A Mixture Model of Discontinuous Development in Heavy Drinking From Ages 18 to 30: The Role of College Enrollment*

STEPHANIE T. LANZA, PH.D.,† AND LINDA M. COLLINS, PH.D.

The Methodology Center, The Pennsylvania State University, 204 E. Calder Way, Suite 400, State College, Pennsylvania 16801

ABSTRACT. Objective: The purpose of this study was to illustrate the use of latent class analysis to examine change in behavior over time. Patterns of heavy drinking from ages 18 to 30 were explored in a national sample; the relationship between college enrollment and pathways of heavy drinking, particularly those leading to adult heavy drinking, was explored. **Method:** Latent class analysis for repeated measures is used to estimate common pathways through a stage-sequential process. Common patterns of development in a categorical variable (presence or absence of heavy drinking) are estimated and college enrollment is a grouping variable. Data were from the National Longitudinal Survey of Youth (N = 1,265). **Results:** Eight patterns of heavy drinking were identified: no heavy drinking (53.7%); young adulthood only (3.7%); young adulthood and adulthood (3.7%); college age only (2.6%); college age, young adulthood, and adulthood (8.7%); high school and college age (4.4%); high school, college age, and young adulthood (6.3%); and persistent heavy drinking (16.9%). **Conclusions:** We found no evidence that prevalence of heavy drinking for those enrolled in college exceeds the prevalence for those not enrolled at any of the four developmental periods studied. In fact, there is some evidence that being enrolled in college appears to be a protective factor for young adult and adult heavy drinking. College-enrolled individuals more often show a pattern characterized by heavy drinking during college ages only, with no heavy drinking prior to and after the college years, whereas nonenrolled individuals not drinking heavily during high school or college ages are at increased risk for adult heavy drinking. (*J. Stud. Alcohol* **67:** 552-561, 2006)

THE EVOLUTION OF PROBLEM BEHAVIOR from high school through emerging adulthood (ages 18-25; Arnett, 2000) shows great variability across individuals. That is, some individuals may consistently exhibit a given problem behavior over time, whereas others never do. Some individuals may show an increasing or decreasing trend in the behavior throughout emerging adulthood. Others may exhibit the behavior only during one period in their lives, which coincides with a particular event or role. Roles that often occur during emerging adulthood include college enrollment, full-time work, marriage, and parenthood. The developmental study of problem behavior such as heavy drinking during emerging adulthood may show that the presence of the behavior at a particular age coincides with various roles that individuals undertake. One role that may be particularly important in the development of heavy drinking is college enrollment.

Quantifying the different pathways, or patterns, of problem behavior that individuals tend to follow throughout emerging adulthood makes it possible to piece together a clear and often simple picture of "types" of individuals that exist in the population. This is not to say that any set of patterns contains the definitive list of the ways in which individuals develop over time but, rather, that a set of patterns presents one way to summarize variability in individual development.

The identification of subgroups of individuals who show similar developmental patterns has become increasingly important in the field of human development. Mixture models, one approach to identifying subgroups, are useful in studying development in both continuous and categorical outcomes. Mixture models for the analysis of change stem directly from latent class theory, a measurement theory for static, categorical latent variables that divide the population into two or more mutually exclusive and exhaustive latent classes (Clogg and Goodman, 1984; Goodman, 1974; Lazarsfeld and Henry, 1968). Latent class theory assumes that there is some underlying grouping variable such that all individuals fall into one and only one of the groups or categories. This variable is latent in the sense that it cannot be directly measured and instead is measured using several fallible indicators. The latent class model divides the population into two or more homogeneous subgroups, and all individuals in a subgroup are expected to have the same

Received: July 25, 2005. Revision: February 3, 2006.

*This research was supported by National Institute on Drug Abuse grants 2-P50-DA-10075 and K05-DA-018206.

†Correspondence may be sent to Stephanie T. Lanza at the above address or via email at: SLanza@psu.edu. Linda M. Collins is with The Methodology Center and Department of Human Development and Family Studies, The Pennsylvania State University, State College, PA.

probability of responding to questionnaire items in a particular way.

There are several practical reasons for identifying a set of developmental patterns in problem behavior such as heavy drinking. One reason to study developmental patterns is to describe, in a concise way, how individuals behave over time. Depending on the patterns that are identified, it may be possible to establish with a fair amount of certainty which pathway an individual is taking early in the process, enabling early intervention. Another motivation is the identification of risk factors related to more problematic pathways; understanding how risk factors relate to various developmental pathways could allow for more targeted intervention and prevention programs.

Several studies have used group-based approaches to identify subgroups of individuals based on their pattern of heavy drinking during emerging adulthood. Various statistical procedures have been used to identify the groups. For example, a study employing cluster analysis revealed seven groups, six labeled chronic, decreased, increased, fling, rare, and never, and a category including all other patterns (Schulenberg et al., 1996). More recently, a number of studies have used general growth mixture modeling (GGMM) to identify heavy-drinking groups that share similar growth curve features. Unlike cluster analysis, which is more exploratory in nature, GGMM provides a model-based approach to identifying groups. For example, Tucker et al. (2003) identified the following five groups based on their heavy-drinking patterns: nonbingers, moderate stables, steady increasers, adolescent bingers, and early highs. Muthén (2001) identified the following three groups: normative drinking, high (but decreasing) drinking, and increasing drinking. Regime switching, an extension of GGMM that allows for transitions between trajectory groups, was used to explore three trajectory groups (light, moderate, and heavy drinking) among regular drinkers (Dolan et al., 2005).

GGMM has been used to identify groups posited to exist in the population, where each group can be summarized by a trend in heavy drinking that follows some functional form (such as quadratic growth) over time. Although regime switching allows individuals to switch between groups at each time point, to our knowledge a group-based approach that captures discontinuous patterns of individual development across multiple time points has not been taken in the examination of heavy drinking during emerging adulthood and adulthood. The application of latent class analysis (LCA) to repeated measures provides an alternative to GGMM. Like GGMM, this approach enables identification of subgroups of individuals who are homogeneous in their pattern of behavior over time in some categorical outcome. Unlike GGMM, in LCA development is not assumed to follow a functional form with respect to time. Each estimated latent class represents one possible pattern of development in the categorical outcome across all time points. Such an approach could be used to identify groups characterized by the absence of the behavior at all times (consistent nonuse), the presence of the behavior at all times (chronic heavy drinking), and anything in between. For example, this approach could reveal homogeneous groups of individuals engaging in heavy drinking during certain developmental periods only, such as during college ages but not during high school, young adulthood, or adulthood.

Grouping variables (such as gender or treatment group) or covariates hypothesized to be related to the prevalence of each developmental pattern can be incorporated in LCA for repeated measures. One role that may have an effect on concurrent or long-term heavy drinking is college enrollment. Heavy drinking in college is known to be a common behavior. Nationally, over 40% of all college students engage in this behavior (e.g., Wechsler et al., 2000). Although this fact has received much attention and suggests that the college population provides an important target for intervention, individuals enrolled in college and those not enrolled in college tend to be similar at these ages in terms of alcohol consumption (Johnston et al., 1996). Further, college enrollment has been shown to be predictive of lower substance use rates among adults (Merline et al., 2004).

Important questions remain regarding continuous and discontinuous patterns of heavy drinking during emerging adulthood and adulthood. In particular, a descriptive analysis of the presence or absence of the behavior across all times will contribute to the understanding of the etiology of heavy drinking from adolescence through adulthood. Also, incorporating college enrollment can illuminate how this role relates to individuals' movement in and out of heavy-drinking states during emerging adulthood and adulthood. The current study will address several research questions using a national sample to better understand heavy-drinking behavior from high school to adulthood and the role played by college enrollment in the development of this behavior. The five research questions we will address are the following:

- Question 1: What patterns of heavy drinking are necessary to represent heterogeneity among individuals from a national sample followed over 12 years (ages 18-30)? What is the prevalence of each pattern?
- Question 2: Does the prevalence of the heavy-drinking patterns from ages 18 to 30 differ for individuals who were enrolled in college?
- Question 3: When does heavy drinking start? Are people enrolled in college more likely to engage in heavy drinking *for the first time* during college ages?
- Question 4: Among college-aged individuals, are those enrolled in college more likely to engage in heavy drinking?
- Question 5: Is there evidence of differential risk of adult heavy drinking for those who enroll in college versus those

who do not? Does this vary as a function of heavy drinking during each developmental period?

Addressing these five questions will provide an illustration of the use of LCA to examine change in behavior over time in repeated measures data.

Method

Participants

The longitudinal panel data in this example are from the 1964 birth cohort of the National Longitudinal Survey of Youth (NLSY). This data set is a nationally representative sample of youth drawn in 1979. The six waves of measurement correspond to individuals of age 18, 19, 20, 24, 25, and 30. These waves of measurement map onto four developmental periods: high school (age 18); post-high school, which will be referred to as college age (ages 19-20); young adulthood (ages 24-25); and adulthood (age 30). The sample size at the first wave of data collection was 1,265.

Measures

The NLSY included an ordinal variable measuring the frequency of six or more drinks at one time in the last month using the following categories: never, 1 time, 2 or 3 times, 4 or 5 times, 6 or 7 times, 8 or 9 times, and 10 or more times. Because heavy drinking was not a normative behavior, the distribution of this variable strongly violated normality. The modal category at each age was "never," and the distribution of the remaining categories was strongly positively skewed.

A binary variable was created for each of the six times of measurement that included the following categories: 0 for missing, 1 for no recent heavy drinking, and 2 for heavy drinking one or more times in the last month. The percentage of participants missing a response to the heavy-drinking item at ages 18, 19, 20, 24, 25, and 30 was 2.8%, 2.9%, 3.7%, 7.5%, 6.3%, and 19.4%, respectively. We speculate that the substantial increase in the amount of missing data at age 30 was due to attrition, as this was a 5-year follow-up. At each time point, an individual could be expected to fall into one of two stages, heavy drinking or no heavy drinking. Each developmental pattern of heavy drinking will be characterized by a unique pattern of stage membership across time.

The NLSY asked individuals at each wave whether they were currently enrolled in college. For purposes of this study, individuals were considered to be enrolled in college during normative college years if they answered "yes" to this question at either age 19 or 20 (normative college ages). Of the 1,265 individuals in this study, 33.1% were enrolled, 62.4% were not enrolled, and 4.5% were missing responses to these items.

LCA for repeated measures

The application of LCA to repeated measures enables identification of common patterns of discontinuous development in a categorical manifest or latent variable. In general, a latent class model with a grouping variable involves several sets of parameters. One set includes estimates of the proportion of individuals in each group (groups can be directly observed or latent; if groups are latent, there is a set of parameters that reflects the correspondence between the groups and the items measuring group membership). Another set includes estimates of the proportion of individuals in each latent class conditional on group (in this study, these parameters reflect the prevalence of each pattern of development conditional on college enrollment). The final set reflects the correspondence between indicators of the outcome at all times and the latent class (these are conditional item-response probabilities). The latent class model assumes that any covariation among indicators of the outcome is accounted for by the latent class variable. This is known as the assumption of local independence. See Lanza (2003) for the mathematical model and parameterization of the latent class model for repeated measures. For a more extensive introduction to latent class models, see Lanza et al. (2003).

In the present study, the outcome is individuals' progression in heavy drinking throughout emerging adulthood and into adulthood. There is a binary indicator of heavy drinking at six time points (corresponding to ages 18, 19, 20, 24, 25, and 30), so there are $2^6 = 64$ different possible patterns representing development. Examples of patterns include the following: no heavy drinking at any time; heavy drinking at all times; heavy drinking during high school and college ages; and heavy drinking during adulthood only. The grouping variable is a composite of college enrollment at ages 19 and 20, as described above. Using LCA, it will be possible to determine which set of patterns is sufficient to describe development and whether college enrollment relates to the prevalence of these patterns.

Parameter restrictions. The restricted latent class model was first presented by Goodman (1974). Parameter restrictions play a crucial role in many LCA models for the following reasons: they aid in model identification, they can help the user to achieve a parsimonious model, they enable tests of whether two parameters are equal (for example, whether the prevalences of two patterns are equal), and they enable confirmatory tests regarding measurement invariance over times and groups. Issues of measurement invariance arise in all latent variable models. Because the item-response probabilities form the basis for interpreting the latent classes, hypothesis tests regarding measurement invariance can be conducted in LCA by imposing equivalence sets on these parameters. (This is conceptually similar to issues in measurement invariance that arise in structural equation modeling. See Meredith [1993] for a discussion of invariance in the common factor model and its extensions.) Lanza et al. (2003) demonstrated how to test latent class model parameter restrictions across groups and times.

Model estimation. Analyses for the current study were conducted using WinLTA (Collins et al., 2002; Collins and Wugalter, 1992; Lanza et al., 2003), a software package designed to estimate latent class and latent transition models. WinLTA employs a full-information maximum-likelihood routine that handles missing data; this procedure assumes that data are missing at random (Little and Rubin, 1989). This approach can handle missing data in the indicators of the dynamic categorical outcome as well as the indicator of the grouping variable.

Model selection. A common measure of fit in a contingency table analysis is the likelihood ratio statistic G^2. This statistic has the advantage that nested models can be compared by a likelihood ratio test, with the resulting statistic distributed as chi-square. The G^2 statistic has the asymptotic property that the distribution is chi-square with degrees of freedom equal to *the number of possible response patterns* minus *the number of parameters estimated minus one.* Ideally, this approach could be used to help determine the appropriate number of latent classes by comparing the fit of two models. Unfortunately, this test is not appropriate for comparing two models with different numbers of patterns because parameters of the simpler model take on boundary values (i.e., probabilities of zero or one) of the parameter space (Everitt, 1988; Rubin and Stern, 1994). Under these conditions, the distribution of the likelihood ratio statistic is undefined.

Various model selection information criteria have been proposed for comparing models with different numbers of classes, including the Akaike Information Criterion (AIC; Akaike, 1987) and the Bayesian Information Criterion (BIC; Schwarz, 1978). The AIC and BIC are penalized log-likelihood test statistics, where the penalty is two times the number of parameters estimated for the AIC and the log of n times the number of parameters estimated for the BIC. Unfortunately, the AIC and BIC serve only to compare the relative fit of several models under consideration but do not help in determining whether a particular model has sufficiently good fit. In practice, it is helpful to examine the G^2 statistic, AIC, and BIC, along with examining the parsimony and interpretability of the model, when choosing the optimal number of latent classes.

Testing hypotheses. WinLTA includes a Bayesian estimation routine for latent class models, called data augmentation (DA), that provides a measure of uncertainty associated with each parameter estimate. This approach is built on the assumption that each parameter has a distribution, and multiple independent draws from this distribution can be used to infer the central tendency and variability of the parameter. There have been several previous applications of DA to latent class and latent transition models (e.g., Lanza et al., 2005; Lanza and Collins, 2002; Rubin and Stern, 1994). In applying this approach to LCA, an individual's latent class membership is treated as missing data and imputed multiple times. These imputed data sets are statistically independent and take into account the uncertainty associated with the fact that an individual's latent class membership is unknown. Parameter estimates and standard errors can be obtained using a set of rules developed by Rubin (1987).

DA allows for tremendous flexibility in hypothesis testing within the latent class framework. This procedure allows for standard hypothesis tests of group differences in single parameter estimates (for example, group differences in the prevalence of one heavy-drinking pattern). In addition, this procedure allows the researcher to define new parameters that map directly onto questions of theoretical interest and test for group differences in the prevalence of these parameters. This approach was described by Lanza et al. (2005).

Results

The ages covered in this birth cohort from the NLSY span four developmental periods. Age 18 falls at the end of high school for most adolescents, whereas ages 19 and 20 fall during the time when many youth attend college (college age). Ages 24 and 25 are designated as young adulthood, and age 30 as adulthood. The fact that the six measurement occasions map onto four developmental periods can aid interpretation. The two measurement occasions during the college ages can be thought of as multiple indicators of heavy drinking during that developmental period. Similarly, the two measurement occasions during young adulthood can be thought of as multiple indicators of heavy drinking during that period. The high school and adulthood periods each have only one indicator. (Note that this does not change the number of possible patterns, which remains 64.) Variability in the number of indicators across developmental periods does not pose a problem for this modeling approach.

Question 1: What patterns of heavy drinking are necessary to represent heterogeneity among individuals from a national sample followed over 12 years (ages 18 to 30)? To answer this, a series of latent class models were compared on the basis of fit indices, and pairs of models that were statistically nested (such as models with the same number of latent classes but different parameter restrictions on the item-response probabilities) were compared using a G^2 difference test. The final model includes a parsimonious and interpretable set of 8 of the 64 possible heavy-drinking patterns. The matrix of item-response probabilities in Table 1 shows how the six questionnaire items defined the eight groups identified by the latent class model. Each number in

TABLE 1. Item-response probabilities for "yes" response to heavy-drinking variable

Latent class	Probability of "yes" response conditional on latent class					
	High school	College age		Young adulthood		Adulthood
	Age 18	Age 19	Age 20	Age 24	Age 25	Age 30
1	.065	.076	.122	.076	.082	.135
2	.065	.076	.122	.853	.761	.135
3	.065	.076	.122	.853	.761	.766
4	.065	.741	.819	.076	.082	.135
5	.065	.741	.819	.853	.761	.766
6	.807	.741	.819	.076	.082	.135
7	.807	.741	.819	.853	.761	.135
8	.807	.741	.819	.853	.761	.766

the table represents the probability of responding "yes" to the heavy-drinking item at that age, conditional on membership in a particular latent class. All probabilities in Table 1 tend to be close to zero or one, showing a strong correspondence between the heavy-drinking item at each time and the latent classes. Remember that each latent class corresponds to a group of individuals with a common heavy-drinking pattern. For example, the first row of the table shows that only 6.5% of individuals in Latent Class 1 reported heavy drinking at age 18, 7.6% reported the behavior at age 19, and so on. This group has a very low probability of reporting heavy drinking at each time. The second latent class consists of individuals who were likely to report heavy drinking at ages 24 and 25 (young adulthood) but not at other ages. In contrast, individuals in Latent Class 8 were likely to report heavy drinking at all six times (high school, college age, young adulthood, and adulthood), suggesting that this is a group of chronic heavy drinkers.

What is the prevalence of each pattern? As discussed above, the results of the LCA suggested that only eight patterns of heavy drinking were necessary to account for heterogeneity in behavior. Table 2 lists, for each heavy-drinking pattern, the expected response to the heavy-drinking question at each time, the prevalence of each pattern, and a label that describes the pattern. All eight patterns of behavior over time are clearly interpretable. Figure 1 depicts the eight heavy-drinking patterns identified in this model. These patterns can be grouped into the following general trends of heavy-drinking behavior: lifetime heavy drinking (16.9%); no heavy drinking (53.7%); increasing trend (12.4%); decreasing trend (10.7%); and short-term heavy drinking (6.3%). The proportion of individuals expected to engage in heavy drinking during college ages only (and not at other times) was 2.6%. Although this proportion is small, analyses incorporating college enrollment will reveal significant group differences for individuals with this pattern of drinking.

The most prevalent pattern of drinking was the no-heavy-drinking class, followed by persistent heavy drinking across all developmental periods. Also interesting is what did not emerge. For example, there was no evidence for a drinking class characterized by heavy drinking during adulthood only, nor was there any class evident involving heavy drinking in adulthood without heavy drinking in the period that immediately preceded it (young adulthood).

Question 2: Does the prevalence of the heavy-drinking patterns from ages 18 to 30 differ for individuals who were enrolled in college? To address this research question, college enrollment was added to the LCA as a grouping variable. The proportion of individuals who were enrolled in college at age 19 or 20 was 34.7% (this was the estimated percentage after taking missing data into account). Parameter restrictions were imposed on the item-response probabilities to ensure that measurement did not vary across groups. This was done so that similar quantities could be compared between individuals who did and did not enroll in college.

Table 3 shows the prevalence of the eight heavy-drinking patterns for each college-enrollment group and the *p* values associated with the hypothesis tests for group differences. The similarities are striking: No heavy drinking and persistent heavy drinking were the most prevalent patterns in both groups. However, there were several interesting differences. Most notable is the proportion of individuals expected to be in the "college-age only" pattern: 8.1% of those who were enrolled in college exhibited this developmental

TABLE 2. Eight patterns of heavy drinking along with expected proportion and assigned label

Heavy drinking expected during developmental period?					
High school	College	Young adult	Adult	Class proportion	Label
No	No	No	No	.537	No heavy drinking
No	No	Yes	No	.037	Young adulthood only
No	No	Yes	Yes	.037	Young adulthood and adulthood
No	Yes	No	No	.026	College age only
No	Yes	Yes	Yes	.087	College age, young adulthood, and adulthood
Yes	Yes	No	No	.044	High school and college age
Yes	Yes	Yes	No	.063	High school, college age, and young adulthood
Yes	Yes	Yes	Yes	.169	Lifetime heavy drinking

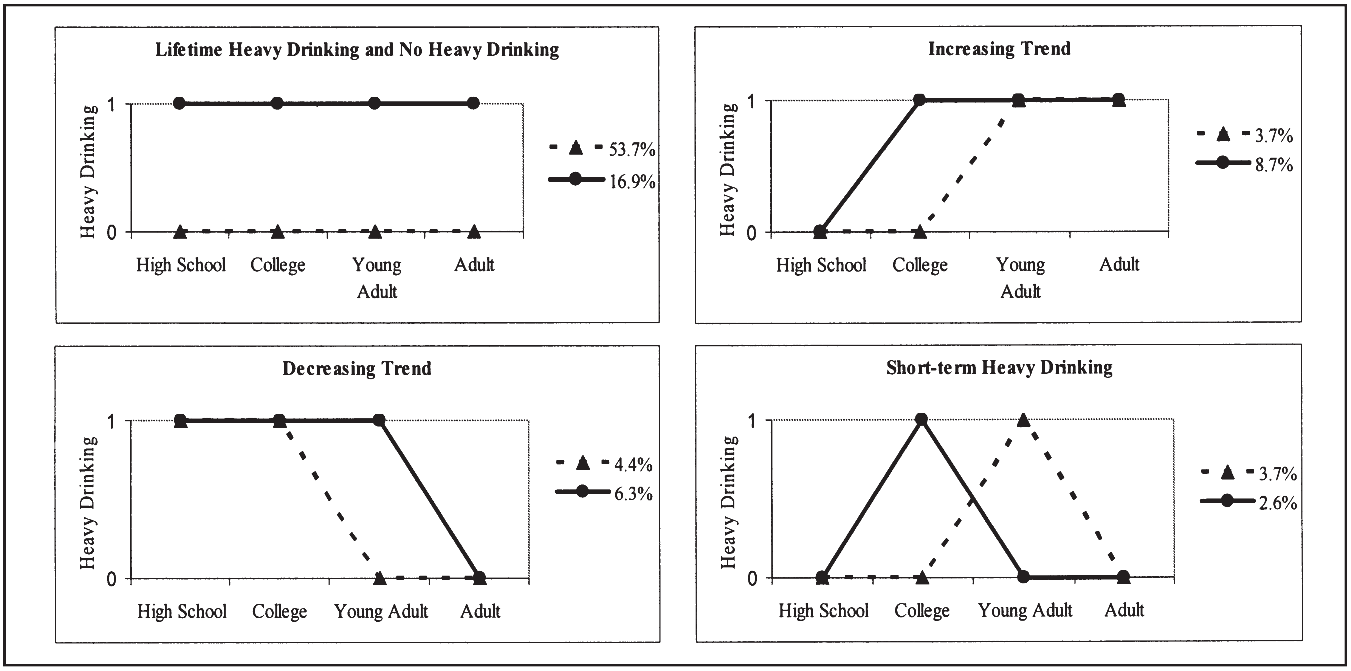

FIGURE 1. Eight developmental patterns of heavy drinking. Top-left panel depicts the drinking patterns over time for "lifetime heavy drinking" and "no heavy drinking"; top-right panel shows two drinking patterns characterized by onset during college ages or young adulthood; bottom-left panel depicts drinking patterns showing desistance of behavior during young adulthood or adulthood; bottom-right panel shows short-term drinking patterns characterized by heavy drinking during college ages or young adulthood only. The prevalence of each pattern is shown in the legends.

pattern, compared with no one in the nonenrolled group (p = .004). The pattern characterized by heavy drinking in young adulthood and adulthood tended to be more common among nonenrolled individuals (5.4% vs 1.6%), although the effect was just short of significant (p = .06).

Question 3: When does heavy drinking start? Are people enrolled in college more likely to engage in heavy drinking for the first time during college ages? While in high school, 26.8% of individuals who later enrolled in college reported heavy drinking, compared with 29.7% of those who did not enroll (p = .27). As can be seen in Table 3, there was a nonsignificant trend for individuals enrolled in college to be more likely to engage in heavy drinking for the first time during college years and for those not enrolled to begin during young adulthood. First occurrence of heavy drinking occurred during college ages for 13.4% of the enrolled group, compared with 9.1% for the nonenrolled group (p = .08). Individuals not enrolled in college were nearly three times more likely than those in the enrolled group to initiate heavy drinking in young adulthood (9.9% vs 3.9%, p = .09). The latent class model indicated that no one initiated heavy drinking during adulthood.

Question 4: Among college-aged individuals, are those enrolled in college more likely to engage in heavy drinking? One straightforward way to address this question is to see whether rates of reported heavy drinking at ages 19 and 20 varied according to college-enrollment status. Chi-square analyses revealed that college-enrollment status was not related to heavy drinking at age 19 (χ^2 = 0.4, 1 df, p = .5) or age 20 (χ^2 = 0.8, 1 df, p = .4).

We also used the prevalence of the developmental patterns to explore group differences in heavy drinking during each developmental period. For each college-enrollment group, the overall prevalence of heavy drinking for each developmental period was calculated by summing across patterns that were characterized by heavy drinking during that period. Table 3 shows the prevalence of drinking at each of the four developmental periods for each group. The overall prevalence of heavy drinking during high school (26.8% for enrolled and 29.7% for nonenrolled individuals) and college (40.2% for enrolled and 38.8% for nonenrolled individuals) ages was not significantly different across groups. However, the prevalence peaked during college ages for the enrolled group, whereas the rate of heavy drinking continued to escalate into young adulthood for the nonenrolled group. The difference in heavy drinking during young adulthood was highly significant (33.3% for enrolled vs 43.3% for nonenrolled individuals, p = .005), an effect that persisted into adulthood (22.2% for enrolled vs 32.8% for nonenrolled individuals, p = .02). So, among college-aged individuals, those enrolled in college were not any more likely to engage in heavy drinking. However, those not enrolled in college were more likely to engage in this risky behavior during young adulthood and adulthood.

Question 5: Is there evidence of differential risk of adult heavy drinking for those who enroll in college versus those

TABLE 3. Hypothesis tests for differences between enrolled and not enrolled groups in heavy drinking

Hypothesis test	Prevalence: Enrolled	Prevalence: Not enrolled	*p* value
Question 2: Prevalence of heavy-drinking patterns			
No heavy drinking	.559	.513	.45
Young adulthood only	.023	.045	.59
Young adulthood and adulthood	.016	.054	.06
College age only	.081	.000	.004
College age, young adulthood, and adulthood	.053	.091	.35
High school and college age	.027	.054	.34
High school, college age, and young adulthood	.088	.060	.32
Lifetime heavy drinking	.153	.183	.38
Question 3: First heavy drinking			
High school	.268	.297	.27
College age	.134	.091	.08
Young adulthood	.039	.099	.09
Adulthood	.000	.000	NA
Question 4: Prevalence of heavy drinking at each developmental period			
High school	.268	.297	.27
College age	.402	.388	.67
Young adulthood	.333	.433	.005
Adulthood	.222	.328	.02
Question 5: Adult heavy drinking given…			
High school heavy drinking	.571	.616	.61
No high school heavy drinking	.094	.206	.002
College age heavy drinking	.512	.706	.06
No college age heavy drinking	.027	.088	.05
Young adulthood heavy drinking	.667	.758	.20
No young adulthood heavy drinking	.000	.000	NA

who do not? As stated above, heavy drinking during adulthood was significantly less likely among the enrolled group (22.2% vs 32.8%, $p = .02$). *Does this vary as a function of heavy drinking during each developmental period?* The probability of heavy drinking during adulthood given that one engaged in the behavior during high school was no different for the college-enrollment groups (57.1% for enrolled vs 61.6% for nonenrolled individuals). However, among individuals who did *not* report heavy drinking in high school, those who did not enroll in college were more than twice as likely to engage in heavy drinking during adulthood (20.6% vs 9.4%, $p = .002$; see Table 3).

Among individuals who drank heavily during college ages, those who were enrolled in college tended to be less likely to continue heavy drinking into adulthood (51.2% for enrolled vs 70.6% for nonenrolled individuals, $p = .06$), although the effect was just short of significant. In addition, for individuals *not* reporting heavy drinking during college ages, those enrolled in college were significantly less likely to drink heavily in adulthood (2.7% vs 8.8%, $p = .05$).

Among individuals reporting heavy drinking during young adulthood, those enrolled and not enrolled in college were equally likely to engage in heavy drinking during adulthood (66.7% for enrolled vs 75.8% for nonenrolled individuals). Among individuals *not* reporting heavy drinking during young adulthood, the latent class model suggested there was little risk of heavy drinking during adulthood for either college-enrollment group.

Discussion

Our data show no evidence that the prevalence of heavy drinking at any of the four developmental periods studied for those enrolled in college exceeds the prevalence for those not enrolled. For those enrolled in college, there is an increase in the incidence of heavy drinking during the college years; however, this increase merely brings the heavy drinking of this group up to the level of those not enrolled in college. Moreover, there is some evidence that being enrolled in college appears to be a protective factor for young adult and adult heavy drinking. For example, a substantial number of college-enrolled individuals drink heavily during college ages only but desist by young adulthood; this pattern of behavior is less likely to be exhibited by the nonenrolled individuals. In addition, the probability of adult heavy drinking is higher for those who did not enroll in college regardless of whether they drank heavily or not in high school, college ages, and young adulthood.

For individuals with early and chronic alcohol problems, heavy drinking during high school may be associated with poor school performance, which may close off some options for higher education. This would tend to keep these individuals in the no-college group, accounting for much of the apparent protective effect associated with college enrollment. However, one striking finding was that, as compared with enrolled individuals, nonenrolled individuals who reported no heavy drinking during high school had twice the risk of transitioning to heavy drinking by adulthood. This suggests that a mechanism other than self-selection out of college for alcohol-related reasons accounts for at least part of the increased risk.

Based on a national sample, we found that more than half of individuals never engage in heavy drinking. Unfortunately, nearly 17% engage in chronic heavy drinking from adolescence through adulthood regardless of college enrollment. Results from this study suggest that, although heavy drinking does begin during the college years for many individuals, it is two to three times more likely to begin in high school. Thus the college years, though important, appear not to be the most important developmental period for the onset of heavy drinking. Instead, the high school years appear to be a critically important time for prevention efforts.

Given that those who do not enroll in college appear to be at risk of young adult and adult heavy drinking at a level surpassing those who are enrolled in college, prevention and harm-reduction efforts should be developed for this population. This is particularly important considering that this group may be more likely to be engaged in activities that could lead to harm to themselves or others if done while drinking. For example, those who are not enrolled in college during the college years may be more likely to drive while drinking, because they are more likely to live in an apartment or house as opposed to a college residence hall. They may also be more likely to have employment that requires them to operate machinery or perform other tasks that are dangerous under the influence of alcohol.

LCA to study heavy drinking over time

Recently there has been an increase in the number of studies describing heterogeneity in change over time by identifying unique groups of individuals distinguished by their growth trajectories. Two common modeling approaches, GGMM (Muthén and Sheddon, 1999) and semiparametric group-based modeling (Nagin, 2005), provide considerable flexibility in terms of the function of growth for each trajectory group. Although these approaches have proven to be very useful for describing development, they typically characterize the growth process as a smooth function of time. Some research questions, such as those having to do with the presence or absence of a particular behavior, may be better addressed by allowing development over time to be discontinuous. As a direct comparison, a growth mixture model of heavy drinking in the NLSY resulted in three latent trajectory classes (normative, high, and increasing drinking classes; Muthén, 2001). The current study, in which we treated the indicators of heavy drinking as categorical rather than continuous, produced a more detailed description of the heterogeneity of development in the sample that included eight developmental patterns. In particular, the following two patterns emerged that were characterized by heavy drinking during only one developmental period: heavy drinking during college ages only and heavy drinking during young adulthood only. In some instances, fitting a smooth quadratic growth curve to individual drinking patterns may obscure subtle but important differences in patterns that define subgroups of individuals. For example, identification of the fairly uncommon pattern characterized by drinking during college ages only made it possible to detect a strong relation between college enrollment and that developmental pattern.

Several longitudinal models have been proposed to handle discontinuities in change over time. One method, latent transition analysis (Lanza et al., 2005), can be used to estimate transitions in a categorical latent variable between two consecutive times. This method can be useful when the latent class structure at each time is fairly complex; LCA for repeated measures can become unwieldy with too many categories in the outcome at each time. However, LCA for repeated measures can be used to identify subgroups characterized by particular patterns of transitions between states across three or more times. Regime switching has also been proposed to allow for discontinuities in behavior, in this case by allowing individuals to switch from one continuous growth curve to another at each time (Dolan et al., 2005). This approach is ideal when the outcome can be modeled as a function of time, and it makes sense to model discontinuities in behavior by allowing individuals to switch between growth trajectories. However, for many applications it is useful to model development by identifying subgroups of individuals who show a common pattern of transitions over time between several qualitatively different states. LCA for repeated measures can provide a good summary of this type of developmental process.

An advantage of LCA for repeated measures is its flexibility. For example, in this approach the categorical outcome can be measured by varying numbers of indicators at each time of measurement. In this study, heavy drinking during high school and adulthood had one indicator each, whereas heavy drinking during college ages and young adulthood had two indicators each. Because the NLSY spanned such a long period of individual development, mapping times of measurement onto distinct periods of human development helped to simplify the interpretation of the results. The fact that LCA can handle varying numbers of indicators creates the possibility for handling other features of longitudinal designs as well. For example, one could use measures from multiple sources at different time points in the study.

The application of LCA to repeated measures could be used to examine many developmental processes. For example, one could model an individual's progression through the stages of change model corresponding to quitting smoking, allowing for relapse episodes. Another potentially important application of LCA for repeated measures would be to model psychological diagnosis over time to identify subgroups characterized by particular patterns of stability or change in diagnosis.

Limitations and future directions

A main limitation of the current study was that heavy drinking at each time point was assessed using a recent timeframe (i.e., heavy drinking during the past 30 days). This means that we could miss the measurement of heavy-drinking episodes that occur outside of these 30-day windows. It is possible that the current measurement of heavy drinking resulted in underestimation of the probability of heavy drinking at each of the six waves of measurement. However, we feel that a 1-month window of behavior is a reasonable period of recall for reporting heavy-drinking behavior and do not expect high variability from month to month in heavy drinking within each developmental period. In fact, the item-response probabilities were very similar across the pairs of items measuring heavy drinking at ages 19 and 20 (two times of measurement during college ages) and at ages 24 and 25 (during young adulthood), suggesting that the behavior was fairly stable within developmental periods.

Another limitation of this study, as with any LCA analysis, is that violations of local independence might affect the number of latent classes detected. It is possible that violations of the local independence assumption could result in overextraction of classes. As discussed earlier, this LCA analysis resulted in eight developmental patterns, whereas a GGMM analysis on similar data resulted in only three. However, treating the heavy-drinking indicators as binary items in the current study allowed us to identify unique groups characterized by discontinuities in behavior over time. Models have been proposed to extend LCA to allow for dependence among indicators (see, for example, Yang and Becker [1997] for the latent class marginal model). Such models may be worth considering when theory strongly suggests the presence of a certain number of latent classes or in diagnostic models in which the number of classes is known (typically the presence or absence of some disease). The current study was exploratory in nature, and the number of discontinuous heavy-drinking patterns was unknown, so a traditional LCA approach was used. It is worth remembering that all models, including latent class models, at best provide a good approximation of true processes that occur in the population. In this particular study, the eight latent classes provided a good summary of common developmental patterns of heavy drinking. Violation of the local independence assumption is a topic that deserves greater attention in future research.

Conclusions

The current study highlights the potential usefulness of a latent class approach to analyzing longitudinal data on alcohol use and other behaviors. This modeling approach is ideal for identifying subgroups characterized by different pathways through a stage-sequential process over three or more times. The results indicate that, although for those enrolled in college, heavy drinking does tend to increase during the college years, it does not exceed rates of heavy drinking exhibited by those not enrolled in college. In fact, at no developmental period examined in this study does the college-enrolled group exhibit a significantly greater rate of heavy drinking than the nonenrolled group. Nonenrolled individuals are associated with greater risk for pathways leading to adult heavy drinking. Together, these results suggest that high school students and college-age individuals who are not enrolled in college can benefit from universal, selective, and indicated prevention efforts aimed at reducing heavy drinking.

Acknowledgments

This work benefited from comments provided by our colleagues at The Methodology Center.

References

Akaike, H. Factor analysis and AIC. Psychometrika **52:** 317-332, 1987.

Arnett, J.J. Emerging adulthood: A theory of development from the late teens through the twenties. Amer. Psychol. **55:** 469-480, 2000.

Clogg, C.C. and Goodman, L.A. Latent structure analysis of a set of multidimensional contingency tables. J. Amer. Stat. Assoc. **79** (388): 762-771, 1984.

Collins, L.M., Lanza, S.T., Schafer, J.L., and Flaherty, B.P. WinLTA User's Guide, Version 3.0, University Park, PA: The Methodology Center, Penn State, 2002.

Collins, L.M. and Wugalter, S.E. Latent class models for stage-sequential dynamic latent variables. Multivar. Behav. Res. **27:** 131-157, 1992.

Dolan, C.V., Schmittmann, V.D., Lubke, G.H., and Neale, M.C. Regime switching in the latent growth curve mixture model. Struct. Equat. Model. **12:** 94-119, 2005.

Everitt, B.S. A Monte Carlo investigation of the likelihood ratio test for number of classes in latent class analysis. Multivar. Behav. Res. **23:** 531-538, 1988.

Goodman, L.A. Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika **61:** 215-231, 1974.

Johnston, L.D., O'Malley, P.M., and Bachman, J.G. National Survey Results on Drug Use from the Monitoring the Future Study, 1975-1994, Vol. 2, NIH Publication No. 96-4027, Washington: Government Printing Office, 1996.

Lanza, S.T. Latent Stage Sequence Analysis, Technical Report No. 03-55, University Park, PA: The Methodology Center, Penn State, 2003.

Lanza, S.T. and Collins, L.M. Pubertal timing and the onset of substance use in females during early adolescence. Prev. Sci. **3:** 69-82, 2002.

LANZA, S.T., COLLINS, L.M., SCHAFER, J.L., AND FLAHERTY, B.P. Using data augmentation to obtain standard errors and conduct hypothesis tests in latent class and latent transition analysis. Psychol. Meth. **10:** 84-100, 2005.

LANZA, S.T., FLAHERTY, B.P., AND COLLINS, L.M. Latent class and latent transition analysis. In: SCHINKA, J.A. AND VELICER, W.F. (Eds.) Handbook of Psychology, Vol. 2: Research Methods in Psychology, Hoboken, NJ: John Wiley & Sons, 2003, pp. 663-685.

LAZARSFELD, P.F. AND HENRY, N.W. Latent Structure Analysis, Boston, MA: Houghton Mifflin, 1968.

LITTLE, R.J.A. AND RUBIN, D.B. The analysis of social science data with missing values. Sociol. Meth. Res. **18:** 292-326, 1989.

MEREDITH, W. Measurement invariance, factor analysis and factorial invariance. Psychometrika **58:** 525-543, 1993.

MERLINE, A.C., O'MALLEY, P.M., SCHULENBERG, J.E., BACHMAN, J.G., AND JOHNSTON, L.D. Substance use among adults 35 years of age: Prevalence, adulthood predictors, and impact of adolescent substance use. Amer. J. Publ. Hlth **94:** 96-102, 2004.

MUTHÉN, B. Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class-latent growth modeling. In: COLLINS, L.M. AND SAYER, A.G. (Eds.) New Methods for the Analysis of Change, Washington, DC: American Psychological Assn, 2001, pp. 291-322.

MUTHÉN, B. AND SHEDDON, K. Finite mixture modeling with mixture outcomes using the EM algorithm. Biometrics **55:** 463-469, 1999.

NAGIN, D.S. Analyzing developmental trajectories: A semiparametric, group-based approach. Psychol. Meth. **4:** 139-157, 1999.

RUBIN, D.B. Multiple Imputation for Nonresponse in Surveys, New York: John Wiley & Sons, 1987.

RUBIN, D.B. AND STERN, H.S. Testing in latent class models using a posterior predictive check distribution. In: VON EYE, A. AND CLOGG, C.C. (Eds.) Latent Variables Analysis: Applications for Developmental Research, Thousand Oaks, CA: Sage, 1994, pp. 420-438.

SCHULENBERG, J., O'MALLEY, P.M., BACHMAN, J.G., WADSWORTH, K.N., AND JOHNSTON, L.D. Getting drunk and growing up: Trajectories of frequent binge drinking during the transition to young adulthood. J. Stud. Alcohol **57:** 289-304, 1996.

SCHWARZ, G. Estimating the dimension of a model. Ann. Stat. **6:** 461-464, 1978.

TUCKER, J.S., ORLANDO, M., AND ELLICKSON, P.L. Patterns and correlates of binge drinking trajectories from early adolescence to young adulthood. Hlth Psychol. **22:** 79-87, 2003.

YANG, I. AND BECKER, M.P. Latent variable modeling of diagnostic accuracy. Biometrics **53:** 948-958, 1997.