Using Data Augmentation to Obtain Standard Errors and Conduct Hypothesis Tests in Latent Class and Latent Transition Analysis

Stephanie T. Lanza, Linda M. Collins, Joseph L. Schafer, and Brian P. Flaherty
Pennsylvania State University

Latent class analysis (LCA) provides a means of identifying a mixture of subgroups in a population measured by multiple categorical indicators. Latent transition analysis (LTA) is a type of LCA that facilitates addressing research questions concerning stage-sequential change over time in longitudinal data. Both approaches have been used with increasing frequency in the social sciences. The objective of this article is to illustrate data augmentation (DA), a Markov chain Monte Carlo procedure that can be used to obtain parameter estimates and standard errors for LCA and LTA models. By use of DA it is possible to construct hypothesis tests concerning not only standard model parameters but also combinations of parameters, affording tremendous flexibility. DA is demonstrated with an example involving tests of ethnic differences, gender differences, and an Ethnicity × Gender interaction in the development of adolescent problem behavior.

Keywords: data augmentation, latent class analysis, latent transition analysis, multiple imputation, latent variable model

Many variables of interest in psychology, including temperament, psychological diagnoses, attitudes, and health behaviors, are latent. Most researchers in psychology are familiar with the factor model, which features continuous latent variables measured by means of multiple fallible indicators. Somewhat less familiar, but used increasingly often, is the latent class model for categorical latent variables and a variation designed especially for longitudinal data, called the latent transition model. The main objective of this article is to introduce a highly flexible approach to estimation and hypothesis testing in latent class and latent transition models, an approach known as data augmentation (DA).

A Brief Introduction to Latent Class and Latent Transition Models

This article begins with a very brief introduction to latent class and latent transition models. More complete introductions can be found in Collins and Wugalter (1992), Lanza, Flaherty, and Collins (2003), and McCutcheon (1987).

The latent class model posits an underlying categorical latent variable that is measured by multiple fallible categorical indicators. Observed responses to the categorical indicators are crossed in a multiway contingency table, which in latent class analysis (LCA) plays much the same role that the observed variance–covariance matrix plays in factor analysis. In contrast to the continuous normal distribution of factor scores, a categorical latent variable has a discrete distribution. In other words, it divides the population into mutually exclusive and exhaustive groups called latent classes. This means that the latent class model is one type of mixture model in which the overall population is a mixture of individuals from the various discrete latent classes. Most latent class models make the assumption of local independence, which specifies that the observed variables are independent within a given latent class. This is conceptually similar to the assumption often made in factor analysis of no covariances among the uniquenesses.

LCA was originally developed with static, or unchanging, latent variables in mind. In longitudinal studies, the re-
searcher is often interested in latent variables that are changing in systematic ways over time (e.g., Collins & Cliff, 1990). LCA can be extended to situations in which change in categorical latent variables is occurring over time in a manner that is conceptually similar to that of latent growth modeling (e.g., Singer & Willett, 2002). Much of this work has involved examining change measured by a single manifest variable at two or more times, an approach known as latent Markov modeling (Langeheine, 1994; van de Pol & Langeheine, 1990; van der Heijden, Dessens, & Bockenholt, 1996; Vermunt, Langeheine, & Bockenholt, 1999).

Latent transition analysis (LTA), which is similar to multiple indicator Markov modeling (e.g., van de Pol & Langeheine, 1990), can be used to fit latent class models in longitudinal panel data by use of multiple manifest indicators (e.g., Collins & Flaherty, 2002; Collins, Graham, Long, & Hansen, 1994; Collins, Graham, Rousculp, & Hansen, 1997; Collins, Hyatt, & Graham, 2000; Collins & Wugalter, 1992; Graham, Collins, Wugalter, Chung, & Hansen, 1991; Hyatt & Collins, 2000; Lanza & Collins, 2002; Lanza et al., 2003).

Since the early fundamental conceptual developments provided by Anderson (1954) and Lazarsfeld and Henry (1968), work by Dempster, Laird, and Rubin (1977) and Goodman (1974) that made estimation feasible, and the availability of increasingly powerful computing, LCA and LTA have been applied in many different behavioral research environments. LCA has been applied in survey research (Biemer, Woltman, Raglin, & Hill, 2001), education (Bergan, 1983), psychology (Lanza et al., 2003), and medical diagnosis (Kendler, Karkowski, & Walsh, 1998). Examples of the use of LTA include modeling yearly change in risk behavior among injection drug users (Posner, Collins, Longshore, & Anglin, 1996), exploring risk factors of adolescent substance use onset (Lanza & Collins, 2002), and assessing transitions between the stages of smoking behavior in the transtheoretical model of behavior change (Velicer, Martin, & Collins, 1996). Given appropriate data, LTA could also be used to estimate and test classic stage theories of development, such as Piaget’s (1973) stages of psychosocial development and Erikson’s (1950) stages of ego identity. More recently, LCA and LTA have been incorporated into structural equation modeling (SEM) software (Muthén & Muthén, 2004).

LCA and LTA present some advantages to the researcher. First, the error estimation that is an integral part of LCA provides a principled way of handling contingency table cells that contain measurement error. This approach, which treats error as an inherent feature of empirical data and incorporates it into the statistical model, is vastly preferable to the often used ad hoc alternative of attempting to identify data that are subject to error and discarding or revising them. Second, just as procedures such as SEM can help make sense of complex patterns of relationships in large covariance matrices, LCA and LTA can help to uncover orderly relationships in a large contingency table that may be impossible to discern by other means. Third, LCA and LTA can be used either in an exploratory fashion or to fit explicit, theory-driven models in empirical data.

Estimation and Hypothesis Testing

An important objective of LCA and LTA is to use observed sample data to estimate and test hypotheses concerning several specific kinds of population quantities or parameters. LCA estimates two sets of parameters. One set consists of the marginal probabilities of membership corresponding to each of the latent classes. The other set consists of the probabilities of each possible response for each observed variable, given latent class membership. LTA estimates these parameters plus an additional set consisting of the incidence of transitions from one latent class to another between two consecutive times.

Frequently, hypothesis testing in LCA and LTA involves combinations of parameters. For example, in both LCA and LTA one may want to compare parameters across groups, such as men and women, in a manner analogous to a multiple-groups analysis in SEM. Testing hypotheses about group differences in the value of a parameter requires estimation of one type of parameter combination, the group difference, as well as the standard error of the difference. Hypothesis tests involving other combinations of parameters and their standard errors may be necessary to address specific research questions or desired to obtain greater statistical power as compared with tests of individual parameters. However, the most common approach to estimation in LCA and LTA, the expectation-maximization (EM) algorithm (Dempster et al., 1977), does not routinely produce standard errors. Standard errors can be obtained on the basis of the likelihood function, but they are accurate only when the likelihood has a smooth, quadratic shape. For complex latent class models, this is rarely the case. In addition, when parameters are estimated at or near a boundary (i.e., a probability estimated close to zero or one) the associated standard errors may be a poor reflection of inferential uncertainty or may be unavailable (Chung, 2003).

DA, a Bayesian approach (see A Brief Introduction to the Logic of DA), provides a highly flexible way of obtaining estimates and standard errors of any desired combination of LCA or LTA parameters. Some readers may be familiar with the use of DA to handle incomplete data problems via multiple imputation (Collins, Schafer, & Kam, 2001; Schafer, 1997; Schafer & Graham, 2002). DA treats latent variables as variables with 100% missing values. When applied to LCA and LTA, DA produces actual imputed values for the latent variables, and these imputed data sets are used to obtain the point estimate and standard error for each parameter. As we discuss in more detail below, the multiple
imputation approach used by DA is similar in some ways to an approach known as Gibbs sampling (Hoijtink & Molenaar, 1997), and in fact the two yield the same answer when applied in an equivalent manner (Schafer, 1997, pp. 135–145). However, DA is usually more efficient, particularly when estimates and standard errors of combinations of parameters are desired. We have found that the EM approach remains the preferred way to carry out initial model selection in LCA and LTA. After a satisfactory model has been identified, DA provides an effective and computationally feasible means to an end, in which the end is traditional hypothesis testing.

A Brief Introduction to the Logic of DA

There is a fundamental difference between the classical perspective and the Bayesian perspective (e.g. Gelman, Carlin, Stern, & Rubin, 2004; Rubin, 1984). According to the classical perspective, sample-based estimates vary about a fixed population parameter. For example, sample means have a distribution about the fixed population mean \( \mu \). By contrast, in the Bayesian perspective the parameters are viewed as having a distribution about the fixed sample-based estimate. However, this distribution reflects not sampling variability but uncertainty in our state of knowledge about the parameter’s value. Despite this important difference, Bayesian procedures can provide a useful framework for many classical hypothesis testing problems. Gelman et al.’s (2004) book provides a thorough introduction to Bayesian statistics.

DA (Schafer, 1997; Tanner & Wong, 1987) can be thought of as a Bayesian analog of the EM algorithm. Both EM and DA approaches treat latent variables as missing data. Whereas EM computes maximum-likelihood estimates of model parameters, DA simulates random draws of parameters from their posterior distribution given the observed data. Beginning with user-specified starting values for the parameters, the EM algorithm alternately predicts the missing data (in the expectation [E] step) and reestimates the parameters (in the maximization [M] step) until convergence is reached. In DA the user again specifies starting values for the parameters; an excellent choice for these starting values is the parameter estimates from a previous EM procedure. In addition, as in all Bayesian procedures, the user of DA must specify a prior distribution for the parameters. Because contingency table frequencies follow a multinomial distribution, it is most convenient to apply a prior distribution from the Dirichlet family (the definition and properties of the Dirichlet distribution are reviewed by Schafer, 1997, chap. 7). Several sample forms of the Dirichlet distribution are illustrated in Figure 1.

The DA algorithm alternates between the following steps, which are analogous to the E and M steps of EM.

Imputation (I) step: Simulate the missing data (i.e., the latent variables) given the current parameter estimates and the observed data. This produces a complete data set, identical to the observed data with the addition of the imputed latent variables.

Posterior (P) step: Using the complete data set from the I step, the Dirichlet prior distribution, and the model for the complete data (in this case, the latent class model), draw a new set of parameters.

Alternating between these two steps creates a Markov chain, a sequence of random variables in which the distribution of each element depends on that of the previous one. The random parameters from the first P step are correlated with the starting values. As the iterations continue, however, the dependence on the starting values eventually burns off and becomes negligible, at which point the process is said to

![Figure 1](image-url) Two examples of Dirichlet densities with three parameters. The left panel shows the density for the Dirichlet with parameter \( \alpha = (5, 3, 5) \), and the right panel shows the density for the Dirichlet with \( \alpha = (1, 3, 2) \). Because one of the three parameters is redundant (i.e., in a three-class model the probability of membership in each class sums to one, so only two parameters must be estimated), the densities are plotted as functions of two parameters. Note that the mode lies inside the parameter space in the left panel (a necessary condition for latent class analysis and latent transition analysis priors) and on the boundary in the right panel.
have achieved stationarity. Stationarity also implies that the
distribution of the parameters at iteration \( k + 1 \) is no
different from their distribution at iteration \( k \), although these
successive draws are correlated with each other (Challis &
Kitney, 1991). Once stationarity has been achieved, the
result from any P step is a random draw from the Bayesian
posterior distribution of the parameters given the observed
data; moreover, the result from any I step is a random draw
from the posterior predictive distribution of the missing data
given the observed data. The latter is the distribution from
which multiple imputations must be drawn (Rubin, 1987).

DA is just one family of Markov chain Monte Carlo
(MCMC) algorithms. MCMC is a fairly new body of tech-
niques for simulating draws from otherwise intractable
probability distributions (Gilks, Richardson, & Spiegelhal-
ter, 1996). Another general class of MCMC algorithms
called Gibbs sampling simulates a multivariate quantity by
drawing from the conditional distributions of each element
given all the others in turn. DA can be regarded as a special
case of Gibbs sampling for incomplete-data problems
(Schafer, 1997, Section 3.4). Indeed, Hoijtink and Molenaar
(1997) presented the I- and P-step procedure for LCA
models described above as a Gibbs sampler. We have cho-
sen to call it DA rather than Gibbs sampling because the
former name is more descriptive of the method; the ob-
served data are being augmented by latent variables at each
iteration. A second reason we call it DA is to emphasize the
fact that we collect and summarize the results in a different
way. Hoijtink and Molenaar retained a long sequence of
random parameters drawn at the P steps and used it to
directly estimate means and quantiles of the posterior dis-
tribution. In contrast, we retain the drawn values from just
a few I steps and use them as multiple imputations in the
manner of Rubin (1987), as detailed below.

The number of DA iterations (\( k \)) required to achieve
stationarity varies from one application to the next, and for
each model and data set, it is important to acquire a feel for
how many iterations are needed. Generally speaking, more
iterations are needed as the complexity of the model grows
or if the measurement of the latent variables is weak. If the
starting values for DA were provided by a preliminary run
of EM, then it is almost always the case that DA achieves
stationarity in fewer iterations than it took EM to converge.
Nevertheless, rather than simply guessing what \( k \) might be,
one should always examine diagnostic plots that are based
on the series of random parameters resulting from the DA
run. The value of \( k \) must be large enough for the lag-\( k \)
autocorrelation in every parameter to die down to zero. In
practice, we also find it useful to examine the autocorrela-
tion in the worst linear function of LCA or LTA parameters.
This scalar function is the linear combination of parameters
for which EM is observed to converge most slowly (Scha-
typically be revealed by examination of time-series and
autocorrelation plots of the worst linear function; this re-
duces the need to examine plots for individual parameters.
It is better to overestimate \( k \) rather than underestimate it;
choosing a value larger than necessary means only that the
computer will have to run a bit longer.

Once we have selected a value for \( k \), the output stream of
simulated parameters from the P step is no longer needed
and may be discarded. Instead, we base our inferences on
just a few imputed data sets retained from the I step. The
imputed data sets from successive iterations are correlated,
but imputations spaced \( k \) or more cycles apart are not.
Therefore, we retain the product of the I step at iterations \( k,
2k, \ldots, Mk \) and use these data sets as \( M \) independent
multiple imputations. Summarizing the output of DA
through these imputations rather than through the simulated
parameters offers two advantages to the investigator. First,
it is more efficient in terms of data storage. For typical
missing-data problems Rubin (1987) showed that efficient
inferences can be achieved with a number of imputations as
small as \( M = 5 \) or even \( M = 3 \). With LCA or LTA models,
the rates of missing information tend to be somewhat
higher, so we often use a value of \( M \) between 10 and 25. In
contrast, obtaining stable estimates directly from the stream
of P steps would require us to retain thousands or even tens
of thousands of simulated parameter values. Second, with
multiple imputation it is not necessary to specify in advance
which combinations of parameters are of interest. Estimates
and standard errors of any desired combination of param-
ters can be produced quickly and easily, because in these \( M \)
imputed data sets, latent class membership is no longer
missing; standard complete-data techniques can be applied
to each data set, yielding a point estimate and within-
imputation variance for each parameter and for any desired
combination of parameters. The results are then combined
across imputations to arrive at final estimates of parameters
and standard errors. This procedure is described in detail
below and illustrated in the empirical example. The result-
ning parameter estimates are not maximum likelihood;
rather, they are the means of the marginal posterior param-
eter distributions, and the standard errors are the standard
deviations of these distributions. More discussion on the
relative merits of using multiple imputations versus simu-
lated parameters from DA is provided by Schafer (1997,
Section 4.5).

Obtaining LCA and LTA Parameter Estimates and
Standard Errors by Means of DA

Each of the \( M \) imputed data sets resulting from the DA
procedure contains randomly different, plausible values for
the latent class membership variables. These data sets pro-
vide the basis for obtaining parameter estimates and stan-
dard errors. The estimates and standard errors are calculated
within each of the $M$ imputed data sets and then are combined across data sets.

**Obtaining Within-Imputation Point Estimates and Standard Errors for Combinations of Parameters**

Several types of combinations of parameters might be of interest in LCA and LTA. These include sums of parameters, such as the prevalence of a set of latent classes; differences of parameters, such as a group difference in the probability of membership in a particular latent class; interactions, such as the interaction between two covariates (e.g., gender and ethnicity) in the prediction of latent class membership; and more complex combinations of parameters, such as stability over time in latent class membership. Parameter estimates for sums, differences, interactions, and other more complex combinations are found within each imputed data set by use of standard complete-data methods. For example, the estimate for the probability of membership in a particular latent class is found by computing the proportion of individuals in the sample imputed to be in that class. The prevalence of a set of latent classes is found by summing the probabilities of membership for each of the latent classes composing the set. The variance associated with each combination of parameters can also be calculated by use of commonly known formulas. For a sum of parameters the variance is estimated by use of the usual expression for the variance of a proportion, $p(1 - p)/n$, where $p$ is the probability associated with the sum (e.g., the prevalence of a set of latent classes).

The difference of proportions falls between $-1.0$ and $1.0$. A difference of zero suggests that the groups have identical conditional distributions, and thus, there is no relationship between group membership and latent class membership. The variance of a difference of proportions within each imputed data set is defined as

$$ U = \frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2} $$

(Agresti, 1990, p. 55), where $p_1$ and $p_2$ are the estimates in Group 1 and Group 2, respectively, and $n_1$ and $n_2$ are the numbers of individuals in Group 1 and Group 2, respectively. For example, if $p_1$ and $p_2$ refer to the probability of membership in a particular latent class for men and women, then $n_1$ is the number of men in the sample and $n_2$ is the number of women.

Estimating the variance of an interaction is a direct extension of the method presented above for the variance of the difference of proportions. For example, one interaction might reflect a different effect of ethnicity for men and women in the prevalence of a latent class. This combination of parameters can be calculated by subtracting the prevalence of that class in White men from the prevalence in White women, subtracting the prevalence of that class in African American men from the prevalence in African American women, and taking the difference of these two numbers. Then, the variance of the interaction is the sum of the four $p(1 - p)/n$ terms for each of the corresponding four groups.

More complex combinations of parameters and their associated standard errors can be estimated by using an approach similar to that for sums of parameters. For example, stability over time can be estimated in each imputed data set by summing up the number of individuals imputed to be in the same latent class at both times. It follows that the variance of this stability parameter can be calculated by $p(1 - p)/n$.

**Combining Point Estimates and Standard Errors Across Imputations**

The multiple imputation combination rule of Rubin (1987) is used to average across the imputations. This rule, which is the same as that used in multiple imputation for missing data, combines the parameter estimates (or combinations of parameters) and their associated within-imputation variances across the $M$ imputed data sets to determine the average point estimate, overall amount of variance, and 95% confidence interval for each statistic. This allows for the statistical comparison of group differences. The final inference is a reflection of the information carried by the observed data only; the missing data have been averaged out across the $M$ imputations (for a thorough discussion of DA and the multiple imputation combination rules, see Schafer, 1997, chap. 4). This averaging across imputations can be done in any data processing software package (see Additional Resources below).

Given $M$ imputations, we can calculate $M$ different estimates (e.g., of a model parameter or a combination of parameters). Let us refer to these estimates as $\hat{Q}$ and their respective variance estimates as $U$. The final point estimate based on DA is the average of the $M$ complete-data estimates,

$$ \hat{Q} = \frac{1}{M} \sum_{m=1}^{M} \hat{Q}^m, $$

where $m$ indexes the data sets, $m = 1, 2, \ldots, M$. The variance estimate $U$ associated with $\hat{Q}$ has two components. The **within-imputation variance** is the average of the $M$ complete-data variances calculated within each data set,

$$ \hat{U} = \frac{1}{M} \sum_{m=1}^{M} U^m. $$

The **between-imputation variance**, $B$, is the variance across the imputations,
The within-imputation and between-imputation variances can then be combined to obtain an estimate of the total variance, $V$, defined as

$$
B = \frac{1}{M-1} \sum_{n=1}^{M} (\hat{Q}^n - \bar{Q})^2.
$$

(4)

The within-imputation and between-imputation variances can then be combined to obtain an estimate of the total variance, $V$, defined as

$$
\bar{V} + (1 + M^{-1})B.
$$

(5)

The standard errors can be used to establish confidence intervals, which reflect the degree of uncertainty associated with each parameter estimate in the model. However, because LCA and LTA parameter estimates are probabilities, they are bounded by zero and one. Symmetric confidence intervals may extend beyond these bounds, especially for parameter estimates near the boundary. It makes sense in this situation to transform point estimates to the logit metric and construct confidence intervals, CIs, on this scale by use of

$$
\text{CI} = \bar{Q} \pm t_{u,1-\alpha/2}\sqrt{V}.
$$

(6)

where $\bar{Q}$ refers to the average point estimate (now in the logit metric), $V$ refers to the total variance based on the multiple imputation combination rules, $\alpha$ is the significance level, and the degrees of freedom, $u$, are defined as

$$
(M - 1) \left[ 1 + \frac{\bar{U}}{1 + M^{-1}}B \right]^{-2}.
$$

(7)

The endpoints of the confidence intervals can then be transformed back to the probability scale by exponentiating them (Rubin & Schenker, 1987). The distribution is $t$ distributed in the logit metric so that standard confidence intervals can be constructed and then transformed back to probabilities for easier interpretation. This produces asymmetric confidence intervals in the probability metric that are bounded by zero and one.

Applying DA to a Latent Transition Model of Adolescent Problem Behavior

Elliott, Huizinga, and Ageton’s (1985) model of problem behavior suggests that there are various qualitatively unique episodes of problem behavior that individuals pass into and out of during adolescence. The different types of problem behavior are characterized by the use of various substances (e.g., alcohol or marijuana), minor acts of delinquency such as property damage, and index offenses (more serious delinquent acts, e.g., use of a weapon). Using Elliott et al.’s model as a starting point, we model adolescent problem behavior using LTA. We then illustrate the use of DA to test hypotheses involving the effects of gender, ethnicity, and their interaction on combinations of LTA parameters representing various aspects of problem behavior and its development over time.

Participants

Participants (N = 5,388) were African American and White adolescents drawn from Waves I (April–December 1995) and II (April–August 1996) of the National Longitudinal Study of Adolescent Health, known as Add Health (Udry, 2003). The sample used in the current study consists of members of the core sample, a nationally representative subset of the full sample of the Add Health study, who were in the 9th, 10th, or 11th grades at Wave I. Within the sample, approximately 50% were male adolescents and 50% were female adolescents, and 80% were White and 20% were African American. School was the primary sampling unit in the Add Health study, although design effects due to the use of a cluster sampling design were not adjusted in the current study.

Measures

This example involves a dynamic latent variable—problem behavior—and variables indicating gender and ethnicity. The manifest indicators of problem behavior are listed in Table 1. They include alcohol use in the past 12 months, delinquency in the past 12 months, marijuana use in the past 30 days, an index offense in the past 12 months, and polydrug (e.g., cocaine, inhalants, ecstasy) use in the past 30 days. Each of these five indicators was transformed from a frequency variable to a nominal variable to indicate the presence or absence of that recent behavior. The Add Health study assessed recent marijuana and polydrug use using a 30-day window and recent alcohol use, delinquency, and index offense using a 12-month window. We assumed that the behaviors assessed with the 12-month window were fairly stable during that period and combined measurement of all five behaviors in the model of problem behavior despite the inconsistent time windows. All variables were coded 0 for missing, 1 for no recent behavior, and 2 for any recent behavior.

Model to Be Tested

The model of problem behavior fit here was developed in a largely exploratory and iterative fashion. Elliott et al.’s (1985) model suggests that the most common ordering in the development of problem behavior is minor delinquency, followed by alcohol use, then marijuana use, then an index offense, and finally hard drug use. This sequence indicates the likely presence of six latent classes characterized, respectively, by no problem behavior; minor delinquency only; minor delinquency and alcohol use; minor delinquency with alcohol and marijuana use; minor delinquency and an index offense with alcohol and marijuana use; and
minor delinquency and an index offense with alcohol, marijuana, and hard drug use. The model also suggests that several alternative sequences in behavior occur over time, indicating the need for additional latent classes of problem behavior. With this model in mind we examined the cell frequencies of the data for each time individually to gain a rough sense of whether the types of problem behavior appearing in Elliott et al.’s model were occurring in the data and whether any frequently occurring cells appeared that would not be well accounted for by Elliott et al.’s model. We then fit a series of models including six classes, seven classes, and so on, until adding another class did not yield an additional, meaningful latent class with a nonzero probability of membership.

The result was a model with the following 11 classes of problem behavior: none; alcohol only; delinquency only; alcohol and marijuana use; delinquency and index offenses; alcohol and marijuana use and delinquency; alcohol use, delinquency, and index offenses; alcohol, marijuana, and polydrug use and delinquency; alcohol and marijuana use, delinquency, and index offenses; and alcohol, marijuana, and polydrug use, delin-

### Table 1

**Manifest Variables and Code Schemes for Problem Behavior**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Code scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol use</td>
<td>During the past 12 months, on how many days did you drink alcohol?</td>
</tr>
<tr>
<td></td>
<td>1 = never in past 12 months</td>
</tr>
<tr>
<td></td>
<td>2 = ever in past 12 months</td>
</tr>
<tr>
<td>Delinquency</td>
<td>During the past 12 months, how often did you...</td>
</tr>
<tr>
<td></td>
<td>paint graffiti or signs on someone else’s property or in a public place?</td>
</tr>
<tr>
<td></td>
<td>deliberately damage property that didn’t belong to you?</td>
</tr>
<tr>
<td></td>
<td>take something from a store without paying for it?</td>
</tr>
<tr>
<td></td>
<td>drive a car without its owner’s permission?</td>
</tr>
<tr>
<td></td>
<td>sell marijuana or other drugs?</td>
</tr>
<tr>
<td></td>
<td>steal something worth less than $50?</td>
</tr>
<tr>
<td></td>
<td>act loud, rowdy, or unruly in a public place?</td>
</tr>
<tr>
<td></td>
<td>1 = none of these in past 12 months</td>
</tr>
<tr>
<td></td>
<td>2 = one or more of these in past 12 months</td>
</tr>
<tr>
<td>Marijuana use</td>
<td>During the past 30 days, how many times did you use marijuana?</td>
</tr>
<tr>
<td></td>
<td>1 = never in past 30 days</td>
</tr>
<tr>
<td></td>
<td>2 = ever in past 30 days</td>
</tr>
<tr>
<td>Index offense</td>
<td>During the past 12 months, how often did you...</td>
</tr>
<tr>
<td></td>
<td>steal something worth more than $50?</td>
</tr>
<tr>
<td></td>
<td>go in to a house or building to steal something?</td>
</tr>
<tr>
<td></td>
<td>use or threaten to use a weapon to get something from someone?</td>
</tr>
<tr>
<td></td>
<td>pull a knife or gun on someone?</td>
</tr>
<tr>
<td></td>
<td>shoot or stab someone?</td>
</tr>
<tr>
<td></td>
<td>1 = none of these in past 12 months</td>
</tr>
<tr>
<td></td>
<td>2 = one or more of these in past 12 months</td>
</tr>
<tr>
<td>Polydrug use</td>
<td>During the past 30 days, how many times did you...</td>
</tr>
<tr>
<td></td>
<td>use cocaine?</td>
</tr>
<tr>
<td></td>
<td>use inhalants?</td>
</tr>
<tr>
<td></td>
<td>use any other type of illegal drugs, such as LSD, PCP, ecstasy, mushrooms,</td>
</tr>
<tr>
<td></td>
<td>speed, ice, heroin, or pills, without a doctor’s prescription?</td>
</tr>
<tr>
<td></td>
<td>take an illegal drug using a needle?</td>
</tr>
<tr>
<td></td>
<td>1 = none of these in past 30 days</td>
</tr>
<tr>
<td></td>
<td>2 = used one or more in past 30 days</td>
</tr>
</tbody>
</table>
frequency, and index offenses. This model corresponds closely to Elliott et al.’s model; for example, the six latent classes corresponding to the most common progression in problem behavior are represented in this set of latent classes. Five additional latent classes emerged, corresponding to alternative combinations of behavior predicted by Elliott et al.’s model (such as alcohol only).

The latent transition model involves estimating problem behavior at two times, with gender and ethnicity (and their interaction) as covariates. A conceptual picture of this model is shown in Figure 2. The latent variables Problem Behavior 1 and Problem Behavior 2 represent the latent class variable measured at each time with the five indicators described in Table 1. The arrows linking the indicators of problem behavior to the latent variables represent the probability of endorsing the variable given latent class membership (rather than regression coefficients) and make up the measurement part of the model. The structural part of the model consists of the arrow from Problem Behavior 1 to Problem Behavior 2 and the paths from the covariates to these latent variables. The bold arrow from Problem Behavior 1 to Problem Behavior 2 represents the set of transition probabilities in the latent class of problem behavior from Time 1 to Time 2. The paths from gender, ethnicity, and the Gender × Ethnicity interaction to Problem Behavior 1 indicate that the prevalence of each latent class of problem behavior at Time 1 is conditioned on these terms. The paths from the covariates to the bold arrow indicate that the transition probabilities are conditioned on these terms as well. Note that paths also could be included from the covariates to the indicators of problem behavior; this would allow for tests of measurement invariance across groups.

**The LTA Model and Notation**

Let \( c = 1, \ldots, C \) index the groups. In this example, \( C = 4 \) (White male, African American male, White female, African American female). The grouping variable in the current study is assumed to be measured without error; however, a latent grouping variable with multiple indicators can be modeled (see Lanza & Collins, 2002, for an example modeling pubertal timing as a latent grouping variable). The dynamic latent variable is measured at \( T \) times, where \( t = 1, 2, \ldots, T \). There are \( S_t \) latent classes at time \( t \), where \( s_t = 1, 2, \ldots, S_t \). In this example, \( T = 2 \), and \( S_1 = S_2 = 11 \), corresponding to the 11 latent classes listed above.

Let \( j = 1, 2, \ldots, J \) index the indicators of the dynamic latent variable. Here \( J = 5 \), corresponding to the five problem behavior indicators. The response categories of the observed indicators of the dynamic latent variable are designated as follows: For variable \( j \) measured at time \( t \), \( r_{jt} = 1, 2, \ldots, R_{jt} \). Let \( W \) refer in general to the cells of the contingency table made by cross-tabulating all measured variables, which include the indicator of the grouping variable and the \( J \) indicators of the dynamic latent variable, measured at \( T \) times. Let \( w = (c, r_{11}, r_{21}, \ldots, r_{JT}) \) represent a particular cell of the cross-tabulation. Then the LTA model is as follows:

*Figure 2.* The latent transition model for this study. Eleven latent classes of problem behavior were estimated from five binary indicators at each time. Gender, ethnicity, and their interaction are included as covariates predicting the probability of membership in problem behavior latent classes at Time 1 and the transition probabilities from Time 1 to Time 2. Alc = alcohol; Mar = marijuana; Poly = polydrug; Del = delinquency; Ind = index offense.
\[ P(W = w) = \sum_{c} \sum_{s_1} \sum_{s_2} \gamma_c S_1 S_2 \delta_{s_1s_2} \prod_{t=2}^{T} \tau_{s_1s_{t-1}} \prod_{j=1}^{J} \prod_{t=1}^{T} \rho_{jtr, j,c} \]  

where

\[ \gamma_c = \text{the probability of membership in group } c \text{ (e.g., the proportion in White male group)}. \]

\[ \delta_{s_1s_2} = \text{the probability of membership in latent class } s_1 \text{ at Time 1, conditional on membership in group } c \text{ (e.g., the probability of membership in the alcohol latent class conditional on membership in the White male group)}. \]

\[ \tau_{s_1s_{t-1}} = \text{the probability of membership in latent class } s_t \text{ at time } t, \text{ conditional on membership in latent class } s_{t-1} \text{ at time } t - 1 \text{ and membership in group } c \text{ (e.g., the probability of membership in the alcohol and delinquency latent class at Time 2, conditional on membership in the alcohol latent class at Time 1 and being a White male adolescent).} \]

\[ \rho_{jtr, j,c} = \text{the probability of response } r_p \text{ to manifest indicator } j \text{ of the dynamic latent variable at time } t, \text{ conditional on membership in latent class } s_t \text{ and group } c \text{ (e.g., the probability of responding "yes" at Time 1 to the question about having tried alcohol, conditional on membership in the alcohol and delinquency latent class and being a White male adolescent).} \]

The \( \rho \) parameters can be considered the measurement part of the model because they relate the observed categorical variables to the latent variable. These are analogous to the factor loadings in factor analysis or SEM and, like factor loadings, provide a basis for interpretation of the latent variable. The other parameters, which reflect latent class membership over time and its relation to covariates, can be referred to as the structural part of the model.

**Maximum Likelihood (ML) Estimation, Model Fit, and Parameter Restrictions**

Overall model selection was conducted by applying the EM algorithm to the combined sample; then the grouping variable was included prior to running the DA procedure and conducting hypothesis tests. A standard implementation of the EM algorithm (Dempster et al., 1977) was used to obtain ML estimates. The algorithm incorporated a standard maximum-likelihood routine (Schafer, 1997) for handling missing data in contingency tables.

One alternative for assessing overall model fit in LCA and LTA, as in other contingency table methods, is the omnibus \( G^2 \) test statistic. This test assumes that the \( G^2 \) is distributed as a chi-square. However, under conditions of sparseness, in which the number of contingency table cells is large relative to the sample size, the \( G^2 \) is not distributed as a chi-square, and the true distribution is unknown (Read & Cressie, 1988). This makes assessing absolute model fit on the basis of the \( G^2 \) difficult. When several alternatives are to be compared, the Bayesian information criterion (BIC; Schwartz, 1978) provides a better basis for model selection in LCA.

As part of the model fitting procedure in LCA, it is possible for the user to specify restrictions on the estimation of parameters (in much the same way as this is done in SEM). Parameter restrictions can be used to specify certain theoretically or practically determined aspects of a model or to reduce the number of parameters estimated in order to avoid underidentification. In fact, because of the sheer number of parameters involved, virtually all LTA models involve some user-specified parameter restrictions. Parameters may be fixed to a prespecified value, or equivalence sets may be designated in which all parameters are estimated at the same value. We specified and compared four variations of the 11-class model to arrive at the most parsimonious model that remained conceptually appealing and maintained a reasonable fit to the data. In general, parsimony is increased by simplifying the model, that is, by adding parameter restrictions.

In this empirical example, we did not have strong a priori ideas about latent class prevalence or changes between latent classes over time. For this reason, in the four models tested here all parameters in the structural part of the model were estimated without restrictions. In other words, the prevalences of problem behavior latent classes were estimated at Time 1, and transitions from all latent classes to any other latent class over time were estimated (all elements of the \( \tau \) matrix were estimated), implying that in this model, individuals can advance over time from any latent class to any other latent class. This resulted in estimation of 10 \( \delta \) parameters (because the latent classes are mutually exclusive and exhaustive at each time, the \( \delta \) parameters must sum to one, so one \( \delta \) can be obtained by subtraction) and 110 \( \tau \) parameters (again because the latent classes are mutually exclusive and exhaustive, each row of the \( \tau \) matrix sums to one, so one element of each row is obtained by subtraction). Parameter restrictions in the \( \tau \) matrix can be used to specify other types of change over time, for example, development when no regression occurs.

Parsimony in latent class and latent transition models can often be gained by imposing restrictions on the \( \rho \) parameters. Of particular interest in latent transition models is measurement invariance across time, that is, an equivalent mapping of the manifest items onto the categorical latent
variable at each time. Just as one may equate factor loadings over time in SEM to establish equivalent units in the continuous latent variables, one may wish to equate $\rho$ parameters at each measurement occasion to establish equivalent measurement across time. When measurement invariance is imposed in SEM and LTA, the meaning of latent variables remains constant over time, and group differences may be identified in factor means and $\delta$ parameters, respectively. Lanza et al. (2003) demonstrated an empirical test of measurement invariance over time in the measurement of latent classes of depression during adolescence.

Evaluation of the measurement model for the 11-class solution was conducted by examining four models with progressively increasing restrictions on the $\rho$ parameters. In a completely unrestricted version of the 11-class model measured by use of dichotomous variables, there would be $110$ $\rho$ parameters estimated ($2 \times 11 \times 5$; Measurement Occasions $\times$ Latent Classes $\times$ Observed Variables). In Model 1, the $\rho$ parameters were constrained to be equal across time, reducing the number of estimated $\rho$ parameters to 55 and resulting in a BIC of 2,039.8.

The restrictions imposed in the remaining three models are based on the idea of sorting the $\rho$ parameters according to the specific kinds of measurement error they represent and using this as a basis for arriving at parameter restrictions that are parsimonious yet conceptually appealing. Recall that the $\rho$ parameters are the probabilities of observed responses conditional on latent class membership. For each latent class, there are possible observed responses that are consistent with the latent class’s meaning and other observed responses that are inconsistent. The inconsistent responses can be designated as negative and positive errors. Consider the alcohol latent class, which according to the model is characterized by alcohol use but no other problem behavior. An example of a negative error for this latent class is a “no” response to the alcohol use indicator. An example of a positive error for this latent class is a “yes” response to the index offenses indicator. Each latent class is associated with $\rho$ parameters corresponding to negative and positive errors.

In Model 2, additional restrictions were imposed on Model 1 so that the $\rho$s corresponding to negative errors for any latent class formed an equivalence set for each indicator, and the $\rho$s corresponding to positive errors for any latent class formed another equivalence set. This resulted in two $\rho$ parameters estimated per indicator for a total of 10 $\rho$ parameters. Model 2 had a BIC of 1,802.4. Model 3 added the further restriction that all $\rho$ parameters corresponding to negative errors for all indicators formed an equivalence set, and all $\rho$ parameters corresponding to positive errors formed another equivalence set. Model 3, which estimated only two $\rho$s, had a BIC of 1,741.2. Finally, Model 4, the most parsimonious model, estimated just one $\rho$ parameter by forming a single equivalence set corresponding to the probability of a negative or positive error for any indicator. Model 4 had a BIC of 2,016.6.

Because Model 3 had the smallest BIC, indicating that it was the most efficient among these models, it was selected as the final model. The likelihood ratio statistic, $G^2$, is 676.7 with 901 degrees of freedom. Degrees of freedom in latent class models are computed as the number of cells in the contingency table formed by the observed indicators minus the number of estimated parameters minus one. The data set involves 5 variables measured at 2 times for a total of 10 binary variables; thus the number of cells in the table equals $2^{*}10 = 1,024$. The number of parameters estimated in Model 3 is $122$ ($10$ $\delta$s, $110$ $\tau$s, and $2$ $\rho$s). Therefore, the degrees of freedom $= 1,024 - 122 - 1$, or 901. A $G^2$ of 676.7 suggests a reasonably good fit. Because, as discussed above, the $G^2$ provides only a rough rule of thumb, it is a good idea to examine residuals. In this case, the residuals were consistent with a good model fit. The two $\rho$ parameter estimates in Model 3 corresponding to negative and positive errors were .006 and .087, respectively. For example, members of the alcohol latent class are estimated to have a .006 probability of responding “no” to the alcohol use indicator and a probability of .087 of responding “yes” to the index offenses indicator. For each latent class, the estimates of the $\rho$s are consistent with the latent class labels in Table 2 (e.g., for the alcohol latent class, the probability of responding “yes” is high for the alcohol use indicator and low for all other indicators).

**Turning to the DA Procedure**

Once a satisfactory model has been identified and fit by use of EM, DA can be applied to conduct hypothesis tests. The user specifies starting values for latent class model parameters and the parameters of the prior distribution. Here the maximum-likelihood estimates from the EM analysis described above were used as starting values, and a diffuse prior was used. A diffuse prior, which is sometimes referred to as a noninformative or flat prior, is a distribution that plays a minimal role in the posterior distribution and therefore in inference (Gelman et al., 2004, p. 61). We recommend the use of diffuse priors for most applications (see Additional Resources below). In our experience, varying the diffuse prior within a reasonable range has relatively little impact on results.

After the DA procedure has been run and before results across imputations have been combined, it is essential to examine diagnostic plots to determine whether a sufficiently large value has been chosen for $k$. This will ensure that the imputations are independent draws and that it is therefore appropriate to combine them and proceed with computation of parameter estimates and standard errors. The left panel in each row of Figure 3 shows plots of the autocorrelations for the worst linear function of the $\delta$, $\tau$, and $\rho$ parameters for the
empirical example. Each worst linear function represents, for a subset of LTA parameters, the linear combination that converged most slowly in the EM algorithm. These plots, which indicate whether there is serial dependence among the DA steps, have the lag on the abscissa and the autocorrelation on the ordinate. The dotted lines on the plot form an envelope corresponding to the 95% confidence interval about zero. When the autocorrelation drops to near zero, the series has converged. If the plot of any set of parameters does not drop to near zero by lag \( k \), this suggests that the random draws are not independent. Then a larger value of \( k \) should be specified, and DA should be rerun. On the basis of Figure 3, \( k = 100 \) appears to be sufficiently large for our data.

The right panel in each row of Figure 3 shows the time series plots for the worst linear combination of the parameters for our example. For each set of parameters, the worst linear function is plotted against the iteration number. Note that the scale of the y-axis differs across the time series plots, because the range of the worst linear combination varies considerably across parameter sets. After the \( k \)th iteration has been reached, these plots should appear to be horizontal bands fluctuating randomly around some value. There are two diagnostic indicators to look for in the time series plots. One is drift—a slow, discernable upward or downward trend. Most DA time series plots show some drift in the first several iterations and then level off once stationarity has been reached. A value of \( k \) should be chosen so as to take the first draw after this leveling-off point. The plots in Figure 3 show that a \( k \) of about 100 or greater is sufficient for the empirical example. By contrast, the top panel of Figure 4 illustrates a time series plot in which drift is in evidence much further into the time series, calling for a \( k \) of 350 or greater. The other diagnostic indicator is mode switching, a sudden jump up or down in the time series. There is no evidence of this in the plots in Figure 3, but the bottom panel of Figure 4 illustrates clear mode switching.

Sometimes an iteration number can be determined beyond which mode switching is no longer in evidence, and a value of \( k \) can be chosen at least equal to this number. More often, mode switching and/or drift that does not resolve itself can be indications of an identification problem. Such problems can usually be dealt with by imposing additional parameter restrictions and rerunning both the EM and DA procedures.

### Hypothesis Tests of Selected Combinations of Parameters

DA produces estimates and standard errors for all LTA parameters as well as for any desired combination of parameters. In this section, to illustrate the possibilities offered by the DA procedure, we present selected results involving several combinations of parameters.

Table 2 shows estimates of Time 1 prevalence rates for alcohol use, delinquency, marijuana use, index offenses, and polydrug use. These were obtained by use of a sum of the appropriate \( \delta \) parameters. For example, to arrive at prevalence rates for polydrug use, \( \delta \) parameters corresponding to both latent classes involving polydrug use were summed: alcohol, delinquency, marijuana, and polydrug use + alcohol, delinquency, marijuana, index, and polydrug use. Thus, in contrast to the latent classes, the prevalence categories do not constitute mutually exclusive groups.

Table 2 also shows the prevalence estimates for three subgroups of interest, characterized by different kinds of change over time. First, the elements on the diagonal of the \( \tau \) matrix were combined to identify a subgroup called *same latent class at both times*. In addition, prevalences were computed for two additional subgroups of interest, representing evolving specialization in different kinds of problem behavior. These prevalences represent a more complex combination involving \( \delta \) and \( \tau \) parameters. The substance use only pathway is characterized by membership in a latent class involving substance use only (either alcohol or alcohol

### Table 2

**Overall Prevalence of Problem Behaviors and Transitions Over Time, Conditional on Gender and Ethnicity**

<table>
<thead>
<tr>
<th>Parameter combination</th>
<th>White Male</th>
<th>White Female</th>
<th>African American Male</th>
<th>African American Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time 1 prevalence of problem behavior</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol use in past 12 months</td>
<td>.627</td>
<td>.623</td>
<td>.442</td>
<td>.461</td>
</tr>
<tr>
<td>Delinquency in past 12 months</td>
<td>.729</td>
<td>.645</td>
<td>.675</td>
<td>.645</td>
</tr>
<tr>
<td>Marijuana use in past 30 days</td>
<td>.185</td>
<td>.179</td>
<td>.187</td>
<td>.110</td>
</tr>
<tr>
<td>Index offense in past 12 months</td>
<td>.177</td>
<td>.068</td>
<td>.244</td>
<td>.128</td>
</tr>
<tr>
<td>Polydrug use in past 30 days</td>
<td>.068</td>
<td>.066</td>
<td>.017</td>
<td>.008</td>
</tr>
</tbody>
</table>

**Transitions over time**

<table>
<thead>
<tr>
<th></th>
<th>White</th>
<th>African American</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same latent class at both times</td>
<td>.499</td>
<td>.514</td>
</tr>
<tr>
<td>Substance use only specialization</td>
<td>.056</td>
<td>.078</td>
</tr>
<tr>
<td>Delinquency only specialization</td>
<td>.113</td>
<td>.070</td>
</tr>
</tbody>
</table>

**Notes.** Table entries reflect combinations of latent transition analysis parameter estimates.
and marijuana) at both times. This combination of parameters represents the sum of the following quantities: the probability of membership in the alcohol latent class at both times ($\delta$ for alcohol multiplied by $\tau$ for alcohol at Time 2 given alcohol at Time 1), the probability of membership in the alcohol latent class at Time 1 followed by the alcohol and marijuana latent class at Time 2 ($\delta$ for alcohol multiplied by $\tau$ for alcohol and marijuana at Time 2 given alcohol at Time 1), and the probability of membership in the alcohol and marijuana latent class at both times ($\delta$ for alcohol and marijuana multiplied by $\tau$ for alcohol and marijuana at Time 2 given alcohol and marijuana at Time 1). Similarly, the delinquency only pathway is characterized by membership in a latent class involving delinquent behavior only (delinquency or delinquency and index) at both times.

Table 3 shows gender and ethnicity effects and the Gender $\times$ Ethnicity interaction for the prevalence estimates of individual problem behaviors and the three subgroups. To test hypotheses of main effects and interactions, we needed to arrive at estimates of the effects (for example, the effect of gender) and the corresponding standard errors. This was done by calculating linear combinations of parameters within each imputation. The main effect for ethnicity was expressed as a difference of proportions: the prevalence for Whites — the prevalence for African Americans. Similarly, the main effect for gender was expressed as the prevalence for male adolescents — the prevalence for female adolescents. An interaction was present if the magnitude of gender differences varied by ethnicity. Thus the interaction effect was defined as the difference of the differences, that is, (White male adolescents — White female adolescents) — (African American male adolescents — African American female adolescents). We performed the hypothesis tests reported here by means of 95% confidence intervals. We tested the null hypothesis that the group differences were zero; when the confidence interval did not contain zero, the null hypothesis was rejected.

Several significant findings emerged. For example, Whites showed higher prevalence than did African Americans on alcohol use, marijuana use, and polydrug use, whereas African Americans showed a higher prevalence of index offenses. The interaction was significant for marijuana use, with a larger gender difference for African Americans than for Whites. Tests related to the subgroups revealed that White adolescents were more likely to be in the same latent class of problem behavior at both times. Male adolescents tended to specialize more often than did female adolescents in delinquency-related behavior, whereas female adolescents tended to specialize more often in substance use. Also, White adolescents were more likely than African Americans to specialize in substance use, whereas African American adolescents were more likely than Whites to specialize in delinquency-related behavior.

**Discussion**

This study has demonstrated that DA, a Bayesian procedure previously applied to missing-data problems, is a flexible and efficient way of testing hypotheses related to individual parameters and combinations of parameters in LCA and LTA. The use of DA to test hypotheses concerning several varieties of combinations of parameters, including group differences, interactions, combinations representing prevalence at a single time, and combinations representing subgroups undergoing particular kinds of change over time, was illustrated. Using this approach to estimation, we reduced complex combinations of parameters to a simple sum or difference within each imputed data set and were combined across imputations in a straightforward manner. Thus the simplicity by which estimates of combinations of parameters and their standard errors can be obtained is an advantage of DA over other methods. DA provides an estimation and hypothesis testing procedure that is very responsive to the scientific questions driving the analysis.

A way to test group differences in proportions that was not illustrated here is the relative risk. The relative risk for a parameter is defined as the probability of an event in one group divided by the probability of the event in another group. For example, the relative risk of membership (say, male adolescents to female adolescents) in a particular latent class can be calculated for each imputed data set by dividing the proportion of male adolescents assigned to that class by the proportion of female adolescents assigned to that class. As with the combinations of parameters demonstrated in this article, any relative risk and its associated standard error can be calculated by use of the imputed data sets. This may be a better approach than tests based on the difference of proportions in cases in which the proportions are very near zero and one (for example, when comparing latent class membership probabilities of .01 and .03 as opposed to comparing probabilities of .41 and .43). An application of DA to test group differences on the basis of the relative risk appears in Lanza and Collins (2002).

The variance estimate of any parameter or combination of parameters is a weighted sum of the within-imputation and between-imputation variances. DA heavily relies on the assumption that draws of imputed data are independent; otherwise, the between-imputation variance, and therefore the total variance estimate, will be too small. In addition, the procedure fails when drift or mode switching occurs between imputations. It falls to the user to diagnose problems related to independence of draws, drift, and mode switching. Fortunately, the autocorrelation functions and time series plots provide simple diagnostic tools for testing these assumptions (see Additional Resources below).

Using DA to test hypotheses based on carefully constructed combinations of parameters, rather than testing all
possible hypotheses based on individual parameters, can reduce the overall number of hypothesis tests. This helps to address an issue that arises in hypothesis testing in LCA as well as in SEM: With many nonindependent hypothesis tests, the possibility of capitalizing on chance can be large, leading to an increase in the overall Type I error rate. Hypothesis tests based on combinations of parameters can also help to decrease the Type II error rate. For certain conditional parameters, such as the \( \tau \) parameters that are conditional on previous latent class membership, standard errors of individual parameters can be quite large. The standard error associated with a combination of parameters is typically considerably smaller than the standard error of any of the individual parameters making up the composite. For example, to test for group differences in the probability of membership in the same latent class at both times, we recommend creating a parameter that reflects the probability of falling on the diagonal of the \( \tau \) matrix, as illustrated above, as a more powerful alternative to looking at each of the individual diagonal elements.

Limitations, Additional Considerations, and Future Directions

All Bayesian procedures, including MCMC approaches to estimating standard errors in latent class models, require the selection of a prior distribution. Although this might seem like a limitation, the choice of a prior rarely has a noticeable effect on the estimates. The choice of a prior should be of increasing concern for smaller sample sizes in conjunction with more complex models. Further, in a case in which substantial prior information is known, for example when class membership probabilities can be approximated on the basis of previous research, incorporating this prior information into the DA procedure may improve the accuracy and stability of parameter estimation. Even adding just a small amount of information can be beneficial in terms of estimation. In the case of sparse data, which can be common in contingency tables used to estimate LCA models and in particular LTA models, even a diffuse prior can improve estimation and convergence (Schafer, 1997). Indeed, in the current study based on a fairly sparse contingency table, the
use of a diffuse prior resulted in fast and smooth convergence. To date, we tend to use diffuse prior information for LTA models. The use of informative priors may prove extremely useful for estimating complex LCA and LTA models with small samples, an area we are actively pursuing.

One important consideration when applying DA to LCA and LTA is the so-called label switching problem (Chung, Loken, & Schafer, 2004). Label switching occurs when the ordering of latent classes, which is always arbitrary, changes across iterations of the DA procedure. It should be watched for particularly when sample sizes are small, when the fit of the model is poor, and when a number of parameters are closer to the middle of the range of possible values rather than to zero and one. Estimates from the $M$ imputations of the DA procedure should be combined only if the class labels are consistent. In some cases it suffices simply to switch the labels of the classes. For example, if there are two classes, A and B, one could identify which class is Class A for each imputed data set and combine the appropriate estimates using the multiple imputation combination rule. In some LCA and LTA applications in which parameter restrictions are used, this label switching problem cannot be dealt with by simply reassigning labels to the classes. Chung et al. (2004) reviewed several alternative ways of dealing with label switching.

Above we mentioned the difficulties associated with relying on the $G^2$ for assessing model fit. DA provides a Bayesian framework for an empirical approach to assessing model fit. The posterior predictive check distribution (Gelman et al., 2004; Rubin, 1984; Rubin & Stern, 1994) is an alternative to the $G^2$ or BIC that is conceptually similar to the parametric bootstrap. It provides an empirical, simulation-based determination of the $p$ value for any relevant test statistic. We plan to explore this approach to model fit in future work.

### Table 3

**Hypothesis Testing for Prevalence Rates and Transitions Over Time: Effects of Gender and Ethnicity and Interaction of Gender and Ethnicity**

<table>
<thead>
<tr>
<th>Parameter combination</th>
<th>Gender effect$^a$</th>
<th>Ethnicity effect$^b$</th>
<th>Interaction$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time 1 prevalence of problem behavior</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol use in past 12 months</td>
<td>.004</td>
<td>.172$^*$</td>
<td>.023</td>
</tr>
<tr>
<td>Delinquency in past 12 months</td>
<td>.073$^*$</td>
<td>.028</td>
<td>.053</td>
</tr>
<tr>
<td>Marijuana use in past 30 days</td>
<td>.022</td>
<td>.037$^*$</td>
<td>−.071$^*$</td>
</tr>
<tr>
<td>Index offense in past 12 months</td>
<td>.109$^*$</td>
<td>−.058$^*$</td>
<td>−.007</td>
</tr>
<tr>
<td>Polydrug use in past 30 days</td>
<td>.005</td>
<td>.055$^*$</td>
<td>−.007</td>
</tr>
<tr>
<td>Transitions over time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same latent class at both times</td>
<td>−.018</td>
<td>.056$^*$</td>
<td>.023</td>
</tr>
<tr>
<td>Substance use only pathway</td>
<td>−.024$^*$</td>
<td>.034$^*$</td>
<td>.016</td>
</tr>
<tr>
<td>Delinquency only pathway</td>
<td>.030$^*$</td>
<td>−.036$^*$</td>
<td>.050$^*$</td>
</tr>
</tbody>
</table>

$^a$ Main effect of gender, based on the difference of proportions (probability in male adolescents − probability in female adolescents). $^b$ Main effect of ethnicity, based on the difference of proportions (probability in Whites − probability in African Americans). $^c$ Gender × Ethnicity interaction, parameterized as (White male adolescents − White female adolescents) − (African American male adolescents − African American female adolescents). $^* p < .05$.

### Additional Resources

Software, additional applications, and technical references for using DA for parameter estimation and hypothesis testing in LCA and LTA are available on our Web site at [http://methodology.psu.edu](http://methodology.psu.edu). This site includes the software used to perform the analyses presented here, WinLTA, and an associated software manual (Collins, Lanza, & Schafer, 2002) that includes more detail about the use of parameter restrictions as well as numerous empirical examples. The software is designed to be friendly to researchers who are novices to DA; it uses a diffuse prior by default and automatically produces diagnostic autocorrelation and time series plots. The Web site also includes a variety of technical reports and utilities of interest to LCA and LTA users. All downloads are free of charge. Appendix 1 includes SAS Version 8.2 code to calculate point estimates for combinations of parameters, and Appendix 2 includes SAS Version 8.2 code to calculate differences of proportions and 95% confidence intervals for combinations of parameters. These appendixes are available at [http://dx.doi.org/10.1037/1082-989X.10.1.84.supp](http://dx.doi.org/10.1037/1082-989X.10.1.84.supp).

### References


Received August 3, 2002
Revision received August 16, 2004
Accepted August 16, 2004 ■

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at http://watson.apa.org/notify/ and you will be notified by e-mail when issues of interest to you become available!