Published in final edited form as:

Struct Equ Modeling. 2012 ; 19(2): 227–249. doi:10.1080/10705511.2012.659622.

Not quite normal: Consequences of violating the assumption of normality in regression mixture models

M. Lee Van Horn, Jessalyn Smith, and Abigail A. Fagan

University of South Carolina

Thomas Jaki

Lancaster University

Daniel J. Feaster

University of Miami

Katherine Masyn

Harvard University

J. David Hawkins

University of Washington

George Howe

George Washington University

Abstract

Regression mixture models are a new approach for finding differential effects which have only recently begun to be used in applied research. This approach comes at the cost of the assumption that error terms are normally distributed within classes. The current study uses Monte Carlo simulations to explore the effects of relatively minor violations of this assumption, the use of an ordered polytomous outcome is then examined as an alternative which makes somewhat weaker assumptions, and finally both approaches are demonstrated with an applied example looking at differences in the effects of family management on the highly skewed outcome of drug use. Results show that violating the assumption of normal errors results in systematic bias in both latent class enumeration and parameter estimates. Additional classes which reflect violations of distributional assumptions are found. Under some conditions it is possible to come to conclusions that are consistent with the effects in the population, but when errors are skewed in both classes the results typically no longer reflect even the pattern of effects in the population. The polytomous regression model performs better under all scenarios examined and comes to reasonable results with the highly skewed outcome in the applied example. We recommend that careful evaluation of model sensitivity to distributional assumptions be the norm when conducting regression mixture models.

Regression mixture models are a potentially powerful approach to understand differential effects, differences across individuals in the relationship between a predictor(s) and an outcome(s). A better understanding of differential effects is important for understanding the complexities of human behavior and can inform our ability to prevent the development of problem behaviors. While theories of individual behavior have tended to focus on direct relationships between independent and dependent variables, they increasingly recognize the need for more nuanced and detailed explanations of the linkages between cause and effect. Given that individuals are likely to react differently to circumstances, theoretical

perspectives that can explain these different effects can offer more realistic and specific explanations of human behavior. In the social sciences, developmental theories tend to take this perspective. They acknowledge that individuals are embedded in multiple contexts—communities, schools, family, and peer groups—and emphasize that to fully understand human behavior, one must understand how these contexts influence one another (Bronfenbrenner, 2005; Elder, 1998; Patterson, DeBaryshe, & Ramsey, 1989; Sampson & Laub, 1993). Similarly, modern biological theories seek to understand the environments in which individual or biological factors are most likely to manifest themselves (Caspi et al., 2003; Rowe & Osgood, 1984). Rather than assuming that a biological trait has a direct impact on behavior for all individuals or under all conditions, biological theories are increasingly concerned with differential effectiveness such as assessing gene-by-environment interactions (Bechtel & Churchman, 2002).

Theories that better specify individual differences in the causes of human behavior offer the potential for preventing the development of problem behaviors. Effective prevention programs and practices focus on changing the risk factors that are related to the targeted outcome (Coie et al., 1993; Kellam, Koretz, & Moscicki, 1999). To date, most have done so by assuming that influences are uniform across individuals and circumstances; they do not typically incorporate differential effects into program designs. However, as the importance of establishing the generalizability of prevention strategies increases there are more efforts to identify the individuals for whom programs are most effective (Flay et al., 2005; Kellam et al., 1999) and even more importantly, to find those likely to experience iatrogenic effects (Dishion, McCord, & Poulin, 1999; Ellickson & Bell, 1990; Sloboda et al., 2009). A better theoretical understanding of differential effectiveness can directly inform the development of prevention programs, increase their potential for achieving positive changes in behaviors, and reduce the likelihood of iatrogenic effects on individuals (Kumpfer, Smith, & Franklin Summerhays, 2008; Spoth, Shin, Guyll, Redmond, & Azevedo, 2006).

While understanding differential effects in substantive theories and in designing interventions is important, there are remarkably few studies that test for differential effects beyond gender, age, and ethnicity. One reason for this mismatch between theory and empirical investigation are the limitations in existing methods for finding differential effects which include the necessity of an a priori identification of moderators and limited power to test for multiple interaction terms (Boyce et al., 1998). An alternative approach for examining differential effects is the use of regression mixture models, a type of finite mixture model in which the effects of a predictor variable are allowed to differ between classes and in which these differences determine class membership. However, as with other mixture models (e.g. Bauer & Curran, 2004), regression mixtures come at the cost of a strong within-class distributional assumption. The aims of this paper are to examine the impacts of violating this assumption on model results, to illustrate an alternative model specification which relaxes this assumption, and to demonstrate both approaches in an applied analysis testing differences in the effects of family risk factors on substance use.

Methodological overview

The use of finite mixture models in the social sciences has increased greatly in the last 15 years. In many areas they are now standard practice, such as in the longitudinal analysis of the development of aggression in the form of general growth mixture models (Muthén, 2006; Muthén, Collins, & Sayer, 2001; Muthén et al., 2002) or semi-parametric models (Nagin, 2005; Nagin, Farrington, & Moffitt, 1995). The strength of these models is their potential to identify groups or classes of individuals who behave similarly. The ability to identify qualitatively different groups empirically is intuitively easy to grasp and fits with many theories of human development. Typically this distinction is made based on

individuals within groups having similar patterns on a set of response variables that differ between groups, such as different kinds of drug use (Van Horn et al., 2008) or, in the case of longitudinal data, intercepts and slopes describing changes over time (Schaeffer, Petras, Ialongo, Poduska, & Kellam, 2003).

Regression mixture models (Desarbo, Jedidi, & Sinha, 2001; Van Horn et al., 2009; Wedel & Desarbo, 1994) are an application of finite mixtures which are still quite new to the social sciences. They differ from the typical finite mixture model in that groups of individuals are identified based in part on having a similar pattern of relationships between two or more variables, rather than being identified based on a common pattern of response to a single variable. For example, in a general growth mixture model, groups are characterized by having different means and possibly different variances on the parameters of a growth curve model (i.e., intercepts and slopes). In a regression mixture model groups are characterized by differences in the effects of a predictor variable on outcomes (such as the slopes and intercepts of a growth curve). The important distinction is that regression mixtures allow differences in *effects* between groups, whereas more traditional mixture models allow differences in *levels* and *variance* of the outcome between groups. Thus, regression mixture models are related to the common definition of moderation, that the effects of a predictor vary across level of a moderating variable (Baron & Kenny, 1986), although this perspective modifies that definition: moderation is the presence of differences in the population in the effects of the predictor on the outcome (Van Horn et al., 2009). These models can have quite powerful implications since they allow for the identification of differential effects without the need for a moderating variable which opens the door to identifying moderated effects empirically without pre-specification of the traits or characteristics that are associated with the moderating process.

Regression mixture models

This section includes a brief overview of regression mixture models and a detailed description of regression mixture models for categorical response variables, which is one focus of this paper.¹ The objective of regression mixture models is to estimate the unknown population distribution of either univariate or multivariate outcomes via latent classes that are empirically estimated and thereby model the heterogeneity in the response. While the traditional finite mixture model investigates only the outcome variable, regression mixture models allow classes to differ in the effects of predictor variables on the outcome and hence allow the investigation of classes representing differential effects.

For the general form of a multivariate regression mixture model we consider a sample of n individuals for which m outcome variables, $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)$ and P observed covariates, $\mathbf{X} = (X_1, X_2, \dots, X_P)$ are measured. The observed values of Y_m and X_p for subject i are denoted y_{im} and x_{ip} respectively. The multivariate conditional probability density function of $\mathbf{Y}$ given $\mathbf{X}$, $f(\mathbf{y} | \boldsymbol{\phi}, \mathbf{X})$, where $\boldsymbol{\phi}$ is the parameter vector, is modeled as a weighted sum of a finite number of probability densities, $f_k(\mathbf{y} | \boldsymbol{\theta}_k, \mathbf{X})$, corresponding to the distributions for K classes with $\boldsymbol{\theta}_k$ denoting the class distribution specific parameters. Class membership is represented by a latent categorical variable, C , where $C = 1, 2, \dots, K$, and most models require K to be specified a priori. The contribution of each class to the overall density is denoted by $\pi_1, \pi_2, \dots, \pi_K$ which are treated as model parameters and consequently are estimated as well. Using this construction we can write the joint distribution of $\mathbf{Y}|\mathbf{X}$ as

¹See Van Horn et al. (2009) for a complete description of the finite mixture models and regression mixture models for Gaussian response variables and a behavioral science application.

$$f(\mathbf{y}|\varphi, \mathbf{X}) = \sum_{k=1}^K \pi_k f_k(\mathbf{y}_k|\theta_k, \mathbf{X}), \quad (1)$$

where $\varphi = (\boldsymbol{\pi}, \boldsymbol{\Theta})$ denotes the vector of all unknown parameters to be estimated; $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_K)$; and $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_K)$. So far we have not made any explicit assumptions about the distribution of $\mathbf{Y}|\mathbf{X}$ within group. In the following section, we briefly describe the most commonly used assumption of a multivariate normal distribution.

Normal outcomes

Assuming that the conditional distribution of $\mathbf{Y}|\mathbf{X}$ within class k is multivariate normal, we have

$$Y_i|\mathbf{X}_i, k = \beta_{0,k} + \sum_{p=1}^P \beta_{pk} \mathbf{X}_{ip} + \varepsilon_{ik}, \varepsilon_{ik} \sim \mathcal{N}(0, \Sigma_k), \quad (2)$$

Where $\boldsymbol{\beta}_{0k}$ is the vector of intercepts (which simplify to means when there are no covariates), Σ_k is the residual variance/covariance matrix for class k , and $\boldsymbol{\beta}_{pk}$ is the vector of regression coefficients for X_p in latent class k ; these class specific coefficients identify this as a regression mixture model.

Categorical outcomes

An alternative approach that does not assume normal errors is to recode the outcome to be an ordered categorical variable. For simplicity of notation we will only consider the case of a single categorical outcome variable, Y , that has S different levels. This outcome variable could be assumed to be an ordered response at the cost of an additional assumption, such as the proportional odds assumption that we use in this paper. This assumption specifies that one $\boldsymbol{\beta}$ is sufficient for describing the relationship of each covariate on the outcome across all levels of s (Kutner, Nachtsheim, Neter, & Li, 2005; Whitehead & Jaki, 2009). In this situation we can take $f_k(\mathbf{y} | \boldsymbol{\theta}_k, \mathbf{X})$ to be the pdf of a multinomial distribution and model the probability of level s in class k , θ_s^k as

$$\theta_s^k = P(Y_i = s|k, \mathbf{X}) = \frac{e^{\beta_{0s}^k + \sum_{p=1}^P \beta_{ps}^k \mathbf{X}_{ip}}}{\sum_{u=1}^S e^{\beta_{0u}^k + \sum_{p=1}^P \beta_{pu}^k \mathbf{X}_{ip}}} \quad s=1, \dots, S \quad (3)$$

or alternatively in the linear model form as

$$\text{logit}(\theta_s^k) = \text{logit}(P(Y_i = s|k, \mathbf{X})) = \beta_{0s}^k + \sum_{p=1}^P \beta_{ps}^k \mathbf{X}_{ip} \quad s=1, \dots, S \quad (4)$$

Note that model identification requires constraints on the parameters. Here we use the last class as a reference group and estimate the thresholds β_{0s}^k for classes 1 through $S-1$.² No specific distributional assumption is made; however, the model does assume that θ_s^k is independent conditional on $\mathbf{X}$.

²A thorough discussion of the identifiability constraints for the multinomial mixture model is given in (Grün&Leisch, 2008).

Model Estimation

The estimation of the parameters is usually achieved using the EM-algorithm (Dempster, Laird, & Rubin, 1977) and is described in detail in McLachlan & Peel (2000) for normal outcomes, and in Wedel and Desarbo (1995) for generalized mixture models such as the multinomial mixture model described above.

Distributional Assumption

One characteristic of finite mixture models is that the assumed within class distributions and model parameters are dependent on each other (Bauer & Curran, 2003, 2004). The shape of any continuous distribution can be well approximated by mixing a finite number of separate distributions that each take on a specific shape. If each of those separate distributions is assumed to be normal with its own mean and variance, then given a specific number of distributions, the class specific means and variances can be estimated such that they best approximate the shape of the unconditional observed distribution. In finite mixtures each of these distributions is a separate 'latent class,' and the class specific mean and variance as well as the proportion of individuals in each class (called mixing proportions) are the model parameters. Parameter estimates in finite mixtures are closely tied to the assumed within class distribution. This approach differs from many other statistical models (such as OLS regression), in that for most models, distributional assumptions affect inference but do not greatly impact parameter estimates (Cohen, Cohen, West, & Aiken, 2003).

For users of these models it is thus important to understand what factors impact the unconditional observed distribution of the outcome, as these factors will also influence the model results. Mixture models, assuming normal errors, attribute non-normality in the outcome to latent classes and assume normality for the within class distribution of the outcome conditional on covariates. When there are no covariates in the model this simplifies to the assumption that the distributions of the outcome components are normal within class. It is instructive to examine this assumption in the case of regression mixtures which include class specific covariates. The within class distribution of the outcome (from equation 2) is a function of β_0 , $\beta_p X_p$, and e_i . The intercept, β_0 impacts the center, not the shape of the distribution. The factors which do influence the within class distribution are therefore $\beta_p X_p$ and e_i . It is the distribution of e_i which impacts model results as the effect of $\beta_p X_p$ is derived from the relationship between X and Y and is already incorporated into the model. The size of beta determines the strength of the influence of the distribution of X on the distribution of Y.

From a practical perspective this means that evaluating the results of regression mixture models requires attention to the factors likely to impact the within class errors. One factor is the covariates not included in the model which are related to the outcome and do not themselves conform to the assumed distribution of the error. A second factor is anything that causes the random error to deviate from the assumed distribution. For instance, floor or ceiling effects will impact the distribution of random responses. Another example common in the social sciences is having outcomes which are a combination of ordinal items. In this case many features of the response options can impact the distribution of 'random errors,' such as the tendency of some respondents to choose particular responses to all items. A final point is that because class membership is unknown the unconditional distribution within each class is unobservable and direct tests of model assumptions such as those used in OLS regression are impossible.

Research Questions

The impact of distributional assumptions in regression mixture models has not yet been explored. Because regression mixtures involve estimating a class specific slope for the relationship between two variables, rather than just a mean and variance, there is reason to believe that the impacts of violating the assumed error distribution would be greater than in traditional finite mixtures. Our first research question assesses the impact of non-normal error terms in regression mixture models with a continuous outcome where the errors are assumed to be normal; the simulated data is drawn from two populations that differ in the effects of a covariate on the outcome and in the intercept and residual variance of the outcome. We hypothesize that increasing skewness in one class will result in the selection of additional classes and bias in parameter estimates. Because additional classes can capture non-normality, we hypothesize that when only one class is skewed, the class with a non-normal error term will be represented by two classes in the final solution and that those two classes will be characterized by parallel regression lines and different intercepts. We also hypothesize that having a skewed error term in both classes will result in more dramatic impacts on both model selection and parameter estimates with additional classes being selected and the model not reliably identifying the differential effects in the population.

The second research question examines an alternative model parameterization in which the outcome is an ordered categorical variable. The same data used in the first aim will be used for this aim, but here the continuous outcome variable will be categorized, making it an ordinal variable. We hypothesize that the 'true' two class model will be selected over a one or three class model, and that the classes will represent differences in the effects of the covariate on the outcome similar to those in each of the populations from which the data was generated.

The third aim is to apply this technique to an applied dataset in which the error term within classes is expected to be non-normal. To this end, regression mixtures are used to examine the relationship between parental management and drug use in a sample of 10th grade students. Using the ordinal outcome we expect to find at least two classes which differ in the effects of parenting on drug use, and with the continuous outcome we expect that the latent classes will be driven by non-normality in the outcome, thus failing to provide clear evidence of differential effects.

Methods

This study utilized Monte Carlo simulations for the first two aims and data from a community-based survey of 10th grade students for aim three, which is described in detail later.

Monte Carlo Simulations

Data were generated in R (R Development Core Team, 2008), and the models fit using Mplus version 5.21 (Muthén & Muthén, 2008). The main interface used was R from which Mplus was called in batch mode. Five hundred simulations were performed for each scenario reported. An exception is the Bootstrapped Likelihood Ratio Test (BLRT; McLachlan & Peel, 2000) which for the categorical outcomes, was carried out only for 100 of the 500 simulations due to high computational costs. Unless otherwise indicated, analyses were conducted using a robust maximum likelihood estimator.

The challenge in Monte Carlo simulations is to match a focused research question to a manageable set of conditions under which data will be simulated. In all analyses we selected a simple two class model with 3000 respondents in each class for differential effects, that is,

a model where two distinct groups are present. In each case one predictor, X , and one outcome variable, Y , was used. In the population X had a normal distribution with a mean of 0 and a standard deviation of 1. In the first class the outcome, Y , was created using the equation: $Y = .50 + .70X + e$, because the sd of X was 1 and e was scaled so that the sd of Y would be 1; this resulted in a correlation of .70 between X and Y in this class. In the second class Y was created with the equation: $Y = .20X + e$, the sd of e was again scaled so that the total sd of Y was 1, thus, the correlation between X and Y was .20. We choose these population values because they represent the types of differential effects we expect to find in real data, which should also be detectable with regression mixtures. Of note, the residual variance was less in the group with the stronger effect of the covariate because the presence of a stronger predictor should reduce error variance. The mean parameter estimates are reported across all simulations; the standard errors are estimated standard deviations of the parameter estimates across simulation runs.

This simple condition allowed us to test the effects of different distributions for the residual variances using five different scenarios, as follows: residual variables were normally distributed in both classes, residuals in class 1 were normally distributed while in class 2 they had a positive skew of either .15, .35, or .49, and in the last condition residuals in both classes had a skew of .49. Non-normality was created by adding a constant to the error to eliminate negative values, and using a power transformation (either a power of 1, 1.3, 1.7, or 2.0 were used in different simulations). The resulting error was then rescaled so that the mean was 0 and the sd of Y was constant across different conditions. This transformation also impacts higher order moments including kurtosis, but the primary impact was on skewness. While there are many other ways that the distributional assumptions can be violated, this power transformation is one of the simpler ways and provides a good starting point for understanding how these models perform. The values for the power transformations were chosen as they represent minor violations of distributional assumptions, consistent with violations that would be difficult to detect in real data situations.

Three additional simulations are conducted after the outcome variable was categorized into six levels; one where there is no skew in either class, one with a skew of .49 in only one class, and one with a skew of .49 in both classes. The categorization was done by splitting the distance from the 5th to the 95th percentiles into six equally spaced bins and placing each subject into their respective bin, with the low and high ends including all values below the 5th or above the 95th percentiles.

Figure 2 includes contour plots of the joint distribution when the skew is 0 in class 1 and .49 in class 2, to demonstrate the effect of the transformations used on the observed, unconditional, distribution of X and Y across both classes. When skew increases, the joint density has an elliptical shape with a more pronounced ridge in the lower right quadrant. In both figures, the ellipses indicate that there is an overlap of normal distributions with differing means. These contour plots are useful because they contain all the information available to estimate model results.

Initial results focus on latent class enumeration, including the proportion of simulations for which each class was chosen according to each criterion. Among other measures, the percentage of model non-convergence is reported for categorical outcomes. Convergence was not a problem for continuous outcomes. Non-convergence is defined as solutions where the log-likelihood value was not replicated in 100 starts or where no solution was obtained. It should be noted that when the true number of classes was specified, all models converged.

One of the problems when using simulations with finite mixtures is label switching (McLachlan & Peel, 2000; Sperrin, Jaki, & Wit, 2010), in which the class labeled 1 in some simulations receives the label 2 in other results. In order to report the model parameters for each class, it is therefore necessary to sort the results so that the first and second class are always substantively the same. Classes were sorted such that the class with the stronger effect of X on Y was always class 1. This creates a potential bias if, across simulations, the distributions of classes with the high slope and low slope overlap. While alternative strategies for relabeling exist, this approach was tested and found sufficient in these analyses due to large differences in the regression weights between classes.

Applied Data Example

Data for the applied analyses come from the Diffusion Project, a naturalistic study of the ways in which states and local communities plan, implement, and monitor drug prevention programs. Data were collected in 2002 from 40 communities in seven states that were participating in the study administered the Communities That Care Youth Survey (CTC; Arthur, Hawkins, Pollard, Catalano, & Baglioni, 2002). The sample included all 10th grade students in public schools in these communities.

Surveys were administered during one classroom period and ensured the anonymity and confidentiality of students' responses. Screening criteria were used to guard against dishonest or biased answers, which included student reports of honesty in completing the survey, the inclusion of a fictitious drug, and examination of inconsistencies in patterns of reported substance use and delinquency. Using these criteria, approximately 5% of the students were excluded from the analyses. Based on school enrollment figures, the average participation rate in the survey was 66% across communities. The sample included 7755 10th grade students (mean age of 15.7 years); 52% were female and 48% were male. About 78% of the students described themselves as Caucasian, 8.0% as Hispanic, 4% as African American, and 10% as from another ethnic background.

The CTC Youth Survey measures risk and protective factors and self-reported rates of substance use and delinquency using scales that have good evidence for reliability, validity, and invariance across ethnicity and sex (Arthur et al., 2002; Glaser, Van Horn, Arthur, Hawkins, & Catalano, 2005). In the current study, regression mixtures assess the relationship between family management and substance use, which are both self-reported by students on the CTC Youth Survey. Eight items measured poor family management. Three items measured the perceived likelihood of students being caught by parents for engaging in delinquent behavior (e.g., *skipping school*, *carrying a handgun*, or drinking beer, wine or liquor without parents' permission). Five items measured parental monitoring skills and expectations: Would your parents know if you did not come home on time? When I am not at home, one of my parents knows where I am and who I am with. The rules in my family are clear. My family has clear rules about alcohol and drug use. My parents ask if I've gotten my homework done. Respondents indicated agreement with each statement using a 4-point Likert scale in which higher scores indicated worse family management. A latent variable with the eight survey items as indicators was then derived. Students reported their lifetime use of five substances: smokeless tobacco, cigarettes, alcohol (beer, wine, or hard liquor), inhalants, and marijuana. Frequency of use was assessed on a 5-point scale for smokeless tobacco and cigarettes (from never to regular use) and a 7-point scale for the other drugs (from zero to 40 or more occasions). Because we wanted a highly skewed but continuous outcome to demonstrate the two approaches to using regression mixtures with this data, we created factor scores for substance use using the parameter estimates from a CFA model using a mean- and variance-adjusted weighted least squares estimator (Flora & Curran, 2004) where the items were treated as ordinal indicators. The resulting variable was highly skewed, with a high proportion of responses at zero indicating no substance use.

Results

How sensitive are regression mixture models to non-normal errors?

The first research question was assessed with simulations with varying degree of skew in the residuals. Table 1 presents the latent class enumeration across the different conditions; values for the AIC, BIC, adjusted BIC, and BLRT are the proportion of simulations choosing k over $k-1$ classes. With no skew in the data, most criteria select the correct 2 class model. To be sure that 4 classes were not being selected over 2 classes, we also compared 2 and 4 classes for each criterion and found that for the zero skew case, the 2 class solution was chosen over the 4 class solution in all cases according to the BIC and adjusted BIC. The AIC, however, chose the 3 class model even when skew was zero, demonstrating that it tends to select too many classes when used in these models. As this result has been noted elsewhere (Nylund, Asparuhov, & Muthen, 2007), we will not further discuss the AIC in this paper. As skew increased to .49 in one class, the probability of choosing 3 over 2 classes increased for all criteria, and approached unity in all except the BIC. Thus, even this moderate amount of skew results in additional classes being identified. When skew reached .49 in the second class, the 4 class model was often selected over the 3 class model. When skewed error was present in both classes, all criteria choose 3 classes and most choose 4. Note that there were no convergence problems with these models and the size of the smallest class was quite small when skew was very small, but increased enough to be considered a viable group when skew was .49. Regarding entropy, the values reported in Table 1 are poor according to the standards of most mixture models. Figure 2 illustrates why is the case. Entropy measures how well individuals are classified, and because the distributions of the two classes overlap almost completely, it was expected that classification of any one individual would be very poor. However, this does not negate the value of these models for finding the population parameters for these classes. Rather it illustrates that the models should not be used to classify individuals in the absence of covariates which can better predict class membership.

Average model parameter estimates for the two class models are reported in Table 2. Because the two intermediate skew conditions show the same trends, we focus only on the skew of 0 and skew of .49 conditions. Results showed that if the correct two class solution were accepted, the model parameters were very good when there is no skew; model parameters were reasonable but included some bias in the intercepts when the error was moderately skewed in class two. One interesting result is that while estimates of the slopes were still fairly reasonable when there was skew in both classes (at least the same substantive conclusion would be drawn), the intercepts were quite biased in this case. While the intercepts should be .50 in the class with the strongest effect and 0 in the other class, in some cases the direction was even reversed such that the intercepts were greater in the class with the lower slope. Additionally the residual variances were poorly estimated in these models.

It was hypothesized that when the error of class 2 was skewed, 3 classes would be selected and two of those classes would jointly make up the skewed class, having roughly parallel regression weights and differing in their intercepts. The parameter estimates for the 3 classes in the case in which the population model had a skew of .49 for the error of one class are presented in Table 3. As hypothesized, one class contained close to 50% of the sample, while the other two classes were smaller, had slopes within 1 standard error of the slope of class 2 in the population (.70), and had intercepts that differed and averaged close to the population intercept of .50 in class 2. It appears that the class with the skewed error in the population is being recovered as two classes in these analyses. In order to verify this, we included a variable indicating whether each individual was in class 1 or class 2 when the data were generated. We then used draws from the posterior probabilities of class

membership to calculate differences between groups in the means of a the true latent class, and found that those in the two classes with parallel slopes had a mean for class two of 57%, versus a mean of 43% for being in the class with a lower slope. This difference was significant and in the correct direction, it is however far from perfect. The reason for this is that the means are estimated using posterior probabilities which are poorly estimated in these models as indicated by the low entropy values.

Finally, the 3 and 4 class solutions were examined for the model in which the population had moderate skew in both classes. Results in Table 3 indicate that the three class solution did not find two classes with similar slopes; these results seem unrelated to the population model under which the data were generated. It is reasonable to expect that the four class model in this case would work well, with two classes having roughly parallel slopes around .20, and two having slopes around .70. However, in this set of simulations, no consistent results were obtained across simulated datasets; a visual inspection of results showed that only a few simulations (approximately 5%) supported this hypothesis. A summary across simulations cannot be reported because there is no reliable pattern with which to sort classes to be comparable across simulations.

Model performance when the outcome is transformed to an ordered categorical variable

The same data were used for this aim as used in the first aim, but now the Y values were recoded into integers 1 through 6. The model assumes that Y is an ordered outcome and that the relationship between X and Y is proportional across levels. Results shown in Table 5 demonstrate that in contrast to the results where Y was a continuous variable with errors assumed to be normally distributed, when Y is an ordered categorical variable, the true 2 class model is selected over the 1 class model and the 3 class model by the BIC and adjusted BIC nearly all of the time. The BLRT appears problematic in this table as it selects the 3 class model when there is either no skew or when both classes are skewed. Upon further inspection, we found that for the 3 class solution across all simulations the best log likelihood value was not replicated 65% of the time, an outcome that did not change when the number of bootstrap draws was increased. Given that most of the bootstrap draws did not converge to replicable solutions, these results should be interpreted with great caution; the use of the BLRT with categorical outcomes needs further investigation. We choose the 2 class solution based on the BIC and adjusted BIC. Because the 3 class model was not selected, the 4 class model was only conducted for the case where there was skew in both classes and a theoretical rational for 4 classes. For the simulations with skew in both classes, the 2 class model was selected over the 4 class model 98% of the time by the BIC and 100% of the time by the adjusted BIC criteria. The BLRT was not conducted for the 4 class model due to computational demand (based on initial runs we estimated it would take about 30 days to generate the results on a 24 core Xenon server).

Given that the 2 class model is selected, the next question concerns the model parameter estimates. Because the model is now parameterized quite differently, with 5 thresholds representing the log odds of the probability of being at the next highest level of Y conditional on X, the results cannot be directly compared to the parameters used to generate the data. Figure 3 shows the results for one model in which the Y-axis is the predicted response category for an individual for each given level of X and the relationship between X and Y is a step function where there are certain thresholds for X above which individuals have a greater than 50% probability of responding at the next higher level of Y. Population parameters (see Table 5) were estimated by regressing Y on X for data from all simulations combined using a proportional odds model separately within each class; thus the population values are estimated using 1,500,000 subjects for each class. Although these results look quite different than those for when Y is continuous, the same substantive conclusions are reached: classes differ in their overall levels of Y as well in the effect of X on Y, the

regression weight of .20 in the continuous model translated into a regression weight of .34 to .35 in the categorical model, the .70 regression weight translated into a weight of 1.73 to 1.88 for the categorical model. The reason why this differed across levels of skew is that the cut points for the ordinal variable were calculated based on the observed values of the outcome without considering class membership (since this is unknown before model estimation). Changing the shape of the distribution results in changes in the cut points and slightly impact all parameter estimates. Although comparing the parameters of the ordinal logistic regression model to the parameters when the outcome is continuous is not straightforward, the substantive conclusion from this model is correct in that there are two classes of respondents with relatively large differences in the effects of X on Y. In Table 5 it can be seen that the standard errors are fairly large in some cases and there are some biased parameter estimates, this is especially true for the extreme thresholds, thus, this method has limitations. Additionally, while the classes are substantively close to the true model, the proportion of subjects in each class is also off, especially in the case of skewness in just one of the two classes where the proportions were 15% off of the true population model. Note that the choice for the number of categories for Y was ad-hoc in this case, something which may impact the results. Nevertheless, using this approach is more likely to lead to the selection of the correct number of classes and to the correct interpretation of those classes than the approach in which Y is treated as continuous; the use of the ordered logistic regression model has potential for latent class enumeration.

Using regression mixtures to examine differential effects of parenting on substance use

In this final aim both the continuous and ordinal outcome approaches were used to examine differential effects of parenting on substance use, a variable known to have a non-normal distribution and for which we expect the errors will be non-normally distributed. For continuous outcomes, when estimating regression mixture models using the standard maximum likelihood estimator which utilizes the expectation of the matrix of second derivatives to estimate the asymptotic covariance matrix, only the 1 class model converged. Using an (MLF) Maximum Likelihood estimator with standard errors approximated using the first order approximation of the asymptotic covariance matrix (Kaplan, 1999), the 1 through 3 class models did converge, but the 4 class model failed to converge. All criteria selected a 2 class over the 1 class model and the 3 class over both the 1 and 2 class models. The MLF two class estimates included one class with 40% of the sample, in which the effects of family management was .12 (se = .01), the intercept was $-.62$ (se = .01); the second class with 60% of the sample had an effect of family management of .37 (se = .02) with an intercept of .33 (se = .02). The three class model included a class with 30% of the sample with an effect of family management of 0 (se = .00) and an intercept of $-.76$ (se = .01), a second class with 36% of the sample had an effect of family management of .25 (se = .02) and an intercept of $-.02$ (se = .04), the last class with 34% of the sample had an effect of family management of .30 (se = .02) and an intercept of .59 (se = .11). In both models the classes show differences in the effect of family management, although, the results are not very stable across estimators or between the 2 and 3 class solutions, and the size of the differences in intercepts between the classes is quite large at about 1 standard deviation. In this case the regression lines representing differential effects do not come close to crossing within the observed data, making it apparent that the difference in intercepts is driving the results more so than differential effects.

When the substance use outcome was converted into a 6-level ordinal variable, the BLRT suggested the presence of 4 classes, but the BIC and adjusted BIC select a 3 class model (see Table 7). There were also convergence issues with the 3 class model, only 7 of 1000 starts successfully converged and in no case was the best log likelihood value replicated; the smallest class contained 5% of the respondents and parameter estimates in the 5% class were

also inadmissible. In simulation results this has been typical when too many classes are extracted from the data. It has been suggested that fewer than 10% of respondents in a class is too low to be reliably identified via regression mixtures (Leisch & Gruen, 2010); thus in this case we interpret the 2 class solution. The results for the 2 class ordinal outcomes model (see Table 8 and Figure 4) showed a large class where there was a moderate effect of family management on the outcome, and a smaller class (about a quarter of the sample) distinguished by what seems to be a threshold effect for poor family management below which students report very low levels of substance use and above which they report very high levels. Figure 4 indicates that in class 2 there is a non-linear relationship between family management and drug use.

Discussion

Regression mixture models are a potentially powerful method for assessing differential effects. Results show that, when model assumptions are met, regression mixture models can reliably find differential effects given data only on a predictor and an outcome. However, our study indicates that, as with other mixture models, they are quite sensitive to distributional assumptions. One dramatic result of this study is the finding that small departures from model assumptions can have substantial impacts on results in terms of the number of classes identified, and to a somewhat lesser extent, on model parameter estimates. While in some cases investigators will know a priori that model assumptions are likely to be violated (with substance use for example), in these simulations we deliberately chose mild violations of model assumptions which would be quite difficult to predict or detect in applied data. These findings indicate that users of regression mixture models should investigate the possibility that results are driven by violations of model assumptions.

The second important finding is that recoding the outcome to an ordinal variable and using a proportional odds model showed promise for detecting differential effects in the presence of relatively small violations of model assumptions. These models worked very well in finding the correct number of latent classes. They also performed reasonably well in estimating most model parameters, especially differences in the slopes of the predictor on the outcome which is the point of regression mixtures. The models estimation of the proportion of individuals in each class was, however, not great, and the more extreme thresholds tended to have some bias in the parameter estimates. There are some obvious costs to using the proportional odds model: many additional parameters need to be estimated; information in the outcome variable is lost; model results do not directly correspond to the original scale of the outcome variable; and given the larger number of parameters estimated we expect that these models require larger sample sizes. The fact that the proportional odds model did an admirable job at model selection and the continuous outcome model did a reasonable job at recovering differential effects when the correct number of classes was specified suggests an interesting alternative; analyses could be run both ways, with the proportional odds model used for latent class enumeration, and the continuous outcome model used to describe the differential effects that emerge. While this seems to have some potential, we do not believe that there is enough information at this point to broadly recommend this approach. Future research should examine these models under additional conditions to determine whether this generalizes. This research should clearly include examination of variations in both the degree and the complexity of the difference in the conditional expectation across classes, and also different types of and more extreme violations of the distributional assumptions.

The applied example in this paper was selected to increase the violation of normal errors further than in the simulations using an outcome which is clearly highly skewed, allowing for the identification of additional problems that might happen when these models are used in other situations (e.g., under 'real world' conditions). The results showed that the

proportional odds model was able to identify two groups for whom the effect of family management differed. In the smaller group there was a strong effect at moderate levels of family management, such that once children in these families experienced more than a certain degree of poor family management practices, their odds of engaging in serious drug use went from near 0 to near 1 very quickly. In the other group, there was a consistent positive relationship between poor family management and drug use. While the proportional odds model identified two classes, the continuous outcomes model suggested at least three classes (models beyond this number failed to converge); results were not very stable and appeared to be driven largely by differences in means between classes. The differential effects that were identified only partially overlapped with the results of the proportional odds model. They did show two groups where one had a stronger effect than the other, however, they appeared to be driven more by mean differences in outcomes, and the regression lines for the effect of family management did not overlap in the continuous outcome model. Further, the three class solution suggested a large class where there is no effect of family management, and this would be a very surprising result given the consistent effects that are seen in the literature. The results are thus more ambiguous in terms of the presence of differential effects. This could be due to the large floor effect pushing the distribution further than was reasonable. Another interesting result is that because of class specific threshold estimates, it is clear that the proportional odds model can sometimes capture non-linear relationships between the predictor and outcome. This finding brings to light a strength of this approach which was not immediately obvious.

This study has presented some issues to consider when using regression mixture models but is by necessity limited in scope. Additional research is needed to examine issues further, including: other approaches to dealing with non-normality (such as allowing for additional classes within each differential effects group that allow for non-normal errors) and using other forms for the error distributions (Stasinopoulos, Rigby, & Akantziliotou, 2010); how these models work with more complex differential effects; the consequences of violating the assumptions of the proportional odds model; performance of the proportional odds model with greater levels of non-normal errors; and model stability across different sample sizes. While no simulation study is all-inclusive, our goal in this study was to use simulations to illustrate how these assumptions can quickly lead to problems with the performance of regression mixtures, and not to explore a comprehensive set of conditions under which the model may fail. Another limitation with this approach is that even these very simple regression mixture models can be very computationally intensive evidenced by a single model taking hours to converge in some cases. As the use of finite mixture models in the social sciences continues to increase, and as demands for new methods to examine individual differences in effects increases, we believe that regression mixtures are a powerful tool that should be considered. However, the results of this study and of other critiques of finite mixtures illustrate some of the trade-offs. While these models can provide compelling and intuitive results, it isn't obvious whether the common interpretation of these results reflect the underlying mechanisms of interest or whether they are driven by other features of the data. Increased and further collaboration between methodologists and substantive researchers is needed to develop new methods to address these issues and identify clear standards for the appropriate use of these techniques.

Acknowledgments

This research was supported by grant #R01 HD054736, M. Lee Van Horn (PI), funded by the National Institute of Child Health and Human Development. Applied data come from the Diffusion Study, J. David Hawkins (PI), funded by grant #R01 DA10768 from the National Institute on Drug Abuse (NIDA). We are grateful for the support and feedback from our colleagues in the Prevention Science and Methodology Group, supported by NIDA through grant # R01 MH4085, Hendricks Brown (PI). We also appreciate the assistance of Melisa Ward George in

reviewing and editing the manuscript. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

Citations

- Arthur MW, Hawkins JD, Pollard JA, Catalano RF, Baglioni AJ Jr. Measuring risk and protective factors for substance use, delinquency, and other adolescent problem behaviors: The Communities That Care Youth Survey. *Evaluation Review*. 2002; 26(6):575–601. [PubMed: 12465571]
- Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*. 1986; 51:1173–1182. [PubMed: 3806354]
- Bauer DJ, Curran PJ. Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*. 2003; 8:338–363. [PubMed: 14596495]
- Bauer DJ, Curran PJ. The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods*. 2004; 9:3–29. [PubMed: 15053717]
- Bechtel, R.; Churchman, A., editors. *Handbook of environmental psychology*. Wiley; New York: 2002.
- Boyce WT, Frank E, Jensen PS, Kessler RC, Nelson CA, Steinberg L, et al. Social context in developmental psychopathology: Recommendations for future research from the MacArthur Network on Psychopathology and Development. *Development and Psychopathology*. 1998; 10:143–164. [PubMed: 9635219]
- Bronfenbrenner, U., editor. *Making human beings human: Bioecological perspectives on human development*. Sage Publications; Thousand Oaks, CA: 2005.
- Caspi A, McClay J, Moffitt TE, Mill J, Martin J, Craig IW, et al. Role of genotype in the cycle of violence in maltreated children. *Science*. 2003; 297:851–859. [PubMed: 12161658]
- Cohen, J.; Cohen, P.; West, SG.; Aiken, LS. *Applied multiple regression/correlation analysis for the behavioral sciences*. Third ed.. Lawrence Erlbaum Associates; Mahwah, NJ: 2003.
- Coie JD, Watt NF, West SG, Hawkins JD, Asarnow JR, Marman HJ, et al. The science of prevention: A conceptual framework and some directions for a national research program. *American Psychologist*. 1993; 48(10):1013–1022. [PubMed: 8256874]
- Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society Series B*. 1977; 39:1–38.
- Desarbo WS, Jedidi K, Sinha I. Customer value analysis in a heterogeneous market. *Strategic Management Journal*. 2001; 22:845–857.
- Dishion TJ, McCord J, Poulin F. When interventions do harm: Peer groups and problem behavior. *American Psychologist*. 1999; 54(9):755–764. [PubMed: 10510665]
- Elder GH. The Life Course Developmental Theory. *Child Development*. 1998; 69(1):1–12. [PubMed: 9499552]
- Ellickson PL, Bell RM. Drug prevention in junior high: A multi-site longitudinal test. *Science*. 1990; 247(4948):1299–1305. [PubMed: 2180065]
- Flay BR, Biglan A, Boruch RF, Castro FG, Gottfredson D, Kellam S, et al. Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*. 2005; 6(3):151–175. [PubMed: 16365954]
- Flora DB, Curran PJ. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*. 2004; 9(4):466–491. [PubMed: 15598100]
- Glaser RR, Van Horn ML, Arthur MW, Hawkins JD, Catalano RF. Measurement properties of the Communities That Care Youth Survey across demographic groups. *Journal of Quantitative Criminology*. 2005; 21:73–102.
- Grün; Leisch. Identifiability of Finite Mixtures of Multinomial Logit Models with varying and fixed effects: Technical University Munich Nr 24o. 2008. Document Number
- Kaplan, D. *Structural Equation Modeling: Foundations and extensions*. Sage Publications; Thousand Oaks, CA: 1999.

- Kellam SG, Koretz D, Moscicki EK. Core elements of developmental epidemiologically based prevention research. *American Journal of Community Psychology*. 1999; 27(4):463–482. [PubMed: 10573831]
- Kumpfer KL, Smith P, Franklin Summerhays J. A wakeup call to the prevention field: Are prevention programs for substance use effective for girls? *Substance Use and Misuse*. 2008; 43:978–1001. [PubMed: 18649225]
- Kutner, MH.; Nachtsheim, CJ.; Neter, J.; Li, W. *Applied linear statistical models*. Fifth ed.. McGraw-Hill; New York: 2005.
- Leisch, F.; Gruen, B. *flexmix: Flexible Mixture Modelings*. Version 2.6-2. 2010.
- McLachlan, G.; Peel, D. *Finite Mixture Models*. John Wiley & Sons, Inc; New York: 2000.
- Muthén B. The Potential of Growth Mixture Modelling. *Infant and Child Development*. 2006; 15(6): 623–625.
- Muthén, B.; Collins, LM.; Sayer, AG. Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class-latent growth modeling. *American Psychological Association*; Washington, DC, US: 2001.
- Muthén BO, Brown CH, Masyn K, Jo B, Khoo S, Yang C, et al. General growth mixture modeling for randomized prevention trials. *Biostatistics*. 2002; 3:459–475. [PubMed: 12933592]
- Muthén, LK.; Muthén, BO. *Mplus*. Version 5.2. Muthén & Muthén; Los Angeles: 2008.
- Nagin, DS. *Group Based Modeling of Development*. Harvard University Press; Cambridge, MA: 2005.
- Nagin DS, Farrington DP, Moffitt TE. Life-Course Trajectories of Different Types of Offenders. *Criminology*. 1995; 33(1):111–139.
- Nylund KL, Asparauhov T, Muthen BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*. 2007; 14(4):535–569.
- Patterson GR, DeBaryshe BD, Ramsey E. A developmental perspective on antisocial behavior. *American Psychologist*. 1989; 44(2):329–335. [PubMed: 2653143]
- R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing; Vienna, Austria: 2008.
- Rowe DC, Osgood DW. Heredity and sociological theories of delinquency: A reconsideration. *American Sociological Review*. 1984; 49:526–540.
- Sampson, RJ.; Laub, JH. *Crime in the Making: Pathways and Turning Points Through Life*. Harvard University Press; Cambridge, MA: 1993.
- Schaeffer CM, Petras H, Ialongo N, Poduska J, Kellam S. Modeling Growth in Boys' Aggressive Behavior Across Elementary School: Links to Later Criminal Involvement, Conduct Disorder, and Antisocial Personality Disorder. *Developmental Psychology*. 2003; 39(6):1020–1035. [PubMed: 14584982]
- Sloboda Z, Stephens RC, Stephens P, Grey SF, Teasdale B, Hawthorne RD, et al. The Adolescent Substance Abuse Prevention Study: A randomized field trial of a universal substance abuse prevention program. *Drug and Alcohol Dependence*. 2009; 102:1–10. [PubMed: 19332365]
- Sperrin M, Jaki T, Wit E. Probabilistic relabelling strategies for the label switching problem in Bayesian mixture models. *Statistics in Computing*. 2010; 20:357–366.
- Spoth RL, Shin C, Guyll M, Redmond C, Azevedo K. Universality of effects: An examination of the comparability of long-term family intervention effects on substance use across risk-related subgroups. *Prevention Science*. 2006; 7(2):209–224. [PubMed: 16791523]
- Stasinopoulos, M.; Rigby, B.; Akantziliotou, C. *gamlss: Generalized Additive Models for Location Scale and Shape*. Version 4.0-0. 2010.
- Van Horn ML, Fagan AA, Jaki T, Brown EC, Hawkins JD, Arthur MW, et al. The use of multilevel mixture models for the identification of differential intervention effects in a community randomized trial. *Multivariate Behavioral Research*. 2008; 43(2):289–326.
- Van Horn ML, Jaki T, Masyn K, Ramey SL, Antaramian S, Lemanski A. Assessing Differential Effects: Applying Regression Mixture Models to Identify Variations in the Influence of Family Resources on Academic Achievement. *Developmental Psychology*. 2009; 45:1298–1313. [PubMed: 19702393]

- Wedel, M.; Desarbo, WS. A review of recent developments in latent class regression models. In: Bagozzi, RP., editor. *Advanced Methods of Marketing Research*. Blackwell Publishers; Cambridge: 1994. p. 352-388.
- Wedel M, Desarbo WS. A mixture likelihood approach for generalized linear models. *Journal of Classification*. 1995; 12:21–55.
- Whitehead J, Jaki T. One- and two-stage design proposals for a phase II trial comparing three active treatments with control using an ordered categorical endpoint. *Statistics in Medicine*. 2009; 28(5): 828–847. [PubMed: 19061203]

Figure 1.
Histograms for the residual variance of Y under the different simulation conditions.

Figure 2.
Contour plots of the joint dataset (classes 1 and 2 combined)

Figure 3.
Results for ordinal outcomes with differing amounts of skew

Figure 4.
Differences in the effects of family management on ordinal drug use

Table 1

Latent class enumeration for continuous outcome with varying degrees of skew in the errors

Classes	No Skew				Skew of .15 in class 2				Skew of .35 in class 2				Skew of .49 in class 2				Skew of .49 in both classes			
	2	3	4		2	3	4		2	3	4		2	3	4		2	3	4	
Loglikelihood	−7973	−7967	−7962		−7984	−7978	−7951		−7988	−7976	−7961		−7992	−7974	−7966		−7943	−7860	−7839	
Entropy	0.14	0.41	0.51		0.14	0.39	0.48		0.13	0.27	0.41		0.13	0.23	0.37		0.29	0.27	0.36	
AIC	1.00	0.76	0.54		1.00	0.80	0.68		1.00	0.99	0.65		1.00	1.00	0.60		1.00	1.00	1.00	
BIC	1.00	0.00	0.34		1.00	0.00	0.00		1.00	0.10	0.27		1.00	0.55	0.38		1.00	1.00	0.67	
Adj BIC	1.00	0.04	0.38		1.00	0.05	0.04		1.00	0.49	0.31		1.00	0.91	0.41		1.00	1.00	0.96	
BLRT	1.00	0.08	0.04		1.00	0.12	0.07		1.00	0.67	0.19		1.00	0.95	0.21		1.00	1.00	0.98	
Smallest Class*	39%–50%	0%–17%	0%–3%		36%–49%	0%–21%	0%–3%		32%–49%	0%–27%	0%–10%		34%–49%	5%–28%	0%–11%		24%–37%	14%–27%	0%–14%	

Note:

* the 95% range of the proportion of cases in the smallest class across all simulations; loglikelihood and entropy values are medians across all simulations

Table 2

Parameter estimates for 2 class results

Parameters	No Skew		Skew of .49 in Class 2		Skew of .49 in both classes			
	Class 1 (50%)	B (SE)	Class 2 (50%)	B (SE)	Class 1 (49%)	B (SE)	Class 2 (51%)	Class 1 (68%) Class 2 (32%)
Intercept	0.00	(0.05)	0.50	(0.03)	0.11	(0.07)	0.38	0.33 0.13 (0.04)
Slope	0.20	(0.04)	0.70	(0.03)	0.19	(0.07)	0.70	0.34 0.70 (0.06)
Residual Variance	0.96	(0.04)	0.51	(0.04)	1.06	(0.06)	0.49	1.10 0.21 (0.07)

Table 3

Parameter estimates for 3 class models with skew of .49 in error terms

Parameters	Skew of .49 in Class 2						Skew of .49 in both classes					
	Class 1 (45%)		Class 2 (36%)		Class 3 (23%)		Class 1 (28%)		Class 2 (38%)		Class 3 (34%)	
	B	(SE)	B	(SE)	B	(SE)	B	(SE)	B	(SE)	B	(SE)
Intercept	-0.02	(0.09)	0.20	(0.08)	0.90	(0.28)	-0.44	(0.12)	0.20	(0.04)	0.87	(0.08)
Slope	0.16	(0.05)	0.64	(0.14)	0.70	(0.10)	0.13	(0.03)	0.68	(0.03)	0.46	(0.04)
Residual Variance	0.96	(0.09)	0.39	(0.23)	0.60	(0.15)	0.44	(0.07)	0.25	(0.04)	0.97	(0.06)

Table 4

Latent class enumeration for ordinal outcome with varying degrees of skew in the errors

Classes	Skew 0			Skew .49			Skew .49 (both classes)			
	2	3		2	3		2	3	4	
Loglikelihood	-10113	-10109		-10113	-10094		-10066	-10052	-10042	
Entropy	0.18	0.39		0.18	0.26		0.14	0.31	0.41	
AIC	1.00	0.45		1.00	0.00		1.00	0.94	0.71	
BIC	0.95	0.00		1.00	0.00		0.97	0.00	0.00	
Adj BIC	1.00	0.00		1.00	0.00		1.00	0.15	0.04	
BLRT	1.00	0.65		1.00	0.12		1.00	1.00	NA	
Smallest Class *	27%–49%	3%–19%		21% – 47%	6% – 28%		31%–49%	5%–25%	4%–12%	
Non-convergence	0%	40%		0%	32%		0%	6%	35%	

Note:

* the 95% range of the proportion of cases in the smallest class across all simulations; loglikelihood and entropy values are medians across all simulations

Table 5

Results for the two class models with ordinal outcomes

Parameters	Population Values					
	Class 1		Class 2		Class 1	
No skew					40%	60%
T1	-1.26	-3.00	-1.25	(0.13)	-2.61	(0.26)
T2	-0.32	-1.64	-0.24	(0.15)	-1.45	(0.21)
T3	0.41	-0.58	0.59	(0.21)	-0.52	(0.15)
T4	1.15	0.47	1.52	(0.35)	0.40	(0.12)
T5	2.17	1.78	3.34	(2.86)	1.59	(0.10)
Slope	0.34	1.80	0.24	(0.13)	1.46	(0.17)
Skew of .49 in Class 2					35%	65%
T1	-1.24	-3.08	-1.34	(0.14)	-2.41	(0.20)
T2	-0.33	-1.60	-0.43	(0.16)	-1.16	(0.12)
T3	0.38	-0.49	0.31	(0.21)	-0.20	(0.10)
T4	1.12	0.57	1.13	(0.26)	0.71	(0.10)
T5	2.15	1.88	2.35	(0.36)	1.83	(0.10)
Slope	0.34	1.73	0.12	(0.17)	1.45	(0.19)
Skew of .49 in both classes					41%	49%
T1	-1.18	-3.31	-1.51	(0.08)	-2.76	(0.37)
T2	-0.25	-1.75	-0.60	(0.09)	-1.31	(0.14)
T3	0.46	-0.55	0.06	(0.09)	-0.08	(0.17)
T4	1.15	0.63	0.73	(0.08)	1.09	(0.27)
T5	2.09	2.05	1.69	(0.09)	2.45	(0.40)
Slope	0.35	1.88	0.29	(0.12)	1.96	(0.35)

Table 6

Model selection for Poor Family Mangement preditiong Substance Use

Classes	Ordinal Outcome			
	1 Class	2 Class	3 Class	4 Class
Loglikelihood	-12903	-12859	-12816	-12800
Entropy	1.00	0.42	0.52	0.54
AIC	25819	25744	25672	25654
BIC	25860	25834	25811	25841
Adj BIC	25841	25793	25747	25756
BLRT	NA	p<.05	p<.05	p<.05
Smallest Class	NA	23%	5%	3%

Note: The solution choosen by each criteria is in bold

Table 7

Results for the two class model of family management predicting ordinal substance use

Parameters	Class 1 (23%)		Class 2 (77%)	
	B	(SE)	B	(SE)
T1	-1.42	(0.22)	-0.59	(0.05)
T2	-0.53	(0.18)	0.09	(0.05)
T3	-0.16	(0.23)	0.83	(0.07)
T4	0.57	(0.22)	1.63	(0.11)
T5	0.99	(0.33)	3.40	(0.40)
Family Management	3.25	(0.54)	0.84	(0.07)