Latent Transition Analysis: Benefits of a Latent Variable Approach to Modeling Transitions in Substance Use

Stephanie T. Lanza[1], **Megan E. Patrick**[2], and **Jennifer L. Maggs**[3]

[1]The Methodology Center, The Pennsylvania State University

[2]Institute for Social Research, University of Michigan

[3]Department of Human Development and Family Studies, The Pennsylvania State University

Abstract

We apply latent transition analysis (LTA) to characterize transitions over time in substance use behavior profiles among first-year college students. Advantages of modeling substance use behavior as a categorical latent variable are demonstrated. Alcohol use (any drinking and binge drinking), cigarette use, and marijuana use were assessed in a sample (*N*=718) of college students during the fall and spring semesters. Four profiles of 14-day substance use behavior were identified: (1) Non-Users; (2) Cigarette Smokers; (3) Binge Drinkers; and (4) Bingers with Marijuana Use. The most prevalent behavior profile at both times was the Non-Users (with over half of the students having this profile), followed by Binge Drinkers and Bingers with Marijuana Use. Cigarette Smokers was the least prevalent behavior profile. Gender, race/ethnicity, early onset of alcohol use, grades in high school, membership in the honors program, and friendship goals were all significant predictors of substance use behavior profile.

Keywords

latent transition analysis; college students; substance use

The transition to college is associated with increases in heavy alcohol use (White et al., 2006) and marijuana use (Fromme, Corbin, & Kruse, 2008). Students who attend college engage in more binge drinking (i.e., consuming five or more alcoholic drinks in a row in the past 2 weeks) and have a higher prevalence of annual and 30-day alcohol use, but do not evidence elevated levels of cigarette, marijuana, or cocaine use compared to their same-age peers who do not attend college (O'Malley & Johnston, 2002). Approximately 40% of college students engage in binge drinking in a 14-day period (O'Malley & Johnston, 2002), and this behavior is associated with well-documented negative consequences (e.g., Hingson, Heeren, Winter, & Wechsler, 2005; Jackson, Sher, & Park, 2006). In addition, college students' tobacco use continues to be a concern, although smoking is less prevalent among college attenders than among non-attenders (Tercyzk, Rodriguesz, & Audrain-McGovern, 2007). Cannabis (i.e., marijuana) use among college students is also associated with substance use disorders and other negative use-related consequences (Caldeira, Arria, O'Grady, Vincent, & Wish, 2008). However, much less research has considered patterns of college students' use of multiple substances and the public health importance of the intersection of these behaviors. A better understanding of substance use behavior and negative consequences and their predictors among college students requires a more holistic

Corresponding author: Stephanie T. Lanza, Ph.D., The Methodology Center, The Pennsylvania State University, 204 E. Calder Way, Suite 400, State College, PA 16801, 814-865-7095 (pone), 814-863-0000 (fax), SLanza@psu.edu.

treatment of behavior, where use of multiple substances is considered simultaneously. The current study takes a person-centered approach to modeling behavior; we demonstrate the advantages of using latent transition analysis (LTA) to describe behavioral profiles characterized by profiles of alcohol use, binge drinking, cigarette use, and marijuana use across the first year of college. Transitions in substance use behavioral profiles between the fall and spring semesters are examined.

Considering predictors of profiles of substance use is important for understanding the phenomenon of substance use and the individuals who are at greatest risk. Important demographic predictors of substance use include gender and race/ethnicity. Men tend to use alcohol more frequently and in larger quantities than women (Wilsnack & Wilsnack, 2002). Men also exceed women in use of most illegal drugs including marijuana, although gender differences in smoking among adults are much smaller (Johnston, O'Malley, Bachman, & Schulenberg, 2008). The prevalence of substance use and alcohol use disorders in the population differs by race/ethnicity, such that Asian Americans and African Americans tend to engage in less drug use than European Americans and Hispanic Americans (Huang et al., 2006; Johnston et al., 2008). In addition, available studies have consistently demonstrated that earlier initiation of alcohol use is a risk for more alcohol use, abuse, and dependence in adulthood (Hawkins et al., 1997; Humphrey & Friedman, 1986; Labouvie, Bates, & Pandina, 1997; York, Welte, Hirsch, Barnes, & Hoffman, 2004), as well as earlier initiation of and dependence on other drugs (Agrawal et al., 2006; Clark, Cornelius, Kirisci, & Tarter, 2005; Hingson, Heeren, & Edwards, 2008). In addition, better academic performance is associated with lower rates of daily smoking, marijuana use, and heavy drinking (Bachman et al., 2008). Finally, having important social goals has been shown to predict greater planned drinking (Rhoades & Maggs, 2006). In addition to examining these covariates, a multivariate (i.e., multiple substances) and developmental (i.e., transitions over time in behavior) approach can provide a more nuanced portrait of substance use across this pivotal developmental period. LTA provides an ideal approach to model the phenomena of substance use in this important population of first-year college students.

Latent Transition Analysis (LTA)

The LTA approach demonstrated in this paper is a longitudinal extension of latent class analysis (LCA). LCA is a multivariate statistical model that is based on a measurement theory which posits that an underlying grouping variable (i.e., a latent class variable) is not observed but can be inferred from a set of categorical indicators (Goodman, 1974; Lazarsfeld & Henry, 1968). Often the latent class variable is used to organize multiple dimensions of behavior, such that individuals in each latent class share common behavior patterns. This measurement model lends itself well to the study of substance use behavior. For example, Cleveland, Collins, Lanza, Greenberg, and Feinberg (in press) found six latent classes of substance use among high school seniors: (1) Non-Users; (2) Alcohol Experimenters; (3) Alcohol, Tobacco, and Other Drug (ATOD) Experimenters; (4) Current Smokers; (5) Binge Drinkers; and (6) Heavy Users (characterized by recent alcohol, cigarettes, and marijuana use). Importantly, LCA can be extended to model longitudinal data, where transitions over time in latent class membership are estimated, in a model called latent transition analysis (LTA). In LCA, latent classes represent stable sets of characteristics or states of behavior. However, in LTA, individuals may change membership in latent classes over time. Thus, instead of using the term "latent classes," we will use "latent statuses" to refer to the behavior subgroups, to reflect the fact that subgroup membership is not assumed to be stable over time.

Three sets of parameters are estimated in LTA. First, **latent status membership probabilities** are estimated at $t > 1$ times. For example, when modeling substance use over

time as a categorical latent variable, these probabilities reflect the proportion of individuals expected to belong in each substance use latent status at each time period. Second, **transition probabilities** reflect the probability of transitioning from a particular latent status at time t to another latent status at time $t+1$. Together, these probabilities reflect the amount of change over time in the outcome. For example, a model of substance use over time might include the probability of membership in a Heavy Users latent status at Time 2 conditional on membership in a Binge Drinkers latent status at Time 1. Third, a set of **item-response probabilities** reflects the correspondence between the observed indicators of the latent variable at each time period and latent status membership, in much the same way that factor loadings link observed indicators to latent variables in factor analysis. For example, an item-response probability might reflect the probability of reporting marijuana use conditional on membership in a Polydrug Use latent status at Time 1. This set of probabilities provides information on how differentiated the latent statuses are, as well as on how to label each latent status. For example, a latent status characterized by very low probabilities of endorsing any substance use items might be labeled Non-Users, whereas a latent status characterized by a high probability of endorsing an item on binge drinking but not on cigarette smoking or illicit drug use might be labeled Binge Drinkers. Although only differentiated by one probability (that of endorsing the binge drinking item), these two latent statuses are clearly unique.

Predictors of latent status membership probabilities and transition probabilities can be incorporated directly in LTA using logistic regression, as can grouping variables (e.g., gender, race/ethnicity). Appendix A presents more technical detail on the mathematical model for LTA. A more complete introduction to LCA and LTA, including empirical demonstrations, can be found in Collins and Lanza (in press), Lanza et al. (2007), Lanza, Flaherty, and Collins (2003), and Lanza and Collins (2008).

In LTA, multiple aspects of behavior assessed across two occasions can be used to jointly indicate an individual's behavior status over time. Within each time period, behavior can be modeled as a multivariate phenomenon. For example, Lanza and Collins (2008) used LTA to model dating and sexual risk behavior over time, where the behavior was indicated by the following four items assessed at each time point: number of dating partners, past-year sexual intercourse, number of past-year sexual partners, and potential exposure to a sexually transmitted infection (i.e., sex without a condom at least one time in the past year). The five behavior statuses identified at each time point were: Non-Daters, Daters, Monogamous, Multi-Partner Safe, and Multi-Partner Exposed. Each of these latent statuses reflects a clearly interpretable and logical intersection of the various dimensions of dating and sexual behavior. In other words, at each time point an individual's behavior status could be characterized in terms of higher-order interactions across various aspects of behavior. In addition, a 5×5 matrix of transition probabilities from Time 1 to Time 2 provided a parsimonious summary of change over time in the behavior, including information about which dating and sexual risk behavior status at Time 1 corresponded to the highest likelihood of transitioning to the Multi-Partner Exposed latent status at Time 2.

LTA can also be applied to identify and describe classes of individuals with distinct characteristics, or profiles, of symptoms. Jackson, O'Neill, and Sher (2006), for instance, modeled the transitions in alcohol dependence among individuals from ages 24–32 years at Time 1 to ages 29–37 at Time 2. Three statuses with increasing severity were described: No Dependence, Mild Alcohol Dependence, and Severe Alcohol Dependence. Across time, marginal rates of dependence were largely stable, although a substantial proportion of individuals transitioned into and out of dependence statuses during this time period. Also using LTA, Chung and Martin (2005) identified the structure of diagnostic symptoms (from the DSM-IV) related to cannabis, hallucinogen, cocaine, and opiate disorders. In their

sample of adolescents referred for addiction treatment, the authors identified Few or No Symptoms, Mild, and Severe latent statuses. Over the year of treatment, adolescents in both inpatient and outpatient treatments were more likely to transition to a less severe status, although inpatient adolescents were more likely to have continuing risk for cannabis disorders, potentially due to the higher incidence of conduct disorder in this group.

In particular, LTA is an appropriate technique to model the stages of use of multiple substances over time (see Collins, 2002). For example, patterns of onset of substance use have been modeled in multiple populations including among Hispanic youth (alcohol, cigarettes, drunkenness, other illicit drugs in Maldonado-Molina et al., 2007) and adolescents in South Africa (alcohol, cigarettes, cannabis, and inhalants in Patrick et al., 2009). Among South African youth, alcohol was most commonly the first substance adolescents had tried, but eighth graders who had only ever used cigarettes in eighth grade were more likely to transition to more advanced statuses involving polydrug use by ninth grade (Patrick et al., 2009). Predictors of the dynamic onset process have also been demonstrated. Lanza and Collins (2002) modeled the relation between early pubertal timing and the stage-sequential process of substance use onset in females (based on indicators of alcohol use, drunkenness, cigarette use, and marijuana use). Eight stages of substance use were identified: (1) No Use; (2) Alcohol Use; (3) Cigarette Use; (4) Alcohol and Cigarette Use; (5) Cigarette and Marijuana Use; (6) Alcohol and Cigarette Use with Drunkenness; (7) Alcohol, Cigarette, and Marijuana Use; and (8) Alcohol, Cigarette, and Marijuana Use with Drunkenness. Early-maturing seventh grade girls (compared to on-time/late maturers) were more likely to be in the most advanced substance use status (Alcohol, Cigarette, and Marijuana Use with Drunkenness) in seventh grade and more likely to transition out of the No Use status between seventh and eighth grades.

In the current study, multivariate substance use behavior profiles and transitions in behavior across the first year of college are examined. The effects of gender, race/ethnicity, early onset of alcohol use, grades in high school, membership in the honors program during college, and friendship goals are explored as possible predictors of substance use behavior profiles and transitions over time in behavior. A comparison is made between using a latent variable approach (LTA) and a manifest variable approach to address the research questions. By modeling substance use as a categorical latent variable over time, profiles of substance use behavior and transitions over time can be summarized in a parsimonious way, allowing for the most important behavior profiles to emerge. The prediction of substance use behavior profile and transitions over time can inform future prevention and intervention efforts directed toward college students who possess particular characteristics or engage in particular profiles of substance use behavior at baseline. Such knowledge might be useful for informing adaptive interventions for college students (Collins, Murphy, & Bierman, 2004).

Method

University Life Study

The University Life Study utilized a longitudinal measurement burst design, with baseline surveys followed by 14 consecutive daily surveys in each of two semesters per academic year. The current analyses include data from Times 1 (Fall 2007) and 2 (Spring 2008). Given the complexity of substance use behaviors, innovative measurement strategies (such as daily reports) are needed to accurately document substance use and variation (e.g., Neal et al., 2006). Aggregate recall measures have been shown to under- or over-estimate alcohol use in comparison to daily reports (Poikolainen, Podkletnova, & Alho, 2002). Therefore, the current analysis utilizes 14 consecutive daily reports of three different substance use behaviors (i.e., alcohol use including binge drinking [4+ drinks for women, 5+ drinks for

men], cigarette smoking, and marijuana use). In this daily design, the behavioral recall window is dramatically reduced as is the subsequent potential for memory errors. Daily data were coded to reflect whether participants reported ever engaging in each of the four behaviors across each of the two 14-day time periods (we will refer to these two time periods as Time 1 and Time 2).

Participants

The current study included *N*=718 first-year college students (49% male, *M* age=18.5 years, *SD*=0.4) who provided data on at least one item measuring substance use at either time. A stratified random sampling procedure was used to achieve a diverse sample of first-year students with respect to gender and race/ethnicity. Eligible first-year students were U.S. citizens or permanent residents, under age 21, and residing within 25 miles of the campus. The students were mailed an informational letter that included a description of the study, a pen, and a $5 cash incentive. Five days later, an e-mail message was sent to each student with an active hyperlink to the Web-based baseline survey. After participants completed the baseline survey, an e-mail message was sent to them the following day inviting them to begin 14 consecutive short daily Web surveys. Once participants completed the first daily Web survey, an e-mail was sent each morning for 13 additional consecutive days, with a link to the daily Web-based survey. A $70 cash incentive was given to students who provided data for all 14 days. In total, 746 students (65.6% response rate) completed the Time 1 baseline survey. Completion rates of the daily surveys were high, with most (86%) of the participants completing at least 12 of the 14 daily surveys, giving a total of 9,482 days of daily data in Semester 1. The sample self-identified as 25.4% Hispanic American, 27.2% European American non-Hispanic (NH), 23.3% Asian American NH, 15.7% African American NH, and 8.4% multiracial NH

Measures

Substance Use—Table 1 presents descriptive statistics of the indicators of substance use during each time period (i.e., 14-days). Each day, participants were asked to answer questions regarding their substance use during the previous day. Alcohol use was assessed with the question, "How many drinks of alcohol did you drink?" using the definition, "By one drink we mean half an ounce of absolute alcohol, for example 12 ounce can or bottle of beer or cooler, 5 ounce glass of wine, or a drink containing 1 shot of liquor or spirits." At each time, responses were coded as 1 = *no drinking* during the 14-day period, 2 = *drinking but not binge drinking*, defined as consuming at least one alcoholic drink during the 14-day period without engaging in binge drinking on any of the days, and 3 = *binge drinking*, defined as consuming 4 or more drinks in a single day for women and 5 or more for men at least once during the 14-day period. Cigarette use was measured by the question, "How many cigarettes did you smoke on [the previous day of the week], if any?" At each time, responses were coded as 1 = *no cigarettes smoked* during the 14-day period and 2 = *one or more cigarettes smoked* during that time period. Finally, participants were asked, "Did you use any illegal drugs on [the previous day of the week]?" If the answer was "yes," participants were invited to check all substances they had used that day from a list containing: "marijuana, hashish, any kind of cocaine (including crack, freebase, or powder), methamphetamines (also called speed, crystal, crank, or ice), other types of illegal drugs (such as LSD, PCP, ecstasy, mushrooms, inhalants, heroin), prescription medicines not prescribed to you, or other illegal drugs." Marijuana use was coded as 1 = *no marijuana use* during that 14-day period and 2 = *any marijuana use* during that time period. All other illicit drug use was too rare to be included in the analysis.

Covariates (Measured at Time 1)—Table 1 provides descriptive statistics for all covariates used in the current study. Race/ethnicity had five categories: Hispanic American,

European American non-Hispanic (NH), Asian American NH, African American NH, and multiracial NH (i.e., reporting more than one race). Race/ethnicity was incorporated in the model as four binary variables, with the largest group, European American NH, as the reference category. Participants reported the importance of social goals, by rating, "Making friends is important to you" on a scale of 0 = *not at all* to 4 = *very important*. Early onset of alcohol use was coded as 1 for *first using alcohol during or before Grade 9* and 0 for *first using alcohol during grade 10 or later* (or has not yet tried alcohol). High school grades were obtained by asking participants, "What were your grades like in your senior year of high school?" Response options were 0 = *Mostly Fs*, 1 = *Mostly Ds*, 2 = *Mostly Cs*, 3 = *Mostly Bs*, and 4 = *Mostly As*. Higher high school grades were coded as 1 = *mostly receiving As* and 0 = *mostly Bs or below*. Honors program in the first semester of college was reported, and coded as 1 = *yes*, 0 = *no*. Friendship goals were coded as 1 for *very important* (ratings of 4), and 0 for *less important* (ratings of 0–3). In addition, grade at first alcohol use was measured with the question, "When if ever did you first try an alcoholic beverage – more than just a few sips?" on a scale of *Never*, *Grade 6 or below*, *Grade 7*, *Grade 8*, *Grade 9*, *Grade 10*, *Grade 11*, *Grade 12*, or *College*.

Analytic Strategy

Manifest Variables Analysis—Two different manifest variables analysis strategies were used to address our research questions. These strategies represent what we believe are the most logical and straightforward methods for analyzing data of this structure. One analysis involved examining the relation between each pair of substance use variables within a particular time period. This was done to determine the extent to which different aspects of behavior co-occurred within individuals. In the other manifest variables analysis, each individual was assigned to an observed behavior pattern at Time 1 and Time 2; there were 12 possible patterns at each time, corresponding to all possible responses to the three substance use items (e.g., "No Alcohol Use, No Cigarette Use, No Marijuana Use"; "Drinking without Binging, No Cigarette Use, No Marijuana Use"; "Binge Drinking, Cigarette Use, Marijuana Use"). The 12-level variable at Time 1 was crossed by the 12-level variable at Time 2 in order to describe how individuals change over time in their profile of substance use behavior. This approach provided a very detailed picture of specific behaviors individuals were engaging in at each time. However, this approach yielded 12×12=144 cells reflecting change over time in substance use behavior patterns. Predicting change over time was not feasible with this many observed patterns.

Latent Variable Analysis—Next, LTA was used to explore whether meaningful latent statuses of substance use could be identified at each measurement occasion. Models with different numbers of latent statuses were compared, and model selection was conducted based on the likelihood-ratio G^2 statistic, Akaike Information Criterion (AIC; Akaike, 1974), Bayesian Information Criterion (BIC; Schwarz, 1978), and interpretability of the latent statuses. Identification of each model under consideration was assessed by fitting the model to the data using multiple sets of random starting values.[1] Presumably, the resultant LTA model would include considerably fewer than 12 latent statuses (substance use behavior profiles) at each time, providing a more parsimonious description of multivariate behavior than was obtained using a manifest variables approach. This model yielded item-response probabilities that characterize the substance use behavior profiles at each time period, prevalence of each of the latent statuses at each time period, and a matrix of

---

[1]A necessary, but far from sufficient, criterion for model identification in LTA is that the degrees of freedom are greater than or equal to one. Because LTA typically is based on a very large contingency table of observed data, degrees of freedom can be positive in models for which the maximum-likelihood solution cannot be sufficiently identified. For this reason, we recommend fitting a particular model to the data using multiple sets of random starting values to see if the solution with the maximum likelihood value can be replicated. For a detailed discussion of this issue see Collins and Lanza (in press).

transition probabilities that describes how the students transition from Time 1 to Time 2 in substance use behavior profiles. Measurement invariance across time was assessed by comparing a model with item-response probabilities freely estimated at each time to a model where the item-response probabilities were constrained to be equal at both times.

Gender was incorporated as a grouping variable so that measurement invariance across men and women could be assessed. To do this, a model with item-response probabilities freely estimated within each gender was compared to a model where these probabilities were constrained to be equal across genders. A difference G^2 test was conducted to test the hypothesis that measurement invariance holds across groups. In addition, gender differences in the prevalence of substance use behavior profiles were assessed, providing information about how male and female first-year college students differed in their level of engagement in various substance use behaviors.

Finally, the following covariates were incorporated as predictors of substance use behavior profiles at Time 1 as well as predictors of transitions in behavior between Time 1 and Time 2: gender,[2] race/ethnicity, early onset of alcohol use, higher grades (mostly As) in high school, membership in the honors program during the first semester of college, and reporting that friendship goals are very important during the first semester of college. Each covariate was entered separately in the LTA model in order to estimate the overall relation between each variable and substance use behavior, although it is possible to include two or more covariates in LTA (Lanza & Collins, 2008). It is worth noting that LTA with covariates does not involve assigning individuals to a latent status at each time period. Rather, latent status membership is estimated and the effects of covariates are estimated simultaneously in a single model, appropriately allowing the uncertainty associated with latent status membership to be taken into account (Collins & Lanza, in press; Lanza & Collins, 2008).

All latent transition models were fit using PROC LTA (Lanza, Lemmon, Schafer, & Collins, 2008); this SAS[3] procedure and its corresponding user's guide are available for download at no cost at http://methodology.psu.edu/. Appendix B includes PROC LTA syntax used in the current study.

Results

Manifest Variables Analysis

First we examined the relation between each pair of substance use variables within a particular time period. This analysis produced six crosstabs (three at each time period), all of which showed clear associations ($p<.0001$). For example, the proportion of college students reporting marijuana use at Time 1 was 0.6% for those who did not report any drinking behavior, 6.7% for those reporting drinking without binging, and 16.6% for those reporting binging. These bivariate analyses are highly descriptive, but they provide no information about the co-occurrence of three or more behaviors within individuals. For example, while these results show an association between marijuana use and drinking at Time 1, they do not shed light on whether cigarette use is more common among individuals who use marijuana, engage in binge drinking, or both. In addition, these analyses cannot be used to describe how individuals are expected to change over time in substance use behavior.

---

[2]Gender was considered first as a grouping variable and then as a covariate for exposition purposes only. Assuming that measurement invariance across groups is plausible, incorporating a binary variable for gender in the model as a grouping variable (with item-response probabilities constrained to be equal across groups) is mathematically equivalent to incorporating a dummy-coded variable for gender as a predictor of the latent status prevalences at Time 1 and the transition probabilities. See Collins and Lanza (in press) for more information on the correspondence between multiple-groups LTA and LTA with covariates.

[3]Copyright 2002–2003 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

Next, we assigned each individual a particular behavior pattern at each time period based on the full set of observed items in order to capture co-occurrence of all behaviors. For example, individuals reporting no alcohol, cigarette, or marijuana use at Time 1 were assigned to Behavior Pattern 1 at that time; individuals reporting drinking but no binge drinking, cigarettes, or marijuana at Time 1 were assigned to Behavior Pattern 2 at that time; and so on. Even in this simple example, where just three behaviors are included, there are 12 possible patterns at each time (3 levels of alcohol use [*none*, *drink but no binge*, *binge drinking*] × 2 levels of cigarette use [*yes*, *no*] × 2 levels of marijuana use [*yes*, *no*]). Table 2 shows the number of individuals with each observed behavior pattern at each time. To assess behavior longitudinally, we crossed the college students' behavior patterns at Time 1 and Time 2. A test of independence could not be reported, however, because of extreme sparseness in the 12-by-12 contingency table.[4] In addition to computational limitations, there was limited potential to draw conclusions from these analyses to describe etiology of use or inform the development of prevention programs because of the large number of cells. A more parsimonious solution would lead to clearer and more specific conclusions or recommendations.

Latent Transition Model of Substance Use Behavior

A series of LTA models with two through six latent statuses of substance use were run, and for each model, identification was assessed. The maximum-likelihood solution for the six-status model could not be identified, suggesting that the model was too complex to be estimated given the data. Table 3 presents model fit information used in selecting the final model of substance use in the current study. The table includes the G^2 likelihood-ratio test statistic, the degrees of freedom, the AIC, and the BIC for models with two through five latent statuses. Note that p-values are not reported for the test statistics in these models because the degrees of freedom (df) are large (df≥99 for each model). Large models suffer from sparseness in the observed data table; when data are sparse it has been shown that the distribution of the G^2 does not follow a chi-square distribution (Koehler, 1986;Koehler & Larntz, 1980). Lower AIC and BIC values reflect an optimal balance between model fit and parsimony. Based on these fit criteria, we narrowed the model choice to the three-status and four-status models. We then compared the three-status and four-status models in terms of conceptual interpretability and chose the four-status model of substance use over time.

An important step in many longitudinal analyses is to consider whether the underlying structure of substance use behavior is the same across time. To assess this, we compared a model with parameter restrictions that constrain the item-response probabilities to be equal at Time 1 and Time 2 (G^2=92.5 with 112 df) to one with independent measurement of substance use at each time period (G^2=76.3 with 96 df). The difference G^2 of 16.2 can be compared to a chi-square table with degrees of freedom equal to 16 (the difference in df between the two models), yielding a p-value of .44. This non-significant p-value indicates that there is no evidence that the underlying structure of substance use behavior differs across time, allowing us to impose the same measurement model over time. This is important both conceptually, as it implies that the nature and meaning of the latent statuses is held constant over time, and for computational purposes because estimation will be more stable for the model based on 16 fewer parameters.

The full set of parameter estimates from the four-status model of substance use is presented in Table 4. The top panel shows the item-response probabilities for each item conditional on latent status membership. Note that these were constrained to be equal across the two times,

---

[4]Of the 12×12=144 possible longitudinal patterns of behavior, 83% had very small (less than five) expected cell counts and 47% (67 cells) had expected cell counts of exactly zero. In addition, Fisher's exact test could not be conducted due to the large amount of computer memory required.

so these probabilities are identical for Time 1 and Time 2. The item-response probabilities together provide a sense of what characterizes the four different substance use behavior profiles among first-year college students. The first latent status was labeled **Non-Users** because individuals in this status had a high probability of reporting no past 14-day alcohol use (.778), no cigarette use (.991), and no marijuana use (.993). The second latent status, labeled **Cigarette Smokers**, was characterized by a very high probability of reporting cigarette use (1.000) but not marijuana use (.042). Interestingly, there was substantial heterogeneity in alcohol use among individuals in this latent status, with approximately half reporting binge drinking. The third latent status was labeled **Binge Drinkers**; these individuals were characterized by a high probability of reporting binge drinking (.975) but a low probability of cigarette use (.059) and marijuana use (.048). Finally, the fourth latent status, labeled **Bingers with Marijuana Use**, was characterized by individuals with a high probability of reporting binge drinking (.878) and marijuana use (.758), although they were heterogeneous in terms of their cigarette use, with approximately half reporting use of this substance. It is interesting to note that the probability of drinking without binge drinking at each time was fairly low (between .046 and .303) for all four latent statuses. Binge drinking characterized the last two latent statuses, whereas cigarette use and marijuana use each strongly characterized just one latent status (the Cigarette Smokers and Bingers with Marijuana Use latent statuses, respectively).

The second panel of Table 4 provides the prevalence of each substance use behavior profile at Time 1 and Time 2. The modal behavior profile was Non-Users at both times; this profile comprised 58.1% of the students at Time 1 and 56.6% of them at Time 2. Prevalence rates of the other three behavior profiles also were fairly stable over time, although the prevalence of these three profiles varied widely. At Time 1, only 4.8% of the students were expected to be characterized as Cigarette Smokers, 29.0% Binge Drinkers, and 8.1% Bingers with Marijuana Use.

Because the prevalence of the four substance use behavior profiles was quite similar at Time 1 and Time 2, we might expect that individuals' behavior tended to be stable over time. The third panel of Table 4, which shows the transition probabilities, confirms this. These parameters reflect the probability of exhibiting a particular behavior profile at Time 2 conditional on Time 1 behavior. Diagonal elements reflect the proportion of individuals with the same behavior profile at both times. For example, Time 1 Non-Users had a probability of .895 of still being classified as a Non-User at Time 2. It is interesting to note that stability in behavior was highest among the Bingers with Marijuana Use; individuals in that latent status at Time 1 had a probability of .938 of remaining in that latent status at Time 2. The Cigarette Smokers at Time 1 were at most risk of advancing to the Bingers with Marijuana Use behavior profile at Time 2 (transition probability=.188).

Gender Differences

A test of measurement invariance across genders suggested that the underlying structure of substance use behavior was not different for male and female college students. This test was conducted by fitting a model with parameter restrictions that constrained the item-response probabilities to be equal across groups (G^2=173.1 with 241 df) to one with no restrictions across groups (G^2 of 151.1 with 225 df). The difference G^2=22.0 can be compared to a chi-square table with degrees of freedom equal to 16 (the difference in df between the two models), yielding a p-value of .14. This non-significant finding suggests that the set of parameter restrictions imposed across groups in the measurement of substance use were plausible (i.e., the meaning of the statuses could be considered to be the same for men and women).

Based on the model with equal measurement across groups, Time 1 gender differences in less advanced statuses of substance use behavior were fairly small, with the proportion of Non-Users equal to 60% for females and 57% for males, the proportion of smokers equal to 6% for females and 3% for males, and the proportion of Binge Drinkers equal to 30% for females and 25% for males. However, a large gender effect was found for the proportion of college students in the Bingers with Marijuana Use stage, which characterized just 4% of females compared to 15% of males (p<.0001). By Time 2, however, this gender difference was smaller because 9% of females were engaging in illicit behavior.

Predicting Substance Use Behavior

In order to determine whether individual characteristics were predictive of substance use behavior profiles at Time 1 or transitions in substance use from Time 1 to Time 2, six predictors were incorporated into the LTA model. Before proceeding with this step, however, a close inspection of Table 4 revealed that five probabilities corresponding to particular transitions were extremely small (≤.002): Non-Users at Time 1 to Bingers with Marijuana Use at Time 2, Cigarette Smokers to Binge Drinkers, Binge Drinkers to Cigarette Smokers, Bingers with Marijuana Use to Cigarette Smokers, and Bingers with Marijuana Use to Binge Drinkers. If a latent status membership probability or a transition probability is estimated to be very close to zero, logistic regression coefficients cannot be estimated and the model with covariates will fail. Therefore, before attempting to predict substance use behavior at Time 1 or transitions in behavior between Time 1 and Time 2, we imposed five additional parameter restrictions to fix these transition probabilities to be equal to zero. This practical solution essentially had no effect on the model (difference G^2=0.1 with 5 degrees of freedom, p>.99) other than to allow us to avoid estimation problems in the prediction of behavior over time.

Table 5 summarizes the findings for the following predictors of substance use behavior profiles at Time 1: gender, race/ethnicity, onset of alcohol use prior to or during grade 9, higher grades (mostly As) in high school, membership in the honors program during the first semester of college, and reporting that friendship goals are very important during the first semester of college. An overall test for the association between race/ethnicity categories and substance use behavior profiles was significant (change in log-likelihood=25.8, 12 df, p=. 01). All other predictors were significantly related to membership in the substance use behavior profiles (p≤.002 for each). The reference group for each multinomial logit model was specified to be the Non-Users, so each odds ratio was interpreted as the effect of the covariate on the odds of membership in a particular behavior profile relative to membership in the Non-Users behavior profile. An odds ratio of 1.0 suggested that individuals at all levels of the covariate had equal odds of belonging to that latent status relative to Users latent status. For binary predictors, an odds ratio greater than 1.0 suggested that having a value of 1 on the predictor placed individuals at increased odds of membership in that particular latent status relative to the Non-Users latent status, compared to individuals with a value of 0 on the predictor. Similarly, an odds ratio less than 1.0 suggested that having a value of 1 on the predictor placed individuals at decreased odds of membership in that particular latent status relative to the Non-Users latent status, compared to individuals with a value of 0 on the predictor.

As Table 5 shows, male college students were less likely than females to be Cigarette Smokers relative to Non-Users, but were 4.5 times more likely than females to be Bingers with Marijuana Use relative to Non-Users. African Americans were less likely than European Americans to be Cigarette Smokers, Binge Drinkers, or Bingers with Marijuana Use relative to Non-Users (*OR*=0.02, *OR*=0.3, *OR*=0.6, respectively). Similarly, Asian Americans were less likely than European Americans to be Binge Drinkers or Bingers with Marijuana Use relative to Non-Users (*OR*=0.3, *OR*=0.2, respectively). However, odds for

European Americans did not differ from those for Hispanic American or multiracial students. Early onset of alcohol use placed individuals at increased risk for membership in all substance use behavior profiles relative to Non-Users (*OR*=18.2 for Cigarette Smokers; *OR*=3.6 for Binge Drinkers; *OR*=9.1 for Bingers with Marijuana Use). Both higher grades in high school and participation in the honors program during the first semester of college were related to substantially decreased odds of membership in the Bingers with Marijuana Use profile (*OR*=0.3 and *OR*=0.02, respectively) relative to the Non-Users profile, although only participation in the honors program was related to decreased odds (*OR*=0.4) of membership in the Binge Drinkers profile relative to the Non-Users profile. Finally reporting that friendship goals during the first semester of college were very important was related to increased odds of membership in all three behavior profiles involving substance use, relative to membership in the Non-Users behavior profile (*OR*=2.1 for Cigarette Smokers; *OR*=3.3 for Binge Drinkers; *OR*=2.0 for Bingers with Marijuana Use).

Hypothesis tests were conducted to assess whether any of the six covariates (effects of gender, race/ethnicity, early onset of alcohol use, high grades in high school, membership in the honors program, and friendship goals during college) were significant predictors of transitions in substance use behavior profiles from Time 1 to Time 2. None of these tests reached statistical significance.

Discussion

The current study demonstrated the advantages of using LTA to model transitions over time in patterns of substance use behavior. This latent variable approach provided a parsimonious yet nuanced summary of the heterogeneity that exists among first-year college students in their engagement in alcohol use (including binge drinking), cigarette use, and marijuana use over a 14-day period. Even in a relatively simple multivariate model, where three measured variables combined to form just 12 possible observed profiles of use at each time period (see Table 2), predicting use at Time 1 and predicting transitions over time in substance use behavior proved insurmountable using a manifest variable approach. Estimating the relation between Time 1 and Time 2 in behavior was not possible using a contingency table method because of sparseness and the large number of cells (144) in the table. In addition, a data reduction strategy where two or more of the 12 observed profiles of use were collapsed would have been somewhat arbitrary, as each indicator reflected a different behavior; different scientists surely would combine the categories in different ways and such an approach would limit the ability to replicate results across studies. LTA is an effective way to organize information about use of multiple substances within individuals in a meaningful way in order to understand the "big picture." Rather than having to consider all 12 *possible* observed behavior profiles, LTA shed light on the *key existing* profiles of behavior that should be considered in this population.

The person-centered perspective may be useful in identifying higher risk profiles for individuals in need of targeted and adaptive intervention approaches, designed to tailor the program to groups of individuals with particular characteristics or behaviors (Collins et al., 2004). Further, understanding what predicts a more advanced substance use pattern (e.g., social goals) is paramount for designing prevention and intervention strategies that will be most salient to students. Being able to model multiple behavior changes in a single model is also of particular interest during developmental transitions—illustrated here by the first year of college—to represent a more holistic approach to modeling development (von Eye & Bergman, 2003). Other important transition points in individuals' development might include the transition from middle school to high school, from school to work, from singlehood to marriage, and from work to retirement. This approach is also relevant for studying other developmental phenomena, including anything from academic achievement

to health status to leisure pursuits. In addition, demographic predictors can be added to describe differences by, for example, gender and ethnicity. In this study, men were more likely than women to evidence a pattern of binge drinking plus marijuana use. African Americans and Asian Americans were less likely than European Americans to be involved in the most problematic patterns of use, although Hispanic Americans and multiracial students did not significantly differ from European Americans.

The substantive contributions of these findings are threefold. First, the prevalence of binge drinking was greater than the prevalence of moderate alcohol use among these underage college students. This pattern reflects the excessive nature of alcohol use in this population, beginning early in the first year, and the prominent party culture on many university campuses (e.g., Maggs, 1997). Second, the longitudinal analyses showed that students who used cigarettes in the fall of their freshman year were the least stable over time in their substance use. Previous research has also demonstrated the unique risk of cigarette smokers to transition more readily to the use of multiple substances, indicating that cigarette smokers may be a small but at-risk group (e.g., Graham, Collins, Wugalter, Chung, & Hansen, 1991; Newcomb & Bentler, 1986; Patrick et al., 2009). Finally, none of the predictors of the transition in substance use from Time 1 to Time 2 was statistically significant. This lack of prediction suggests that the effects of gender, race/ethnicity, early onset of alcohol use, high school grades, membership in the honors program, and friendship goals during college are established early in the college experience. However, other more powerful predictors of changes in use may include those that are more proximal to the ongoing college experience, including college peer group affiliations and activity involvement during the first and second semesters of college life.

The 14 consecutive daily reports of substance use resulted in improved measurement of recent substance use behavior. However, limitations of the current analyses are that we modeled transitions in "snapshots" of behavior aggregated across a 14-day period in each semester and that none of the predictors of change were significant. Predicting transitions in substance use may be more relevant in studies that employ a different design. For example, in studies that assess substance use behavior engaged in at any time during the fall and spring semesters, it would be possible to study the stability of behavior across semesters. Although the daily diary method used to assess recent substance use is a significant strength, we also recognize the implicit limitation that the particular 14-day period during each semester may or may not generalize to the rest of the semester.

Practical Considerations when Using LTA

Although several model selection tools are available in LTA, model selection is best conducted when a great deal of attention is paid to model interpretability. In the current study, if we had gone to five latent classes, the substance use profiles would more closely resemble the observed profiles—in other words, heterogeneity within classes is reduced as classes are added. The five-status model included the four latent statuses reported here, plus an additional (rare) behavior profile involving both binge drinking and cigarette use. However, with that increasing detail comes a loss of parsimony and of generalizability of the model.

Once a model has been selected, naming the latent statuses in LTA is very important. In all applied examples, the names of the latent statuses strongly convey a meaning in all of the results. Unlike in factor analysis, where the latent constructs that must be labeled are of a single dimension, in LTA labels must be assigned to actual subgroups (or types) of individuals based on multiple dimensions. We find that it is worth taking great care to establish labels that both characterize each latent status and help the reader to draw distinctions between the latent statuses.

More research on statistical power in LTA is warranted so that scientists can design studies with sufficient power for identifying and predicting the latent statuses. In addition, the current study provided statistical tests for the association between each covariate and Time 1 latent status membership, although confidence intervals for individual odds ratios are not yet available in PROC LTA.

When LTA is a Good Method to Consider

Given the great popularity of growth curve models (e.g., Bollen & Curran, 2006; Raudenbush & Bryk, 2002), modeling and predicting change over time typically is thought of in terms of mean-level change. The concept of change over time takes on a different meaning in LTA, however, where transition probabilities between qualitatively different states over time summarize behavior change in the population under study. In our example, change over time was characterized by transitions in unique profiles of substance use behavior. It is only appropriate to use LTA to address research questions that conceptually map onto discrete change over time, and it is ideal when the observed data are somewhat categorical in nature. In the current study, we were interested in identifying groups of first-year college students who shared common behavior patterns. Patterns were characterized by use (or no use) of various substances; thus, the indicators of substance use latent statuses were by nature categorical. Modeling latent statuses based on the level of substance use (on continuous scales) would require a different statistical model (e.g., latent profile analysis), and the statuses would be interpreted in terms of their means on each continuous indicator. Of course, many variables involved in studying substance use behavior can have extremely skewed distributions, so caution must be used when treating the variables as continuous. As with any longitudinal analysis, matching the research questions to an appropriate statistical method is paramount.

In sum, LTA is ideally suited for modeling multivariate constructs developmentally. In the current study, this approach allowed us to assess multivariate behavior profiles in the population, and to model multiple substance transitions together. Findings such as these can help to determine which groups of individuals, defined by their profiles of use, are most at-risk for onset of more advanced substance use behavior over time. In addition, predictors of profiles of use and transitions in use can provide important information about individual characteristics that place individuals at risk for poor substance use outcomes. Such information can inform future prevention and intervention efforts, allowing for resources to be targeted to individuals who possess those characteristics or engage in particular behaviors.

Biographies

Stephanie T. Lanza, Ph.D., is Scientific Director of The Methodology Center at Penn State. She currently leads a research team focused on advancing latent class analysis and related methods. Her latest book, with Linda Collins, Ph.D., is Latent Class and Latent Transition Analysis for the Social, Behavioral, and Health Sciences (in press) Wiley.

Megan E. Patrick, Ph.D., is a Faculty Research Fellow at the University of Michigan's Institute for Social Research. Her research surrounds the development of substance use and sexual behaviors during adolescence and the transition to adulthood and the role of motivations in prevention and intervention programs to promote health.

Jennifer Maggs, Ph.D., is a developmental psychologist whose research focuses on motivations for and consequences of alcohol use and other risk behaviors in adolescence and the transition to adulthood. Her projects use varied developmental designs, including daily measurement bursts and long-term longitudinal studies.

Acknowledgments

References

Agrawal A, Grant JD, Waldron M, Duncan AE, Scherrer JF, Lynskey MT, Madden PAF, Bucholz KK, Heath AC. Risk for initiation of substance use as a function of age of onset of cigarette, alcohol and cannabis use: Findings in a Midwestern female twin cohort. Preventive Medicine 2006;43:125–128. [PubMed: 16697036]

Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control 1974;19:716–723.

Bachman, JG.; O'Malley, PM.; Schulenberg, JE.; Johnston, LD.; Freedman-Doan, P.; Messersmith, E. The education-drug use connection: How successes and failures in school relate to adolescent smoking, drinking, drug use, and delinquency. New York: Lawrence Erlbaum Associates; 2008.

Bollen, KA.; Curran, PJ. Latent curve models: A structural equation perspective. Hoboken, NJ: Wiley; 2006.

Caldeira KM, Arria AM, O'Grady KE, Vincent KB, Wish ED. The occurrence of cannabis use disorders and other cannabis-related problems among first-year college students. Addictive Behaviors 2008;33:397–411. [PubMed: 18031940]

Chung T, Martin CS. Classification and short-term course of DSM-IV cannabis, hallucinogen, cocaine, and opioid disorders in treated adolescents. Journal of Consulting and Clinical Psychology 2005;73:995–1004. [PubMed: 16392973]

Clark DB, Cornelius JR, Kirisci L, Tarter RE. Childhood risk categories for adolescent substance involvement: A general liability typology. Drug and Alcohol Dependence 2005;77:13–21. [PubMed: 15607837]

Cleveland MJ, Collins LM, Lanza ST, Greenberg MT, Feinberg ME. In press Does individual risk moderate the effect of contextual-level protective factors? A latent class analysis of substance use. Journal of Prevention and Intervention in the Community.

Collins, LM. Using latent transition analysis to examine the Gateway Hypothesis. In: Kandel, DB., editor. Stages and pathways of drug involvement: Examining the Gateway Hypothesis. New York: Cambridge University Press; 2002. p. 254-269.

Collins, LM.; Lanza, ST. Latent class and latent transition analysis for applications in the social, behavioral, and health sciences. Hoboken, NJ: Wiley; In press

Collins LM, Murphy SA, Bierman KL. A conceptual framework for adaptive preventive interventions. Prevention Science 2004;5:185–196. [PubMed: 15470938]

Fromme K, Corbin WR, Kruse MI. Behavioral risks during the transition from high school to college. Developmental Psychology 2008;44:1497–1504. [PubMed: 18793080]

Goodman LA. Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika 1974;61:215–231.

Graham JW, Collins LM, Wugalter SE, Chung NK, Hansen WB. Modeling transitions in latent stage-sequential processes: A substance use prevention example. Journal of Consulting and Clinical Psychology 1991;59:48–57. [PubMed: 2002142]

Hawkins JD, Graham JW, Maguin E, Abbott R, Hill KG, Catalano RF. Exploring the effects of age of alcohol use initiation and psychosocial risk factors on subsequent alcohol misuse. Journal of Studies on Alcohol 1997;58:280–290. [PubMed: 9130220]

Hingson RW, Heeren T, Edwards EM. Age at drinking onset, alcohol dependence, and their relation to drug use and dependence, driving under the influence of drugs, and motor-vehicle crash involvement because of drugs. Journal of Studies on Alcohol and Drugs 2008;69:192–201. [PubMed: 18299759]

Hingson R, Heeren T, Winter M, Wechsler H. Magnitude of alcohol-related mortality and morbidity among U.S. college students ages 18–24: Changes from 1998 to 2001. Annual Review of Public Health 2005;26:259–279.

Huang B, Grant BF, Dawson DA, Stinson FS, Chou SP, Saha TD, Goldstein RB, Smith SM, Ruan WJ, Pickering RP. Race-ethnicity and the prevalence and co-occurrence of *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*, alcohol and drug use disorders and Axis I and II disorders: United States, 2001 to 2002. Comprehensive Psychiatry 2006;47:252–257. [PubMed: 16769298]

Humphrey JA, Friedman J. The onset of drinking and intoxication among university students. Journal of Studies on Alcohol 1986;47:455–458. [PubMed: 3795959]

Jackson KM, O'Neill SE, Sher KJ. Characterizing alcohol dependence: Transitions during young and middle adulthood. Experimental and Clinical Psychopharmacology 2006;14:228–244. [PubMed: 16756427]

Jackson, KM.; Sher, KJ.; Park, A. Drinking among college students: Consumption and consequences. In: Galanter, M., editor. Alcohol problems in adolescents and young adults: Epidemiology, neurobiology, prevention, and treatment. New York: Springer; 2006.

Johnston, LD.; O'Malley, PM.; Bachman, JG.; Schulenberg, JE. Monitoring the Future national survey results on drug use, 1975–2007: Volume II, College students and adults ages 19–45. Bethesda, MD: National Institute on Drug Abuse; 2008. NIH Publication No. 08-6418B

Koehler KJ. Goodness-of-fit tests for loglinear models in sparse contingency tables. Journal of the American Statistical Association 1986;81:483–493.

Koehler KJ, Larntz K. An empirical investigation of goodness-of-fit statistics for sparse multinomials. Journal of the American Statistical Association 1980;75:336–344.

Labouvie E, Bates KE, Pandina RJ. Age of first use: Its reliability and predictive utility. Journal of Studies on Alcohol 1997;58:638–643. [PubMed: 9391924]

Lanza ST, Collins LM. Pubertal timing in the onset of substance use in females during adolescence. Prevention Science 2002;3:69–82. [PubMed: 12002560]

Lanza ST, Collins LM. A new SAS procedure for latent transition analysis: Transitions in dating and sexual risk behavior. Developmental Psychology 2008;44(2):446–456. [PubMed: 18331135]

Lanza ST, Collins LM, Lemmon DR, Schafer JL. PROC LCA: A SAS procedure for latent class analysis. Structural Equation Modeling 2007;14(4):671–694. [PubMed: 19953201]

Lanza, ST.; Flaherty, BP.; Collins, LM. Latent class and latent transition analysis. In: Schinka, JA.; Velicer, WF., editors. Handbook of psychology: Research methods in psychology. Hoboken, NJ: Wiley; 2003. p. 663-685.

Lanza, ST.; Lemmon, DR.; Schafer, JL.; Collins, LM. PROC LCA & PROC LTA User's Guide Version 1.1.5 beta. University Park: The Methodology Center, Penn State; 2008.

Lazarsfeld, PF.; Henry, NW. Latent structure analysis. Boston, MA: Houghton Mifflin; 1968.

Maggs, JL. Alcohol use and binge drinking as goal-directed action during the transition to postsecondary education. In: Schulenberg, JE.; Maggs, JL.; Hurrelmann, K., editors. Health risks and developmental transitions during adolescence. New York: Cambridge University Press; 1997. p. 289-304.

Maldonado-Molina MM, Collins LM, Lanza ST, Prado G, Ramirez R, Canino G. Patterns of substance use onset among Hispanics in Puerto Rico and the United States. Addictive Behaviors 2007;32:2432–2437. [PubMed: 17493764]

Neal DJ, Fromme K, Del Boca FK, Parks KA, King LP, Pardi AM, Collins RL, Muraven M, Vetter C, Corbin WR. Capturing the moment: Innovative approaches to daily alcohol assessment. Alcoholism: Clinical and Experimental Research 2006;20:282–291.

Newcomb MD, Bentler PM. Frequency and sequence of drug use: A longitudinal study from early adolescence to young adulthood. Journal of Drug Education 1986;16:101–120. [PubMed: 3735021]

O'Malley PM, Johnston LD. Epidemiology of alcohol and other drug use among American college students. Journal of Studies on Alcohol 2002 Suppl. 14:23–39.

Patrick ME, Collins LM, Smiths E, Caldwell L, Flisher A, Wegner L. A prospective longitudinal model of substance use onset among South African adolescents. Substance Use & Misuse 2009;44:647–662. [PubMed: 19360538]

Poikolainen K, Podkletnova I, Alho H. Accuracy of quantity-frequency and graduated frequency questionnaires in measuring alcohol intake: Comparison with daily diary and commonly use laboratory markers. Alcohol and Alcoholism 2002;37:573–576. [PubMed: 12414549]

Raudenbush, SW.; Bryk, AS. Hierarchical linear models: Applications and data analysis methods. 2nd Ed.. Thousand Oaks, CA: Sage; 2002.

Rhoades BL, Maggs JL. Do academic and social goals predict planned alcohol use among college-bound high school graduates? Journal of Youth and Adolescence 2006;35:913–923.

Schwarz G. Estimating the dimension of a model. Annals of Statistics 1978;6:461–464.

Terczyak KP, Rodriguez D, Audrain-McGovern J. High school seniors' smoking initiation and progression 1 year after graduation. American Journal of Public Health 2007;97:1397–1398. [PubMed: 17600259]

von Eye A, Bergman LR. Research strategies in developmental psychopathology: Dimensional identity and the person-oriented approach. Development and Psychopathology 2003;15:553–580. [PubMed: 14582932]

White HR, McMorris BJ, Catalano RF, Fleming CB, Haggerty KP, Abbott RD. Increases in alcohol and marijuana use during the transition out of high school into emerging adulthood: The effects of leaving home, going to college, and high school protective factors. Journal of Studies on Alcohol 2006;67:810–822. [PubMed: 17060997]

Wilsnack SC, Wilsnack RW. International gender and alcohol research: Recent findings and future directions. Alcohol Research & Health 2002;26:245–250. [PubMed: 12875033]

York JL, Welte J, Hirsch J, Barnes G, Hoffman JH. Association of age at first drink with current alcohol drinking variables in a national general population sample. Alcoholism: Clinical and Experimental Research 2004;28:1379–1387.

Appendix

Appendix A. The LTA Mathematical Model

Fundamental Equation for LTA with Multiple Groups and Covariates

Suppose a latent transition model with n_s latent statuses is to be estimated based on a data set including M categorical items measured at each of T times for a total of MT items, a covariate X, and a grouping variable G. Let $Y_i = (Y_{i11}, Y_{i12}, \ldots, Y_{i1M}, Y_{i21}, Y_{i22}, \ldots, Y_{i2M}, Y_{iT1}, Y_{iT2}, \ldots, Y_{iTM})$ represent the vector of individual i's responses for all times $t=1, \ldots, T$ and items $m=1, \ldots, M$, where an individual response Y_{itm} may take on the values 1, 2, …, r_m. Let $s_{1i}=1, 2, \ldots, n_s$ be individual i's latent status membership at Time 1, $s_{2i}=1, 2, \ldots, n_s$ be individual i's latent status membership at Time 2, and so on. Let $I(y = k)$ be the indicator function which equals 1 if response y equals k and 0 otherwise. Suppose also that G_i represents the value of individual i's group membership, X_i represents the value of the covariate X for individual i and that the value of X can relate to the probability of membership in each latent status, δ, and each transition probability, τ. Then the latent transition model can be expressed as:

$$P(Y_i=y|X_i=x, G_i=g)=\sum_{s_1=1}^{n_s} \cdots \sum_{s_t=1}^{n_s} \delta_{s_1|g}(x)\tau_{s_2|s_1,g}(x) \cdots \tau_{s_t|s_{t-1},g}(x)\prod_{m=1}^{M}\prod_{k=1}^{r_m}\prod_{t=1}^{T}\rho_{mk|s_t,g}^{I(y_m=k)}$$

(Lanza & Collins, 2008).

Latent Statuses and Transition Probabilities

$\delta_{s_1|g}(x) = P(S_{1i} = s_1 | X_i = x, G_i = g)$ is a standard baseline-category multinomial logistic model predicting individual i's membership in latent status s_1 at Time 1. For example, with one covariate X the δ parameters are expressed as a function of the β parameters (i.e., the multinomial logistic regression estimates) and X:

$$\delta_{s_1|g}(x) = P(S_{1i} = s_1 | X_i = x, G_i = g) = \frac{\exp\{\beta_{0s_1|g} + x\beta_{1s_1|g}\}}{1 + \sum_{j=1}^{n_s-1} \exp\{\beta_{0j|g} + x\beta_{1j|g}\}}$$

for $s_1 = 1, \ldots, n_s - 1$ with latent status n_s as the reference status in the logistic regression. This enables estimation of the log-odds that an individual falls in latent status s_1 relative to reference status n_s. For example, if latent status 2 is the reference status, the log-odds of membership in latent status 1 relative to latent status 2 for an individual in group 1 with value x on the covariate is:

$$\log\left(\frac{\delta_{1|1}(x)}{\delta_{2|1}(x)}\right) = \beta_{01|1} + \beta_{11|1} x.$$

Exponentiated β parameters are odds ratios. For example, $e^{\beta_{11|1}}$ is an odds ratio reflecting the increase in odds of membership in latent status 1 (relative to reference status n_s) corresponding to a one-unit increase in the covariate, among individuals in group 1.

Similarly, $\tau_{s_2|s_1,g}(x) = P(S_{2i} = s_2 | S_{1i} = s_1, X_i = x, G_i = g)$ is a baseline-category multinomial logistic model estimating the probability of individual i's move to latent status s_2 conditional on current membership in latent status s_1. For example, the probability of individual i transitioning from latent status s_1 at Time 1 to latent status s_2 at Time 2 given membership in group g and covariate value x is:

$$\tau_{s_2|s_1,g}(x) = P(S_{2i} = s_2 | S_{1i} = s_1, X_i = x, G_i = g) = \frac{\exp\{\beta_{0s_2|s_1,g} + x\beta_{1s_2|s_1,g}\}}{1 + \sum_{j=1}^{n_s-1} \exp\{\beta_{0j|s_1,g} + x\beta_{1j|s_1,g}\}}$$

For $s_2 = 1, \ldots, n_s$. (Here latent status n_s is serving as the reference status.) Note that more than one covariate can be included, and different covariates can be specified for δ and for each τ matrix (i.e., Time 1 to Time 2, Time 2 to Time 3, etc.) (Lanza & Collins, 2008).

Appendix

Appendix B. Example Syntax to Fit LTA Models

*Final four-status model of substance use over time (measurement invariance across times);

    PROC LTA DATA=caldar;

        NSTATUS 4;

        NTIMES 2;

        ITEMS alc1 d1cig d1mj

```
                    alc2 d2cig d2mj;

            CATEGORIES 3 2 2;

            MEASUREMENT times;

            SEED 108;

        RUN;

*Model with gender as grouping variable (measurement invariance across gender and
times);

    PROC LTA DATA=caldar;

        NSTATUS 4;

        NTIMES 2;

        ITEMS alc1 d1cig d1mj

                alc2 d2cig d2mj;

        CATEGORIES 3 2 2;

        GROUPS s1gender;

        GROUPNAMES female male;

        MEASUREMENT times groups;

        SEED 108;

    run;

*Predictors of time 1 substance use status and transitions over time;

    %MACRO LTA_COV2(cov);

    PROC LTA DATA=caldar START=start_fixtau RESTRICT=restrict_fixtau;

        NSTATUS 4;

        NTIMES 2;

        ITEMS alc1 d1cig d1mj

                alc2 d2cig d2mj;

        CATEGORIES 3 2 2;

        COVARIATES1 &cov;

        REFERENCE1 2;

        COVARIATES2 &cov;

        REFERENCE2 1 2 3 4;

        MEASUREMENT times;

        STABILIZE;

    RUN;

    %MEND;

    %LTA_COV2(male);

    %LTA_COV1(black hisp asian multi);
```

```
%LTA_COV2(earlyalc);
%LTA_COV2(lowgrade);
%LTA_COV2(s1honors);
%LTA_COV2(fr_imp);
```

Table 1

Descriptive statistics of all variables used in latent transition analysis (*N*=718)

I. Indicators of latent status	Code	Label	Time 1 Frequency (Valid %)	Time 2 Frequency (Valid %)
14-day alcohol use	1	None	324 (45.8)	289 (46.5)
	2	Drink	106 (15.0)	81 (13.0)
	3	Binge	278 (39.3)	252 (40.5)
	.	Missing	10	96
14-day cigarette use	1	No	623 (88.2)	542 (87.1)
	2	Yes	83 (11.8)	80 (12.9)
	.	Missing	12	96
14-day marijuana use	1	No	654 (92.4)	557 (89.4)
	2	Yes	54 (7.6)	66 (10.6)
	.	Missing	10	95

II. Grouping variable	Code	Label	Frequency (Valid %)
Gender	1	Female	367 (51.1)
	2	Male	351 (49.9)

III. Covariates	Code	Label	Frequency (Valid %)
Race/ethnicity	1	European American	195 (27.2)
	2	African American	113 (15.8)
	3	Hispanic American	182 (25.4)
	4	Asian American	167 (23.3)
	5	Multiracial	60 (8.4)
	.	Missing	1
Early onset of alcohol use	0	No	538 (75.0)
	1	Yes (by grade 9)	179 (25.0)
	.	Missing	1
Higher grades in high school	0	No	192 (26.9)
	1	Yes (mostly As)	523 (73.1)
	.	Missing	3
Honors in first semester college	0	No	665 (92.9)
	1	Yes	51 (7.1)
	.	Missing	2
Friendship goals	0	Less important	258 (36.1)
	1	Very important	457 (63.9)
	.	Missing	3

Table 2

Frequency of observed behavior patterns at each time

Behavior Pattern	Alcohol Use	Cigarette Use	Marijuana Use	Time 1 Frequency	Time 2 Frequency
1	No	No	No	313	271
2	Drink	No	No	89	67
3	Binge	No	No	193	166
4	No	Yes	No	8	8
5	Drink	Yes	No	11	12
6	Binge	Yes	No	38	32
7	No	No	Yes	2	6
8	Drink	No	Yes	3	1
9	Binge	No	Yes	23	31
10	No	Yes	Yes	0	4
11	Drink	Yes	Yes	3	1
12	Binge	Yes	Yes	23	23

Table 3

Model fit information used in selecting the LTA model

Number of Latent Statuses	Likelihood-Ratio G^2	Degrees of Freedom	AIC	BIC
2	243.6	132	265.6	315.9
3	138.6	123	178.6	270.1
4	**92.5**	**112**	**154.5**	**296.3**
5	76.67	99	164.7	366.0

Note: Bold entries reflect selected model.

Table 4

Four-status model of recent substance use for first-year college students

	Latent Status			
	Non-Users	Cigarette Smokers	Binge Drinkers	Bingers with Marijuana Use
Probability of Response to Indicator of Substance Use[1]:				
14-day alcohol use				
None	.778	.157	.000	.075
Drink	.149	.303	.125	.046
Binge	.073	.540	.875	.878
14-day cigarette use				
No	.991	.000	.941	.466
Yes	.009	1.000	.059	.534
14-day marijuana use				
No	.993	.958	.953	.242
Yes	.007	.042	.048	.758
Prevalence of Latent Status:				
Time 1 (fall semester)	58.1%	4.8%	29.0%	8.1%
Time 2 (spring semester)	56.6%	5.5%	27.6%	10.3%
Transition probabilities (Rows for Time 1, Columns for Time 2):				
Non-Users	**.895**	.042	.062	.002[2]
Cigarette Smokers	.165	**.645**	.002[2]	.188
Binge Drinkers	.115	.000[2]	**.827**	.058
Bingers with Marijuana Use	.062	.000[2]	.000[2]	**.938**

Note. Transition probabilities in bold font correspond to membership in the same latent status at both times.

[1] Item-response probabilities constrained to be equal at time 1 and time 2.

[2] Transition probabilities fixed to exactly .000 for subsequent analyses involving covariates.

Table 5

Odds ratios reflecting effects of predictors on time 1 latent statuses of recent substance use behavior among first-year college students

		Latent Status			
Predictor	p-value[1]	Non-Users[2]	Cigarette Smokers	Binge Drinkers	Bingers with Marijuana Use
Male	.0002	1.0	0.6	0.9	4.5
African American	<.0001	1.0	0.02	0.3	0.6
Hispanic American	.53	1.0	1.5	0.9	1.5
Asian American	<.0001	1.0	0.9	0.3	0.2
Multiracial	.49	1.0	2.2	0.7	0.9
Early onset of alcohol use	<.0001	1.0	18.2	3.6	9.1
High grades in high school	.002	1.0	0.5	1.0	0.3
Honors in first semester college	.005	1.0	0.9	0.4	0.02
Very important friendship goals	<.0001	1.0	2.1	3.3	2.0

[1] p-value reflects 3 degrees of freedom test of overall relation between covariate and latent status membership at time 1.

[2] Non-Users served as reference group in multinomial logistic regressions; membership in all other latent statuses is compared to membership in this status.

Note. Overall test of significance of race/ethnicity categories is significant (change in log-likelihood=25.8, 12 df, p=.01).