

# Latent-Class Logistic Regression: Application to Marijuana Use and Attitudes Among High-School Seniors

Hwan Chung, Brian P. Flaherty, and Joseph L. Schafer \*

## Abstract

Analyzing the use of marijuana is challenging in part because there is no widely accepted single measure of individual use. Similarly, there is no single response variable that effectively captures attitudes toward its social and moral acceptability. One approach is to view the joint distribution of multiple use and attitude indicators as a mixture of latent classes. Pooling items from the annual Monitoring the Future surveys of American high-school seniors from 1977 to 2001, we find that marijuana use and attitudes is well summarized by a four-class model. Secular trends in class prevalences over this period reveal major shifts in use and attitudes. Applying a multinomial logistic model to the latent response, we investigate how class membership relates to demographic and lifestyle factors, political beliefs and religiosity over time. Inferences about the parameters of the latent class logistic model are obtained by a combination of maximum likelihood and Bayesian techniques.

**KEY WORDS:** Categorical data; Data augmentation; Finite mixture; MCMC; Multiple imputation

---

\*Hwan Chung is Research Associate (E-mail: hwan@stat.psu.edu), Brian P. Flaherty is Research Associate, The Methodology Center, The Pennsylvania State University, University Park, PA 16802, Joseph L. Schafer is Associate Professor, Department of Statistics, The Pennsylvania State University, University Park, PA 16802. This research was supported by the National Institute on Drug Abuse, grant 1-P50-DA10075.

## 1. INTRODUCTION

Many theories of substance use and dependence describe behavior in terms of population classes or developmental stages. For example, the acquisition of nicotine dependence is often depicted as a behavioral sequence that includes initial trying of tobacco, experimentation and regular smoking prior to dependence (Leventhal and Cleary 1980; Flay 1993; Mayhew, Flay and Mott 2000). Additional types of smokers have been identified such as non-dependent irregular smokers (called “chippers”) (Shiffman, Kassel, Paty and Gnys 1994). Stage-sequential models have been used to describe substance-use onset (Lanza and Collins 2002) and to test the gateway hypothesis, which regards marijuana as a potential conduit to more harmful substances such as heroin or cocaine (Graham, Collins, Wugalter, Chung and Hansen 1991; Collins 2002). A common theme of these theories is that, at any moment, individuals are placed into distinct categories or groups rather than along a continuum.

Latent-class (LC) analysis (Goodman 1974; Clogg and Goodman 1984) is ideally suited to theories of this type. In LC analysis, relationships among categorical variables are explained by positing the existence of an unobserved or latent classifier that makes them conditionally independent. If an LC model fits well, it may provide a parsimonious and intuitively appealing summary of the cell frequencies in a high-dimensional contingency table and reveal features that are not apparent from an item-by-item analysis. Another advantage of an LC analysis is that, rather than taking each response at face value, the items are treated as fallible indicators of unseen true states. Logical inconsistencies—e.g. a subject claiming use in the last 30 days but no lifetime use—are allowed by the model and do not need to be edited out of the data. In recent years, LC modeling has become an increasingly popular strategy for quantifying measurement error in substance use self-reports (Collins, Flaherty and Colby 2002). Indeed, Biemer and Wiesen (2002) have demonstrated that LC models may yield plausible estimates of the rates of misreporting comparable to what one would get from an expensive process of reinterview. Recent extensions to the traditional LC model allow covariates to predict class membership through binary or polytomous logistic regression (Bandeem-Roche, Miglioretti, Zeger and Rathouz 1997).

In this article, we apply LC logistic regression to items on self-reported marijuana use and attitudes in representative samples of American high-school seniors from 1977 to 2001. The data were drawn from

Monitoring the Future (MTF), an ongoing survey that explores changes values, behaviors, and lifestyle orientations of contemporary American youth (Johnston, Bachman and O'Malley 2001). Analyses of MTF show that rates of marijuana use among seniors have fluctuated dramatically over the last three decades. Prevalence rose during the late 1960s and throughout most of 1970s, declined steadily and substantially throughout the 1980s, and began to rise again during the 1990s (Johnston, O'Malley and Bachman 2002). The use of marijuana is known to be strongly correlated with individuals' attitudes toward its social and moral acceptability (Johnston 1982; Johnston 1985). If attitudinal measures are treated as exogenous predictors, interpretation of the estimated relationships is clouded by the strong possibility that patterns of use may influence attitudes. Rather than attempting to disentangle use and attitudes, we use LC analysis to examine how they have jointly co-occurred and changed over time.

Our work reveals that throughout this period the population has been dominated by four groups: non-users who disapprove of use, non-users who are somewhat approving, experimenters who disapprove of regular use, and regular users who approve of regular use. Previous analyses from MTF revealed significant relationships between marijuana use and a variety of other factors including demographic characteristics, lifestyle and beliefs (Bachman, Johnston and O'Malley 1998). Accordingly, we use these as covariates to jointly predict the prevalence of group membership. We show how the rates of group membership have historically changed and examine group composition in terms of the covariates.

In most applications of LC models to date, parameters have been estimated by maximum likelihood (ML). ML may lead to reasonable estimates but often fails to provide useful measures of uncertainty. Even with a large sample and well fitting model, ML estimates for some quantities may lie on or near a boundary of the parameter space, rendering the usual Hessian-based standard errors ineffective (Chung 2003). As an alternative to ML, Bayesian analysis by Markov chain Monte Carlo (MCMC) has been applied to LC models by Hoijtink (1998), Garrett and Zeger (2000) and Garrett, Eaton and Zeger (2002). We extend the work of these authors by showing how to incorporate covariates and account for missing items. Moreover, we demonstrate how the technique of multiple imputation (Rubin 1987) can be applied to latent-class indicators to simplify aspects of modeling, inference, and diagnosis.

## 2. A LATENT-CLASS LOGISTIC REGRESSION MODEL

Let  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iM})$  be a vector of  $M$  survey items for  $i$ th individual, where variable  $Y_{im}$  takes possible values  $1, 2, \dots, r_m$ . The basic idea of the LC model is that associations among items are assumed to arise because the population is composed of different unseen classes. Let  $l_i = 1, 2, \dots, L$  be the latent-class membership of the  $i$ th individual, and let  $I(y = k)$  denote the indicator function which takes the value 1 if  $y$  is equal to  $k$  and 0 otherwise. If  $l_i$  were observed, the joint probability that  $i$ th individual belongs to class  $l$  and provides responses  $\mathbf{y}_i = (y_{i1}, \dots, y_{iM})$  would be

$$\Pr(\mathbf{Y}_i = \mathbf{y}_i, l_i = l) = \gamma_l \prod_{m=1}^M \prod_{k=1}^{r_m} \rho_{mk|l}^{I(y_{im}=k)},$$

where  $\gamma_l = \Pr(l_i = l)$  denotes the marginal rate of  $l$ th class membership in the population, and  $\rho_{mk|l} = \Pr(Y_{im} = k \mid l_i = l)$  represents the probability of response  $k$  to  $m$ th item given class membership in  $l$ . Therefore, the marginal probability of item responses without regard for the unseen class membership is

$$\Pr(\mathbf{Y}_i = \mathbf{y}_i) = \sum_{l=1}^L \gamma_l \prod_{m=1}^M \prod_{k=1}^{r_m} \rho_{mk|l}^{I(y_{im}=k)}. \quad (1)$$

Here we have assumed that the items  $Y_{i1}, \dots, Y_{iM}$  are conditionally independent or unrelated within each class of  $l_i$ . This assumption, called “local independence” by Lazarsfeld and Henry (1968), is the crucial feature of the LC model that allows us to draw inferences about the unseen class variable.

A natural way to extend the LC model in (1) is to include stratification or grouping variables and investigate whether a common latent-class structure holds across groups. Clogg and Goodman (1984) introduced an observed variable  $g_i = 1, 2, \dots, G$  and allowed the  $\gamma$  and/or  $\rho$ -parameters to vary across groups. This approach, which represents a first attempt to incorporate covariates into LC analysis, is obviously limited to situations where the grouping variable is discrete and the number of groups is small. Dayton and Macready (1988) proposed a more general way to incorporate subject specific covariates, allowing them to influence  $\gamma$  through a logistic link. They computed ML estimates by a computationally intensive simplex method, whereas Bandeen-Roche et al. (1997) applied an EM algorithm. Pfeiffermann, Skinner and Humphreys (1998) described a generalization where not only  $\gamma$  but also  $\rho$  is functionally related to covariates. ML routines for these models have been implemented in the software packages Mplus (Muthén and Muthén 1998) and Latent GOLD (Vermunt and Magidson 2000). Allowing  $\rho$  to depend on covariates can be problematic, because the

covariates may introduce associations among the items within a latent class, violating local independence. In such a model, the latent class structure changes as covariates are added or deleted. If  $\rho$  is allowed to covary with predictor variables, then the composition of the latent classes is no longer constant over the population, and the meaning of  $\gamma$  becomes unclear (Clogg and Goodman 1984).

We apply a generalized version of the Dayton and Macready (1988) model in which a grouping variable  $g_i$  is allowed to affect either the values of  $\rho$ ,  $\gamma$  or both, but  $\rho$  is not allowed to change with covariates. Let  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})'$  denote a vector of covariates for  $i$ th individual. Implicitly conditioning on  $\mathbf{x}_1, \dots, \mathbf{x}_n$  and  $g_1, \dots, g_n$ , our model is

$$\Pr(\mathbf{Y}_1 = \mathbf{y}_1, \dots, \mathbf{Y}_n = \mathbf{y}_n) = \prod_{g=1}^G \prod_{i \in g} \sum_{l=1}^L \gamma_{l|g}(\mathbf{x}_i) \prod_{m=1}^M \prod_{k=1}^{r_m} \rho_{mk|lg}^{I(y_{km}=k)}, \quad (2)$$

where  $\prod_{i \in g}$  denotes the product over the set of individuals in subgroup  $g$ . Rates of class membership are related to covariates by

$$\begin{aligned} \gamma_{l|g}(\mathbf{x}_i) &= \Pr(l_i = l \mid \mathbf{x}_i, g_i = g) \\ &= \frac{\exp\{\mathbf{x}_i' \boldsymbol{\beta}_{l|g}\}}{1 + \sum_{j=1}^{L-1} \exp\{\mathbf{x}_i' \boldsymbol{\beta}_{j|g}\}} \end{aligned} \quad (3)$$

for  $l = 1, \dots, L-1$ , with  $\gamma_{L|g}(\mathbf{x}_i)$  available as  $1 - \gamma_{1|g}(\mathbf{x}_i) - \dots - \gamma_{L-1|g}(\mathbf{x}_i)$ . In (3),  $\boldsymbol{\beta}_{l|g} = (\beta_{1l|g}, \dots, \beta_{pl|g})'$  is a  $p \times 1$  vector of logistic-regression coefficients for subgroup  $g$  influencing the log-odds that an individual falls into class  $l$  relative to the baseline class  $L$ .

We compute ML estimates for the unknown  $\rho$  and  $\beta$ -parameters in (2) by a hybrid procedure which combines iterations of EM with steps of Newton-Raphson (Chung 2003). This algorithm is being incorporated into the next version of the free software package WinLTA (Collins, Lanza, Schafer and Flaherty 2002), which will be released in 2005. Our algorithm allows missing values to occur among the items in  $\mathbf{y}_i$ , provided that they are missing at random (Rubin 1976). Upon convergence, standard errors for the estimated parameters are obtained by inverting the Hessian matrix of the observed-data loglikelihood. In many cases, however, this procedure fails because the loglikelihood is not concave. If any of the estimated  $\rho$ 's is close to zero or one—which happens when the item in question is a strong predictor of class membership—then we cannot obtain standard errors either for the  $\rho$ 's or the  $\beta$ 's. Even when the loglikelihood is concave at the solution, the Hessian-based covariance matrix may be a poor summary of the data's evidence about the parameters.

If the sample is not large enough, or if the variation in  $\rho$ -parameters across the latent classes is not strong enough, the loglikelihood can be nearly flat in certain directions, again rendering Hessian-based standard errors unreliable. This stems from the fact that loglikelihood function from this finite mixture is invariant with respect to the  $L!$  possible reorderings of the class labels. Confidence regions sometimes extend into portions of the parameter space where the class labels have permuted from their order at the current mode, making the interpretation of the region unclear (Chung 2003; Chung, Loken and Schafer 2004). Because of these common difficulties with ML, we have also developed Bayesian alternatives, which we now describe.

### 3. BAYESIAN APPROACHES BASED ON MCMC

Bayesian methods for LC models have been described by Hoijtink (1998), Garrett and Zeger (2000) and Lanza, Collins, Schafer and Flaherty (in press). Without covariates, a simple Markov chain Monte Carlo (MCMC) algorithm may be implemented as an iterative two-step procedure which can be regarded as a form of data augmentation (Tanner and Wong 1987) or Gibbs sampling (Gelfand and Smith 1990). Covariates introduce a minor complication because there is no simple conjugate prior family for the coefficients of a multinomial logistic model. We overcome this difficulty by embedding steps of a Metropolis algorithm (Metropolis, Rosenbluth, Rosenbluth, Teller and Teller 1953) for the  $\beta$ 's into the Gibbs sampler (Robert and Casella 2004).

Let  $\mathbf{z}_i$  be a  $L$ -dimensional vector of binary variables indicating the latent class to which  $i$ th individual belongs. That is, if individual  $i$  belongs to class  $l$ , then  $\mathbf{z}_i$  contains a 1 in position  $l$  and 0 in every other position. In the first step of our MCMC procedure—the Imputation or I-step—we generate a random draw for each  $\mathbf{z}_i$  given the observed data  $\mathbf{y}_i$  and current parameter guesses. In the second step—the Posterior or P-step—we draw new random values for the parameters from the augmented-data posterior distribution which regards the latent-class membership indicators  $\mathbf{z}_i$  as known. Repeating this two-step procedure creates a sequence of iterates converging to the stationary observed-data posterior distribution. Details of this two-step procedure are given below.

In the I-step, given current simulated parameter values, we calculate the posterior probabilities of class

membership using

$$\begin{aligned}\delta_{il|g} &= \Pr(l_i = l \mid \mathbf{Y}_i = \mathbf{y}_i, g_i = g) \\ &= \frac{\gamma_{l|g}(\mathbf{x}_i) \prod_{m=1}^M \prod_{k=1}^{r_m} \rho_{mk|lg}^{I(y_{im}=k)}}{\sum_{j=1}^L \gamma_{j|g}(\mathbf{x}_i) \prod_{m=1}^M \prod_{k=1}^{r_m} \rho_{mk|jg}^{I(y_{im}=k)}}\end{aligned}\quad (4)$$

for  $l = 1, \dots, L$ . Then we draw  $\mathbf{z}_i$  from Multinomial( $1, \delta_{i|g}$ ) independently for all individuals in group  $g$ , where  $\delta_{i|g} = (\delta_{i1|g}, \dots, \delta_{iL|g})$ .

Once class membership has been imputed, the augmented-data likelihood factors into independent likelihood functions for the  $\rho$  and  $\beta$ -parameters. Let  $\mathbf{B}$  represent coefficients vectors  $(\beta_{1|1}, \dots, \beta_{L-1|G})$ , and let  $\rho$  denote the vectorized item-response probability containing all  $\rho$ -parameters. The augmented data-posterior given class membership may be expressed as

$$\Pr(\mathbf{B}, \rho \mid \mathbf{y}, \mathbf{z}) \propto \left[ \Pr(\mathbf{B}) \prod_{g=1}^G \prod_{i \in g} \prod_{l=1}^L \gamma_{l|g}(\mathbf{x}_i)^{z_{il}} \right] \left[ \Pr(\rho) \prod_{g=1}^G \prod_{l=1}^L \prod_{m=1}^M \prod_{k=1}^{r_m} \rho_{mk|lg}^{n_{mk|lg}} \right], \quad (5)$$

where  $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$ ,  $\mathbf{z} = (\mathbf{z}_1, \dots, \mathbf{z}_n)$ ,  $n_{mk|lg} = \sum_{i \in g} I(y_{im} = k, z_{il} = 1)$ , and  $\Pr(\mathbf{B})$  and  $\Pr(\rho)$  are the priors for  $\mathbf{B}$  and  $\rho$ , respectively.

In the P-step, we draw new random values for  $\rho$  and  $\mathbf{B}$  independently from (5). Applying a Jeffreys prior to  $\rho_{m|lg} = (\rho_{m1|lg}, \dots, \rho_{mr_m|lg})$ , new random values for  $\rho_{m|lg}$  are drawn from Dirichlet( $n_{m1|lg} + 1/2, \dots, n_{mr_m|lg} + 1/2$ ) independently for  $m = 1, \dots, M$ ,  $l = 1, \dots, L$ , and  $g = 1, \dots, G$ . A draw from Dirichlet distribution can be obtained by normalizing a vector of  $r_m$  independent gamma variates (Kennedy and Gentle 1980).

For the  $\beta$ -parameters, we apply an improper uniform prior and generate  $\beta_g = (\beta_{1|g}, \dots, \beta_{L-1|g})$  by the following Metropolis algorithm. At iteration  $t$ , a candidate for the next  $\beta_g$  is sampled from a multivariate Student's  $t$  proposal distribution  $t(\beta_g^{(t)}, c^2 \Sigma)$  with 4 degrees of freedom, where  $\beta_g^{(t)}$  is the parameter value at iteration  $t$ . Following the advice of Gelman, Carlin, Stern and Rubin (2004) for an efficient Metropolis jumping rule, we set  $c \approx 2.4\sqrt{p(L-1)}$ . For our guess of  $\Sigma$ , we invert the  $\beta_g$ -submatrix of the Hessian evaluated at the ML estimates  $\hat{\mathbf{B}}$  and  $\hat{\rho}$ ,

$$\Sigma = - \left( \frac{\partial^2 \ell}{\partial \beta_g \partial \beta_g'} \right)^{-1} \Bigg|_{\mathbf{B}=\hat{\mathbf{B}}, \rho=\hat{\rho}}, \quad (6)$$

where  $\ell$  is the logarithm of the likelihood function defined in (2). (This inverse typically exists even if  $\hat{\rho}$  lies

on a boundary, because we are ignoring the cross-derivatives with respect to  $\mathbf{B}$  and  $\rho$ .) Expressions for the elements of the Hessian are similar to those given by Bandeen-Roche et al. (1997). The candidate point  $\beta_g^c$  is then accepted with probability

$$\alpha(\beta_g^{(t)}, \beta_g^c) = \min \left( 1, \prod_{i \in g} \prod_{l=1}^L \left[ \frac{\gamma_l(\mathbf{x}_i) | \beta_g^c}{\gamma_l(\mathbf{x}_i) | \beta_g^{(t)}} \right]^{z_{il}} \right).$$

Extending this procedure to a dataset with missing items in  $\mathbf{y}_i$  is straightforward. In the I-step, given the current parameter guesses, the posterior probabilities are calculated only from the observed part of  $\mathbf{y}_i$ ,

$$\begin{aligned} \delta_{il|g}^{obs} &= \Pr(l_i = l | \mathbf{Y}_{obs_i} = \mathbf{y}_{obs_i}, g_i = g) \\ &= \frac{\gamma_{l|g}(\mathbf{x}_i) \prod_{m \in obs_i} \prod_{k=1}^{r_m} \rho_{mk|lg}^{I(y_{im}=k)}}{\sum_{j=1}^L \gamma_{j|g}(\mathbf{x}_i) \prod_{m \in obs_i} \prod_{k=1}^{r_m} \rho_{mk|jg}^{I(y_{im}=k)}}, \end{aligned}$$

where  $obs_i$  denotes the sets of items responded to by  $i$ th individual. The class membership  $z_i$  is drawn from Multinomial( $1, \delta_{i|g}^{obs}$ ) independently for all individuals in group  $g$ . In the I-step, we also generate the missing part of each  $\mathbf{y}_i$  as follows. Suppose the  $m$ th item is missing for individual  $i$  who belongs to group  $g$  and class  $l$ ; then  $y_{im}$  is randomly drawn from Multinomial( $1, \rho_{m|lg}$ ), where  $\rho_{m|lg}$  is a set of sample from the previous iteration. In the P-step, the updated  $\mathbf{B}$  and  $\rho$  are simulated in the manner described above, treating the imputed values for the missing elements of  $\mathbf{y}_i$  as known.

In a typical application of MCMC, the analyst runs the algorithm for a burn-in period to eliminate dependence on the starting values. After the burn-in, averaging the output stream of simulated parameters produces estimates for the posterior means and variances (Tierney 1994). Various methods for choosing the length of the series and the burn-in period have appeared in the literature (Geweke 1992; Roberts 1992; Gelman and Rubin 1992; Best, Cowles and Vines 1995). We use time-series plots and autocorrelation functions to visually monitor the behavior of output values from MCMC and to confirm our choice for the length of the burn-in period. After burn in, we typically use 10,000 or more cycles of MCMC to estimate posterior means and variances.

As an alternative to averaging over the simulated parameters, the output from MCMC may also be summarized through multiple imputation (MI) (Rubin 1987). In MI, we retain a few (say, 10–20) simulated draws of the latent class variables  $l_1, \dots, l_n$  and missing parts of  $\mathbf{y}_1, \dots, \mathbf{y}_n$  from the I-steps, spacing them enough cycles apart to ensure that they are essentially independent. Treating these imputed values as



known, we compute point and variance estimates for the  $\rho$ 's and  $\beta$ 's by standard complete-data methods for proportions and logistic regression coefficients. The results are then combined using Rubin's (1987) rules for scalar estimands. Imputations for the latent variables and missing items are quite handy for performing exploratory analyses and testing hypotheses about parameters that may not be readily available from the output stream. With MI, for example, we can directly examine the (no longer latent) classes to see how they differ with respect to the distributions of key covariates. We can also examine odds ratios among the manifest items within classes to diagnose departures from the assumption of local independence. The relative merits of MI versus direct simulation of parameters by MCMC are discussed by Schafer (1997, Sec. 4.5.1).

A potentially troublesome aspect of MCMC for LC models is label switching. As mentioned previously, the likelihood function from an LC model is invariant to permutations of the class labels. If the priors are also symmetric with respect to class labels, the posterior distribution will be similarly invariant. This invariance makes the output stream for parameters difficult to interpret if the class labels switch during the MCMC run. Time-series plots of the output stream must be examined for evidence of label switching. Label switching did not occur in our analyses of data from Monitoring the Future, presumably because our sample was very large. A variety of strategies for resolving problems related to label switching are reviewed by Chung et al. (2004).

#### 4. APPLICATION TO MONITORING THE FUTURE

##### 4.1 Data

Monitoring the Future (MTF) has been conducted annually since 1975. Each year, a sample of 14,000–18,000 students is selected from 125–140 schools to represent a cross-section of high-school seniors in the 48 contiguous United States. Each participant receives one of five different questionnaire forms with items pertaining to drug use, attitudes toward government, social institutions, race relations, changing roles of women, educational aspirations, occupational aims, and marital and family plans.

In MTF, marijuana use is retrospectively measured by self-reported frequency of use over various periods of time. Subjects report how often they have used marijuana in their lifetimes (*Life*), over the last twelve months (*12Mo*) and over the last 30 days (*30Da*). Responses fall into ordered categories ranging from none to heavy use. For these items, we suspect that recall difficulty may introduce substantial rates of misreporting

at the high end. To help mitigate the effects of response error, we reduced each of these items to a binary indicator, denoting use for a given period by 1 and non-use by 2. A fourth item asks how likely it is that the subject will use marijuana in the next twelve months (*Nxt12Mo*), with four possible responses ranging from “definitely will” to “definitely will not”; this item was recoded to 1 for “will use” and 2 for “will not use.”

Like marijuana use itself, no single item effectively summarizes an individual’s attitudes toward use. MTF’s Form 3 included three items for attitudes. Subjects were asked whether they disapprove of people trying marijuana once or twice (*TryMJ*), smoking marijuana occasionally (*OccUse*) and smoking marijuana regularly (*RegUse*). Once again, the possible responses to these items were simplified to binary values of 1 = don’t disapprove or 2 = disapprove.

Our logistic LC regression models are based on these seven dichotomized use and attitude items. Our covariates included sex (male, female), race (white, non-white), political beliefs (conservative, moderate, liberal), importance of religious beliefs (not important, important, very important), number of skipped classes, grades and number of evenings out per week. Dummy variables were created for sex, race, political and religious beliefs. Including an intercept term, the full covariate vector  $x_i$  for each individual had length 10.

Because of differences in questionnaire formats in the first two years of MTF (1975 and 1976), we omitted those years and began our analyses with 1977. Year ( $1 = 1977, \dots, 25 = 2001$ ) was specified as a grouping variable  $g$ , and the logistic coefficients  $\beta_{l|g}$  were estimated separately for each year  $g = 1, \dots, 25$ . In some cases we allowed the  $\rho$ -parameters to vary by year, but in other cases we constrained them to be equal across years.

Our estimation procedures allow for missing values in the use and attitude items but not in the covariates. After combining the samples from 1977 to 2001 and removing individuals with missing covariates, our pooled dataset contained 40,690 individuals. Deleting subjects may introduce bias if those with missing covariates are systematically different from the complete cases (Little and Rubin 2002). These biases may be reduced by reweighting the complete cases by estimated inverse probabilities of response (Robins, Rotnitzky and Zhao 1994). In this case, however, a dearth of strong predictors for nonresponse would make the effect of weighting adjustments inconsequential.

Table 1: Fit statistics and degrees of freedom for a series of unconstrained latent-class models without covariates

Number of classes	$-2$ loglikelihood	$G^2$	df
2	222066.89	28571.91	2800
3	207345.49	13850.51	2600
4	197589.88	4094.90	2400
5	196395.72	2900.74	2200

#### 4.2 Unconstrained LC Analyses

The first and most crucial step in an LC regression analysis is to choose an appropriate class structure. As shown by Bandeen-Roche et al. (1997), the model (2) has an appealing marginalization property: averaging over an arbitrary distribution for the covariates  $\mathbf{x}_i$  produces a conventional LC model with the same number of classes and identical values for the  $\rho$ 's. Therefore, we do not need to consider covariates when selecting the number of latent classes or interpreting them. In the absence of strong prior beliefs, the number of classes is usually chosen to strike a balance among parsimony, fit and interpretability. If the number is too small, the model may give poor fit to the joint distribution of observed item responses. If the number is too large, we may find that the  $\rho$ -parameters for some classes are too similar to attach substantively different meanings to them, or that some classes have estimated prevalence rates that are nearly zero.

Because we have pooled samples over a 25-year period, there is a strong possibility that the number of latent classes or their  $\rho$ -parameters could vary over time. To allow for that possibility, we fit a sequence of unconstrained LC models without covariates that allow the  $\rho$ -parameters to vary by year. We started with two classes and increased to three, four, and five. Loglikelihood values, deviance statistics ( $G^2$ ) and degrees of freedom for these models are shown in Table 1. Fit statistics such as  $G^2$  must be interpreted carefully when the data contain missing values. As noted by Little and Rubin (2002, Ch. 13), fit statistics are aggregated over the cross-classified contingency tables for all missingness patterns appearing in the dataset. Models that fit well may have large values of  $G^2$ , because this statistic also detects departures from the (usually implausible) hypothesis of missing completely at random. To overcome this difficulty, we adjusted each  $G^2$

statistic in Table 1 by removing the portion corresponding to the saturated model. Details of this adjustment are described by Schafer (1997, Sec. 8.5.2).

Evaluating the significance of these  $G^2$  statistics by the usual  $\chi^2$  approximation is not appropriate because, despite the large sample size, estimated expected counts for many of the  $25 \times 2^7 = 3,200$  cells are close to zero. Differences in  $G^2$  should not be compared to  $\chi^2$  distributions either, because likelihood-ratio statistics for testing the fit of an  $L$ -class model against an  $(L + 1)$ -class alternative do not have limiting chisquare distributions (Lindsay 1995). Models with different numbers of classes are typically compared using penalized likelihood criteria such as AIC (Akaike 1987) and BIC (Schwarz 1978) or with posterior predictive checks (Rubin and Stern 1994; Garrett and Zeger 2000). Nevertheless, Table 1 shows large drops in  $G^2$  as the the number of classes increases from two to three and four, but a much smaller drop from four classes to five. As suggested by Garrett and Zeger (2000), we also compared the estimated probabilities of major response patterns under the four and five-class models by a posterior predictive check distribution and found that five classes did not fit much better than four.

Another reason for preferring the four-class model is that its  $\rho$ -parameters are more stable over time. Temporal stability of these parameters is necessary to assign a clear interpretation to trends in class prevalence and composition. If they vary strongly over time, then the meaning of a statement such as “Membership in Class 1 increased from 40% in 1980 to nearly 70% by 1992” becomes dubious, because “Class 1” may have very different meanings in 1980 and 1992. Plots of the estimated  $\rho$ 's by year for each use and attitude item are shown for the four-class model in Figure 1 and for the five-class model in Figure 2. The five-class versions show large fluctuations from one year to the next. Even more troubling are the many places where the plotted lines cross each other, because the interpretation of the classes is largely derived from their ordering with respect to these parameters. In contrast, the plots for the four-class model show less variation and fewer crossings.

Examining Figure 1 closely, however, we do see a few troublesome features. Consider the response probability for the regular use item (*RegUse*) in Class 4. Members of Class 4 have a high probability (almost 80%) of endorsing regular use at the beginning of the study period, but it drops to nearly 40% by 1990 and then rises again to about 60% by the end of the study. The item measuring attitude toward occasional use

Figure 1: Estimated item-response probabilities under the unconstrained four-class model.

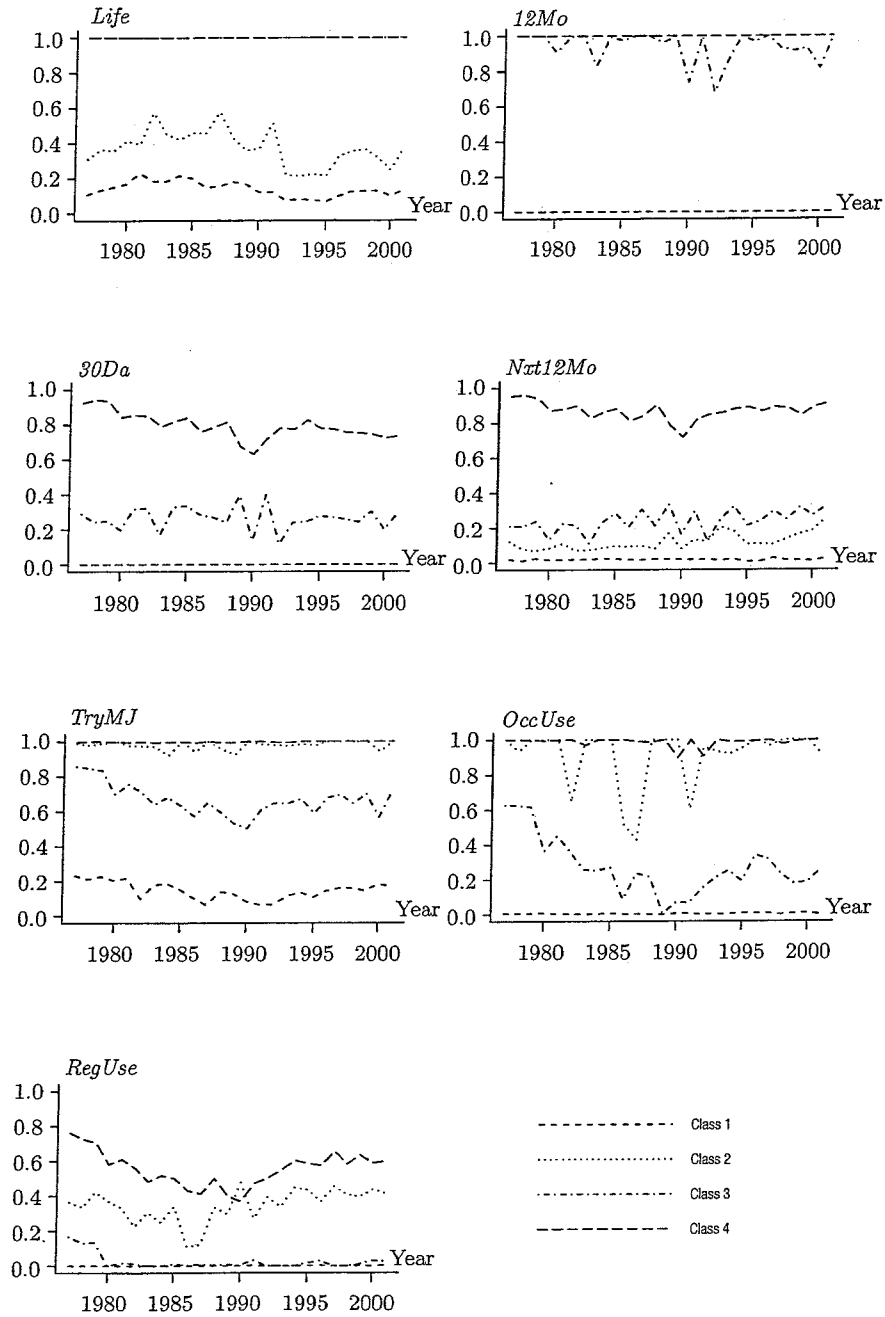
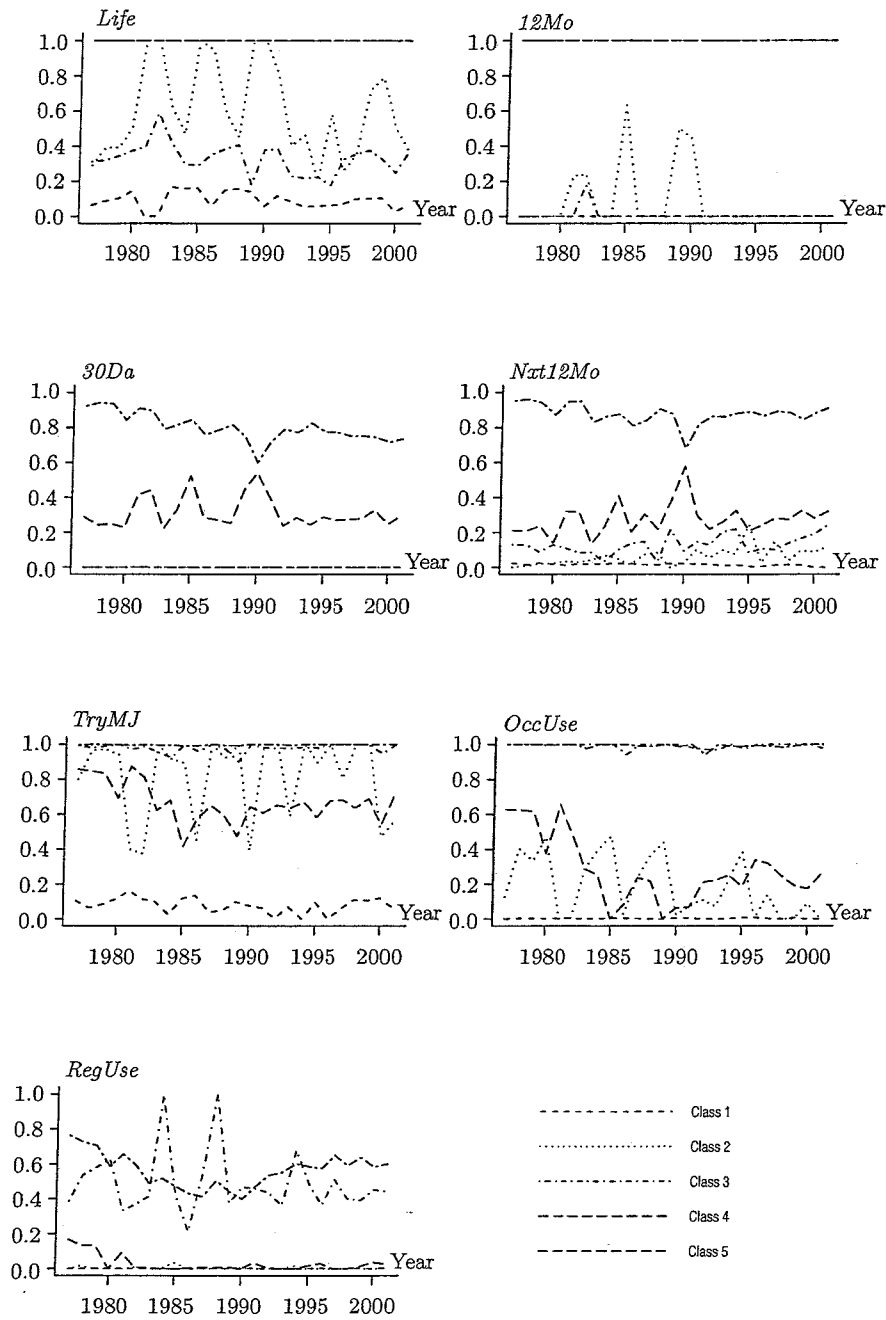


Figure 2: Estimated item-response probabilities under the unconstrained five-class model.



(*OccUse*) has two unusual aspects. First, the probability of endorsing occasional use in Class 3 drops from about 60% in 1977 to almost zero by 1990 before rising again. Second, the rate of endorsement in Class 2 is generally high, but there are four years in which it suddenly drops. Aside from these few difficulties, the overall picture that emerges from Figure 1 is that the four-class model is rather well behaved.

#### 4.3 Four-class Model with Equality Constraints

To smooth out the temporal fluctuations, we fit another four-class model in which the  $\rho$ -probabilities are constrained to be equal for all 25 years. The  $G^2$  for this constrained model is 5981.59 with 3072 df. Comparing this to the unconstrained four-class model reported in Table 1, we find a  $G^2$  difference value of 1886.69 with 672 df. Although this drop appears to be statistically significant—which is not surprising, given the large sample size—the relative stability of the parameter estimates gives us confidence that this constrained model captures the most essential features of the class structure over the 25-year period. For example, Figure 3 displays for one of the items (*Nxt12Mo*) the point and 95% interval estimates for the unconstrained  $\rho$ 's for all years superimposed over the constrained estimates. For each class, the constrained estimate and interval endpoints appear as horizontal lines, and the dots with vertical lines represent the unconstrained estimates and intervals. A large majority of the intervals from the unconstrained model overlap the estimates from the constrained model. Using a constrained model, even if it ignores some of the fine details of the item distributions for some years, it is a sensible way to strengthen inferences about the other model parameters and facilitate comparisons of prevalence rates and covariate effects across years. Although we acknowledge some decrement in fit, we gain a great deal in terms of interpretability.

Estimates and 95% intervals for all of the  $\rho$ -parameters from the constrained four-class model are shown in Table 2. These were computed by three methods: maximum-likelihood using a combination of Newton-Raphson and EM (ML), our data augmentation MCMC algorithm taking long-run averages of the simulated parameters (DA), and multiple imputation of the latent classes and missing items (MI). Point estimates from the three methods are nearly indistinguishable. Standard errors are not available for the ML method; the Hessian matrix could not be inverted, because some of the estimates lie at or near boundaries. The Bayesian intervals from DA and MI are nearly identical and very narrow, due to the large sample size. The value of these intervals seems limited, given that there is some fluctuation in the  $\rho$ 's over time. But the point

Figure 3: Constrained and unconstrained estimates and 95% intervals for item-response probabilities for *Nxt12Mo* in (a) Class 1, (b) Class 2, (c) Class 3, and (d) Class 4.

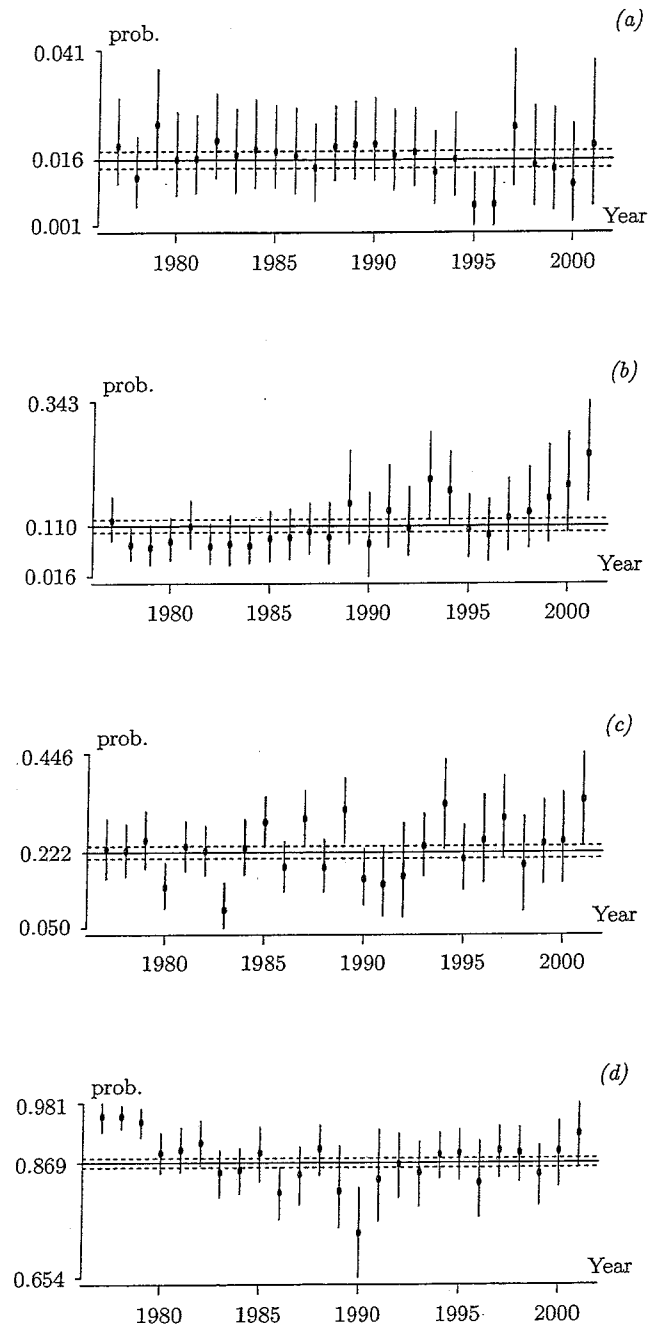




Table 2: Estimated item-response probabilities and 95% confidence intervals of “1=any use” and “1=don’t disapprove” from maximum likelihood (ML), data augmentation (DA), and multiple imputation (MI).

	Class	Use of Marijuana				Attitudes toward Marijuana		
		Life	12Mo	30Da	Nxt12Mo	TryMJ	OccUse	RegUse
ML	1	.128	.000	.000	.016	.131	.003	.001
	2	.350	.000	.000	.110	.980	.890	.338
	3	1.00	.867	.245	.225	.682	.277	.009
	4	1.00	1.00	.818	.869	.998	.996	.586
DA	1	.127 (.121, .133)	.000 (.000, .000)	.000 (.000, .000)	.016 (.014, .018)	.131 (.124, .138)	.004 (.002, .005)	.001 (.000, .001)
	2	.345 (.325, .365)	.001 (.000, .003)	.000 (.000, .001)	.110 (.099, .122)	.981 (.974, .988)	.897 (.854, .934)	.343 (.319, .367)
	3	1.00 (1.00, 1.00)	.860 (.841, .880)	.242 (.228, .255)	.222 (.208, .236)	.682 (.669, .695)	.277 (.261, .293)	.009 (.006, .012)
	4	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)	.817 (.808, .827)	.869 (.860, .877)	.998 (.996, .999)	.995 (.991, .997)	.584 (.573, .595)
MI	1	.127 (.121, .133)	.000 (.000, .000)	.000 (.000, .000)	.016 (.014, .018)	.131 (.124, .138)	.004 (.002, .005)	.001 (.000, .001)
	2	.345 (.325, .365)	.001 (.000, .003)	.000 (.000, .001)	.110 (.099, .122)	.981 (.974, .988)	.897 (.854, .934)	.344 (.319, .367)
	3	1.00 (1.00, 1.00)	.860 (.841, .880)	.242 (.228, .255)	.222 (.208, .236)	.682 (.669, .695)	.277 (.261, .293)	.009 (.006, .012)
	4	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)	.817 (.808, .827)	.869 (.860, .877)	.998 (.996, .999)	.995 (.991, .997)	.584 (.573, .595)

estimates clearly reveal the nature of each of the four classes.

Examining the estimates from Class 1, we see that this group is comprised of non-users who strongly disapprove of marijuana use. Although a small minority of them has tried marijuana at some point, they have not done so recently, do not intend to do so, and do not approve of it. Class 2 contains those who apparently do not use marijuana themselves but tend to approve of others doing so on a more or less experimental basis. Class 3 consists of individuals who do use marijuana on occasion but do not approve of regular use. Members of Class 4 are regular users who generally approve of use. Speaking broadly, Classes 1 and 2 are non-users of marijuana, whereas Classes 3 and 4 are users. In terms of attitudes, however, Classes 1 and 3

are the most disapproving, and Classes 2 and 4 are more approving.

#### 4.4 Historical Trends in Class Prevalence

Although this four-class model imposes constraints on the  $\rho$  parameters, the class prevalence rates freely vary by year. Estimated prevalences for the classes over time are plotted in Figure 4. Figure 4 (a) shows the prevalence for each class separately, and Figure 4 (b) shows the combined prevalence for users (3+4) and approvers (2+4). From these plots, we see that marijuana use declined steadily from 1980 to 1992 due to very strong growth in Class 1 and shrinking of Class 4. After 1992, use rose again until 1999. Approval of marijuana use, as seen in the combined membership of Classes 3 and 4, declined throughout the 1980s and rose through the early and mid-1990s. Interestingly, Figure 4 (b) shows that shifting trends in attitudes anticipated the trends in use by two to three years. Approval of marijuana use began to rise after 1990, two years before use itself began to rise; approval then leveled off after 1996, three years before use stopped rising. The most stable class in terms of membership rate has been Class 2 (non-users who approve of experimentation), while the most vigorous fluctuations over time occurred in the behaviorally and attitudinally polarized groups (Classes 1 and 4).

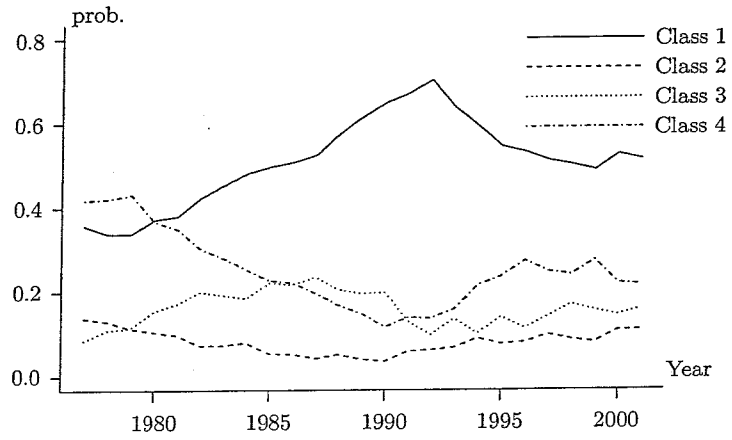
The estimates in Figure 4 were generated from a model that includes the covariate vector  $\mathbf{x}_i$  for each individual as described in Section 4.1. To obtain prevalence rates from a model with covariates, we can recommend several approaches. One approach is to average the ML estimates of the subject-specific class probabilities over the sample,

$$\bar{\gamma}_{lg} = \frac{1}{n} \sum_{i=1}^n \left( \frac{\exp\{\mathbf{x}'_i \hat{\beta}_{lg}\}}{1 + \sum_{j=1}^{L-1} \exp\{\mathbf{x}'_i \hat{\beta}_{jlg}\}} \right).$$

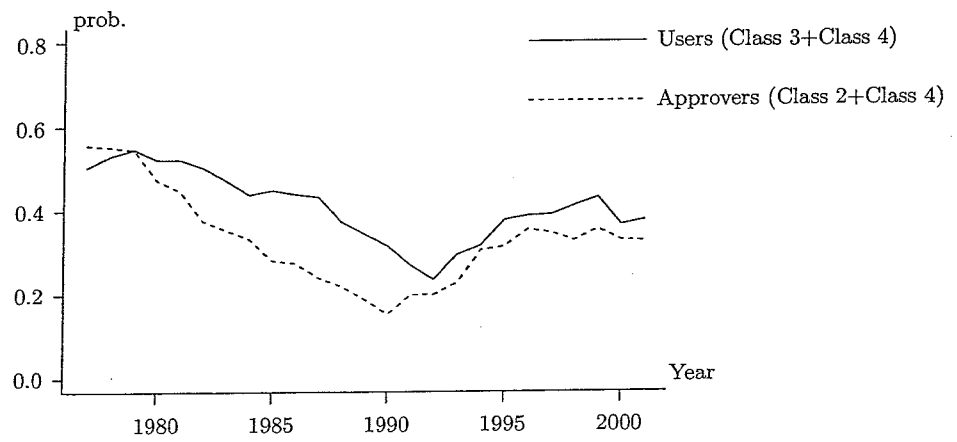
If the sample was obtained from an unequal-probability design, then a weighted average over the sample should be taken, with weights proportional to the inverse probabilities of selection. (Although MTF does have an unequal-probability design, its samples are quite representative of the overall population of high-school seniors in the 48 contiguous states, and taking weighted rather than unweighted averages over the sample had almost no discernible effect in any of our analyses.) When using MCMC rather than ML, we would average the simulated values of  $\gamma_{lg}(\mathbf{x}_i)$  over the sample at each cycle using the simulated values for  $\beta_{lg}$ , and then average these over the cycles of MCMC. A third approach is to retain multiple imputations

Figure 4: Estimated class prevalence by year.

(a)



(b)



of the latent class variables  $l_1, \dots, l_n$  for all subjects and estimate the proportions directly from these. With a large sample, the differences among these methods tend to be negligible.

#### 4.5 Covariates

Our four-class model estimates 30 regression coefficients for each of the 25 years. Due to space limitations, we will summarize these results only briefly. Effects of gender on class membership were small, but the effects of race were quite substantial. Across the years, the odds of membership in Class 4 relative to Class 1 were 2–4 times as high for non-whites relative to whites, controlling for the other covariates. This confirms a well known result that rates of marijuana use among black youths tend to be lower than among their white counterparts.

Lifestyle factors tended to be related to class membership in directions that one would expect. High school seniors who have better grades, less truancy and spend fewer evenings out for fun each week were less likely to use marijuana and to approve of use. Interestingly, these covariates are more strongly related to actual use than to attitudes. Figure 5 shows the odds ratios comparing Class 1 with each of the other classes for grades in school and number of evenings out. Neither of these covariates has much power to distinguish Class 2 (non-users who approve of experimentation) from Class 1 (non-users who disapprove of use), but they consistently and significantly distinguish Class 3 (experimenters) and Class 4 (regular users) from Class 1.

Religiosity, on the other hand, seems to correlate more strongly with attitudes than behavior. Designating those who regard their religious beliefs as “very important” as the baseline, we created dummy indicators for those who responded “not important” and those who responded “important.” The odds ratios associated with these dummy indicators are plotted in Figure 6. Those who responded “not important” or “important” rather than “very important” had dramatically higher odds of falling into Class 2 rather than Class 1, which are distinguished by attitudes but not behavior. The less religiously inclined groups also had dramatically higher odds of falling into Class 4 versus 1, which reflects a difference in both attitudes and behavior. But the odds ratios comparing Class 3 (experimenters who disapprove) with Class 1 (non-users who disapprove), although significantly different from 1.0, are not as large as the other odds ratios, suggesting that religion has a smaller effect in discouraging marijuana use among those who already disapprove of it.

Figure 5: Estimated odds ratios (—) and 95% confidence intervals (- - -) by year for grades and number of evenings out.

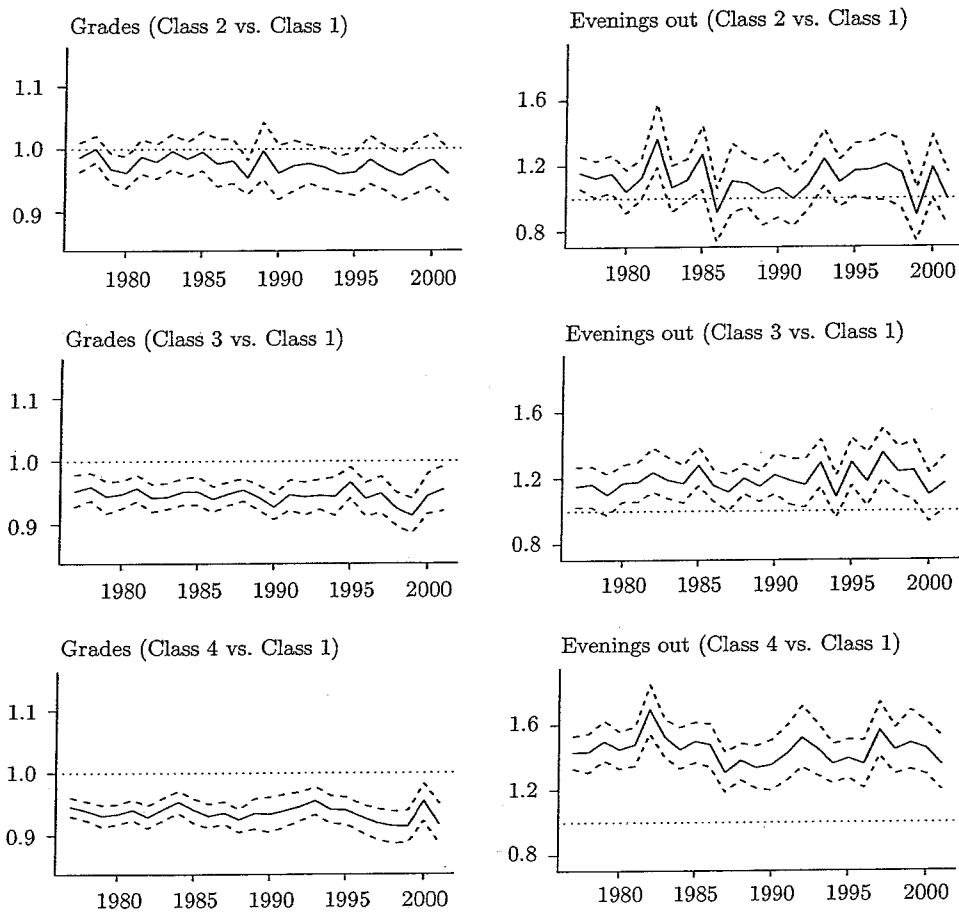
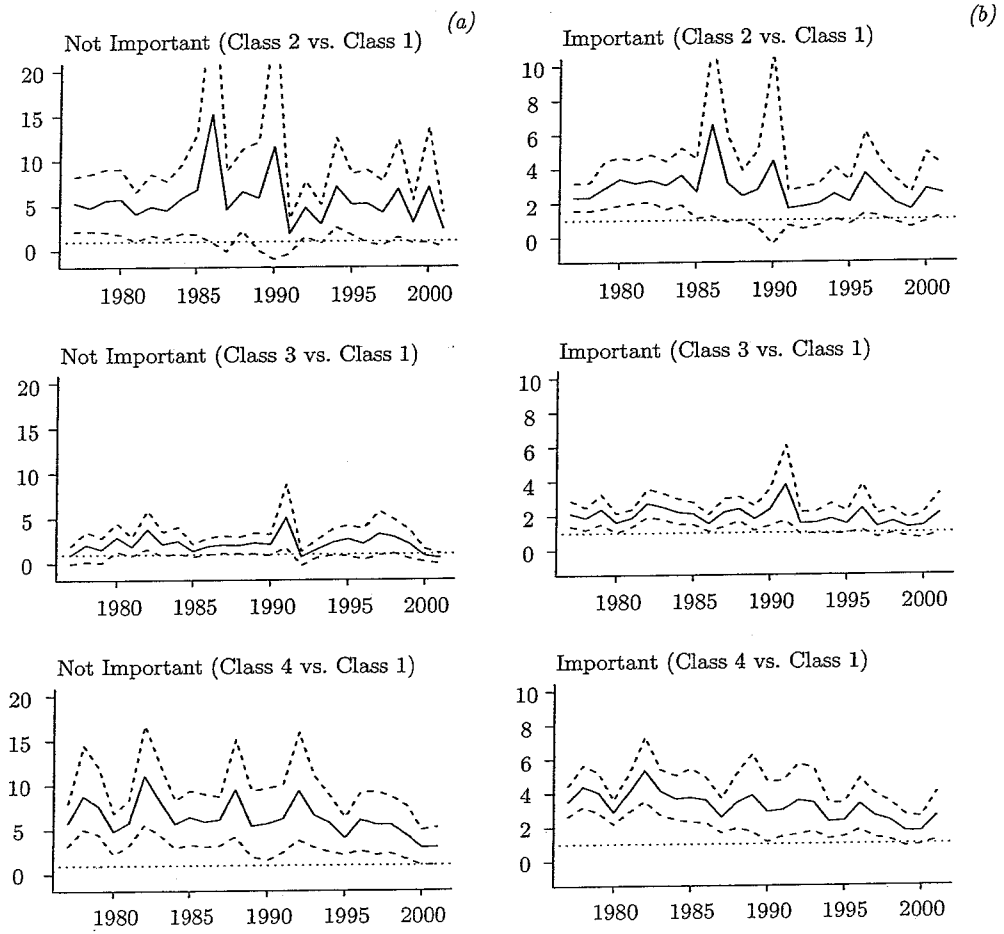


Figure 6: Estimated odds ratios (—) and 95% confidence intervals (---) for religious importance: (a) not important; and (b) important versus very important.



Many questions still remain about why marijuana use and attitudes have shifted over the years. The  $\beta$ -parameters of the LC regression model, which relate the probabilities of class membership to demographic, lifestyle and belief factors, do not directly reveal how the composition of the classes themselves may have changed due to the influx or exodus of various types of individuals. With multiple imputation (MI), however, we can address these questions in a very straightforward manner. After imputing the unseen class variables  $l_1, \dots, l_n$  for the sampled individuals, we can divide each year's sample into classes and examine their composition with respect to any covariate.

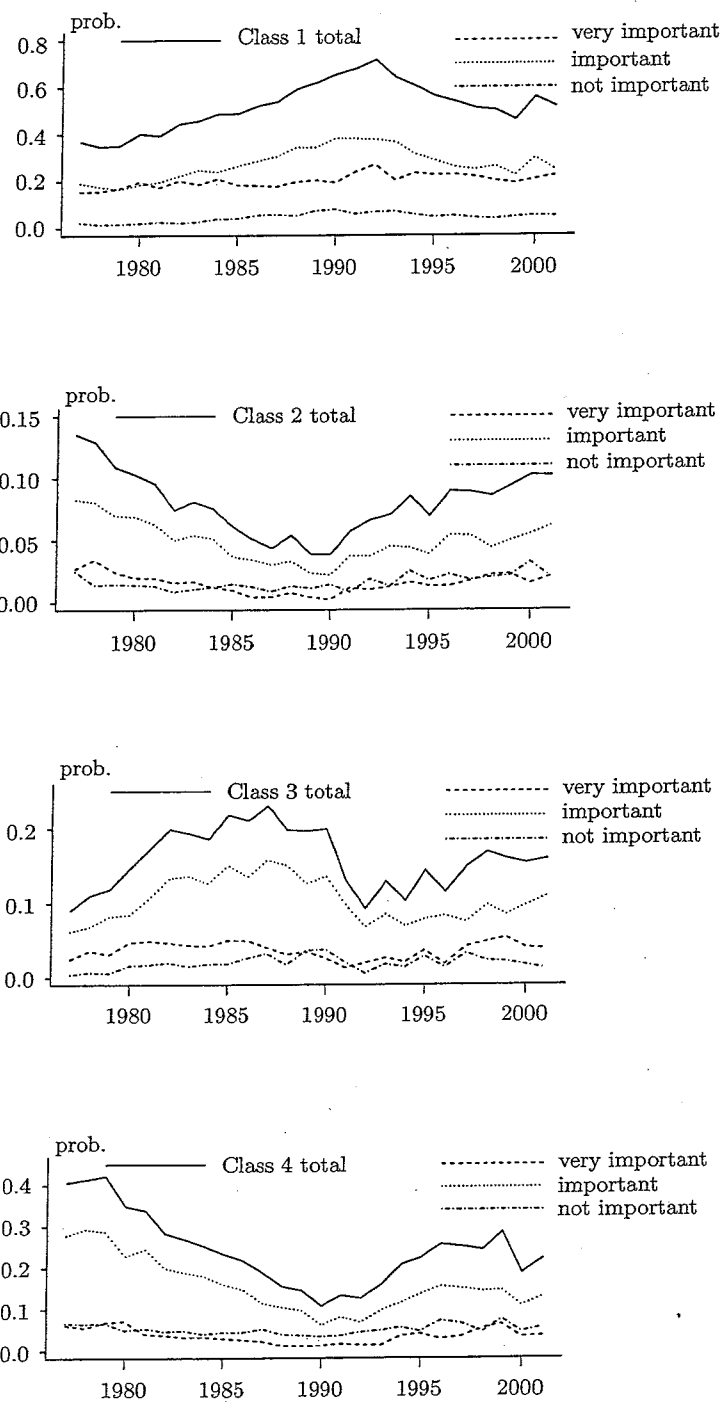
To illustrate, we imputed the latent variables 20 times. In each imputed dataset, we subdivided each of the four classes by categories of each covariate, and then averaged the resulting proportions across the imputed datasets. The composition of each class with respect to the three categories of religious belief is plotted in Figure 7. In each of the four plots in Figure 7, the total class prevalence (denoted by a solid line) reproduces the total prevalence shown in Figure 4 (a). Each solid line is also the sum of the three lines below it. These plots show that the prevalences of the highly religious ("very important") and irreligious ("not important") youth in each of the four classes has been remarkably stable over the entire 25 years. The moderately religious ("important") youth, however, show dramatic shifts that closely parallel the national trends. That is, the national trends have been largely driven by changes in attitude and behavior among the moderately religious. In one sense, this is not surprising, because the moderately religious are the largest of the three groups. On the other hand, the relative intransigence of the "not important" and "very important" groups does suggest that efforts to influence their attitudes and behavior toward marijuana may prove to be less effective.

We examined plots similar to those shown in Figure 7 for the other covariates and did not see any dramatically different trends across the response categories. In general, for all other covariates, the prevalences for each category of the covariate tended to rise and fall with the overall class prevalence.

## 5. DISCUSSION

LC analysis has long been used by social scientists to summarize the joint distribution of attitudinal items and identify population subgroups that might not be apparent in an item-by-item analyses (McCutcheon 1987). Advances in computing power and new computational algorithms now allow us to fit richer models to larger

Figure 7: Joint prevalence of class and religious importance by year.





datasets. Traditional applications of LC models focused on identifying and validating the latent structure. With LC regression, we are now able to relate a latent structure to large numbers of covariates at once.

Because ML routines for LC regression are available in two software packages (Latent GOLD and Mplus), we expect the popularity of these models to grow. Properties of the likelihood have also been explored to a limited extent. For example, identifiability conditions have been described by Bandeen-Roche et al. (1997). Because we are accustomed to thinking ML as the default method, it is tempting to believe that reliable inferences for LC models are simply not possible with usual way when the shape of the likelihood is abnormal (e.g., when estimates lie on a boundary. We have found, however, that Bayesian inference by MCMC is an attractive alternative to ML. Simulations by Chung (2003) have shown that Bayesian methods usually perform as well as or better than ML by the traditional frequentist criteria of bias, variance, interval coverage and width. Multiple imputation can be used in tandem with Bayesian methods to perform many innovative analyses that would be difficult if we limited ourselves to simply averaging simulated parameters over the MCMC output stream.

In our analyses of MTF, we were able to identify a plausible latent structure from survey items related to both attitudes and behaviors. Supporting evidence for this latent structure was provided not only by quantitative measures (e.g. fit statistics) but by its interpretability and relative stability over a long period of time. This will not always be the case. For example, MTF has several items that measure the perceived risks of marijuana use. When we applied LC analyses to the perceived risk and use items, we could not identify a plausible latent structure that fit well over a long period of time.

#### REFERENCES

- Akaike, H. (1987), "Factor Analysis and AIC," *Psychometrika*, 52, 317-332.
- Bachman, J. G., Johnston, L. D., and O'Malley, P. M. (1998), "Explaining Recent Increases in Students' Marijuana Use: Impacts of Perceived Risks and Disapproval, 1976 through 1996," *American Journal of Health*, 88, 887-892.
- Bandeen-Roche, K., Miglioretti, D. L., Zeger, S. L., and Rathouz, P. J. (1997), "Latent Variable Regression for Multiple Discrete Outcomes," *Journal of the American Statistical Association*, 92, 1375-1386.

- Best, N., Cowles, M., and Vines, S. (1995), Coda: Convergence Diagnostics and Output Analysis Software for Gibbs Sampler Output, Version 0.3., Technical report, MRC Biostatistics Unit.
- Biemer, P. P., and Wiesen, C. (2002), "Measurement Error Evaluation of Self-Reported Drug Use: A Latent Class Analysis of the US National Household Survey on Drug Abuse," *Journal of the Royal Statistical Society, Ser. A*, 165, 97–119.
- Chung, H. (2003), Latent-Class Modeling with Covariates, PhD thesis, The Pennsylvania State University, University park.
- Chung, H., Loken, E., and Schafer, J. L. (2004), "Difficulties in Drawing Inferences with Finite-Mixture Models: A Simple Example with a Simple Solution," *The American Statistician*, 58, 152–158.
- Clogg, C. C., and Goodman, L. A. (1984), "Latent Structure Analysis of a Set of Multidimensional Contingency Tables," *Journal of the American Statistical Association*, 79, 762–771.
- Collins, L. M. (2002), "Using Latent Transition Analysis to Examine the Gateway Hypothesis," in *Stages and Pathways of Drug Involvement: Examining the Gateway Hypothesis*, ed. D. B. Kandel, Cambridge: Cambridge University Press, pp. 254–269.
- Collins, L. M., Flaherty, B. P., and Colby, S. M. (2002), "Measurement and Design Issues in Tobacco and Drug Use Research (Editorial)," *Drug and Alcohol Dependence*, 68, S1–S2.
- Collins, L. M., Lanza, S. T., Schafer, J. L., and Flaherty, B. P. (2002), *WinLTA User's Guide*, The Methodology Center.
- Dayton, C. M., and Macready, G. B. (1988), "Concomitant-Variable Latent-Class Models," *Journal of the American Statistical Association*, 83, 173–178.
- Flay, B. R. (1993), "Youth Tobacco Use: Risks, Patterns, and Control," in *Nicotine Addiction*, eds. C. Orleans, and J. Slade, New York: Oxford University Press, pp. 360–384.

- Garrett, E. S., Eaton, W. W., and Zeger, S. L. (2002), "Methods for Evaluating the Performance of Diagnostic Tests in the Absence of a Gold Standard: A Latent Class Model Approach," *Statistics in Medicine*, 21, 1289–1307.
- Garrett, E. S., and Zeger, S. L. (2000), "Latent Class Model Diagnosis," *Biometrics*, 56, 1055–1067.
- Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004), *Bayesian Data Analysis, Second edition*, Chapman & Hall: London.
- Gelman, A., and Rubin, D. B. (1992), "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457–511.
- Geweke, J. (1992), "Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments," in *Bayesian Statistics 4*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford: Oxford University Press, pp. 169–193.
- Goodman, L. A. (1974), "Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models," *Biometrika*, 61, 215–231.
- Graham, J. W., Collins, L. M., Wugalter, S. E., Chung, N. K., and Hansen, W. B. (1991), "Modeling Transitions in Latent-Sequential Processes: A Substance Use Prevention Example," , .
- Hojtink, H. (1998), "Constrained Latent Class Analysis Using the Gibbs Sampler and Posterior Predictive P-values: Applications to Educational Testing," *Statistica Sinica*, 8, 691–711.
- Johnston, L. D. (1982), *A Review and Analysis of Recent Changes in Marijuana Use by American Young People, Marijuana: The National Impact on Education*, New York: American Council on Marijuana.
- Johnston, L. D. (1985), "The Etiology and Prevention of Substance Use: What Can We Learn from Recent Historical Changes?," in *Etiology of Drug Abuse: Implications for Prevention*, eds. C. L. Johnes, and R. J. Battes, Washington: US Govt Printing Office.

- Johnston, L. D., Bachman, J. G., and O'Malley, P. M. (2001), *Monitoring the Future: A Continuing Study of American Youth (12th-Grade Survey), 2000*, Ann Arbor: Inter-university Consortium for Political and Social Research.
- Johnston, L. D., O'Malley, P. M., and Bachman, J. G. (2002), *Monitoring the Future National Survey Results on Drug Use, 1975-2001, Volume I: Secondary School Students*, Bethesda: National Institute on Drug Abuse.
- Kennedy, W. J., and Gentle, J. E. (1980), *Statistical Computing*, New York: Marcel Dekker.
- Lanza, S. T., and Collins, L. M. (2002), "Pubertal Timing and the Onset of Substance Use in Females During Early Adolescence," *Prevention Science*, 3, 69-82.
- Lanza, S. T., Collins, L. M., Schafer, J. L., and Flaherty, B. P. (in press), "Using Data Augmentation to Obtain Standard Errors and Conduct Hypothesis Tests in Latent Class and Latent Transition Analysis," *Psychological Methods*, .
- Lazarsfeld, P. F., and Henry, N. W. (1968), *Latent Structure Analysis*, Boston: Houghton Mifflin.
- Leventhal, H., and Cleary, P. D. (1980), "The Smoking Problem: A Review of the Research and Theory in Behavioral Risk Modification," *Psychological Bulletin*, 88, 370-405.
- Lindsay, B. G. (1995), *Mixture Models: Theory, Geometry and Applications*, NSF-CBMS Regional Conference Series in Probability and Statistics Volume 5, Hayward: Institute of Mathematical Statistics.
- Little, R. J. A., and Rubin, D. B. (2002), *Statistical Analysis with Missing Data, Second edition*, New York: Wiley.
- Mayhew, K. P., Flay, B. R., and Mott, J. A. (2000), "Stages in the Development of Adolescent Smoking," *Drug and Alcohol Dependence*, 59, S61-S81.
- McCutcheon, A. L. (1987), "Sexual Morality, Pro-life Values, and Attitudes Toward Abortion: A Simultaneous Latent Structure Analysis for 1978-1983," *Sociological Methods and Research*, 16, 256-275.

- Metropolis, M., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), "Equations of State Calculations by Fast Computing Machine," *Journal of Chemical Physics*, 21, 1087–1091.
- Muthén, L. K., and Muthén, B. O. (1998), *Mplus User's Guide*, Muthén & Muthén.
- Pfeffermann, D., Skinner, C., and Humphreys, K. (1998), "The Estimation of Gross Flows in the Presence of Measurement Error Using Auxiliary Variables," *Journal of the Royal Statistical Society, Ser. A*, 161, 13–32.
- Robert, C. P., and Casella, G. (2004), *Monte Carlo Statistical Methods, Second edition*, New York: Springer.
- Roberts, G. O. (1992), "Convergence Diagnostics of the Gibbs Sampler," in *Bayesian Statistics 4*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford: Oxford University Press, pp. 775–782.
- Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994), "Estimation of Regression Coefficients When Some Regressors Are Not Always Observed," *Journal of the American Statistical Association*, 89, 846–866.
- Rubin, D. B. (1976), "Inference and Missing Data," *Biometrika*, 63, 581–592.
- Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: Wiley.
- Rubin, D. B., and Stern, H. S. (1994), "Testing in Latent Class Models Using a Posterior Predictive Check Distribution," in *Latent Variables Analysis: Applications for Developmental Research*, eds. A. von Eye, and C. C. Clogg, Thousand Oaks: Sage, pp. 420–438.
- Schafer, J. L. (1997), *Analysis of Incomplete Multivariate Data*, London: Chapman & Hall.
- Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464.
- Shiffman, S., Kassel, J. D., Paty, J. A., and Gnys, M. (1994), "Smoking Typology Profiles of Chippers and Regular Smokers," *Journal of Substance Abuse*, 6, 21–35.
- Tanner, W. A., and Wong, W. H. (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the Royal Statistical Society, Ser. B*, 62, 795–809.

Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions (with discussion)," *Annals of Statistics*, 22, 1701–1762.

Vermunt, J. K., and Magidson, J. (2000), *Latent GOLD User's Guide*, Statistical Innovations.