



Published in final edited form as:

Dev Psychol. 2008 March ; 44(2): 446–456. doi:10.1037/0012-1649.44.2.446.

A New SAS Procedure for Latent Transition Analysis: Transitions in Dating and Sexual Risk Behavior

Stephanie T. Lanza, Ph.D. and

The Methodology Center, The Pennsylvania State University

Linda M. Collins, Ph.D.

Department of Human Development and Family Studies and The Methodology Center, The Pennsylvania State University

Abstract

The set of statistical methods available to developmentalists is continually being expanded, allowing for questions about change over time to be addressed in new, informative ways. Indeed, new developments in methods to model change over time create the possibility for new research questions to be posed. Latent transition analysis, a longitudinal extension of latent class analysis, is a method that can be used to model development in discrete latent variables, for example, stage processes, over two or more times. The current article illustrates this approach using a new SAS procedure, PROC LTA, to model change over time in adolescent and young adult dating and sexual risk behavior. Gender differences are examined, and substance use behaviors are included as predictors of initial status in dating and sexual risk behavior and transitions over time.

Many constructs in psychology, such as temperament and parenting style, suggest that individuals can be classified into groups based on some underlying characteristic. Often this dimension is difficult to observe directly, and instead is indicated by several variables. When an individual's classification can change over time, development occurs in a stage-sequential fashion. Latent class theory (Goodman, 1974; Lazarsfeld & Henry, 1968) provides a framework for measuring categorical latent variables, describing stage-sequential development, and predicting initial status and transitions over time.

Latent class analysis (LCA) is a latent variable model that is used to identify underlying (unobserved) subgroups in a population. The model posits that each individual belongs to one of a set of mutually exclusive and exhaustive latent classes. This framework is a categorical analog to the factor model, which is used to measure continuous latent variables. Like the factor model, LCA estimates and removes measurement error. In traditional LCA, two sets of parameters are estimated: class membership probabilities, which are analogous to factor scores, and item-response probabilities conditional on class membership, which are analogous to factor loadings. LCA has been used increasingly often in the social and behavioral sciences. For example, latent subgroups have been modeled for temperament (Stern, Arcus, Kagan, Rubin,

Correspondence concerning this article should be addressed to Stephanie T. Lanza, The Methodology Center, The Pennsylvania State University, 204 E. Calder Way, Suite 400, State College, Pennsylvania 16801 (phone: 814-865-7095). Electronic mail may be sent via Internet to SLanza@psu.edu.

Publisher's Disclaimer: The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at www.apa.org/pubs/journals/dev

& Snidman, 1995), depression (Lanza, Flaherty, & Collins, 2003), teaching style (Aitkin, Anderson, & Hinde, 1981), poverty (Dewilde, 2004), and multidimensional alcohol use (Auerbach & Collins, 2006; Jackson, Sher, Gotham, & Wood, 2001; Lanza, Collins, Lemmon, & Schafer, in press).

The traditional LCA model has been extended in several useful ways. For example, multiple-groups LCA allows for an exploration of invariance of item-response parameters across groups (in a latent class analog to examining measurement invariance), and latent class membership probabilities can be estimated for each group (Clogg & Goodman, 1984). Latent class regression analysis (LCA with covariates) includes predictors of latent class membership (Bandein-Roche, Miglioretti, Zeger, & Rathouz, 1997; Dayton & Macready, 1988).

An important longitudinal extension of LCA called latent transition analysis (LTA) allows latent class membership to change over time; in this model, change is quantified in a matrix of transition probabilities between two consecutive times. Often, developmental theory can suggest models of stage-sequential development that can be tested using LTA. For example, the Transtheoretical Model of health behavior change posits that individuals move through five discrete stages of behavior change: precontemplation, contemplation, preparation, action and maintenance (Prochaska & Velicer, 1997).

We will reserve the term 'latent classes' for subgroups in which individuals do not change membership over time, and use the term 'latent statuses' to denote subgroups in which individuals' membership can change over time. The most basic LTA model involves three sets of parameters: latent status membership probabilities at Time 1, transition probabilities between latent statuses over time, and item-response probabilities conditional on latent status membership and time. LTA models with covariates involve an additional set of parameters that are multinomial logistic regression coefficients linking predictors to latent status membership and transitions over time (Chung, Park, & Lanza, 2005). A grouping variable may be introduced, allowing parameters to be estimated conditional on group membership. LTA has been used to examine change over time in a variety of behavioral outcomes, including substance use onset (Guo, Collins, Hill, & Hawkins, 2000; Lanza & Collins, 2002), smoking cessation (Velicer, Martin, & Collins, 1996), and alcohol problems among adolescents (Chung & Martin, 2001).

The LCA and LTA approaches have several advantages. First, they can be a good way to represent multidimensional latent variables, in other words, variables that cannot be represented by a single quantitative dimension. For example, Auerbach and Collins (2006) demonstrated that latent statuses of alcohol use in emerging adulthood were distinguished by three characteristics. The characteristics were frequency of use, quantity of use, and whether or not there was heavy episodic drinking. The latent statuses differed along all three dimensions, but in different ways, so any one dimension would have been insufficient to represent the latent structure.

Second, LTA can be an excellent way to model change over time that is in some sense discrete, and to investigate predictors of this change over time. Graham, Collins, Chung, Wugalter, and Hansen (1991) used LTA to model the very early part of the substance use onset process in a sample of adolescents who had been in an experimental trial of a school-based substance use prevention program. They compared the incidence of latent status transitions occurring between Grade 7 and Grade 8 in the treatment group with those occurring in the control group. They found that for the latent statuses characterized by either no substance use; having tried alcohol only; having tried alcohol and tobacco; or having tried alcohol and tobacco and having been drunk at least once, adolescents assigned to the intervention condition were less likely to transition out of their latent status than those in the control condition. However, adolescents

who in Grade 7 were in a latent status characterized by having tried tobacco only, and not alcohol, appeared to be unaffected by the intervention program. Moreover, these adolescents appeared to be on an accelerated onset trajectory. This nuanced finding about differential program effects would have been difficult to uncover using a continuous latent variable model.

Third, LCA and LTA can help make a large contingency table interpretable. For example, the data analyzed by Graham et al. (1991) involved four dichotomous variables at two times, forming a contingency table consisting of 256 cells. It would have been extremely difficult to discern trends in the data by inspection of such a large table.

The Current Study: Development of Dating and Sexual Risk Behavior

This article illustrates how to use PROC LTA to fit and interpret models in empirical data on adolescent and young adult dating and sexual risk behavior. The goals of the study are to explore dating and sexual risk behavior longitudinally in this population, explore whether substance use behaviors (cigarette use, drunkenness and marijuana use) are predictive of dating and sexual risk behavior, and examine gender differences in dating and sexual risk behavior, as well as gender differences in the effects of substance use on this behavior.

Various dimensions of sexual risk behavior, including sexual intercourse, the number of sexual partners and inconsistent condom use, have been related to the use of alcohol, cigarettes and marijuana (Lowry, et al., 1994; Poulin & Graham, 2001; Tapert, Aarons, Sedlar, & Brown, 2001). In addition, several reviews have established gender differences in the relation between these constructs (e.g., Cooper, 2002; Perkins, 2002). Sexual risk behavior that can result in acquisition of sexually transmitted diseases (STD's) involves the intersection of several different dimensions of behavior, although few studies have attempted to identify particular sexual risk behavioral patterns that may increase risk of exposure. Exceptions include a study by Newman and Zimmerman (2000) that used cluster analysis to empirically identify four subgroups based on condom use and number of partners, and one by Beadnell et al. (2005) that employed latent profile analysis to identify latent classes based on condom use, number of partners, and frequency of sex. An advantage of these approaches is that important predictors of high-risk subgroups can be identified.

A model that takes into account dating and sexual risk behavior over time would extend the utility of this approach. Dating may be important to consider when modeling risky behavior because it is confounded with the number of sexual partners, which is a commonly used indicator of risk. For many adolescents, dating activity is a precursor to sexual activity. If certain patterns of dating behavior correspond to concurrent or later sexual risk behavior, dating behavior could be used to indicate adolescents who are at heightened risk for future sexual risk behavior. Including a longitudinal aspect to such an investigation, as well as predictors of behavior and behavior change, is imperative to understanding which individuals are expected to transition to more risky behavior in the future (as well as which individuals may transition to less risky behavior).

The following three research questions will be addressed in the current study:

- Research Question 1: Can a model of the development of dating and sexual risk behavior be identified? How does the probability of latent status membership differ by gender?
- Research Question 2: How does substance use behavior predict dating and sexual risk behavior at Time 1? How does this relation differ by gender?
- Research Question 3: How does past-year drunkenness predict change over time in dating and sexual risk behavior?

Method

Participants

Data used in the current study are from Rounds 2, 3, and 4 of the National Longitudinal Survey of Youth 1997 study (NLSY97; Bureau of Labor Statistics, 2005). The NLSY97 survey is sponsored and directed by the U.S. Bureau of Labor Statistics and conducted by the National Opinion Research Center at the University of Chicago, with assistance from the Center for Human Resource Research at The Ohio State University. The sample used in the current study consists of 2937 adolescents who were assessed at age 17 or 18 in 1998, then assessed again in 1999 and 2000. Adolescents in this age range who reported being married at any of the three time points were not included in this sample. The final sample was comprised of 51% boys and 49% girls, with 56% white, 29% African American, 2% Asian or Pacific Islander, 13% other race. The mean household income for the sample was \$48,442 ($SD = \$44,166$); 22% of residential mothers had less than a twelfth-grade education, 37% had a twelfth-grade education, and 41% had some level of education beyond high school.

Measures

Four categorical variables were used as indicators of dating and sexual risk behavior: number of dating partners in the past year (0, 1, 2 or more), past-year sex (yes, no), number of sexual partners in the past year (0, 1, 2 or more), and whether or not the participant reported at least one instance of intercourse without use of a condom, in other words, possible exposure to STD's, in the past year (yes, no). These four indicators were measured at each of the three times. Three additional binary variables were used as covariates to predict latent status membership or transitions. These were: whether participants had used cigarettes, been drunk, or used marijuana in the past year. The covariates were measured at Time 1 only. In addition, gender was used as a grouping variable. Table 1 shows the distribution of all variables used in this study.

Model Specification

As mentioned previously, the following sets of parameters are estimated in LTA: latent status membership probabilities at Time 1, called δ (delta) parameters; probabilities of transitions between latent statuses over time, called τ (tau) parameters; and item-response probabilities conditional on latent status membership and time, called ρ (rho) parameters. The ρ parameters express the correspondence between the observed items and the latent statuses, and form the basis for interpretation of the latent statuses.

Multiple-groups LTA—A multiple-groups analysis can be conducted in order to explore group differences in latent status membership, transition probabilities, and item-response probabilities. It is often helpful to test measurement invariance across groups before making conclusions about group differences in status membership or transition probabilities. This can be done by fitting two models: one with item-response probabilities estimated freely in each group, and one with these probabilities constrained to be equal across groups. The difference in the G^2 statistics is distributed as a chi-square with degrees of freedom equal to the difference in degrees of freedom for these two models. If the G^2 difference is not significant, there is strong support for measurement invariance across groups. If the difference is significant, the item-response probabilities should be examined closely to see how widely they vary across groups, and whether it is reasonable to simplify the model by equating measurement.

LTA with Covariates—Covariates can be incorporated in the latent transition model using a logistic link function. One or more variables can be specified as covariates, or predictors, of latent status membership at Time 1. The logistic regression coefficients expressing these relations can be used to address research questions related to initial status in the developmental

process. In addition, one or more variables can be specified as covariates of transition probabilities from Time 1 to Time 2, Time 2 to Time 3, and so on. Covariates for different sets of transition probabilities can be the exact same variable (e.g., gender) or can vary with time. For example, it might make sense to have behavior at Time 1 predict transitions from Time 1 to Time 2, and the same behavior measured at Time 2 predict transitions from Time 2 to time 3. Note that when a grouping variable is included in LTA with covariates, the logistic regression parameters are estimated for each group.

When one or more covariates are included, two additional sets of parameters may be estimated: a set of β (beta) parameters which are logistic regression coefficients for covariates predicting latent status membership at Time 1, and a set of β parameters which are logistic regression coefficients for covariates predicting transitions over time. When covariates are included, only ρ and β parameters are actually estimated; in this case, the δ and τ parameters are calculated as functions of β parameters and the covariates, and are provided in PROC LTA output. If a grouping variable is included, all sets of parameters ($\delta, \tau, \rho, \beta$) can be conditioned on group.

Both latent class and latent transition models rely on an assumption of local independence, that is, the assumption that within a latent class or status the indicators are independent. This assumption posits that the latent class variable causes any relation among indicators that is observed in the full sample. When categorical variables are used as indicators, no additional assumptions about the distributions of variables are made.

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 $\mathbf{Y}_i = (Y_{i11}, Y_{i12}, \dots, Y_{i1M}, Y_{i21}, Y_{i22}, \dots, Y_{i2M}, Y_{iT1}, Y_{iT2}, \dots, Y_{iTM})$ represent the vector of individual i 's responses for all times $t = 1, \dots, T$ and items $m = 1, \dots, M$, where an individual response Y_{itm} may take on the values $1, 2, \dots, r_m$. Let $s_{1i} = 1, 2, \dots, n_s$ be individual i 's latent status membership at Time 1, $s_{2i} = 1, 2, \dots, n_s$ 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:

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

where $\delta_{s_1|g}(x) = P(S_{1i} = s_1 | X_i = x, G_i = g)$ is a standard baseline-category multinomial logistic model (Agresti, 2002) 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 :

$$\begin{aligned} \delta_{s_1|g}(x) &= P(S_{1i} \\ &= s_1 | X_i \\ &= x, G_i = g) = \frac{\exp\{\beta_{0s_1|g} + x\beta_{1s_1|g}\}}{1 + \sum_{j=1}^{n_s-1} \exp\{\beta_{0j|g} + x\beta_{1j|g}\}} \end{aligned} \tag{2}$$

for $s_j = 1, \dots, n_s - 1$ 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_j 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:

$$\log \left(\frac{\delta_{1|1}(x)}{\delta_{2|1}(x)} \right) = \beta_{0|11} + \beta_{1|11}x. \quad (3)$$

Exponentiated β parameters are odds ratios. For example, $e^{\beta_{1|11}}$ 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, $\tau_{s_2|s_1,g}(x) = P(S_{2i} = s_2 \mid S_{1i} = s_1, X_i = x, G_i = g)$ 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:

$$\begin{aligned} \tau_{s_2|s_1,g}(x) &= P(S_{2i} \\ &= s_2 \mid S_{1i} \\ &= s_1, X_i = x, G_i = g) = \frac{\exp \{ \beta_{0s_2|s_1,g} + x\beta_{1s_2|s_1,g} \}}{1 + \sum_{j=1}^{n_s-1} \exp \{ \beta_{0j|s_1,g} + x\beta_{1j|s_1,g} \}} \end{aligned} \quad (4)$$

For $s_2 = 1, \dots, n_s$. (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.). A more thorough presentation of the mathematical model appears in a user's guide available for download at <http://methodology.psu.edu> (Lanza, Lemmon, Schafer & Collins, 2007).

In PROC LTA, parameters are estimated by maximum likelihood using the Expectation-Maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977), with Newton-Raphson incorporated for models with covariates. Missing data on the latent status indicators are handled in this procedure, with data assumed to be missing at random (MAR). When there are missing values on the indicators, the model given by Equation 1 is modified so that the product over $m = 1, \dots, M$ is replaced by a product over the items observed for that individual. The EM algorithm iterates until either the convergence criterion is achieved or the maximum number of iterations is reached (defaults in PROC LTA are a maximum absolute deviation (MAD) of .000001 and 5000 iterations, respectively).

Coding the Covariates

Covariates are treated as numeric in the statistical model, so each covariate must be either a continuous variable or a dummy-coded variable (or set of dummy-coded variables for categorical covariates with three or more response categories). When continuous covariates are standardized, standardized logistic regression coefficients are produced for the logistic regression coefficients.

Specifying the Model in PROC LTA

Publicly available software for LTA includes WinLTA (Collins, Lanza, Schafer, & Flaherty, 2002) and Mplus (Muthén & Muthén, 1998). The present article introduces PROC LTA, a new SAS procedure for LTA developed for SAS Version 9.1 for Windows¹. PROC LTA can be used to fit a variety of latent transition models, including multiple-groups LTA and LTA with covariates. The software is available for download free of charge at <http://methodology.psu.edu>.

A summary of all statements and options available in PROC LTA appears in Appendix A, and syntax from the series of models specified for this study appears in Appendix B. Details about the syntax required for this procedure are provided in a user's guide (Lanza et al., 2007). In addition, some of the statements and options are discussed and demonstrated in Lanza et al. (in press).

Model Selection

LTA models with different numbers of latent statuses can be compared using several statistics and criteria, including the likelihood-ratio G^2 statistic, Akaike Information Criterion (AIC; Akaike, 1974) and Bayesian Information Criterion (BIC; Schwarz, 1978). It is also important to consider the interpretability of the latent statuses when selecting a model. For example, when two or more latent statuses can be interpreted in essentially the same way, a model with fewer latent statuses should be considered.

Starting Values, Identification, and Parameter Restrictions

In some LTA models, the optimal solution can be difficult to identify if the amount of information provided by the data is small relative to the number of parameters being estimated. As models become more complex, such as models with more latent statuses, groups, or covariates, more information from the data is required for adequate identification. All else being equal, a smaller sample size provides less information. By fitting the same model to a particular data set using a number of different sets of starting values (for example, by varying the seed value), identification problems can be detected. The optimal solution has likely been identified if most of the starting values converge to that solution and it has the smallest log-likelihood value among the solutions obtained using different starting values. In PROC LTA, random starting values can be generated by providing a seed value, or user-defined starting values can be provided (see Appendix C for an example of a SAS data file containing starting values).

Parameter restrictions may be used to simplify a model, either to help achieve an identified solution, or to express or test specific hypotheses about parameter values. Parameters may be fixed to a prespecified value, or placed in an equivalence set. Parameters in an equivalence set can be estimated to be any value, but all parameters in the set are constrained to be equal to each other, in other words, are assigned the same estimated value.

Parameter restrictions for the ρ parameters can be used to improve model identification or to test specific hypotheses about the measurement of the latent variable. For example, measurement invariance across groups is expressed by constraining the ρ parameters to be equal across groups. Similarly, measurement invariance can be imposed across times. In PROC LTA either type of invariance can be imposed using keywords, or user-defined parameter restrictions can be provided. Another important function of parameter restrictions is to specify or test features of development. For example, stationarity of a developmental process with

¹Copyright 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.

three times of measurement can be tested by equating each τ parameter in the Time 1 to Time 2 matrix with the corresponding τ parameter in the Time 2 to Time 3 matrix.

Occasionally parameter restrictions can help to deal with estimation problems. For example, if a τ parameter estimate is very close to zero, trying to predict the transition probabilities from a covariate can lead to estimation problems. In this case, it can be useful to fix very small τ parameters to zero before adding a covariate to the model. More information about starting values and parameter restrictions are available in a freely downloadable user's guide (Lanza et al., 2007).

Results

Research Question 1: Can a model of the development of dating and sexual risk behavior be identified? How does the probability of latent status membership differ by gender?

Models with two, three, four, five and six latent statuses were compared to identify the number of statuses that provides the optimal balance of fit and parsimony. The likelihood-ratio G^2 statistic, degrees of freedom, AIC and BIC for each model appear in Table 2. Based on this table, the models with five or six statuses appear to represent the data best. An examination of the interpretation of the latent statuses in each model suggested that the five-status model was preferable.

Each column of Table 3 shows, for a particular latent status, the item-response probabilities for each response category (these were constrained to be equal across time²), the overall probability of membership in the status at each time, and the transition probabilities given latent status membership at the previous time. The item-response probabilities suggest the following interpretational labels for the five latent statuses: *Non-daters*, *Daters*, *Monogamous*, *Multi-partner safe*, and *Multi-partner exposed*. As Table 3 shows, Non-daters are very likely to report zero dating partners in the past year, no sex in the past year, no sex partners in the past year, and no STD exposure in that year. Daters have a high probability (0.793) of reporting two or more dating partners, but no sex partners, in the past year. Individuals in the Monogamous status are likely to report just one dating partner, and almost certainly report having had sex in the past year with only one sex partner; 60.1% of the individuals in this group reported having unprotected sex, and thus potentially have been exposed to STDs. Those in the Multi-partner safe group are likely to report two or more dating and sexual partners in the past year, but have a high probability (.820) of having used a condom every time they had sex. Finally, individuals in the Multi-partner exposed status almost certainly reported two or more dating and sexual partners in the past year, and have a very high probability (.810) of possible exposure to STDs with at least one sex partner. This high-risk status is not insignificant in size; 18% of 17- and 18-year old students are expected to be in this status, with nearly 25% of the population in it two years later, during early adulthood.

The most common latent status at Time 1 is the Daters latent status (28.9%), followed by the Multi-partner safe latent status (23.1%); the Monogamous latent status is the least prevalent at Time 1 (11.7%). At Time 3, however, the Monogamous latent status has become the most prevalent status (29.0%), followed closely by the high-risk Multi-partner exposed latent status.

Entries along the diagonal of each transition probability matrix (marked in bold font in Table 3) reflect the probability of membership in the same latent status at two consecutive times of

²Because models with item-response probabilities estimated freely over time and constrained to be equal over time are statistically nested, a G^2 difference test can be conducted to compare these models. However, this omnibus test can be quite sensitive to slight differences in item-response probabilities. For example, while the G^2 difference of 109.3 with 60 degrees of freedom was statistically significant, a careful inspection of the ρ parameters suggested that the interpretation of the five latent classes was very consistent over time. Therefore, the more parsimonious model was chosen.

measurement; for example, non-daters at Time 1 have about a 60% chance of being in the non-dater status again at Time 2, whereas individuals in the Multi-partner exposed latent status at Time 1 have an 81% chance of being in that latent status at Time 2. Entries off the diagonal of each matrix reflect the probability of transitioning to a different status one year later. Non-daters who change status are most likely to transition to the Daters status, and Daters who change status tend to move to the Monogamous and the Multi-partner safe statuses. Interestingly, the highest probability of transitioning to the high-risk Multi-partner exposed status is among individuals in monogamous relationships at the previous time. This probability is .141 for Time 1 to Time 2, and .240 for Time 2 to Time 3. (In comparison, the probability of transitioning from the Multi-partner safe to the Multi-partner exposed status is .107 for Time 1 to Time 2 and .149 for Time 2 to Time 3.) This suggests that those in the Monogamous status are at higher risk for future sexual risk behavior. Finally, individuals in the high-risk Multi-partner exposed status who change status membership over time almost certainly transition to the Monogamous status. In other words, although they may not transition to a status involving safe sex, they would be exposed to STDs from just one partner, rather than multiple partners.

Gender differences in dating and sexual risk behavior—Gender was added to the five-status model as a grouping variable in order to compare the prevalence of each latent status for males and females. (Although not reported here, the transition probabilities also were allowed to vary across groups.) Table 4 shows the prevalence of each latent status over time for males and females. Significant gender differences in sexual risk behavior at Time 1 were identified ($G^2 = 71.50$ with 4 *df*, $p < .0001$), with females more likely to belong to the Monogamous status (17.6% female versus 8.0% male) and males more likely to belong to the Multi-partner safe status (15.6% female versus 29.7% male). As might be expected, the proportion of males and females in the Monogamous status increases consistently with time. However, the proportion at every time is considerably larger among females. In contrast, while the proportion of individuals in the Multi-partner safe status decreases with time for both groups, the proportion at every time is smaller among females. Membership in the high-risk Multi-partner exposed status increases with time for both males and females, but the increases occur at a faster rate for males.

Research Question 2: How does substance use behavior predict dating and sexual risk behavior at Time 1? How does this relation differ by gender?

Three substance use behaviors were used to predict membership in dating and sexual risk behavior latent statuses at Time 1: cigarette use, drunkenness, and marijuana use in the past year. The top panel of Table 5 shows the odds ratios (exponentiated logistic regression coefficients) for the effect of each predictor in the full sample. The Non-daters status was specified as the reference group, therefore odds ratios larger than 1.0 indicate an increased risk of membership in a latent status relative to the Non-daters latent status. The effects of all three covariates were highly significant ($p < .0001$ for each), with similar patterns of results; for example, adolescents who reported using these substances are roughly two to three times more likely than nonusers to belong in the Daters or Monogamous latent status relative to the Non-daters latent status, and approximately three times more likely than nonusers to belong in the Multi-partner safe latent status relative to the Non-daters status. However, the effects of the different substances diverge substantially in prediction of the high-risk Multi-partner exposed status. Individuals who reported cigarette use in the past year are 3.16 times more likely than nonusers to be in the Multi-partner exposed status, while those who reported having been drunk or marijuana use are 8.36 and 10.54 times more likely, respectively, to belong to this high-risk status relative to the Non-daters status. This differential effect suggests that drunkenness and marijuana use are stronger predictors of high-risk sexual behavior than cigarette use.

Gender differences in the effects of substance use—Several gender differences have been identified above in the prevalence of the dating and sexual risk behavior statuses, and strong effects of substance use behavior on membership in these statuses were found. By including both the covariates and gender as a grouping variable, differential effects of substance use can be explored for males and females. The bottom panel of Table 5 shows the odds ratios for each gender. Cigarette use appears to be more strongly related to dating and sexual risk behavior among females than among males. Specifically, females who report engaging in cigarette use are more likely than males to belong in the Daters, Multi-partner safe, and Multi-partner exposed statuses relative to the Non-daters status. Although within each gender the association between drunkenness or marijuana use and dating and sexual risk behavior is strong, there do not appear to be large gender differences in the sizes of the associations.

Research Question 3: How does past-year drunkenness predict change over time in dating and sexual risk behavior?

There are a wide variety of research questions that can be addressed by incorporating predictors of transitions in dating and sexual risk behavior. To demonstrate how to address such questions, past-year drunkenness at Time 1 will be included as a covariate to predict transition probabilities from Time 1 to Time 2. For completeness, drunkenness also will be included as a predictor of the δ parameters, or latent status membership probabilities at Time 1.

The first question that will be addressed is whether past-year drunkenness predicts the probability of making each transition over time in dating and sexual risk behavior. For each row of the τ matrix, the reference status is the one on the diagonal; that is, we are predicting the probability of each transition relative of remaining in the same status. Before adding the covariate, however, three transition probabilities were fixed to zero because the estimates were so close to zero (.009 or smaller) that group differences in that pattern of change would be trivial. In addition, such restrictions often can prevent the logistic regression model from failing. The three probabilities fixed to zero were: the probability of transitioning from the Monogamous or Multi-partner exposed statuses to the Multi-partner safe status, and the probability of transitioning from the Multi-partner exposed status to the Non-daters status. Next, past-year drunkenness was added as a predictor; however, when the LTA model was run, the model did not converge. We noticed that the β parameter corresponding to the transition from the Multi-partner safe status to the Multi-partner exposed status was moving toward negative infinity, indicating that among one of the drunkenness groups, no individuals made that particular transition. As can occur with standard multinomial logistic regression (i.e., when the outcome is not latent), the model for that row of the transition probability matrix is not estimable. Because questions about the relation of drunkenness and transitions in dating and sexual risk behavior are still of interest in other rows of the transition probability matrix, the regression was skipped for this row of the matrix.

Results for the multinomial logistic regression are reported in Table 6 (panel A). As expected, the effect of drunkenness on the odds of membership in each latent status at Time 1 relative to the Non-daters status is large (results not shown). Individuals who have been drunk in the past year are 19 times more likely to belong in the Multi-partner exposed latent status than the Non-daters status, relative to those who did not report drunkenness. (Note that the reference group in this part of the model can be specified by the user; for details refer to Lanza et al., 2007). Table 6 shows the odds ratios associated with each transition probability relative to staying in the same status over time (note that the three transition probabilities that were fixed to zero are not included in the multinomial regression equation for that row). Drunkenness is associated with an increased probability of transitioning from the Non-daters and Daters statuses to the Multi-partner exposed status relative to remaining in the same low-risk status over time (OR = 3.59 for Non-daters; OR = 2.98 for Daters). Also, those who have engaged in drunkenness

in the past year are less likely to transition from the high-risk Multi-partner exposed status to the Monogamous status (OR = 0.69).

The second research question that will be addressed is whether, for each initial status in dating and sexual risk behavior, past-year drunkenness predicts transitioning to the high-risk Multi-partner exposed status relative to transitioning to any other status or staying in the same latent status. This is a simplified version of the model described above, as the outcome in each regression model is now binary, rather than a five-category multinomial. For each row of the τ matrix, the reference group is the first four statuses collapsed together. Results for the binomial logistic regression are reported in Table 6 (panel B). For each initial status in dating and sexual risk behavior, past-year drunkenness is associated with an increased odds of membership in the Multi-partner exposed status at Time 2 relative to membership in all other statuses combined. These effects are strongest for individuals who initially are Non-daters or Daters.

Discussion

This illustration shows some of the ways in which LTA provides a unique approach to modeling change over time. The developmental process of increasing involvement in dating, sexual activity and sexual risk is difficult to characterize along a single continuous dimension. Approaches such as repeated measures ANOVA and growth curve modeling are appropriate when development is conceptualized as continuous, but are less appropriate for addressing research questions about multifaceted constructs of behavior like risky sexual behavior, and about patterns of behavioral change over time. LTA provides a way to identify five meaningful discrete, qualitatively distinct behavioral patterns, or latent statuses, across three time points. It also provides a way to characterize development in terms of transitions between these latent statuses, and to model the effect of predictors on latent status membership and transitions between latent statuses. The results of LTA models can be highly descriptive of a multifaceted construct of behavior over time, and may allow for the identification of types of individuals who may be at risk for sexual risk behavior either concurrently or in the future.

A five latent status model of dating and sexual risk behavior provided a concise description of behavioral change over time. While a six latent status model provided a more detailed classification system, moving from five to six statuses essentially resulted in dividing the Monogamous latent status into two Monogamous statuses: one involving safe sex and the other involving possible STD exposure. However, because the current study was focused on sexual risk behavior, and a monogamous relationship without consistent condom use does not necessarily connote high risk, we selected the more parsimonious model.

Risk Based on Status Membership at Time 1

Regardless of prior substance use, individuals who are members of the Monogamous status are at heightened risk for transitioning to the Multi-partner exposed status at the subsequent time. This is possibly due to the fact that most individuals in the Monogamous status are not using condoms consistently; as they transition to having multiple sex partners in a subsequent year, they are not developing the habit of consistent condom use.

Risk Based on Substance Use

Although all three substances were associated with increased sexual risk behavior, the effect was much stronger for drunkenness and marijuana. This is probably because these substances affect judgment. The association with sexual risk appeared to be about the same for males and females. This suggests that both males and females would be important to target for intervention efforts aimed at reducing sexual risk behavior through reducing heavy alcohol and marijuana

use, or through a harm reduction approach aimed at increasing condom use when alcohol and marijuana are being used.

Individuals reporting drunkenness at Time 1 have an increased risk of transitioning from the Non-daters and Daters statuses to the high-risk Multi-partner exposed status, suggesting that this at-risk group is less careful about STD protection in the year following reported drunkenness. Similarly, those reporting drunkenness are less likely than those who did not report the behavior to transition from the high-risk Multi-partner exposed status to the Monogamous status relative to remaining in the Multi-partner exposed latent status over time. During these years of late adolescence and early adulthood, drunkenness is an important predictor of sexual risk behavior both concurrently and over time, possibly due to a lack of judgment caused by heavy use. Together, these findings suggest substance use behavior patterns (particularly drunkenness and marijuana use) and statuses of dating and sexual risk behavior that may indicate targets for risk reduction prevention and intervention efforts. These findings demonstrate research questions about stage-sequential development that can be addressed uniquely with LTA.

Practical Considerations for Applying LTA

Probability weights—When individuals have been sampled with known but unequal probabilities, arriving at unbiased estimates of population parameters requires that probability weights be incorporated. In addition, incorporating weights can be important when modeling complex survey data so that parameter estimates and standard errors can be adjusted. Probability weights cannot be incorporated in the current version of PROC LTA (version 1.1.3).

Sample size considerations—One potential limitation of LTA, as with all categorical models, is the difficulty that can arise in estimation with small- or even medium-sized samples. In the current study, the contingency table formed by crossing all indicators of dating and sexual behavior at three times has 46,656 cells. Even though most cells were empty in the current study, the sample size of 2937 provided enough information to estimate model parameters. However, with small samples, problems such as insufficient model identification arise more frequently. Bayesian methods may provide excellent solutions to these difficulties in estimation. For example, Chung, Lanza and Loken (in press) recently demonstrated that a small amount of data-dependent prior information can dramatically improve estimation in LTA with small samples.

Because Time 1 latent status membership probabilities are based on the full sample size, N , few estimation problems should be encountered for the set of logistic regression coefficients linking predictors to status membership, provided that N is sufficiently large. In contrast, each row of the transition probability matrix contains a set of parameters that are conditional probabilities. For example, the first row of a matrix describing transitions from Time 1 to Time 2 involves a set of probabilities conditioned on membership in the first latent status. When a status membership probability is small, so then is the number of individuals who contribute to the logistic regression model for that row. As in any categorical model, a multinomial logistic model will be unestimable if sparseness is too extreme. This is most likely to occur when all participants who make a particular transition are at one level of the covariate (e.g., if everyone who made the transition from the Multi-partner safe to the Multi-partner exposed latent status had been drunk in the past year, the corresponding logistic regression coefficient is infinite). Several features of PROC LTA, such as allowing the user to collapse across latent statuses for the logistic regression, maximize the developmental questions that can be addressed in this framework. Bayesian LTA also provides an avenue for addressing issues with logistic regression. PROC LTA has an option to apply a data-derived prior to stabilize the estimation of logistic models in LTA models with covariates (see Clogg, Rubin, Schenker, Schultz, and

Weidman (1991) for more information about this prior.) This approach, which does not require any input from the user regarding prior information, can be helpful when a β estimate diverges to infinity due to insufficient information for estimation (i.e., extreme sparseness).

Hypothesis testing—A limitation of the current study is the absence of statistical tests involving particular odds ratios. Although an omnibus test for the overall significance of a covariate on latent status membership is available, more flexible hypothesis testing (e.g., whether a particular odds ratio differs across gender) is not available. Bayesian estimation may provide a good avenue for addressing a variety of hypothesis tests (see Lanza, Collins, Schafer and Flaherty, 2005, for a presentation of hypothesis testing in Bayesian LTA).

Conclusions

Scientists are increasingly using latent class models to identify underlying subgroups of individuals who share important characteristics or behaviors. In addition, latent transition models are gaining popularity as a method for studying processes that can be conceptualized as stage-sequential development. PROC LCA (for latent class analysis) and PROC LTA (for latent transition analysis) provide straight-forward techniques for estimating these models in the SAS environment. Both procedures are available for download free of charge at <http://methodology.psu.edu>.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by National Institute on Drug Abuse grants P50 DA 10075 and K05 DA 018206. We would like to thank Joseph L. Schafer for his technical advice, David R. Lemmon for providing programming expertise, and Bethany Cara Bray for assistance with data preparation.

References

- Agresti, A. *Categorical Data Analysis*. 2nd. New York: Wiley; 2002.
- Aitkin M, Anderson D, Hinde J. Statistical modeling of data on teaching styles. *Journal of the Royal Statistical Society – A* 1981;144:419–461.
- Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 1974;19:716–723.
- Auerbach KJ, Collins LM. A multidimensional developmental model of alcohol use during emerging adulthood. *Journal of Studies on Alcohol* 2006;67:917–925. [PubMed: 17061010]
- Bandeem-Roche K, Miglioretti DL, Zeger SL, Rathouz PJ. Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association* 1997;92:1375–1386.
- Beadnell B, Morrison DM, Wilsdon A, Wells EA, Murowchick E, Hoppe M, Gillmore MR, Nahom D. Condom use, frequency of sex, and number of partners: Multidimensional characterization of adolescent sexual risk-taking. *The Journal of Sex Research* 2005;42:192–202.
- Bureau of Labor Statistics, U.S. Department of Labor. National Longitudinal Survey of Youth 1997 cohort, 1997-2003 (rounds 1-7) [Data file] Produced by the National Opinion Research Center, the University of Chicago and distributed by the Center for Human Resource Research. Columbus, OH: The Ohio State University; 2005.
- Chung H, Lanza ST, Loken E. Latent transition analysis: Inference and estimation. *Statistics in Medicine*. in press.
- Chung H, Park Y, Lanza ST. Latent transition analysis with covariates: Pubertal timing and substance use behaviours in adolescent females. *Statistics in Medicine* 2005;24:2895–2910. [PubMed: 16134129]

- Chung T, Martin CS. Classification and course of alcohol problems among adolescents in addictions treatment programs. *Alcoholism: Clinical and Experimental Research* 2001;25:1734–1742.
- Clogg CC, Goodman LA. Latent structure analysis of a set of multidimensional contingency tables. *Journal of the American Statistical Association* 1984;79:762–771.
- Clogg CC, Rubin DB, Schenker N, Schultz B, Weidman L. Multiple imputation of industry and occupation codes in census public-use samples using Bayesian logistic regression. *Journal of the American Statistical Association* 1991;86:68–78.
- Collins, LM.; Lanza, ST.; Schafer, JL.; Flaherty, BP. WinLTA User's Guide Version 3.0. University Park: The Methodology Center, Penn State; 2002.
- Cooper ML. Alcohol use and risky sexual behavior among college students and youth: Evaluating the science. *Journal of Studies on Alcohol* 2002:101–117.
- Dayton CM, Macready GB. Concomitant-variable latent-class models. *Journal of the American Statistical Association* 1988;83:173–178.
- Dempster AP, Laird NM, Rubin DB. Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B* 1977;39:1–38.
- Dewilde C. The multidimensional measurement of poverty in Belgium and Britain: A categorical approach. *Social Indicators Research* 2004;68:331–369.
- 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]
- Guo J, Collins LM, Hill KG, Hawkins JD. Developmental pathways to alcohol abuse and dependence in young adulthood. *Journal of Studies on Alcohol* 2000;61:799–808. [PubMed: 11188485]
- Jackson KM, Sher JJ, Gotham HJ, Wood PK. Transitioning into and out of large-effect drinking in young adulthood. *Journal of Abnormal Psychology* 2001;100:378–391. [PubMed: 11502081]
- Lanza ST, Collins LM. Pubertal timing and the stages of substance use in females during early adolescence. *Prevention Science* 2002;3:69–82. [PubMed: 12002560]
- Lanza ST, Collins LM, Lemmon D, Schafer JL. PROC LCA: A new procedure for latent class analysis. *Structural Equation Modeling*. in press.
- Lanza ST, Collins LM, Schafer JL, Flaherty BP. Using Data Augmentation to Obtain Standard Errors and Conduct Hypothesis Tests in Latent Class and Latent Transition Analysis. *Psychological Methods* 2005;10:84–100. [PubMed: 15810870]
- Lanza, ST.; Flaherty, BP.; Collins, LM. Latent class and latent transition models. In: Schinka, JA.; Velicer, WF., editors. *Handbook of Psychology: Vol 2: Research Methods in Psychology*. Hoboken, NJ: Wiley; 2003. p. 663-685.
- Lanza, ST.; Lemmon, D.; Schafer, JL.; Collins, LM. PROC LCA & PROC LTA User's Guide Version 1.1.3 beta. University Park, PA: The Pennsylvania State University, The Methodology Center; 2007.
- Lazarsfeld, PF.; Henry, NW. *Latent structure analysis*. Boston, MA: Houghton Mifflin; 1968.
- Lowry R, Holtzman D, Truman BI, Kann L, Collins JL, Kolbe LJ. Substance use and HIV-related sexual behaviors among US high school students: Are they related? *American Journal of Public Health* 1994;84:1116–1120. [PubMed: 8017535]
- Muthén, LK.; Muthén, BO. *Mplus User's Guide*. Fourth. Los Angeles, CA: Muthén & Muthén; 19982006.
- Newman PA, Zimmerman MA. Gender differences in HIV-related sexual risk behavior among urban African American youth: A multivariate approach. *AIDS Education and Prevention* 2000;12:308–325. [PubMed: 10982121]
- Perkins HW. Surveying the damage: A review of research on consequences of alcohol misuse in college populations. *Journal of Studies on Alcohol* 2002:91–100. [PubMed: 11925064]
- Poulin C, Graham L. The association between substance use, unplanned sexual intercourse and other sexual behaviours among adolescent students. *Addiction* 2001;96:607–621. [PubMed: 11300964]
- Prochaska JO, Velicer WF. The Transtheoretical Model of health behavior change. *American Journal of Health Promotion* 1997;12:38–48. [PubMed: 10170434]
- Schwarz G. Estimating the dimension of a model. *Annals of Statistics* 1978;6:461–464.

- Stern HS, Arcus D, Kagan J, Rubin DB, Snidman N. Using mixture models in temperament research. *International Journal of Behavioral Development* 1995;18:407–423.
- Tapert SF, Aarons GA, Sedlar GR, Brown SA. Adolescent substance use and sexual risk-taking behavior. *Journal of Adolescent Health* 2001;28:181–189. [PubMed: 11226840]
- Velicer WF, Martin RA, Collins LM. Latent transition analysis for longitudinal data. *Addiction* 1996;91:S197–S209. [PubMed: 8997793]

Table 1

Descriptive Statistics (N=2937)

Variable in Model		Code	Label	Frequency at Time 1 (Valid %)	Frequency at Time 2 (Valid %)	Frequency at Time 3 (Valid %)
<i>Indicators of latent status:</i>						
Past-year number of dating partners	1	0	0	521 (18.0)	354 (13.3)	323 (12.6)
	2	1	1	455 (15.7)	642 (24.2)	729 (28.3)
	3	2 or more	2 or more	1927 (66.4)	1659 (62.5)	1520 (59.1)
Past-year sex	.	Missing	Missing	34	282	365
	1	No	No	1349 (47.3)	981 (37.2)	735 (28.9)
	2	Yes	Yes	1504 (52.7)	1657 (62.8)	1812 (71.1)
Past-year number of sexual partners	.	Missing	Missing	84	299	390
	1	0	0	1391 (50.0)	993 (38.4)	752 (29.9)
	2	1	1	568 (20.4)	757 (29.3)	877 (34.9)
Exposed to STD in past year	3	2 or more	2 or more	825 (29.6)	835 (32.3)	886 (35.2)
	.	Missing	Missing	153	352	422
	1	No	No	2017 (75.0)	1736 (68.8)	1422 (58.8)
2	Yes	Yes	672 (25.0)	786 (31.2)	998 (41.2)	
.	Missing	Missing	248	415	517	
<i>Covariates:</i>						
Past-year cigarette use	0	No	No	1671 (57.1)		
	1	Yes	Yes	1256 (42.9)		
	.	Missing	Missing	10		
Past-year drunkenness	0	No	No	2211 (75.9)		
	1	Yes	Yes	703 (24.1)		
	.	Missing	Missing	23		
Past-year marijuana use	0	No	No	2093 (71.8)		
	1	Yes	Yes	822 (28.2)		
	.	Missing	Missing	22		

Variable in Model	<i>Code</i>	<i>Label</i>	<i>Frequency (%)</i>
<i>Grouping variable:</i>			
Gender	1	Male	1498 (51.0)
	2	Female	1439 (49.0)

Table 2
Comparison of Models

Number of Statuses	Likelihood-Ratio G^2	Degrees of Freedom	AIC	BIC
2	5403.4	46,638	5437.4	5539.2
3	3556.9	46,623	3620.9	3812.4
4	3171.1	46,604	3273.1	3578.3
5	2565.3	46,581	2713.3	3156.2
6	2360.4	46,554	2562.4	3166.9

Note: Bold font indicates the selected model.

Table 3
Item-response Probabilities (Probability of Item Response Given Latent Status), Prevalence of latent Statuses, and Transition Probabilities in Latent Status Membership

	Latent Status				
	Non-Daters	Daters	Monogamous	Multi-Partner Safe	Multi-Partner Exposed
<i>Item-Response Probabilities:</i>					
Number of dating partners in past year					
0	.762	.005	.099	.050	.020
1	.179	.202	.657	.025	.053
2 or more	.059	.793	.244	.926	.927
Had sex in past year					
No	.976	.994	.000	.000	.000
Yes	.024	.006	1.000	1.000	1.000
Number of sexual partners in past year					
0	1.000	1.000	.001	.021	.001
1	.000	.000	.969	.335	.086
2 or more	.000	.000	.031	.644	.913
Exposed to STD in past year					
No	1.000	1.000	.399	.820	.190
Yes	.000	.000	.601	.180	.810
<i>Prevalence of Statuses at:</i>					
Time 1	.186	.289	.117	.231	.177
Time 2	.134	.234	.215	.210	.206
Time 3	.114	.178	.290	.169	.249
<i>Transitions from Time 1 (rows) to Time 2 (columns):</i>					
Non-daters	.612	.183	.083	.089	.034
Daters, No Sex	.010	.570	.161	.206	.053
Monogamous	.050	.042	.678	.090	.141
Multi-partner Safe	.041	.108	.208	.536	.107

	Latent Status				
	Non-Daters	Daters	Monogamous	Multi-Partner Safe	Multi-Partner Exposed
Multi-partner Exposed	.012	.033	.146	.000	.809
<i>Transitions from Time 2 (rows) to Time 3 (columns):</i>					
Non-daters	.636	.154	.151	.059	.000
Daters, No Sex	.037	.529	.194	.165	.075
Monogamous	.035	.051	.664	.010	.240
Multi-partner Safe	.041	.095	.142	.574	.149
Multi-partner Exposed	.016	.011	.252	.000	.720

Note: Item-response probabilities constrained to be equal at all three time points. Entries in bold font indicate membership in the same latent status at two consecutive times.

Table 4

Prevalence of Latent Statuses by Gender

	Latent Status			
	Non-Daters	Daters	Monogamous	Multi-Partner Exposed
<i>Males:</i>				
Time 1	.167	.280	.080	.176
Time 2	.121	.240	.160	.222
Time 3	.123	.166	.223	.267
<i>Females:</i>				
Time 1	.197	.303	.176	.167
Time 2	.141	.235	.292	.182
Time 3	.096	.197	.383	.223

Table 5
Odds Ratios for Predictors of Stage Membership at Time 1

	Latent Status at Time 1			
	Non-daters	Daters	Monogamous	Multi-Partner Exposed
<i>Overall effect of covariate:</i>				
Past-year cigarette use	---	1.99	2.80	3.50
Past-year drunkenness	---	3.36	3.68	3.53
Past-year marijuana use	---	1.69	2.53	2.58
<i>Effect for males:</i>				
Past-year cigarette use	---	1.57	3.30	2.40
Past-year drunkenness	---	3.72	5.46	3.42
Past-year marijuana use	---	1.61	2.37	2.49
<i>Effect for females:</i>				
Past-year cigarette use	---	2.54	2.67	6.78
Past-year drunkenness	---	2.89	3.20	3.22
Past-year marijuana use	---	1.75	3.20	2.19

Note: Dashes indicate the reference class.

Table 6
Odds Ratios Reflecting the Effects of Past-year Drunkenness on Transitions from Time 1 to Time 2

	Latent Status				
	Non-daters	Daters	Monogamous	Multi-Partner Safe	Multi-Partner Exposed
A. Multinomial logistic regression coefficients for effect of drunkenness					
Non-daters	---	0.91	1.58	1.10	3.59
Daters, No Sex	0.60	---	1.96	1.63	2.98
Monogamous	0.89	0.98	---	1.0 ^f	0.97
Multi-partner Safe	1.0 ^{ne}	1.0 ^{ne}	1.0 ^{ne}	1.0 ^{ne}	1.0 ^{ne}
Multi-partner Exposed	1.0 ^f	1.30	0.69	1.0 ^f	---
B. Binomial logistic regression coefficients for effect of drunkenness					
Non-daters	---	---	---	---	3.90
Daters, No Sex	---	---	---	---	3.29
Monogamous	---	---	---	---	1.56
Multi-partner Safe	1.0 ^{ne}	1.0 ^{ne}	1.0 ^{ne}	1.0 ^{ne}	1.0 ^{ne}
Multi-partner Exposed	---	---	---	---	1.59

Note: Dashes indicate the reference class.

^f Transition probability was trivially small, and thus fixed to zero. This status not included in logistic regression for that row of matrix.

^{ne} For at least one level of the covariate, one cell in row of transition probability matrix was empty. Logistic regression model not conducted for this row.