standard normal-theory statistical models. In this article, the authors present several currently available modeling options, all of which make appropriate distributional assumptions for the observed categorical data. Three are based on the generalized linear model: a hierarchical generalized linear model, a growth mixture model, and a latent class growth analysis. They also describe a longitudinal latent class analysis, which requires fewer assumptions than the first 3. Finally, they illustrate all of the models using actual longitudinal adolescent alcohol-use data. They guide the reader through the model-selection process, comparing the results in terms of convergence properties, fit and residuals, parsimony, and interpretability. Advances in computing and statistical software have made the tools for these types of analyses readily accessible to most researchers. Using appropriate models for categorical data will lead to more accurate and reliable results, and their application in real data settings could contribute to substantive advancements in the field of development and the science of prevention.

Keywords: longitudinal analysis, generalized linear models, mixture models, drinking trajectories, adolescent alcohol use

Preventing mental health disorders and problem behaviors, such as delinquency, risky sexual behaviors, and substance use, in childhood and adolescence is critically important to the well-being of young people, and ultimately, our society. Problem behaviors often occur in tandem with one another (Donovan & Jessor, 1985) and are associated with concurrent difficulties, such as family dysfunction, academic failure, and poor peer relationships during childhood and adolescence (e.g., Hawkins, Catalano, & Miller, 1992; Wiesner & Windle, 2004). Later, as these behaviors continue into emerging adulthood, avenues toward a successful life course may be shut off. This can lead to adulthood failures in areas such as work and education, as well as physical and emotional disorders, all of which are costly to the individual and to society as a whole (Hill, White, Chung, Hawkins, & Catalano, 2000; Marmorstein & Iacono, 2005; Wiesner & Silbereisen, 2003). Understanding the etiology of childhood and adolescent problem behaviors, in part through optimal statistical modeling of developmental trajectories, may help point the way toward more successful approaches to, and timing of, interventions (Shaw, Gilliom, Ingoldsby, & Nagin, 2003).

Problem behaviors are problematic to study for a number of reasons. Because psychopathology is a developmental process, proper understanding requires longitudinal data and analyses (Cicchetti & Toth, 1998; Sameroff, 2000). In addition, if researchers want to generalize their results to the population at large, they need to study population-based samples.[1] But problem behaviors are, by definition, rare in the general population, and the data generally, often egregiously, violate assumptions of standard normal-theory linear models. They are frequently measured on a categorical scale, and the categories are generally unevenly spaced (e.g., "During the past year, how often did you drink alcohol?" 1 = *never*, 2 = *a couple of times*, 3 = *1–3 times a month*, 4 = *1 or more times a week*, 5 = *daily*). The data tend to have large masses in the lowest category, which signifies an absence of the behavior; and to the extent that people do display the behavior, the distributions tend to be quite skewed. Ordinal data like these are most often treated as though they were continuous and normally distributed. Although some researchers have suggested that categorical variables with four or more categories could reasonably be

[1] Results from analyses of clinical samples can only be generalized to that population and may not apply to the general public, and use of other types of nonrepresentative samples may also have problems with respect to inference (Hernán, Hernández-Diaz, & Robins, 2004).

analyzed using models that assume normal distributions (e.g., Bentler & Chou, 1987), other research suggests that this is not the case. Rather, Dolan (1994) found that this business-as-usual approach can lead to biased estimates, incorrect standard errors, and incorrect fit statistics even when the data are symmetric; and these problems get worse to the extent that the distribution departs from symmetry (also see, e.g., DiStefano, 2002; Feldman & Masyn, 2008; West, Finch, & Curran, 1995). Treating the data as censored normal (censored at zero) is a popular way of accounting for the asymmetry caused by the pileup of zeros, but it still assumes that the data above zero are continuous, so it is subject to many of the same problems that occur when the data are treated as normally distributed.

When these ordinal variables are modeled as inherently categorical, a lack of symmetry is not a problem, because the proportions in each category are explicitly modeled. Additionally, the mass at the bottom of the scale is only a problem to the extent that it may, with a limited sample size, be responsible for empty cells at the high end of the scale; and too many empty cells can cause estimation difficulties. Longitudinal or clustered categorical-data models have been available for a number of years and used in fields such as medicine (Harville & Mee, 1984; Hedeker & Gibbons, 1994), economics (Butler & Moffitt, 1982), and education (Bock & Lieberman, 1970); until recently, however, software that estimated these models was relatively difficult to use, and the estimation process was computationally arduous. Because of advances in computers and statistical software, appropriate analysis of longitudinal categorical data has become much more straightforward for applied researchers.

In this article, we demonstrate some of the methods appropriate for modeling trajectories with longitudinal categorical observed data. Our goal is to place these techniques into a coherent framework. Such a framework will, we believe, help developmental researchers feel more confident in selecting and using the methods. A second important contribution this article makes is to offer a systematic approach to choosing from among competing models for longitudinal categorical data. Because there is no generally available test of overall fit for most of these models, nor any single, widely accepted criterion for model selection, we present several techniques for assessing the models. We use adolescent alcohol-use data to illustrate the process of model specification, selection, and interpretation; and we guide the reader through the process of choosing an optimal model for the data from among those investigated by comparing results with respect to fit (using comparative fit statistics and residual analysis), parsimony, and substantive considerations.

Models for Longitudinal Categorical Data

The Generalized Linear Model

In the sections that follow, we describe several models for characterizing individual differences in change or growth over time, given categorical data. Appendix A lists the models we discuss here along with their distributional assumptions, and alternative names and acronyms sometimes seen in the literature.

To introduce models for categorical data, we begin with the most basic and general form of a generalized linear model,[2] in which the *expected value* (or mean), μ_i, of individual i's response,

y_i (given covariate values x_{0i}, x_{1i}, ..., x_{mi}), is related to the covariates through a *linear predictor*, η_i. η_i has a linear relationship with the covariates that is quantified by regression coefficients β_0, β_1, ... β_m (Agresti, Booth, Hobert, & Caffo, 2000; Skrondal & Rabe-Hesketh, 2004):

$$\eta_i = \beta_0 x_{0i} + \beta_1 x_{1i} + \ldots + \beta_m x_{mi}; \tag{1}$$

where, usually, $x_{i0} = 1$ (making β_0 the intercept of the equation), and there are m covariates. η_i is related to μ_i (given individual i's values on covariates) through some *link function*, g:

$$g(\mu_i) = \eta_i. \tag{2}$$

The type of link function used depends on the conditional distribution of the data.[3] For instance, when the distribution is normal (Gaussian), g is typically an identity function, and the equation is a standard linear regression (with an identity link, the estimate, $\hat{\mu}_i$, can be thought of as $\hat{y}_i$):

$$\eta_i = \mu_i = \beta_0 x_{1i} + \beta_1 x_{2i} + \ldots + \beta_m x_{mi}. \tag{3}$$

But if y_i has a different distribution, an identity link may not be appropriate. For example, if y_i is binary, then its distribution is Bernoulli, and the expected value, μ_i, is the conditional probability of giving a positive (or correct) response: $pr(y_i = 1|x_1, \ldots, x_m) = \mu_i$. Using an identity link in this case is improper because it would allow the model-predicted response, $\hat{\eta}_i$, to go below zero or above one—outside of the range of probabilities. Because of this, either a logit (or logistic) link or a probit link is usually used with dichotomous data. In the logit-link model for binary data, the natural log of the odds of giving a positive response (often referred to as the log-odds or logit) is modeled as a linear function of the covariates:

$$g(\mu_i) = \ln\left(\frac{\mu_i}{1 - \mu_i}\right) = \eta_i, \text{ and} \tag{4}$$

$\eta_i = \beta_0 x_{0i} + \beta_0 x_{0i} + \ldots + \beta_m x_{mi}$, as in Equation 3.

Therefore:

$$\ln\left(\frac{\mu_i}{1 - \mu_i}\right) = \beta_0 x_{0i} + \beta_1 x_{1i} + \ldots + \beta_m x_{mi}, \tag{5}$$

or, equivalently,

$$\mu_i = \frac{\exp[\beta_0 x_{0i} + \beta_1 x_{1i} + \ldots + \beta_m x_{mi}]}{1 + \exp[\beta_0 x_{0i} + \beta_1 x_{1i} + \ldots + \beta_m x_{mi}]}. \tag{6}$$

In this article, the logit-link model is used for ordinal data—an extension of the binary model referred to as the *ordered logit* or *cumulative logit* model. For an ordinal variable with $j = 1, 2, \ldots$

---

[2] For extended discussions of generalized linear models, including methods for various types of non-Gaussian distributions, see, for example, Agresti (2002), McCullagh and Nelder (1999), and Skrondal and Rabe-Hesketh (2004).

[3] All of these models assume that the distribution of the observed response variable is a member of the exponential family of distributions; most commonly seen among discrete distributions: multinomial (ordered or unordered categorical data), Poisson (count data), or Bernoulli (dichotomous data); and continuous distributions: gamma and Gaussian (or normal).

J categories, the ordered logit model represents the probability of scoring in category j or above, versus any category below j (see, e.g., Agresti, 2002). In discussing ordered logit models, it is helpful to think of the ordinal observed outcome as a coarse categorization of an underlying (latent) continuous variable, y_i^* (see Figure 1), for which the residual errors are assumed to have a logistic distribution (Agresti, 2002; Gurland, Lee, & Dahm, 1960; Hedeker & Gibbons, 1994; Skrondal & Rabe-Hesketh, 2004).[4] In this characterization, referred to by Skrondal and Rabe-Hesketh (2004) as a *latent response* formulation, the cutpoints that separate the underlying continuous distribution into categories are referred to as *thresholds*, or τs, and the thresholds define the relationship between the categorical observed y_i (in Figure 1, $y_i = 1, 2, 3$, or 4), and the continuous latent y_i^*, such that $y_i \geq j$ if $y_i^* \geq \tau_j$, and $y_i = j$ if both $y_i^* \geq \tau_j$ and $y_i^* < \tau_{j+1}$; $j = 0, 1, 2, \ldots, J - 1$, $\tau_0 = -\infty$ and $\tau_J = \infty$.

Imagine, for example, that Figure 1 is showing an ordinal measure used to assess frequency of suicidal ideation in the past year in which $0 = $ *no ideation*, $1 = $ *thought of suicide once*, $2 = $ *thought of suicide more than once, but not frequently*, and $3 = $ *contemplated suicide many times*. If individual i's latent suicidal ideation score, y_i^* (latent, because only a category is observed), is greater than or equal to τ_1 (Figure 1), then the observed value for individual i (y_i) is 1, 2, or 3 (this individual has experienced suicidal ideation). However, if y_i^* is greater than or equal to τ_1 and less than τ_2, then $y_i = 1$ (this individual thought of suicide once). Likewise, if y_i^* is less than τ_1, then $y_i = 0$ (not suicidal); and if it is equal to or greater than τ_3, then individual i's response is 3 (chronically suicidal).

The expected value of y_i^*, μ_i^*, is modeled directly with an identity link, as in Equation 3:

$$\mu_i^* = \beta_0 x_{0i} + \beta_1 x_{1i} + \ldots + \beta_m x_{mi}, \quad (7)$$

and

$$y_i^* = \beta_0 x_{0i} + \beta_1 x_{1i} + \ldots + \beta_m x_{mi} + \varepsilon_i, \quad (8)$$

where $\varepsilon_i \sim \text{logistic}\left(0, \frac{\pi^2}{3}\right)$ with fixed mean and variance.

It is important to note that the ordered logit model assumes that the multiplicative effect of a covariate (observed or latent) on the odds of being in a category j is the same for all $j = 0, 1, 2, \ldots J - 1$. This *proportional odds* assumption is necessary to permit the relationship between the ordinal variable and each covariate to be quantified by a single coefficient.[5]

Longitudinal Categorical Data

When researchers gather multiple cross-sectional or longitudinal measures on the same individuals (for example, repeatedly measuring individuals' levels of suicidality over several years), the repeated measures on an individual are correlated with one another. This violates an important assumption of statistical models—that of conditionally independent observations—resulting in standard errors that are too low. Generalized estimating equations (GEEs) were developed to solve this problem with categorical data. GEE uses a working correlation matrix (specified by the data analyst) to represent the correlations between observations, and it uses a special estimator (known as a "sandwich" estimator) to correct the standard errors. However, the GEE approach has a few limitations that hamper its usefulness in modeling psychological data. The estimates that result from GEE are *population averaged* estimates. That is, the estimates are mean effects for the population as a whole rather than for any given individual (one coefficient applies equally to all members of the population). GEE does not allow the researcher to explicitly model the population heterogeneity, or individual differences. For instance, in the suicidality example, GEE would give only an estimate of the average intercept and slope of suicidality in the population. GEE can estimate mean differences based on observed covariates (e.g., the difference

Figure 1. Pictorial representation of how thresholds (τs) in an underlying, continuous latent response variable (Y^*) correspond to the categories in an ordinal variable.

---

[4] In the probit model, the errors are assumed to be normally distributed (Agresti, 2002; Skrondal & Rabe-Hesketh, 2004). For early references to similarly formulated probit models, see, for example, Ashford (1959), and Goldberger (1964).

[5] Other modeling options are possible if the proportional-odds assumption does not hold (see, e.g., Agresti, 2002; Hedeker & Mermelstein, 1998).

between male and female averages) but not, for instance, a specific individual's estimated suicidality level or trajectory (or how it might change on the basis of a covariate value). Nor can it permit individual prediction from the intercept and slopes to a distal outcome.

Hierarchical Generalized Linear Model

A longitudinal model for categorical data that does model the individual differences is the hierarchical generalized linear model (HGLM; Raudenbush & Bryk, 2002), also known as a generalized linear mixed model (GLMM; Skrondal & Rabe-Hesketh, 2004).[6] This approach combines the generalized linear model (GLM) with a hierarchical linear model (HLM; also known as a random-effects, random-coefficients, mixed, or latent growth-curve model), in which repeated measurements on individuals are expressed as a function of time (see Figure 2). Individual differences in the outcome variable when time equals zero and change in the outcome over time are modeled by permitting the intercept and slope coefficients to vary across individuals. The intercept and slope(s) are, therefore, referred to as random coefficients, random effects, or (latent) growth factors. Both time-varying (not discussed in this article) and time-invariant covariates may be used to explain within- and between-person variability, respectively.

In modeling the categorical drinking data, we use the logit link described in Equations 4 to 6. A categorical random-effects model has been described by Hedeker and colleagues (Hedeker & Gibbons, 1994; Hedeker, Gibbons, & Flay, 1994; Hedeker & Mermelstein, 1998, 2000) and others (Agresti, 2002; Agresti et al., 2000; B. Muthén & Asparouhov, 2002; Raudenbush & Bryk, 2002; Skrondal & Rabe-Hesketh, 2004; Vermunt, 2006). In the models that follow, μ_{ti}^* (individual i's expected latent response value at time t; see Equations 7 and 8), is modeled as a linear function of the random coefficients (growth factors) that make up the intercept coefficient, β_{0i} (the expected latent response value for individual i when time $= 0$) and, usually, one or more slope coefficients, β_{si} ($s = 1, 2, \ldots, S$), which are multiplied by powers of time (e.g., $\beta_{1i} a_{ti}, \beta_{2i} a_{ti}^2, \beta_{3i} a_{ti}^3, \ldots, \beta_{Si} a_{ti}^S$). The slope coefficients describe systematic linear or curvilinear change over time in individual i's latent responses.[7] The times of measurement, a_{ti}, are usually centered by subtracting a relevant age or timepoint; often this is the time of or age at the first measurement occasion, as is shown in Figure 2. This makes the intercept, β_{0i}, individual i's expected latent response at the first measurement occasion. The times of measurement may be the same for all individuals, as shown in Figure 2, or may be specified in the model to be unique for each individual (e.g., actual age).

The model that follows has linear slope (higher order polynomials are also common) and is expressed as having two levels. Level 1 characterizes the individual latent responses at each timepoint, and Level 2 characterizes the individual trajectories over time. The generalized linear model (Equations 4–8) is at Level 1, embodied in y_{ti}^*:

$$y_{ti}^* = \beta_{0i} + \beta_{1i} a_{ti} + \varepsilon_{ti}; \tag{9}$$

where $\varepsilon_{ti} \sim \text{logistic}\left(0, \frac{\pi^2}{3}\right)$ with fixed mean and variance and are assumed independent between individuals (i.e., $\text{cov}[\varepsilon_{ti}, \varepsilon_{tj}] = 0$)

and conditionally independent between times (i.e., $\text{cov}[\varepsilon_{ti}, \varepsilon_{vi}] = 0$; referred to as *local independence*).

At Level 2, individual differences in the random coefficients from Level 1 (β_{0i}, β_{1i}) are represented by variability (u_{0i}, u_{1i}) around the mean intercept (γ_{00}) and mean slope (γ_{10}). The individual differences are modeled as a function of an individual-level, time-invariant covariate, x_i (multiple covariates are possible), quantified by regression coefficients γ_{01} and γ_{11} for intercept and slope, respectively. The conditional joint distribution of the intercept and slope is assumed to be multivariate normal:

$$\beta_{0i} = \gamma_{00} + \gamma_{01} x_{1i} + u_{0i},$$

$$\beta_{0i} = \gamma_{00} + \gamma_{01} x_{1i} + u_{0i}; \tag{10}$$

where $\boldsymbol{u} \sim \text{N}(\boldsymbol{0}, \boldsymbol{\Psi})$, $\text{cov}(\varepsilon_t, u_s) = 0$ for all $t = 1, \ldots, T$ and $s = 0, 1$, and $\gamma_{00} = 0$ for identification. This identification restriction is necessary because the scale and location of the intercept are arbitrary when this latent response parameterization is used.[8] The thresholds (τs) are assumed to be constant across time.

If we use the earlier example of suicidality, this hierarchical generalized linear model characterizes individuals' systematic change over time in the continuous latent response. The τs that relate the latent suicidality score to the observed categorical suicidality measures are estimated as part of the model (Equations 5–8); because they are constant across time, we know that the latent suicidality score, despite changing over time, relates to the observed categories of the ordinal variables in the same way at each measurement occasion. The intercept mean is defined to equal zero ($\gamma_{00} = 0$) on the arbitrary scale of the latent suicidal-ideation response (necessary in this parameterization); however, the estimated individual intercepts vary around zero. The individual slopes vary about the estimated average slope, as well, and these individual intercepts and slopes are often correlated.

Growth Mixture Model

Several assumptions have been discussed that are very important for the HGLM: (a) The random coefficients (intercept and slopes) are assumed to have a multivariate normal distribution; (b) change is a smooth function of time (most often linear or curvilinear); (c) the proportional-odds assumption governs the relationships between covariates (or predictors) and the categorical outcomes; and (d) measures are independent, conditional on the random effects (local independence). Any of these assumptions might not hold, especially when modeling problem-behavior data. In the models that follow, we discuss methods that relax the first three assumptions.

To begin, when the data are categorical and highly skewed, it is possible that the individual intercepts and slopes are not normally

---

[6] For extended discussions of hierarchical generalized linear models, see, for example, Agresti et al. (2000); Raudenbush, Bryk, and Congdon (2005); Singer and Willett (2003); Skrondal and Rabe-Hesketh (2004).

[7] For models in which change over time is permitted to be nonlinear, see, for example, Blozis (2004, 2007) and Ferrer and McArdle (2003).

[8] For examples of alternative parameterizations in which different restrictions on the thresholds permit γ_{00} be estimated, see Mehta, Neale, and Flay (2004), Millsap and Yun-Tein (2004), and Skrondal and Rabe-Hesketh (2004).
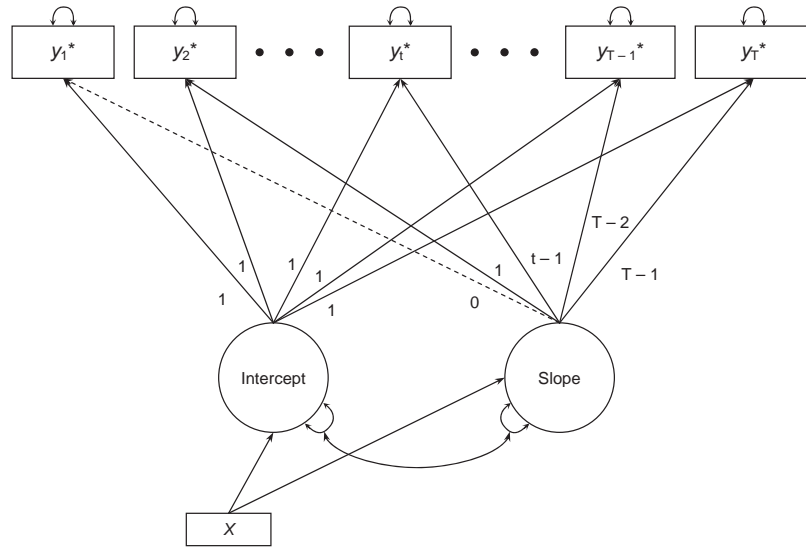
Figure 2. Path diagram of a hierarchical generalized linear model (HGLM) with a covariate, *X*. Model is shown with fixed times of measurement, but individual times of measurement may be specified.

distributed, potentially resulting in biased parameter estimates (B. Muthén & Asparouhov, 2006; Vermunt, 2006). However this assumption is difficult to test (Agresti, 2002, p. 496; Carlin, Wolfe, Brown, & Gelman, 2001; Vermunt, 2006). The Level 2 normality assumption can be relaxed by using a mixture of normal distributions, known as mixing components or classes, to characterize the joint distribution of the random effects. This is pictured in Figure 3A. This model is a type of finite mixture model (see McLachlan & Peel, 2000), which we refer to here as a (generalized linear) growth mixture model (GMM), and it assumes only within-class normality.[9] The classes have different mean structures at Level 2 and can comprise different proportions of the population. The within-class Level 2 variance–covariance matrix can also differ across classes; however, under some circumstances, this may lead to estimation problems resulting from an unbounded likelihood (McLachlan & Peel, 2000). The multinomial latent class variable, c_i, which identifies individual *i*'s class membership, takes on values $k = 1, 2, \ldots K$, where *K* is the number of classes. Here, the τs are assumed to be constant across both time and classes.

Figure 3A shows how the individual differences are, in this model, characterized in part as discrete classes and in part as continuous distributions within those classes. Because of this within-class variability, covariates can be used to predict individual differences within class, as well as to predict the probability of class membership. The HGLM can be seen as a special case of the GMM with just a single class ($K = 1$). When GMM has two or more classes, Level 1 equations are indexed by class:

$$y^*_{kti} = \beta_{k0i} + \beta_{k1i}a_{ti} + \varepsilon_{kti}; \qquad (11)$$

for individual *i* at time *t* in class *k*; ε_{kti} distributed as before.

The Level 2 equations are also indexed by class, and a multinomial logistic equation is added, which expresses the conditional probability of membership in class *k* as a function of a class-specific intercept, α_k, and the covariate, x_i, multiplied by coefficient, λ_k:

$$\beta_{k0i} = \gamma_{k00} + \gamma_{k01}x_i + u_{k0i},$$

$$\beta_{k1i} = \gamma_{k10} + \gamma_{k11}x_i + u_{k1i}, \qquad (12)$$

$$pr(c_i = k|x_i) = \frac{\exp[\alpha_k + \lambda_k x_i]}{\displaystyle\sum_{h=1}^{K} \exp[\alpha_h + \lambda_h x_i]}; \qquad (13)$$

where $\gamma_{K00} = 0$ and $\alpha_K = \gamma_K = 0$ for identification; $\mathbf{u}_{ki}$ distributed as before.

This model has a more complicated interpretation than the HGLM. In the suicidality example, if there are, for instance, three classes, this implies three unique sets of intercept and slope means—one for each class. Within each class, individual intercepts and slopes vary about those means and can be predicted by covariates, as can class membership. The grouping of individuals into classes, however, is seldom perfect. Instead, person *i* has some estimated probability of membership in each suicidality class. These are referred to as *posterior probabilities.* For each individual, the posterior probabilities sum to 1 and are based on that individual's observed outcome responses and covariate values. Thus, person *i*'s estimated trajectory of suicidality is a function of the probability of membership in each of the classes, the class-specific means and covariance structures, and possibly covariates.

Latent Class Growth Analysis

There are several reasons that the within-class normality assumption of the GMM may not be sustained by the data. The primary one is that there may be insufficient variability to estimate

---

[9] Muthén and Asparouhov (2006) also refer to this model as a growth mixture analysis (for more on GMM with continuous outcomes, see, e.g., Muthén, 2004; Muthén et al., 2002).
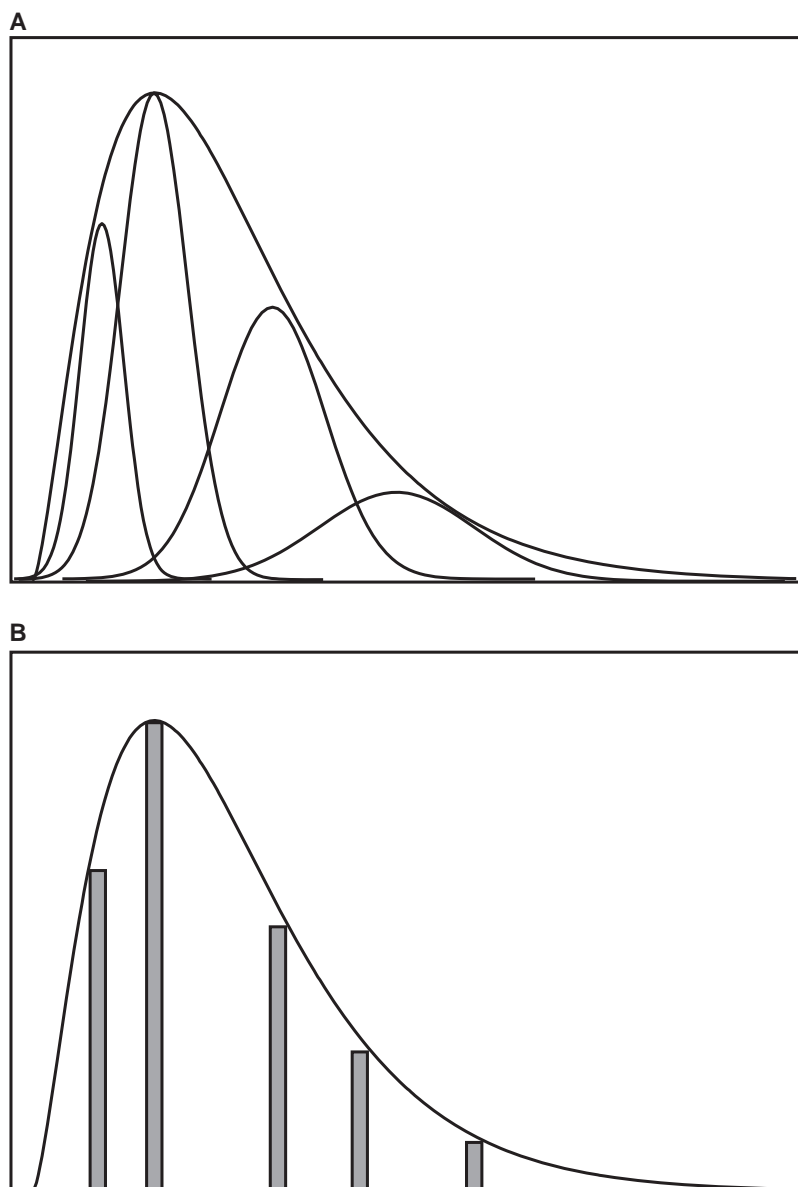
Figure 3. A skewed univariate distribution characterized by (A) a mixture of normal distributions (growth mixture model; GMM), or (B) mass points (latent class growth analysis; LCGA).

the within-class variances and covariances in one or more classes. Two or more classes can be collapsed together, but if the classes are well separated (i.e., the mean trajectories are very different), then collapsing them might not be optimal. In this case, it is possible to constrain one or more elements in the within-class covariance matrix (Ψ_k) to zero. When all variances and covariances in a class are fixed at zero, it means effectively replacing a continuous, normally distributed class with a mass point, as shown in Figure 3B. We refer to this model as a (generalized linear) latent class growth analysis (LCGA) and, like the HGLM, this model is a special case of GMM.[10] The Level 1 equation is the same as that in GMM, but the Level 2 equations change to reflect the elimination of within-class variances (and with no variation to explain, within-class predictors are also dropped):

$$\beta_{k0i} = \gamma_{k00},$$

$$\beta_{k1i} = \gamma_{k10}. \tag{14}$$

When all within-class variances and covariances equal zero, individual differences at Level 2 are being modeled solely by class membership, so it is likely that more classes will be needed to adequately characterize the variability in the random effects distribution (B. Muthén & Muthén, 2000b). The probability of membership in each of the K classes may be regressed upon the covariates as in Equation 13, with the same identification con-

---

[10] This approach is sometimes referred to as semiparametric group-based modeling (Nagin, 1999; see Appendix A).

straints. Figure 4A shows a path diagram of a GMM and an LCGA. The solid lines belong to both models, and dashed lines show the optional parameters that, if any are present, define the model as GMM (for more on the comparisons between these models, see, e.g., Kreuter & Muthén, 2007; Nagin, 2005; Raudenbush, 2005; Reinecke, 2006).

This model has a slightly simpler interpretation than the GMM because all individual differences in estimated suicidality trajectories are characterized by class membership, and those differences can be modeled as a function of predictors of class membership. Within each class, all individuals are assumed to share the same suicidality trajectory. One way to think about the shared within-class trajectory is that it is an approximation of the true within-class individual latent trajectories, similar to the way mass points approximate the true distribution of the random effect shown in Figure 3B. Person i's predicted trajectory of suicidality is a straightforward function of his or her posterior class membership probabilities and the class intercept and slope means.

Longitudinal Latent Class Analysis

Finally, it is possible that development cannot be characterized as a simple function of time, as is assumed in the models we have been discussing, or that the proportional odds assumption does not hold. All of the models thus far are bound by these two assumptions because they model scaled growth (change across time along a single continuous scale). Longitudinal latent class analysis (LLCA; latent class analysis applied to longitudinal data) is a type of mixture model, but it does not belong to the family of growth models we have been discussing. This is because, rather than modeling scaled change, LLCA models patterns of states across time.[11] If the outcome data are multinomial, LLCA is the only appropriate analytic approach, because any changes are necessarily state changes (see Figure 4B for a path diagram of LLCA).

LLCA models the joint distribution of the repeated outcome measures directly with a latent class variable, which characterizes both the within-person variation that had been handled at Level 1 of the previous growth models and the between-person differences that had been handled at Level 2 (see Appendix A). Its only assumption is local independence. In LLCA, the time- and class-specific probability of scoring in or above category j is modeled directly, so the thresholds are indexed by time (t) and class (k), in addition to category:

$$pr(y_{ti} \geq j | x, k) = \frac{\exp[-\tau_{tjk}]}{1 + \exp[-\tau_{tjk}]}, \quad (15)$$

and

$$pr(y_{ti} \geq j | x) = \sum_{k=1}^{K} \left[\left(\frac{\exp[-\tau_{tjk}]}{1 + \exp[-\tau_{tjk}]} \right) pr(c_i = k | x_i) \right]. \quad (16)$$

Class membership is based on groupings of similar patterns of responses over time, and covariates can be used to predict class membership probabilities, as shown in Equation 13. Note that because it is the category thresholds that define the differences both across time and across classes in LLCA, they are all estimated. Thus, although LLCA makes no assumptions about the distribution of the observed variables or the form of change, it

trades fewer assumptions for many more estimated parameters (Vermunt, 2006), in a manner similar to spline or piecewise models (e.g., McArdle, 2004). In addition, the model, as parameterized here, does not permit unique, individually varying times of measurement. For an application of LLCA to drinking patterns across time, see Lanza and Collins (2006).

In an LLCA of the suicidality example, patterns of change are defined by the variable categories, rather than by estimated intercepts and slopes. Like the LCGA, everybody in a class shares the same parameters; however now it is the category thresholds that are shared within class and different between classes. Because thresholds can also change across time, change can be much more irregular or complex in LLCA. For instance, the course of suicidality for individual i might fluctuate or be intermittent. Differences in suicidality could, perhaps, be associated with influences such as puberty, high school graduation, or other shared events; and the covariates can be used to predict the individual's probability of membership in each class.

Adding Covariates

The selection and modeling of covariates in mixture models is a complex issue currently under investigation, and an extended discussion of the issues is beyond the scope of this article. However, early work (Nylund & Masyn, 2008) suggests that the number of classes should be determined using an unconditional model and that substantial changes in parameter values after covariates are included indicate misspecification of their effects (e.g., omitting direct effects of covariates on observed variables). As is the case when estimating any statistical model, it is also important to let substantive theory, along with consideration of parsimony and interpretability, inform the inclusion and specification of covariates. The added complexity of mixture models makes this a particularly critical issue. For example, in a growth mixture model, it is theoretically feasible to allow a covariate to affect class membership as well as one or more growth factor variances in one or more classes. Additionally, all of the standard rules for typical regression and HLM models still pertain (e.g., concern with multicollinearity among covariates, the need to dummy-code nominal predictors).

Applying the Models

Longitudinal mixture models (GMM, LCGA, and LLCA) are often used to find unobserved but distinct groups of individuals (e.g., Lanza & Collins, 2006; B. Muthén & Muthén, 2000a; Shaw, Lacourse, & Nagin, 2005). However, any nonnormal distribution can be approximated by a finite mixture of normal distributions (Bauer & Curran, 2003; McLachlan & Peel, 2000), so the extraction of two or more classes does not necessarily indicate that distinct groups exist in the population. It is possible that finding multiple classes is an indication of a nonnormal distribution of the

---

[11] LLCA differs from latent transition analysis (LTA) because LTA models state changes across consecutive timepoints, rather than patterns that span several timepoints (see, e.g., Collins, Graham, Rousculp, & Hansen, 1997; Graham, Collins, Wugalter, Chung, & Hansen, 1991). As such, it answers different types of questions than LLCA and the other models discussed here, just as autoregressive models answer different types of questions than latent growth curves.
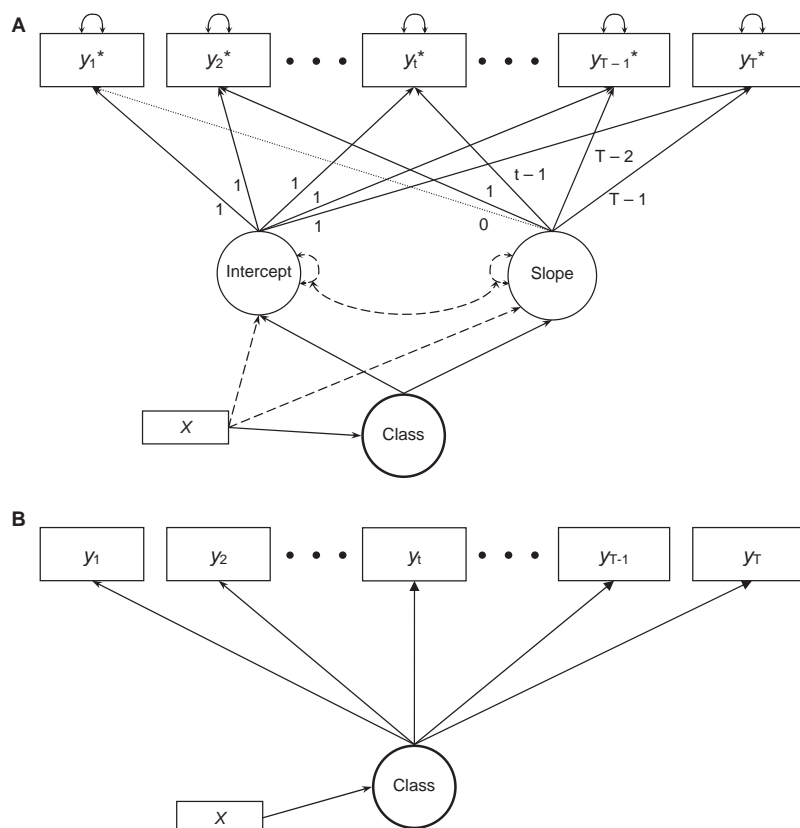
Figure 4. A: Path diagram for growth mixture model (GMM) and latent class growth analysis (LCGA). Solid lines are parameters present in both models, and dashed lines represent parameters that, if present, define the model as GMM. B: Path diagram of longitudinal latent class analysis (LLCA). Both models are shown with covariate, *X.*

random effects or potentially some other violation or model mis-specification (Bauer, 2005). In addition, class membership is seldom unequivocal. Rather, as mentioned earlier, the individual has an estimated probability of belonging to each class.

In this article, we use actual alcohol-use data, collected from a sample of adolescents when they were in 7th through 12th grades, to illustrate the similarities and differences between these analytic approaches and to motivate the process of model selection (see Appendix B for annotated Mplus syntax). The value of these models is enhanced to the extent that childhood events and circumstances may be used to help predict later pathological processes, as this can potentially allow targeted interventions to take place before such processes begin. In addition, under some circumstances, covariates can aid in model selection. Thus, we include a covariate that has been found to predict adolescent alcohol use—association with alcohol-using peers (e.g., Curran, Stice, & Chassin, 1997; Hawkins et al., 1992)—to demonstrate how a covariate may be specified and interpreted in the models.

Method

Participants and Procedure

Data come from a longitudinal, community epidemiological study of 451 (52% female) target youths and their families from

two-parent households in the Midwest. Participants in the study were recruited from eight rural counties, and the original sample of families was primarily lower middle or middle class. Due to a very small minority population in this area, all participants are of European heritage. Annually from 1989 (7th grade) through 1992 (10th grade), and again in 1994 (12th grade), the adolescents and participating family members individually completed questionnaires pertaining to subjects such as the demographic characteristics of the family, their own personal characteristics and behavior, and characteristics and behavior of other family members and friends. The retention rate in 1994 (the last wave used in the current study) was 94%. Additional details about the study, which is still ongoing, can be found in Conger and Conger (2002).

The primary outcome for these analyses is the alcohol-use measure from the questionnaire data. Target adolescents were asked each year how often they had consumed beer, wine, or hard liquor during the previous year. The drinking items were coded into a single, ordinal alcohol-use variable with four possible responses: $0 = never$, $1 = less\ than\ weekly$, $2 = once\ or\ twice\ a\ week$, and $3 = three\ or\ more\ times\ a\ week$. Each year the target adolescents were also asked how many of their close friends used alcohol during the previous year. Answers ranged from $0 = none\ of\ them$ to $4 = all\ of\ them$. The adolescent's answer to this question in seventh grade was used as a covari-

ate, predicting the intercept and change in the target adolescent's own drinking.[12]

Statistical Models

We began the unconditional analyses with a hierarchical generalized linear model (HGLM; Figure 2), using both linear and quadratic slopes and using the logit link and model shown in Equations 9 and 10. This model has the simplest interpretation, and it gives an overview of the form of change. Following HGLM, we tested the three mixture models, beginning with a growth mixture model (GMM; Equations 11–13). GMM comprises a very large set of potential models. For example, a quadratic model with three classes has 14 possible ways of specifying the random effects if they are constrained to be equal across classes and dozens more if they are permitted to vary across classes. To limit the number of models we were testing, we decided to constrain variances to be equal across classes, as discussed earlier, and to use a systematic approach to relaxing the constraints on the variance–covariance matrix. For each model, we tested the following: (a) intercept variance only; (b) intercept and linear slope variances; and (c) intercept, linear, and quadratic slope variances. We tested each model both with all possible covariances estimated and with all covariances constrained to zero.

All intercept and slope variances were fixed to zero for the latent class growth analysis (LCGA; Equation 14), and finally, we modeled the joint distribution of the repeated drinking outcomes directly with longitudinal latent class analysis (LLCA; Equations 15 and 16). Intercepts for HGLM, GMM, and LCGA are placed at ninth grade—the first year of high school. Entry to high school begins a time of increased risk for developing problem drinking. Incoming students are exposed to older adolescents, many of whom have already begun to drink, and alcohol is much more available in high school (Johnston, O'Malley, Bachman, & Schulenberg, 2006).

For all analyses, we used a full-information maximum likelihood (FIML) estimator with robust standard errors, implemented as MLR in Mplus 4.21 (B. Muthén & Muthén, 2006), to estimate the parameters (the robust standard errors offer protection against inflated alpha values in case of certain types of model or distributional misspecifications). With any longitudinal study, some individuals will miss some assessments or drop out altogether. These models use all of the available data under a missing-at-random assumption (as would multiple imputation), which allows that the missingness may be related to variables included in the analysis (Little & Rubin, 2002). Unless missing data are missing completely at random (i.e., missingness is unrelated to any data being modeled, present or missing), which is a much stronger assumption, FIML yields less biased estimates than other methods of missing-data methods, such as listwise deletion (which also reduces the power of the analysis by reducing sample size) or mean imputation (Schafer & Graham, 2002). FIML does not, however, guarantee unbiased estimates if missingness is related to the missing values themselves, and this risk of bias increases with the proportion of missingness (the maximum for a single wave in this study is 10.5%). However, we feel that, given the broad range of topics under study, missing-at-random is a reasonable assumption (for more on mechanisms of missingness and techniques for dealing with missing data, see, e.g., Foster, Fang, & Conduct Problems

Prevention Research Group, 2004; Little, 1995; Little & Rubin, 2002; Schafer & Graham, 2002). FIML can be computationally demanding with categorical data; however, it offers advantages over methods such as penalized quasi-likelihood and marginal quasi-likelihood, which, although computationally less demanding, may result in biased estimates.

Model Assessment and Selection

Any given data set can potentially be analyzed using a multitude of statistical models; however, there are generally theoretical and statistical reasons that rule out many possible models. For example, we are not considering any models that require an assumption of continuous observed data (e.g., two-part models; see Olsen & Schafer, 2001; Tooze, Grunwald, & Jones, 2002). Furthermore, because we are interested in modeling change across multiple timepoints, rather than pairs of timepoints, we are not considering any autoregressive-type models, such as latent transition analysis (LTA).

For most longitudinal categorical analyses, there are no readily available statistics that assess overall fit of the model to the data. The two categorical goodness-of-fit statistics—the Pearson chi-square and the likelihood chi-square (also called deviance, or G^2)—may not perform well when there are many low expected cell frequencies (Agresti, 2002), and this is often the case when longitudinal categorical data are used (e.g., five waves of a four-category variable results in 4^5 or 1,024 cells, most of which are likely to be empty). Nor is there a method for comparing models that is widely accepted as best (B. Muthén & Asparouhov, 2006; Nylund, Asparouhov, & Muthén, 2007). Thus, when assessing models, researchers generally consider both statistical and substantive criteria (e.g., Jackson, Sher, & Schulenberg, 2005; Tucker, Orlando, & Ellickson, 2003). In this section, we discuss four complementary approaches to model assessment: (a) quality of convergence, (b) comparative fit, (c) residual analysis, and (d) visualization. In addition, issues such as parsimony and, in mixture models, class size should be taken into consideration. For example, when a class comprises a small number of individuals, the parameter estimates may not be very stable or reliable. Often, no single criterion alone is sufficient to select a model, and two or more of the criteria we discuss must be considered simultaneously.

Convergence. The estimation algorithm searches for the global maximum of the likelihood function, but in complex categorical data models and mixture models, algorithms are more likely to converge on local maxima than with continuous data or in less complicated models, so the use of multiple starts from random locations in the parameter space is recommended (Hagenaars & McCutcheon, 2002; Hipp & Bauer, 2006; McLachlan & Peel, 2000; B. Muthén & Asparouhov, 2006; Vermunt, 2006; see Appendix B). In some cases, there may not be a global maximum (McLachlan & Peel, 2000). This can result in nonconvergence, or failure of the algorithm to replicate a maximum log-likelihood

---

[12] Allowing covariates to affect the slope means that the effect of the covariates on the outcome changes across time; however, the covariate is assumed to affect all individuals in the same fashion and is known as a "fixed effect." Alternatively, we could constrain the covariate effect to be time-invariant but allow that effect to vary across individuals (as a random effect).

value over many starting values. That is, each start may end up at a different local maximum, suggesting that the model parameter estimates are untrustworthy. A second indication that adequate convergence was not reached is a failure of the algorithm to generate standard errors. This can be brought about by a singular information matrix, which implies that the model is not identified (L. K. Muthén & Muthén, 2006). A model that has not converged, has resulted in a singular information matrix, or has failed to yield a consistent maximum log-likelihood value is considered a failed model in this article and eliminated from further consideration.

Comparative fit. The standard chi-square difference test (likelihood ratio test; LRT) is not helpful in determining which model is best, because the different models are not necessarily nested. It also cannot help choose the optimal number of classes in a mixture model, because regularity conditions of the test are violated when comparing a k-class model to a $(k - 1)$-class model (McLachlan & Peel, 2000). Often, an information index, such as the Bayesian information criterion (BIC; sometimes called Schwartz information index) is used in model selection. This index and similar ones (e.g., Akaike information criterion) take into account the model log-likelihood (higher is better) and penalize for model complexity (i.e., the number of parameters estimated relative to the sample size). Thus, using them reduces the risk of overfitting the model to a single sample, thereby improving the possibility of replicating the model findings with future samples. In general, a lower value on an information criterion indicates a better model. Mplus reports three information indices, but the most widely used is the BIC (see Hagenaars & McCutcheon, 2002; Nylund et al., 2007; Vermunt, 2006).

Two more statistics, available in Mplus 4.21 (B. Muthén & Muthén, 2006), may be useful for determining the optimal number of classes in a mixture model. The adjusted Lo-Mendell-Rubin test (LMR; B. Muthén, 2004) analytically approximates the LRT distribution, and the bootstrapped LRT (BLRT), suggested by McLachlan and Peel (2000), uses bootstrap samples to empirically derive the sampling distribution of the LRT statistic (Nylund et al., 2007). Both tests, as implemented, compare a k-class model with a $(k - 1)$-class model; in both cases, a statistically significant p value suggests the current model offers improvement over the model with one class fewer.

Also discussed in the literature on mixture models is entropy (B. Muthén, 2004), which, similar to the average posterior probabilities for most likely class membership, serves as a measure of the precision of individual classification. It ranges from 0 (everybody has an equal posterior probability of membership in all classes) to 1 (each individual has posterior probability 1 of membership in a single class and probability 0 of membership in the remaining classes). Equivalently, high entropy indicates clear class separation. Entropy is not a measure of fit, nor was it originally intended for model selection (Ramaswamy, Desarbo, Reibstein, & Robinson, 1993); however, if it is extremely low, that suggests the model may not be useful for some purposes. For instance, if the model is intended to find homogeneous clusters of individuals with distinctive patterns of change (e.g., Nagin, 1999), low entropy indicates that it may be doing a poor job. However, the reverse is not necessarily true; that is, high entropy does not necessarily show the existence of homogeneous clusters of individual trajectories (Feldman & Masyn, 2008).

Residual analysis. The observed and model-predicted cell and marginal proportions can be compared to help assess fit of the model to the data (Carlin et al., 2001; Skrondal & Rabe-Hesketh, 2004). As mentioned earlier, when a model is large and complex, such as a growth process with multiple timepoints, most cells have expected frequencies that are very small or approach zero. Often, however, a few patterns of responses recur relatively frequently. In this case, it may be helpful to inspect standardized Pearson residuals (Haberman, 1973) in the cells with these more common response patterns (B. Muthén & Asparouhov, 2006). In addition, standardized residuals from the univariate margins (i.e., category proportions at each timepoint) and bivariate margins (i.e., proportions in cells of all two-timepoint cross-tabulations) may be inspected (Skrondal & Rabe-Hesketh, 2004). Standardized residuals can be compared with a standard normal distribution (Haberman, 1973); thus, too many standardized residuals greater than two suggest poor fit of the model to the data. All of these observed and expected cell proportions, along with standardized residuals, can be generated in Mplus with the TECH10 output command (see Appendix B). Instructions on how to calculate the expected frequencies when other software is used are available from Betsy J. Feldman.

Plots. Graphical methods can give a clearer picture than numeric output alone of model-predicted trajectories and patterns of change, but using them effectively for model selection depends upon researcher judgment and substantive knowledge. For instance, plots may make it easier to judge whether results of an analysis make interpretive sense or map reasonably onto theory and previous findings. Mixture models, in particular, must be considered in the context of theory, interpretability, and usefulness, because they tend to be used in an exploratory fashion and are innately data-driven.

Linear (and curvilinear) trajectories based on the (within-class) means of the random effects can be plotted, and these give an easy-to-describe picture of the overall mean change; but because the scale of these parameters is arbitrary, the plots can potentially be difficult to interpret. To facilitate interpretation in our line plots, we include the estimated thresholds (τ_j) that divide the categories of the observed data. In addition, we show plots of the time- and class-specific estimated category probabilities, which give a more nuanced picture of how alcohol use changes over time. The predicted-probabilities plots were created in Excel using the TECH7 output from Mplus, and line plots were generated in Excel by using Mplus estimated intercept and slope means multiplied by time (and the model-estimated thresholds, which are fixed across time).

Results

The sample size for these analyses is 451 (236 girls). The observed category proportions at each grade are shown in Figure 5, broken out by gender. Nearly 30% of the target adolescents of both genders reported drinking in 7th grade, and this grew to approximately 65% by 12th grade (Figure 5). Going along with this increase in the likelihood of drinking is an increase in the proportion of youths drinking heavily. By the time the adolescents graduated from high school, approximately 20% of them (17% of girls, 21% of boys) reported drinking weekly or more often (Figure 5). In addition, out of 451 seventh graders, approximately one third
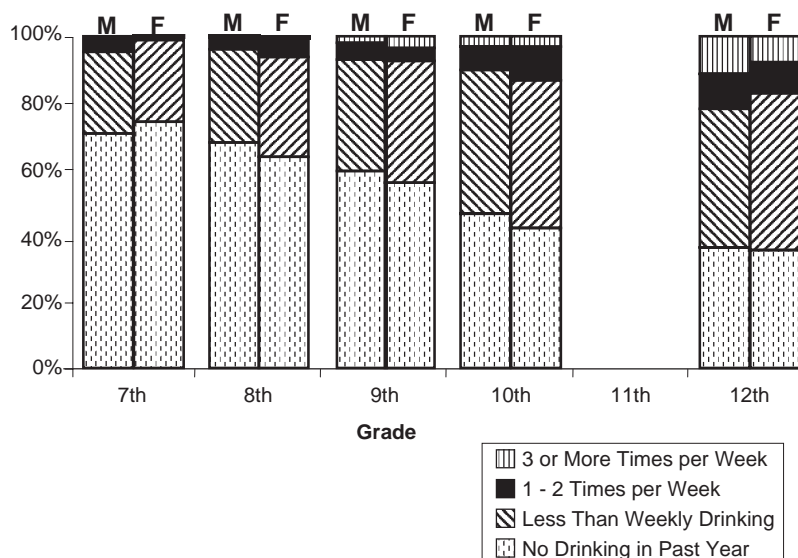
Figure 5. Observed drinking category probabilities by gender and grade.

(27% of girls and 34% of boys) reported having at least one close friend who drank alcohol, and 5% of girls and 7% of boys reported that half or more of their friends drank. Because drinking did not differ by very much between boys and girls (Figure 5), we did not use gender as a covariate in the model.

Model Estimates and Fit Statistics

Fit indices are shown in Table 1 for all models. Bolding indicates the best value for that statistic in the set of models listed in the corresponding row.

HGLM. Two unconditional HGLMs were tested. The first characterized drinking-frequency trajectories over adolescence with an intercept (in 9th grade) and linear slope, and the second added a quadratic slope. The quadratic slope significantly improved the model fit ($\Delta\chi^2 = 49.83$, $\Delta df = 4$, $p < .001$) and was retained. On the basis of this result, the mixture models that follow were also specified with quadratic slopes. The intercept mean was fixed to zero for identification (Equations 9 and 10), and the average linear trend was positive and statistically significant ($\beta_{10} = 0.84$, $p < .001$; see Table 2). The mean quadratic slope was not significant; however, there was considerable individual variability around the intercept and both slopes. In addition, the statistically significant covariances suggest that the frequency with which adolescents were drinking in 9th grade was related to changes in their drinking over the 6-year time span (Table 2). On average, more drinking or more rapidly increasing drinking in 9th grade was associated with trajectories that were flattening out by the end of high school, whereas less frequent drinking in 9th grade was associated with drinking trajectories that were accelerating in 12th grade. The BIC for this model was the lowest of all models tested. The other three statistics in Table 1 do not apply unless there are two or more classes.

GMM. In general, when more within-class variances and covariances were estimated in the growth mixture models, fewer classes were successfully extracted. For example, when all three variances and their covariances were estimated, even the two-class model failed

to converge, but with the covariances fixed to zero, the same two-class model did converge (see Table 1). Increasing the number of classes to three resulted in a singular information matrix.

In the next model, the quadratic variance was constrained to zero, and only intercept and linear slope variances were estimated. The two-class models converged, both with and without the covariance, but there was no significant difference in fit between the two ($\Delta\chi^2 = 0.59$, $df = 1$, $p = .44$), so we constrained the covariance to zero. Here again, the model did not converge with three classes.

In the last set of GMM models, only the intercept variance was estimated, and the remaining variances (and necessarily, all covariances) were fixed at zero. In this case, we were able to estimate up to four classes. The five-class model did not converge (the information matrix was singular; Table 1), and we ruled out the four-class model because one class was too small to estimate reliably (4%, representing approximately 18 individuals). This left plausible models with two and three classes. In comparing the two- and three-class models, the fit statistics did not agree on which was optimal: The LMR preferred two classes, the BIC three classes, and the BLRT no fewer than four classes (Table 1). It is worth noting that entropy for all of the GMM models was much lower than it was for any other mixture models (Table 1). We retained all of the models that converged except the four-class GMM.

LCGA. As mentioned earlier, because all within-class variances are fixed at zero, more classes may be needed in LCGA to characterize the same joint distribution of growth factors. Equivalently, as the number of classes increases, the latent class variable is likely to account for more and more between-individual variability. Although a maximum of only four classes could be estimated with GMM, seven classes were successfully extracted with LCGA. The seven-class LCGA had a class that was too small to estimate well (3%, representing approximately 12 individuals). Nonetheless, the BLRT was still significant at seven classes, suggesting that no fewer than seven classes would be needed to

Table 1
Fit Statistics for All Models

Statistic	1 Class	2 Classes	3 Classes	4 Classes	5 Classes	6 Classes	7 Classes
HGLM							
LL	−1,676.73						
No. of parameters	11						
BIC	3,420.69						
GMM: Intercept, linear, quadratic[a]							
LL		−1,678.06	−1,676.73[b]				
No. of parameters		12	16				
Entropy		.27	—				
BIC		**3,429.46**	—				
BLRT (p)		26.49 (.01)	—				
LMR (p)		27.58 (<.01)	—				
GMM: Intercept, linear[a]							
LL		−1,678.34	−1,670.10[b]				
No. of parameters		11	15				
Entropy		.29	—				
BIC		**3,423.90**	—				
BLRT (p)		29.09 (<.01)	—				
LMR (p)		27.95 (.15)	—				
GMM: Intercept[a]							
LL		−1,697.05	−1,683.17	−1,673.79	−1,665.69[b]		
No. of parameters		10	14	18	22		
Entropy		.34	.38	.52	—		
BIC		3,455.21	**3,451.91**	3,457.59	—		
BLRT (p)		44.65 (<.01)	27.45 (<.01)	18.76 (<.01)	—		
LMR (p)		**42.89** (.04)	26.66 (.06)	18.03 (.14)	—		
LCGA							
LL		−1,737.51	−1,704.75	−1,692.11	−1,682.26	−1,673.20	−1,661.50
No. of parameters		9	13	17	21	25	29
Entropy		.75	.69	.67	.68	.69	.72
BIC		3,530.02	3,488.94	**3,488.11**	3,492.85	3,499.19	3,500.23
BLRT (p)		437.63 (<.01)	65.52 (<.01)	25.28 (<.01)	19.71 (<.01)	18.11 (<.01)	23.41 (<.01)
LMR (p)		420.43 (<.01)	62.95 (<.01)	**24.28** (.01)	18.93 (.12)	17.39 (.54)	22.49 (.01)
LLCA							
LL		−1702.59	−1671.25	−1652.50	−1639.53	−1627.09	−1617.49[c]
No. of parameters		31	47	63	79	95	111
Entropy		.81	.73	.79	.86	.79	—
BIC		**3,594.64**	3,629.74	3,690.02	3,761.87	3,834.76	—
BLRT		481.94 (<.01)	62.68 (<.01)	**37.50** (<.01)	25.94 (.31)	24.89 (.27)	—
LMR		**477.06** (<.01)	62.05 (.95)	37.12 (.23)	25.67 (1.00)	24.64 (1.00)	—

Note. All models estimated with intercept, and linear and quadratic slopes. Dashes indicate that the values could not be obtained (because the model failed to converge). Bolding indicates the best value for that statistic in the set of models listed in the corresponding row. HGLM = hierarchical generalized linear model (all variances and covariances estimated); GMM = growth mixture model; LCGA = generalized linear latent class growth analysis (all variances and covariances estimated); LLCA = longitudinal latent class analysis; LL = log-likelihood; BIC = Bayesian information criterion; BLRT = bootstrapped likelihood ratio test; LMR = Lo-Mendel-Rubin likelihood ratio test.
[a] Variances that were estimated; all covariances constrained to zero. [b] Singular information matrix. [c] No repeated log-likelihood.

model these data (Table 1). The BIC and LMR both selected the four-class LCGA, but the three-class model had a BIC value that was nearly equivalent to that of the four-class model. We retained the three- and four-class models.

LLCA. The LLCA failed to replicate the best log-likelihood when seven classes were specified, and none of the fit statistics selected more than four classes (Table 1). The BLRT suggested that four classes were adequate, and the BIC and LMR both selected the two-class model. Because the LLCA models require so many parameters, it is not surprising that no fit index selected a model with more than four classes, nor that its BIC values were consistently higher than for any other models (as was entropy, suggesting cleaner class separation when more parameters were estimated). We retained the models with two, three, and four classes.

Residuals

We inspected the univariate and bivariate residuals for the HGLM and the best 10 out of the 17 mixture models that had converged properly. Table 3 shows the frequencies of univariate and bivariate residuals greater than two.[13] All models were able to replicate the univariate observed values adequately, but there was more variability among the bivariate residuals (Table 3). The three- and four-class LLCA models fit the bivariate cell counts the

---

[13] Absolute size of standardized residuals can also be taken into consideration if they are very large (e.g., >10), which did not occur in these analyses.

Table 2
Hierarchical Generalized Linear Model Results

Parameter	Estimate	SE	p
Means			
β_{00} Intercept	0.00[a]	—	—
β_{10} Linear slope	0.84	0.10	<.001
β_{20} Quadratic slope	−0.06	0.04	.098
Variances and covariances			
Ψ_{11} Intercept	10.28	1.58	<.001
Ψ_{22} Linear slope	0.79	0.17	<.001
Ψ_{33} Quadratic slope	0.14	0.03	<.001
Ψ_{12} Intercept with linear (r)	0.58 (.21)	0.30	.055
Ψ_{13} Intercept with quadratic (r)	−0.52 (−.44)	0.17	.003
Ψ_{23} Linear with quadratic (r)	−0.18 (−.55)	0.05	<.001
Thresholds			
τ_1	0.53	0.19	.006
τ_2	5.34	0.35	<.001
τ_3	7.35	0.46	<.001

[a] Intercept mean is fixed to zero for identification. Dashes indicate that values are not applicable in this case.

most accurately, and the worst-fitting were the two-class models: LLCA and the three GMMs (Table 3).

Table 4 shows comparisons of observed and predicted cell frequencies for the most commonly observed patterns of responses over the five occasions. Note that, although each of the four categories of the outcome were endorsed by at least a few individuals at every time point, the most commonly observed patterns contained only zeros and ones (representing nondrinkers [0] and infrequent drinkers [1]). Here again, the three- and four-class LLCA models offered the best fit to the observed data. Only the two-class LLCA model did a noticeably worse job than the other models under consideration at reproducing the cell counts. It is also worth noting that the frequency of one response pattern (01111) was not reproduced well by any model except the most highly parameterized LLCAs.

In summary, the bivariate residuals suggested that all of the two-class models fit the data relatively poorly, and the two-class LLCA also failed to adequately match the common response patterns. We dropped the two-class models from further consideration. In contrast, three- and four-class LLCAs both fit the observed frequencies particularly well, but the four-class model required 16 more parameters and had a higher BIC (Table 1). Thus, we rejected the four-class LLCA in favor of the more parsimonious three-class model. We continued to consider the HGLM, the three-class GMM with intercept variance, the three-class LLCA, and three- and four-class LCGAs.

Plots

Plots of model-predicted trajectories and patterns of change can also help a researcher choose the most appropriate or reasonable model from several competing ones, and plots are often used this way in model selection (e.g., Greenbaum, Del Boca, Darkes, Wang, & Goldman, 2005; B. Muthén & Muthén, 2000b). Figure 6 shows a comparison of the estimated mean curves, based on the y^* metric, generated by the three-class GMM (top) and three-class LCGA (bottom). Because the log-odds trajectories are on arbitrary scales, the plots show category thresholds as dashed lines to

facilitate comparison across analyses. In addition, the GMM plot includes *SD* bars above and below the class means at each time-point (based on the intercept variance). These *SD* bars illustrate the overlap of classes that accompanies the extremely low entropy. The GMM finds one unexpected class, comprising slightly over one third of the sample, in which average drinking probabilities are constant across adolescence. This class shows approximately 50% of the individuals reporting alcohol use at each time (the first threshold separates nondrinkers from drinkers). Its other two GMM classes are increasing, one early and one later, which is consistent with the covariances found in the HGLM (Table 2). The LCGA plot shows one class with consistently low drinking probabilities, one with rapidly increasing probabilities, and one with relatively high drinking probabilities across all timepoints. These last two classes are also consistent with the HGLM covariances in Table 2. That is, one class has lower drinking probabilities in

Table 3
Univariate and Bivariate Standardized Residuals

	Value of standardized residual			
Model	2–3	3–4	4–5	Total
Univariate (16 cells)				
HGLM	1	0	0	1
GMM				
Intercept[a]				
2 classes	0	0	0	0
3 classes	1	0	0	1
Intercept, linear slope[a]				
2 classes	2	0	0	2
Intercept, linear, quadratic slopes[a]				
2 classes	2	0	0	2
LCGA				
3 classes	0	0	0	0
4 classes	0	0	0	0
LLCA				
2 classes	0	0	0	0
3 classes	0	0	0	0
4 classes	0	0	0	0
Bivariate (160 cells)				
HGLM	6	0	1	7
GMM				
Intercept[a]				
2 classes	13	3	0	16
3 classes	6	4	0	10
Intercept, linear[a]				
2 classes	14	4	0	18
Intercept, linear, quadratic[a]				
2 classes	12	3	1	16
LCGA				
3 classes	11	0	1	12
4 classes	8	0	1	9
LLCA				
2 classes	12	3	0	15
3 classes	3	0	0	3
4 classes	2	0	0	2

Note. All models estimated with intercept, and linear and quadratic slopes. HGLM = hierarchical generalized linear model (all variances and covariances estimated); GMM = growth mixture model; LCGA = latent class growth analysis (all variances and covariances estimated); LLCA = longitudinal latent class analysis.
[a] Variances that were estimated; all covariances constrained to zero.

Table 4
Observed and Predicted Frequencies of Most Common Observed Response Patterns

Response pattern	Observed frequency	Estimated frequencies									
		HGLM	GMM: Classes				LCGA: Classes		LLCA: Classes		
			Intercept[a]		Intercept, linear[a]	Intercept, linear, quadratic[a]					
			2	3	2	2	3	4	2	3	4
00000	81	74.40	75.33	74.68	82.71	82.61	75.65	78.99	61.51	71.93	73.73
00001	33	41.60	36.22	37.97	35.57	31.87	36.86	38.23	42.45	36.97	35.04
00011	21	20.22	24.86	20.93	14.93	17.06	16.57	18.18	17.34	17.55	22.53
01111	21	11.94	9.84	10.56	12.55	11.56	11.17	11.64	12.96	15.44	16.40
00111	19	16.79	17.53	20.02	14.02	16.51	19.15	19.25	8.00	19.99	18.88
00010	15	11.22	12.39	13.85	11.01	12.59	9.21	15.48	24.36	16.91	14.43
					Standardized residuals						
00000	81	0.86	0.73	0.82	−0.21	0.72	0.69	0.26	2.72	1.19	0.95
00001	33	1.42	0.56	0.85	−0.45	−1.06	−0.67	−0.89	−1.54	−0.69	−0.36
00011	21	0.18	0.80	0.01	1.60	0.51	1.11	0.68	0.90	0.85	−0.33
01111	21	2.67	3.61	3.26	2.43	2.82	2.99	2.79	2.27	1.45	1.16
00111	19	0.55	0.36	0.23	1.36	0.63	−0.03	−0.06	3.93	−0.23	0.03
00010	15	1.15	0.75	0.31	1.22	0.69	1.93	−0.13	−1.96	−0.48	0.15

Note. A response pattern is the pattern of observed responses across all five waves. In this case, the most common pattern was all zeros (i.e., 81 individuals did not report drinking at any wave). All models were estimated with intercept and linear and quadratic slopes. HGLM = hierarchical generalized linear model (all variances and covariances estimated); GMM = growth mixture model; LCGA = latent class growth analysis (all variances and covariances estimated); LLCA = longitudinal latent class analysis.
[a] Variances that were estimated; all covariances constrained to zero.

ninth grade, but those probabilities are increasing rapidly; the other, more likely to be drinking (or drinking more) in ninth grade, is increasing more slowly (and shows more quadratic slope; Figure 6).

The second way to plot the trajectories and patterns of change is to show the predicted category probabilities from each time (see Figure 7). This gives a somewhat richer picture of the change because it illustrates the shifts in proportions at each level of a categorical variable. It complements the plots of y^*, shown in Figure 6, and when the categories are not evenly spaced or have different substantive meanings, this approach can be more informative than a line plot. Figure 7 shows a comparison of the three-class GMM, LCGA, and LLCA (which *only* yields category probabilities). The LLCA and LCGA classes are very similar to one another, both in class-specific patterns of change and size, suggesting that the extra parameters in LLCA are not yielding much more information. Both are consistent with prior literature (e.g., Hix-Small, Duncan, Duncan, & Okut, 2004; White, Johnson, & Buyske, 2000) in which researchers have found classes with low, high, and increasing alcohol-use probabilities across adolescence.

Only GMM found the class with constant moderate drinking probabilities, and Figure 7 shows that this class groups together individuals who responded in all four drinking categories. Because this class fails to differentiate high versus low drinking probabilities (all categories are represented in the class), it may not be very informative. It is also not consistent with prior literature. In addition to these problems, the overlap of classes in GMM makes it difficult to interpret, and a mixture model with such indistinct classes may offer little if any advantage over the more standard

HGLM without classes. For all of these reasons, we eliminated it from further consideration.

The four-class LCGA was selected by both LMR and BIC as the best LCGA model; however, the BICs for the three- and four-class models were very close (Table 1), and overall, the two models did approximately equally well with respect to the standardized residuals (Tables 3 and 4). Figure 8 shows the small added class (12%; labeled "High at Wave 1"), which is largely a subset of the low class from the three-class model, but with higher drinking probabilities in the 7th and 8th grades (such class splitting can be investigated with a cross-tabulation of modal class assignments). By 10th grade, it does not differ substantially from the low class. This type of pattern, in which adolescents drink more in the early and later years, but not in between, is not predicted by theory or prior literature (e.g., Colder, Campbell, Ruel, Richardson, & Flay, 2002; Hill et al., 2000; Orlando, Tucker, Ellickson, & Klein, 2005). It is our feeling that this added class arises as an artifact of a testing effect that is well established among self-report measures in longitudinal studies: Higher levels of negative feelings and behaviors are often reported at the first wave of data collection (see, e.g., Knowles, Coker, Scott, Cook, & Neville, 1996; Twenge & Nolen-Hoeksema, 2002). An inflated 1st-year report could explain the added class, so we decided to eliminate the four-class LCGA.

At this point, on the basis of combinations of fit statistics, residuals, considerations of parsimony, and plots, we felt that all two- and four-class models, as well as all GMMs, could be reasonably ruled out for this data set, leaving the three-class LCGA and LLCA models and the HGLM still under consideration.
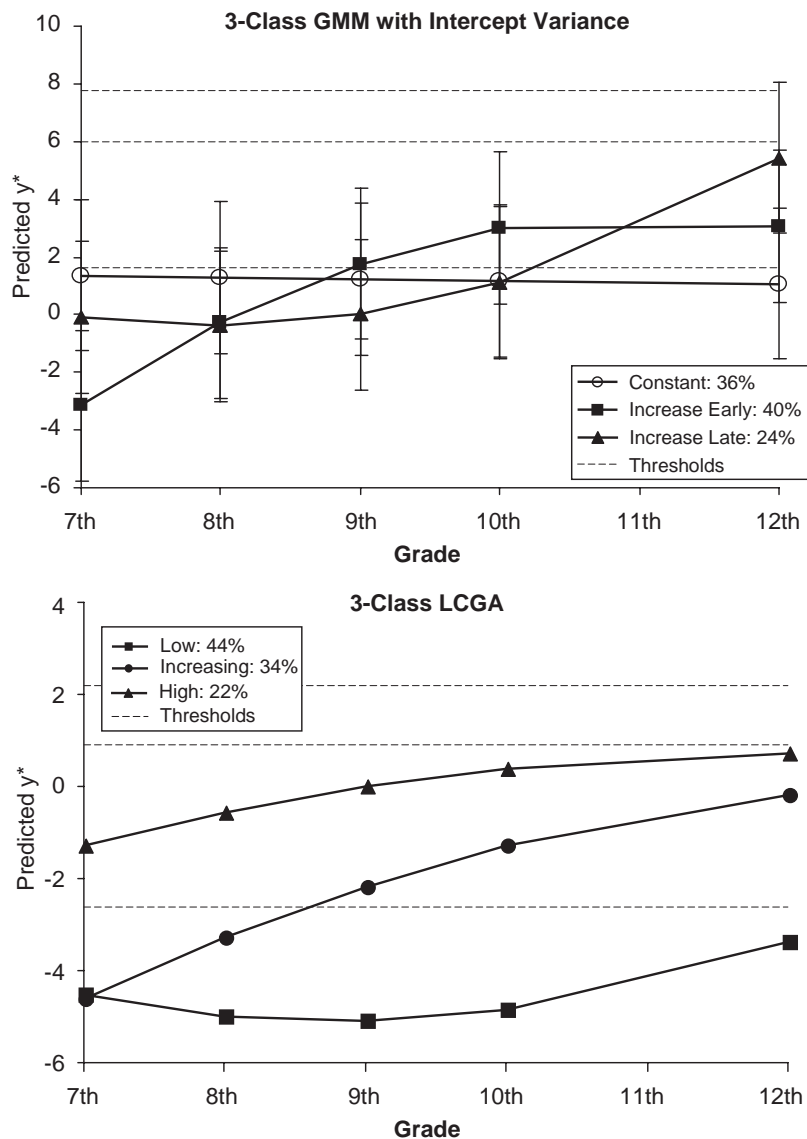
Figure 6. Estimated mean trajectories for three-class growth mixture model (GMM) and latent class growth analysis (LCGA). GMM plot includes standard deviation bars at each time. In both plots, thresholds are shown by dashed lines.

Adding Covariates

After selecting the best unconditional models, we used the adolescents' seventh-grade report of alcohol consumption by close friends to explain individual differences and to help guide model selection. The covariate was allowed to predict class membership in the mixture models and the growth-factor variability in the HGLM. If a GMM were still under consideration, it would be possible to use the covariate to explain within-class variation, as well as the probability of class membership. In the case of GMM, as in the other cases, covariates are used to explain individual differences. When a covariate is permitted to affect both within-class and between-class variability, it is predicting a complex combination of an individual's probability of class membership and that individual's trajectory, relative to the mean trajectory in

each class in which the individual has some non-zero probability of membership. This can make the covariate's effects potentially difficult to interpret.

In the HGLM, targets with more close 7th-grade friends drinking alcohol were themselves predicted to be drinking more in 9th grade ($\beta_{01} = 1.36$, $SE = 0.24$, $p < .001$), but it was also predicted that their drinking would be increasing more slowly in 9th grade ($\beta_{11} = -.19$, $SE = 0.07$, $p = .011$). Despite this effect, however, a target who reported that even a few of his or her friends were drinking in 7th grade (i.e., 1 on the 0 to 4 scale) would still end up, on average, drinking more in 12th grade than an adolescent who reported having no drinking friends in 7th grade (see Figure 9).

The regression coefficient for a single class represents the change in the log-odds of membership in that class versus a reference class for
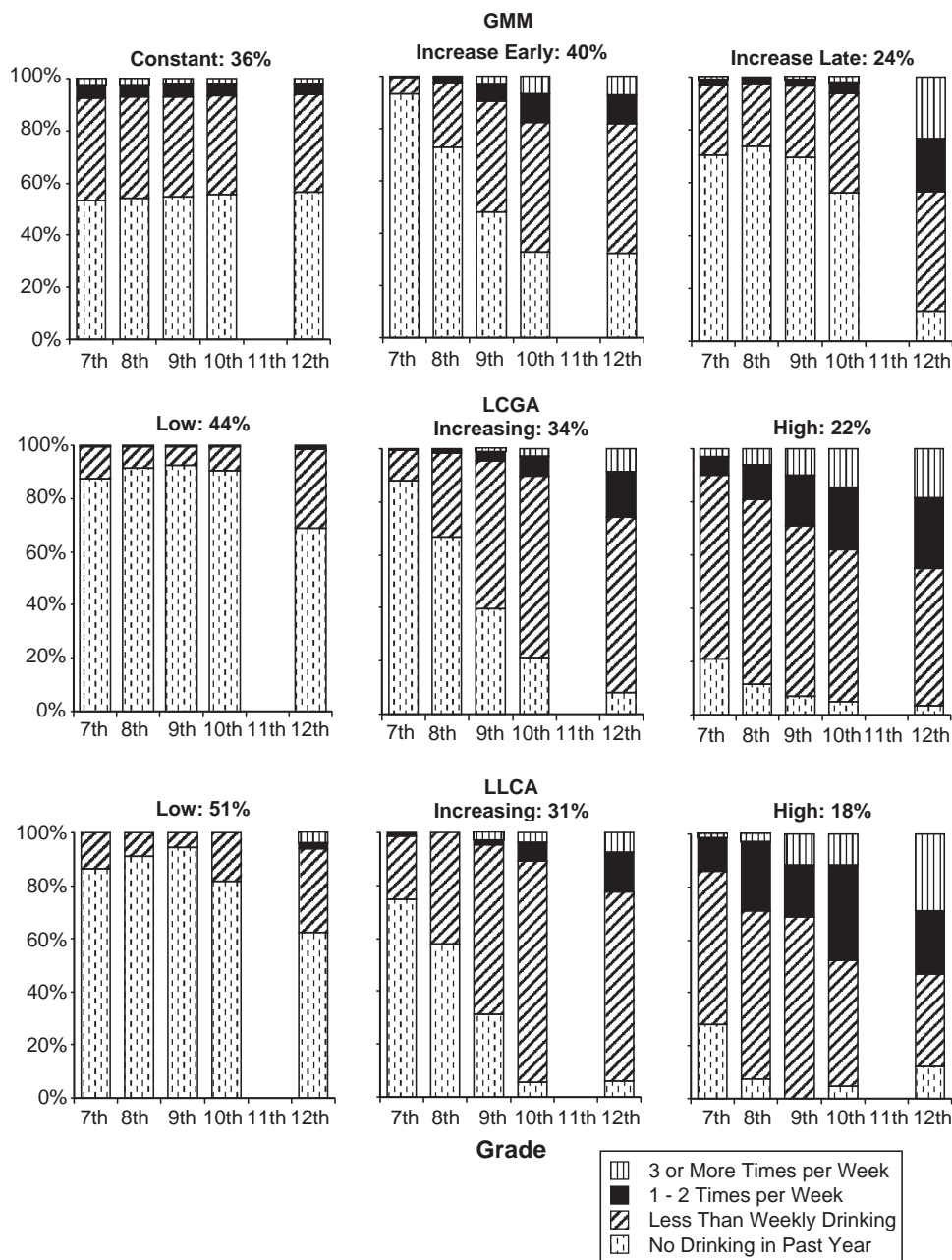
Figure 7. Plots of predicted category proportions for three-class models: growth mixture model (GMM; top), latent class growth analysis (LCGA; middle), and longitudinal latent class analysis (LLCA; bottom).

a one-unit change in the covariate (for members of those two classes). The reference class can be any class, and Mplus 4.21(B. Muthén & Muthén, 2006) gives results with each class, in turn, treated as the reference. Viewing results this way permits comparison of all pairs of classes with respect to the covariate, which may be more useful than simply using a single reference class. Note, however, that as the number of classes increases, testing all possible pairs may not be necessary or useful and could introduce the possibility of an inflated Type-1 error rate. In addition, global tests of association between the latent class variable (with a given number of classes) and one or more

covariates can be conducted using the standard likelihood ratio test for nested models.

Table 5 shows all paired class comparisons in three-class LCGA and LLCA models. The results across the two models are very similar. Members of the high class reported more alcohol-using friends in seventh grade than either of the remaining two classes, which did not differ significantly from one another. In deciding between LLCA and LCGA three-class models, we weighed the better fit of the LLCA against its lack of parsimony compared to the LCGA (the LLCA requires 34 more parameters; Table 1). If the LLCA patterns of

change had shown noticeable signs of nonlinearity, then it would have made sense to choose that model over LCGA. However, it did not (Figure 7), and the substantive stories told by the trajectory plots and the relationships between classes and the covariate are the same for the two models. On the basis of all of these factors, we retained the LCGA and dropped the LLCA. Parameter estimates for the conditional three-class LCGA are shown in Table 6.

Summary

A single set of categorical adolescent-drinking data was analyzed using a HGLM with linear and quadratic slopes and 21 specifications of mixture models. The HGLM showed a statistically significant mean linear increase in drinking (with a nonsignificant mean quadratic slope) and reflected significant variability around the intercept, in ninth grade, and in both linear and quadratic slopes. All of the mixture models also revealed an overall increase in drinking across time. In general, freeing more variances and covariances in the GMMs led to fewer classes being estimable, resulting in a range from one class (equivalent to HGLM), when all variances and covariances were estimated, to seven classes, when all variances were constrained to zero (in the LCGA). On the basis of the convergence properties, fit indices (BIC and LMR proved most useful), and, in two cases, the extraction of unacceptably small classes, we were able to pare our list down to 10 potential models, including the HGLM.

We next considered the fit in terms of model-predicted versus observed category frequencies by looking at standardized residuals for
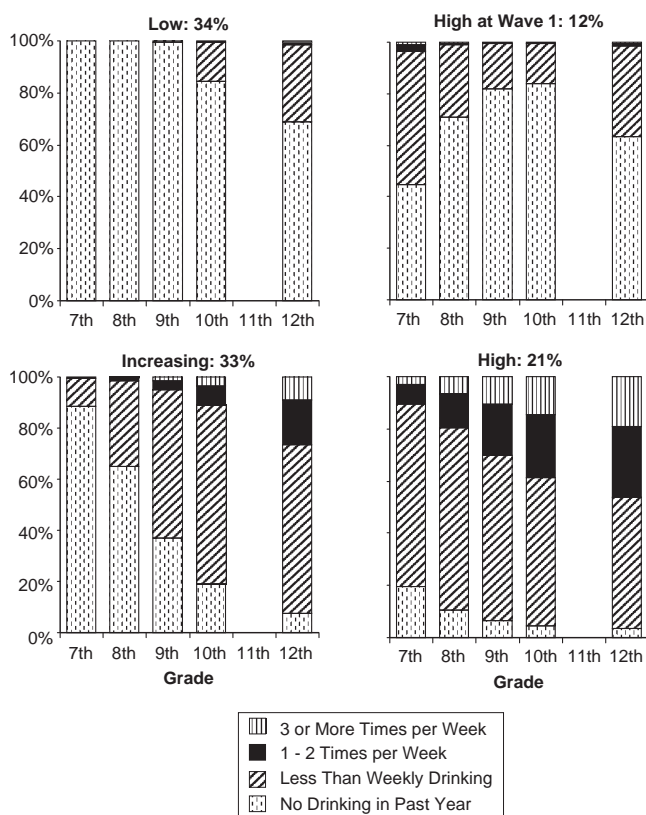

Figure 8. Plots of predicted category proportions for four-class latent class growth analysis (LCGA).
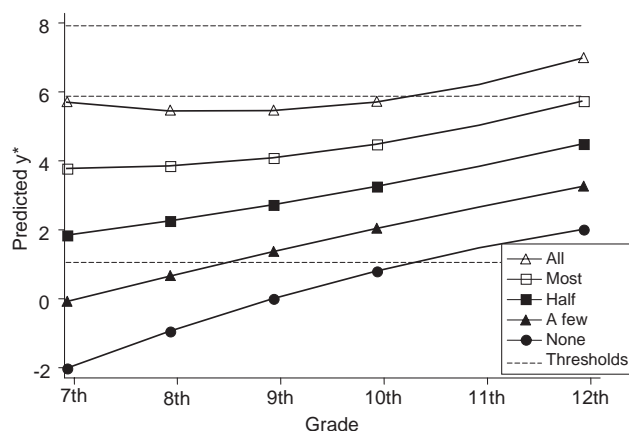

Figure 9. Plot of predicted hierarchical generalized linear model (HGLM) trajectories by levels of the covariate, target's report of the number of alcohol-using friends in 7th grade (range of scale: 0 = *none* to 4 = *all of them*).

the univariate and bivariate marginal frequencies (Table 3) and the most commonly occurring patterns of responses (Table 4). Too many standardized residuals larger than 2 is a sign of poor fit of the model to the data. All models fit the univariate marginal frequencies fairly well, and all but one—the two-class LLCA model—fit the common-pattern frequencies adequately. We eliminated the two-class LLCA and also dropped the remaining two-class models, all of which had 15 or more large bivariate residuals (this cutoff was chosen somewhat arbitrarily). The three- and four-class LLCA models both predicted cell frequencies extremely well, so we eliminated the four-class model, which required 16 more parameters and had a higher BIC.

For the remaining selection decisions, we used plots to help us consider substantive issues: interpretability, usefulness, relationships with covariates, and consistency with theory and previous literature. On the basis of the plots, we rejected the sole remaining GMM. One of its classes was of little use in terms of explaining outcomes, and it lacked a class with consistently low drinking probabilities that is predicted by theory and previous literature (e.g., Colder et al., 2002).[14] Finally, the plots showed how poor the class separation was when the intercept variances were estimated. Given the overlap between classes and the difficulty in estimating and interpreting this model, even with all other things being equal, we would see little reason to select this GMM over the more straightforward and parsimonious HGLM.

We also compared the plots for the three- and four-class LCGAs, which had very similar BICs (Table 1). The added class showed a pattern that was neither predicted by theory nor found in previous studies of adolescent alcohol use. Rather, this new class appeared to be an artifact resulting from test–retest effects often found in longitudinal studies comprising self-reported data. We rejected the four-class model in favor of the three-class.

The three-class LCGA and LLCA models were very similar to one another (Figure 7). Both had one large class with low drinking probabilities across adolescence and a small increase toward the end

---

[14] It is possible in GMM and other mixture models to specify a class with zero probability of drinking at all timepoints if such a class is believed to exist. For details on how this is done, see Kreuter and Muthén (2007).

Table 5
Effects of Seventh Grade Friends Who Drink on Three-Class LCGA and LLCA Trajectory Class Membership

	LCGA				LLCA			
Variable	Est. (*SE*)	OR	95% CI	*p*	Est. (*SE*)	OR	95% CI	*p*
High to low	1.33 (0.48)	3.79	1.48−9.67	.005	1.39 (0.35)	4.03	2.03−7.99	<.001
Increasing to low	0.45 (0.41)	1.57	0.71−3.48	.266	0.32 (0.40)	1.38	0.63−3.01	.423
High to increasing	0.88 (0.31)	2.41	1.31−4.42	.005	1.07 (0.35)	2.93	1.48−5.80	.002

Note. LCGA = latent class growth analysis; LLCA = longitudinal latent class analysis; Est. = estimate; OR = odds ratio.

of high school (labeled "low"), a small class with relatively high drinking probabilities (including more frequent drinking) across adolescence ("high"), and a class that started with low drinking probabilities in 7th grade and then increased fairly rapidly until its drinking probabilities were similar to the "high" class in 12th grade ("increasing").

Finally, we added a covariate. The three remaining models (LCGA and LLCA three-class models, and HGLM) all showed that adolescents who reported having close friends who used alcohol in seventh grade were likely to be drinking more frequently themselves across all of the junior high and high school years. Such similar findings across the three models suggest that all offer reasonable representations of drinking across middle and late adolescence, and given the large difference in parsimony between LCGA and LLCA (13 vs. 47 parameters, respectively), it made sense to reject the LLCA in favor of the LCGA. In situations where plots show that change over time is less continuous and linear or curvilinear (e.g., episodic problems like suicidality) or when the proportional odds assumption is not tenable, LLCA may offer considerable advantage over the other models.

Table 6
Conditional Three-Class Generalized Linear Latent Class Growth Analysis (LCGA) Parameter Estimates

Parameters	Estimate	*SE*	*p*
High			
β_{00} Intercept	5.29	0.43	<.001
β_{10} Linear slope	0.45	0.19	.017
β_{20} Quadratic slope	−0.08	0.06	.199
Increasing			
β_{00} Intercept	3.07	0.43	<.001
β_{10} Linear slope	0.86	0.18	<.001
β_{20} Quadratic slope	−0.08	0.06	.166
Low			
β_{00} Intercept	0.00[a]	—	—
β_{10} Linear slope	0.10	0.15	.492
β_{20} Quadratic slope	0.16	0.06	.001
Covariate effects (class comparisons)			
High vs. low	1.33	0.48	.005
Increasing vs. low	0.45	0.41	.266
High vs. increasing	0.88	0.31	.005
Thresholds			
τ_1	2.51	0.36	<.001
τ_2	5.99	0.37	<.001
τ_3	7.30	0.38	<.001

[a] Intercept mean for one class is fixed to zero for identification. Dashes indicate that values are not applicable because no standard error is estimated when a parameter is fixed to a specific value (zero, in this case).

The two remaining models, HGLM and the three-class LCGA, were equivalent in terms of fit to the data and to theory, and they told essentially the same substantive story, so we did not choose between them. It is possible that, based on further analyses, one of the two would stand out as a clearly better model. For example, differences might emerge when other covariates are used, or one of the two models could prove a better predictor of individual values on a distal outcome.

Discussion

In any data-analysis situation, there is a multitude of decisions that must be made, from determining a statistical model for the distribution of the data to choosing the structural models that best answer the research questions. The study of problem behaviors can be a particular challenge; the behaviors represent developmental processes, requiring longitudinal analyses, and because the behaviors are rare, their distributions are highly skewed. In addition, most data measuring these behaviors are discrete (binary, ordinal, or count). We have presented a selection of alternative models for analyzing longitudinal categorical data. These approaches assume more appropriate distributions for problem-behavior data than the traditional continuous-data models, permitting researchers to have more faith in the results.

Starting with the concepts involved in the generalized linear model (GLM) for categorical data, we discussed longitudinal extensions, including the hierarchical (or longitudinal) version of the generalized linear model (HGLM), and several specifications of mixture models. Although there has been a great deal of excitement recently over using mixture models to find homogeneous unobserved subgroups in the population, we showed how such approaches can also be viewed as extensions of the random-effects model (Bauer & Curran, 2003; B. Muthén & Asparouhov, 2006) that relax some of its strong assumptions. Because of this, they might be the most appropriate choices for certain types of data distributions, irrespective of any question of subpopulations.

Our goal in this article was to present a group of related models for longitudinal categorical data and to offer substantive researchers a useful guide to testing and selecting between the alternative models. This is a particularly difficult issue with categorical data and complicated longitudinal models, such as mixture models, because, as we find here, there are no completely reliable fit statistics for these types of models and no readily available measures of absolute fit. We suggested several considerations to use in model selection. The first have to do with convergence properties, measures of comparative fit, and residual analysis. We dismissed models that did not converge well (including both improper solutions and failure to repeat the best log-likelihood value). From the remaining models, we chose the most

plausible ones on the basis of the fit statistics and then ruled out several more because they failed to adequately reproduce the observed cell frequencies or were equivalent to more parsimonious models with respect to fit. Next, we included substantive considerations in model selection. For instance, we rejected a model because its substantive meaning was both obscure, rendering the model less useful, and not a reasonable fit to theory or to prior research. We also selected one of two similar models, despite poorer fit, because it was more parsimonious. Of course, we tested a limited number of models and only a single covariate. Choice and specification of covariates are important factors in model selection, and differences in these could potentially lead to different model choices.

Through this process of elimination, we selected the two models that offered the most reasonable combination of parsimony, fit to the data, and fit to theory—the HGLM and the three-class LCGA. The two models also use roughly the same number of parameters (11 and 13, respectively), and the substantive findings were similar. Both showed that although adolescents are increasingly likely to drink as they approach the end of high school, those who reported having more friends who were drinking in seventh grade were more likely to be drinking themselves throughout the rest of the secondary school years. The choice between these two models, based on the data we have presented, is, in our opinion, largely a matter of ease of interpretation, although further investigation could turn up substantive or statistical reasons to choose one over the other or even to select a different model over these two.

This study has the strengths and the weaknesses that inevitably result from using actual data as an example. On the one hand, it shows how the models might look in an actual research situation. On the other hand, with simulated data, we could compare the model results with a known population to discover whether there are any systematic biases in the analyses. It is also possible that different approaches to measuring alcohol use could yield different results (Feldman & Masyn, 2008). Finally, there is one more critical piece that warrants further study. Often the purpose of longitudinal studies is to predict what will happen to the young people later in life. The models may not be of much use if they fail in that respect, no matter how well they appear to characterize the data at hand. Future research needs to take distal outcomes into account in assessing the adequacy of the models.

In this article we present and compare a fairly large set of models, but it is not an exhaustive list of potential models. There is an infinite set of possible models, and with empirical data, we cannot know what the true model is. For instance, there are models outside of the GLM framework that could be considered when growth is not linear or curvilinear, as well as two-part models, in which behavior is modeled as two related processes: (a) exhibiting the behavior versus not doing so, and (b) intensity or extent of the behavior (Olsen & Schafer, 2001; Tooze et al., 2002). However, although there remains some uncertainty with respect to model choice, we feel that all of the methods shown here may offer improvement over approaches that treat all data, irrespective of their actual distributions, as continuous and normally distributed.

In addition, because comparing longitudinal categorical models is not always straightforward, we have offered a systematic approach to assessing and selecting models from among several competing, equally appropriate statistical options. It is our belief that adopting this approach will improve researchers' ability to understand trajectories of problem behaviors over childhood and adolescence. With improved models, we can better determine risks and protective factors, and this information could potentially be used for early identification of those most at risk, helping to select the best candidates for targeted preventive interventions.

References

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). Hoboken, NJ: Wiley.

Agresti, A., Booth, J. G., Hobert, J. P., & Caffo, B. (2000). Random-effects modeling of categorical response data. *Sociological Methodology, 30,* 27–80.

Ashford, J. R. (1959). An approach to the analysis of data for semi-quantal responses in biological assay. *Biometrics, 15,* 573–581.

Bauer, D. J. (2005). A semiparametric approach to modeling nonlinear relations among latent variables. *Structural Equation Modeling, 12,* 513–535.

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods, 8,* 338–363.

Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural modeling. *Sociological Methods and Research, 16,* 78–117.

Blozis, S. A. (2004). Structured latent curve models for the study of change in multivariate repeated measures. *Psychological Methods, 9,* 334–353.

Blozis, S. A. (2007). On fitting nonlinear latent curve models to multiple variables measured longitudinally. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(2), 179–201.

Bock, R. D., & Lieberman, M. (1970). Fitting a response model for *n* dichotomously scored items. *Psychometrika, 35*(2), 179–197.

Butler, J. S., & Moffitt, R. (1982). A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica, 50,* 761–764.

Carlin, J. B., Wolfe, R., Brown, C. H., & Gelman, A. (2001). A case study on the choice, interpretation and checking of multilevel models for longitudinal binary outcomes. *Biostatistics, 4,* 397–416.

Cicchetti, D., & Toth, S. L. (1998). The development of depression in children and adolescents. *American Psychologist, 53,* 221–241.

Colder, C. R., Campbell, R. T., Ruel, E., Richardson, J. L., & Flay, B. R. (2002). A finite mixture model of growth trajectories of adolescent alcohol use: Predictors and consequences. *Journal of Consulting and Clinical Psychology, 70,* 976–985.

Collins, L. M., Graham, J. W., Rousculp, S. S., & Hansen, W. B. (1997). Heavy caffeine use and the beginning of the substance use onset process: An illustration of latent transition analysis. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 79–99). Washington, DC: American Psychological Association.

Conger, R. D., & Conger, K. J. (2002). Resilience in Midwestern families: Selected findings from the first decade of a prospective, longitudinal study. *Journal of Marriage and Family, 64,* 361–373.

Curran, P. J., Stice, E., & Chassin, L. (1997). The relation between adolescent alcohol use and peer alcohol use: A longitudinal random coefficients model. *Journal of Consulting and Clinical Psychology, 65,* 130–140.

DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. *Structural Equation Modeling, 9,* 327–346.

Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology, 47,* 309–326.

Donovan, J. E., & Jessor, R. (1985). Structure of problem behavior in adolescence and young adulthood. *Journal of Consulting and Clinical Psychology, 53,* 890–904.

Feldman, B. J., & Masyn, K. E. (2008, May). *The measurement and modeling of alcohol use: A simulation study.* Paper presented at the International Meeting of the Psychometric Society, San Francisco.

Ferrer, E., & McArdle, J. J. (2003). Alternative structural models for multivariate longitudinal data analysis. *Structural Equation Modeling, 10,* 493–524.

Foster, E. M., Fang, G. Y., & the Conduct Problems Prevention Research Group. (2004). Alternative methods for handling attrition: An illustration using data from the Fast Track evaluation. *Evaluation Review, 28,* 434–464.

Goldberger, A. S. (1964). *Econometric theory.* New York: Wiley.

Graham, J. W., Collins, L. M., Wugalter, S. E., Chung, N. K., & Hansen, W. B. (1991). Modeling transitions in latent stage-sequential processes: A substance use prevention example. *Journal of Consulting and Clinical Psychology, 59,* 48–57.

Greenbaum, P. E., Del Boca, F. K., Darkes, J., Wang, C.-P., & Goldman, M. S. (2005). Variation in the drinking trajectories of freshmen college students. *Journal of Consulting and Clinical Psychology, 73,* 229–238.

Gurland, J., Lee, I., & Dahm, P. A. (1960). Polychotomous quantal response in biological assay. *Biometrics, 16,* 382–398.

Haberman, S. J. (1973). The analysis of residuals in cross-classified tables. *Biometrics, 29,* 205–220.

Hagenaars, J. A., & McCutcheon, A. L. (Eds.). (2002). *Applied latent class analysis.* New York: Cambridge University Press.

Harville, D. A., & Mee, R. W. (1984). A mixed-model procedure for analyzing ordered-categorical data. *Biometrics, 40,* 393–408.

Hawkins, J. D., Catalano, R. F., & Miller, J. Y. (1992). Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: Implications for substance abuse prevention. *Psychological Bulletin, 112,* 64–105.

Hedeker, D., & Gibbons, R. D. (1994). A random-effects ordinal regression model for multilevel analysis. *Biometrics, 50,* 933–944.

Hedeker, D., Gibbons, R. D., & Flay, B. R. (1994). Random-effects regression models for clustered data with an example from smoking prevention research. *Journal of Consulting and Clinical Psychology, 62,* 757–765.

Hedeker, D., & Mermelstein, R. J. (1998). A multilevel thresholds of change model for analysis of stages of change data. *Multivariate Behavioral Research, 33,* 427–455.

Hedeker, D., & Mermelstein, R. J. (2000). Analysis of longitudinal substance use outcomes using ordinal random-effects regression models. *Addiction, 95*(Suppl. 3), S381–S394.

Hernán, M. A., Hernández-Diaz, S., & Robins, J. M. (2004). A structural approach to selection bias. *Epidemiology, 15,* 615–625.

Hill, K. G., White, H. R., Chung, I.-J., Hawkins, J. D., & Catalano, R. F. (2000). Early adult outcomes of adolescent binge drinking: Person- and variable-centered analyses of binge drinking trajectories. *Alcoholism: Clinical and Experimental Research, 24,* 892–901.

Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods, 11,* 36–53.

Hix-Small, H., Duncan, T. E., Duncan, S. C., & Okut, H. (2004). A multivariate associative finite growth mixture modeling approach examining adolescent alcohol and marijuana use. *Journal of Psychopathology and Behavioral Assessment, 26,* 255–270.

Jackson, K. M., Sher, K. J., & Schulenberg, J. E. (2005). Conjoint developmental trajectories of young adult alcohol and tobacco use. *Journal of Abnormal Psychology, 114,* 612–626.

Johnston, L. D., O'Malley, P. M., Bachman, J. G., & Schulenberg, J. E. (2006). *Monitoring the future: National results on adolescent drug use. Overview of key findings, 2005* (NIH Publication No. 06–5882). Bethesda, MD: National Institute on Drug Abuse.

Knowles, E. S., Coker, M. C., Scott, R. A., Cook, D. A., & Neville, J. W. (1996). Measurement-induced improvement in anxiety: Mean shifts with repeated assessment. *Journal of Personality and Social Psychology, 71,* 352–363.

Kreuter, F., & Muthén, B. (2007). Longitudinal modeling of population heterogeneity: Methodological challenges to the analysis of empirically derived criminal trajectory profiles. In G. R. Hancock & K. M. Sam-

uelsen (Eds.), *Advances in latent variable mixture models* (pp. 53–75). Charlotte, NC: Information Age.

Lanza, S. T., & Collins, L. M. (2006). A mixture model of discontinuous development in heavy drinking from ages 18 to 30: The role of college enrollment. *Journal of Studies on Alcohol, 67,* 552–561.

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association, 90,* 1112–1121.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: Wiley.

Marmorstein, N., & Iacono, W. G. (2005). Longitudinal follow-up of adolescents with late-onset antisocial behavior: A pathological yet overlooked group. *Journal of the American Academy of Child & Adolescent Psychiatry, 44,* 1284–1291.

McArdle, J. J. (2004). Latent growth curve analyses using structural equation modeling techniques. In D. M. Teti (Ed.), *Handbook of research methods in developmental science* (pp. 340–366). Malden, MA: Blackwell.

McCullagh, P., & Nelder, J. A. (1999). *Generalized linear models* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

McLachlan, G., & Peel, D. (2000). *Finite mixture models.* New York: Wiley.

Mehta, P. D., Neale, M. C., & Flay, B. R. (2004). Squeezing interval change from ordinal panel data: Latent growth curves with ordinal outcomes. *Psychological Methods, 9,* 301–333.

Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research, 39,* 479–515.

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345–368). Newbury Park, CA: Sage.

Muthén, B., & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. *Mplus Web Notes, 4.* Retrieved December 10, 2008, from http://www.statmodel.com/download/webnotes/CatMGLong.pdf

Muthén, B., & Asparouhov, T. (2006). Growth mixture analysis: Models with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Advances in longitudinal data analysis* (pp. 143–165). Boca Raton, FL: Chapman & Hall/CRC Press.

Muthén, B., Brown, C. H., Masyn, K., Jo, B., Khoo, S-T., Yang, C.-C., et al. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics, 3,* 459–475.

Muthén, B., & Muthén, L. K. (2006). Mplus (Version 5.0) [Computer software]. Los Angeles: Muthén & Muthén.

Muthén, B., & Muthén, L. K. (2000a). The development of heavy drinking and alcohol-related problems from ages 18 to 37 in a U.S. national sample. *Journal of Studies on Alcohol, 61,* 290–300.

Muthén, B., & Muthén, L. K. (2000b). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24,* 882–891.

Muthén, L. K., & Muthén, B. (2006). *Mplus user's guide* (4th ed.). Los Angeles: Muthén & Muthén.

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

Nagin, D. S. (2005). *Group-based modeling of development.* Cambridge, MA: Harvard University Press.

Nylund, K. L., 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,* 535–569.

Nylund, K. L., & Masyn, K. E. (2008, May). *Covariates and latent class analysis: Results of a simulation study.* Paper presented at the Society for Prevention Research Annual Meeting, San Francisco.

Olsen, M. K., & Schafer, J. L. (2001). A two-part random-effects model for semicontinuous longitudinal data. *Journal of the American Statistical Association, 96,* 730–745.

Orlando, M., Tucker, J. S., Ellickson, P. L., & Klein, D. J. (2005). Concurrent use of alcohol and cigarettes from adolescence to young adulthood: An examination of developmental trajectories and outcomes. *Substance Use & Misuse, 40,* 1051–1069.

Ramaswamy, V., Desarbo, W. S., Reibstein, D. J., & Robinson, W. T. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science, 12,* 103–124.

Raudenbush, S. W. (2005). How do we study "what happens next"? *Annals of the American Academy of Political and Social Science, 602,* 131–144.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2005). HLM (Version 6) [Computer software]. Lincolnwood, IL: Scientific Software International.

Reinecke, J. (2006). Longitudinal analysis of adolescents' deviant and delinquent behavior: Applications of latent class growth curves and growth mixture models. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 2,* 100–112.

Sameroff, A. J. (2000). Developmental systems and psychopathology. *Development and Psychopathology, 12,* 297–312.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177.

Shaw, D. S., Gilliom, M., Ingoldsby, E. M., & Nagin, D. S. (2003). Trajectories leading to school-age conduct problems. *Developmental Psychology, 39,* 189–200.

Shaw, D. S., Lacourse, E., & Nagin, D. S. (2005). Developmental trajectories of conduct problems and hyperactivity from ages 2 to 10. *Journal of Child Psychology and Psychiatry, 46,* 931–942.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis:*
Modeling change and event occurrence. New York: Oxford University Press.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models.* Boca Raton, FL: Chapman & Hall/CRC.

Tooze, J. A., Grunwald, G. K., & Jones, R. H. (2002). Analysis of repeated measures data with clumping at zero. *Statistical Methods in Medical Research, 11,* 341–355.

Tucker, J. S., Orlando, M., & Ellickson, P. L. (2003). Patterns and correlates of binge drinking trajectories from early adolescence to young adulthood. *Health Psychology, 22,* 79–87.

Twenge, J. M., & Nolen-Hoeksema, S. (2002). Age, gender, race, socioeconomic status, and birth cohort difference on the children's depression inventory: A meta-analysis. *Journal of Abnormal Psychology, 111,* 578–588.

Vermunt, J. K. (2006). Growth models for categorical response variables: Standard, latent-class, and hybrid approaches. Retrieved March 20, 2006, from http://spitswww.uvt.nl/~vermunt/

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56–75). Thousand Oaks, CA: Sage.

White, H. R., Johnson, V., & Buyske, S. (2000). Parental modeling and parenting behavior effects on offspring alcohol and cigarette use: A growth curve analysis. *Journal of Substance Abuse, 12,* 287–310.

Wiesner, M., & Silbereisen, R. K. (2003). Trajectories of delinquent behaviour in adolescence and their covariates: Relations with initial and time-averaged factors. *Journal of Adolescence, 26,* 753–771.

Wiesner, M., & Windle, M. (2004). Assessing covariates of adolescent delinquency trajectories: A latent growth mixture modeling approach. *Journal of Youth and Adolescence, 33,* 431–442.

Appendix A

Types of Generalized Linear Growth Models

Name	Acronym	Alternative name (acronym)	Link	Distributional assumptions	
				Level 1	Level 2 +
Hierarchical linear model	HLM	Random effects model	Identity	Normal	Normal
		Random coefficients model			
		Mixed model			
		Multilevel model (MLM)			
Hierarchical generalized linear model	HGLM	Generalized linear mixed model (GLMM)	May be identity, log, logistic or logit, probit, complementary log-log, etc.[a]	Member of exponential family[b]	Normal
Growth mixture model	GMM	Growth mixture analysis (GMA)	May be identity, log, logistic or logit, probit, complementary log-log, etc.	Exponential family	Multinomial[c] & normal
		Finite mixture model			
Latent class growth analysis	LCGA	Semiparametric group-based model (SGBM)		Exponential family	Multinomial
Latent profile analysis			Identity	Normal	Multinomial
Longitudinal latent class analysis[d]	LLCA	Latent class analysis (LCA)	May be identity, logistic or logit, probit, complementary log-log, etc.	Multinomial	

[a] Depends on the distribution of the observed data. [b] Exponential family of distributions includes normal, exponential, gamma, multinomial, binary or Bernoulli, Poisson, zero-inflated Poisson, negative binomial, etc. [c] Multinomial includes both ordered and unordered categorical. [d] Model lacks latent growth factors.

(*Appendixes continue*)

Appendix B

```
TITLE:
    Unconditional hierarchical generalized linear model (HGLM)

DATA:
    FILE is drk5yr.dàt;
```
> Data file should be in folder with the input file.

```
VARIABLE:
    NAMES are famid sex drk1-drk5
            frdrink;

    USEVARIABLES are drk1-drk5;
    CATEGORICAL are drk1-drk5;
```
> NAMES tells Mplus the names of the variables in the data, USEVARIABLES identifies the variables in the analysis, and CATEGORICAL labels the data as binary or ordinal

```
    MISSING are all .;
```
> This identifies the missing value tag (as .)

```
ANALYSIS:
    TYPE = missing;
    STARTS = 100 10;
    ALGORITHM=integration;
```
> "TYPE = missing" tells Mplus to use all available data (despite missingness); no longer needed in version 5. "STARTS" gives Mplus instructions about how many randomly selected starting values to use, and ALGORITHM = integration tells Mplus how to estimate the model.

```
MODEL:
int slope quad | drk1@-2 drk2@-1 drk3@0
     drk4@1 drk5@2;
```
> Variances and covariances between growth factors are estimated by default.

> Intercept is placed at the third timepoint in this example

```
OUTPUT:  TECH1 TECH4 TECH10;
```
> TECH10 gives residuals for the overall analysis.

```
PLOT:
    TYPE is plot3;
    SERIES is  drk1(0) drk2(1) drk3(2) drk4(3) drk5(5) ;
```
> Plot commands generate an estimated mean (logit) trajectory.

```
SAVEDATA:
    file = HGLM_isq_isqv_uncond.txt;
    save = fscores;
```
> SAVEDATA can be used to save various information from the analysis. This command saves model estimates (individual estimated scores on the growth factors; fscores) in a text file.

Figure B1. Syntax Example 1: Unconditional (i.e., no covariates) hierarchical generalized linear model (HGLM).

```
TITLE: Conditional growth mixture model (GMM) or latent class growth analysis
       (LCGA)

DATA:
    FILE is drk5yr.dat;

VARIABLE:
    NAMES are famid sex drk1-drk5 frdrink;
    USEVARIABLES are drk1-drk5 sex frdrink ;
    CATEGORICAL are drk1-drk5;
```

CLASSES defines the name (c) and number of classes in the mixture model

```
CLASSES = c(2);
```

MISSING identifies the missing value indicator in the data. ID identifies the person identifier and with this command, Mplus will include the variable in output files of individual estimates.

```
MISSING are all .;
ID = famid;

ANALYSIS:
    TYPE = mixture missing ;
    STARTS = 500 70;
    ALGORITHM = integration;
```

"TYPE = mixture" is required for mixture models.

```
MODEL:
%OVERALL%
int slope quad | drk1@0 drk2@1 drk3@2 drk4@3 drk5@5;
```

Intercept is placed at the first timepoint in this example

```
int*;
slope@0;
quad@0;
```

%OVERALL% is a necessary command for a mixture model. The next line is typical Mplus notation for a random effects model of any type. This shows fixed times of measurement, but they may be unique to each individual. int* frees the intercept variance and the linear and quadratic variances are fixed to zero (@0). If all three variances are fixed at zero, the analysis is LCGA.

"int on…" regresses the intercept on the covariates. "c#1 on…" regresses the first class (as compared with the second class) on the covariates

```
int on sex pdrink;
c#1 on sex pdrink;

OUTPUT:  TECH1 TECH4 TECH7 TECH10;

PLOT:
    TYPE is plot3;
    SERIES is drk1(0) drk2(1) drk3(2) drk4(3) drk5(5);

SAVEDATA:
    file = GMM2_isq_iv_pred.txt;
    save = cprob fscores;
```

These last commands save model results in a text file: individual scores on the growth factors (fscores) and individual class posterior probabilities (cprob).

Figure B2. Syntax Example 2: Conditional (i.e., with covariates) growth mixture model (GMM) or latent class growth analysis (LCGA).

(*Appendixes continue*)

```
TITLE:
    Unconditional Longitudinal Latent Class Analysis (LLCA)

DATA: file is drk5yr.dat;

VARIABLE:
    names are famid sex drk1-drk5 frdrink;

    usevariables are drk1-drk5 ;
    categorical are drk1-drk5;

    classes = c(2);
    missing are all .;

ANALYSIS: type = mixture missing ;
    starts = 100 10;
    lrtstarts = 100 10 100 10;
```

> In addition to the random starts (STARTS), this syntax tells Mplus how many random starts to use for the bootstrap LRT (BLRT) test (LRTSTARTS).

```
MODEL:
%overall%
```

> The difference from GMM/LCGA is that no linear model is specified after the %OVERALL% statement

```
OUTPUT:   TECH1 TECH4 TECH7 TECH10 TECH11 TECH14 residual;
```

> TECH11 and TECH14 give the Lo-Mendel-Rubin test (LMR) and Bootstrapped Likelihood Ratio Test (BLRT)

```
PLOT:
    type is plot3;
    series is  drk1(0) drk2(1) drk3(2) drk4(3) drk5(5) ;

SAVEDATA:
    file = LLCA2_uncond.txt;
    save = cprob;
```

> Only the posterior class probabilities are saved (there are no estimated intercepts or slopes in LLCA).

Figure B3.   Syntax Example 3: Longitudinal latent class analysis (LLCA) without covariates.