The Korean Journal of Applied Statistics (2011) DOI: http://dx.doi.org/10.5351/KJAS.2011.24.4.709
24(*4*), *709–719*

Multiple-Group Latent Transition Model for the Analysis of Sequential Patterns of Early-Onset Drinking Behaviors among U.S. Adolescents

Hwan Chung[1]

[1]Department of Statistics, Ewha Womans University

(Received May 2011; accepted June 2011)

Abstract

We investigate the latent stage-sequential patterns of drinking behaviors of U.S. adolescents who have started to drink by age 14 years (seven years before the legal drinking age). A multiple-group latent transition analysis(LTA) with logistic regression is employed to identify the subsequent patterns of drinking behaviors among early-onset drinkers. A sample of 1407 early-onset adolescents from the National Longitudinal Survey of Youth(NLSY97) is analyzed using maximum-likelihood estimation. The analysis demonstrates that early-onset adolescents' drinking behaviors can be represented by four latent classes and their prevalence and transition are influenced by demographic factors of gender, age, and race.

Keywords: Latent stage-sequential process, latent transition analysis, maximum likelihood, under-age drinking.

1. Introduction

According to the 2009 Youth Risk Behavior Survey(YRBS), a national survey of the United States (U.S.), 21% of high school students began to drink alcohol before the age of 13 (Centers for Disease Control and Prevention, 2010). This very early-onset drinking has been a well-known risk factor for health and well-being among adolescents between the ages of 12 and 20. Indeed, numerous studies have found that early-onset drinking is linked to a variety of other risky behaviors that have adverse health consequences. For example, the earlier a youth begins to drink, the more likely it is that the youth will at some point have unintentional injuries (Hingson *et al.*, 2000), drive after drinking (Lynskey *et al.*, 2007), and engage in physical fights (Hingson *et al.*, 2001).

Recently, the idea of latent stage-sequential process has motivated numerous studies on the initiation and progression of alcohol use, since prevention scientists are able to find the intervals that provide the best opportunities to slow the process of alcohol dependence. The current study investigates the latent stage-sequential process of drinking behaviors among U.S. adolescents who have

This research was supported by the National Research Foundation of Korea(KRF) grant award (2011-0004947).

[1]Associate Professor, Department of Statistics, Ewha Womans University, 52 Ewhayeodae-gil, Seodaemun-gu, Seoul 120-750, Korea. E-mail: hchung@ewha.ac.kr

started to drink by age 14 years. We employ a multiple-group latent transition analysis(LTA) with logistic regression. The LTA has been derived from the family of latent class analysis(LCA) (Clogg and Goodman, 1984; Goodman, 1974), where the measurement model at each time point is specified with an LCA, and stage-sequential process can be summarized by the transition probabilities among latent classes over time. The multiple-group LTA estimates the group-specific probability of belonging to any of hypothesized latent classes at each time point. In addition, the model estimates the group-specific transition rates among latent classes over successive measurement occasions. A transition among latent classes is typically represented with a first-order Markov chain, on the assumption that class membership at time t depends only on class membership at time $t-1$.

We provide detailed explanation of a maximum-likelihood(ML) estimation method via Expectation-Maximization(EM) algorithm (Dempster *et al.*, 1977). In terms of demonstration, we suggest some speculative interpretations from the proposed LTA based on substantive findings using data from the National Longitudinal Survey of Youth 1997 (NLSY97, `http://www.bls.gov/nls/nlsy97.htm`). The organization of the rest is as follows. We introduce the LTA and provide estimation strategies using the ML method in Section 2. In Section 3 we apply an LTA to alcohol drinking items drawn from the NLSY97 to assess stage-sequential process of alcohol use among early-onset drinkers. In Section 4 we discuss the advantages and limitations of an LTA application and conclude the paper.

2. Latent Transition Analysis and EM Algorithm

2.1. Model

Suppose we construct a multiple-group LTA model with C classes from a set of M items over T time periods. The group variable G_i represents the group membership for the i^{th} individual ranging from 1 to G, and $\mathbf{C}_i = (C_{i1}, \ldots, C_{iT})$ denotes a vector of the latent class membership variable for the i^{th} individual from initial time $t = 1$ to time T, where variable C_{it} takes possible values $1, \ldots, C$. Let $\mathbf{y}_{it} = (y_{i1t}, \ldots, y_{iMt})'$ be a vector of discrete responses to M items given by the i^{th} individual. These item responses are used to measure latent class membership at time t, where each y_{imt} can take values from 1 to r_m for $m = 1, \ldots, M$. Further, let $\mathbf{x}_{it} = (x_{i1t}, \ldots, x_{ip_t t})'$, measured at time t, denote a $p_t \times 1$ vector of covariates for individual i that may influence class prevalence at time t. The probability that the i^{th} subject provides responses $\mathbf{y}_i = (\mathbf{y}_{i1}, \ldots, \mathbf{y}_{iT})$ conditioned on $G_i = g$ and $(\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT})$ can be obtained by marginalizing the joint probability of $\mathbf{c} = (c_1, \ldots, c_T)$ and $\mathbf{y}_i$ over the class membership $\mathbf{c}$:

$$
\begin{aligned}
L_i &= P(\mathbf{Y}_i = \mathbf{y}_i \mid G_i = g, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}) \\
&= \sum_{c_1=1}^{C} \cdots \sum_{c_T=1}^{C} P(\mathbf{C}_i = \mathbf{c}, \mathbf{Y}_i = \mathbf{y}_i \mid G_i = g, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}) \\
&= \sum_{c_1=1}^{C} \cdots \sum_{c_T=1}^{C} \delta_{c_1|g}(\mathbf{x}_{i1}) \prod_{t=2}^{T} \tau^{(t)}_{c_t|c_{t-1}g}(\mathbf{x}_{it}) \prod_{t=1}^{T} \prod_{m=1}^{M} \prod_{k=1}^{r_m} \rho_{mkt|c_t g}^{I(y_{imt}=k)} \\
&= \sum_{c_1=1}^{C} \cdots \sum_{c_T=1}^{C} L_i^*, \qquad (2.1)
\end{aligned}
$$

where $I(y = k)$ is the usual indicator function which has the value 1 if y is equal to k and 0 otherwise. In likelihood (2.1), the following three sets of parameters are estimated:

- $\delta_{c_1|g}(\mathbf{x}_{i1}) = P(C_{i1} = c_1 \mid G_i = g, \mathbf{x}_{i1})$ represents the probability of the i^{th} individual belong to class c_1 at time 1 given a group membership in g.
- $\tau^{(t)}_{c_t|c_{t-1}g}(\mathbf{x}_{it}) = P(C_{it} = c_t \mid C_{i,t-1} = c_{t-1}, G_i = g, \mathbf{x}_{it})$ represents the transition probability of class membership in c_t in time t given the previous class membership in c_{t-1} and the group membership in g.
- $\rho_{mkt|c_tg} = P(Y_{imt} = k \mid C_{it} = c_t, G_i = g)$ represents the probability of response k to the m^{th} item at time t given a class membership in c_t at time t and a group membership in g.

In (2.1) we have assumed local independence, that is, the items $Y_{i1t}, \ldots, Y_{iMt}$ are conditionally independent given c_t for $t = 1, \ldots, T$. In addition, the sequence $\mathbf{C}_i = (C_{i1}, \ldots, C_{iT})$ is assumed to constitute a first-order Markov chain for $t = 2, \ldots, T$. The marginal probability of the class membership at the initial time $t = 1$ would be

$$\begin{aligned}\delta_{c_1|g}(\mathbf{x}_{i1}) &= P(C_{i1} = c_1 \mid G_i = g, \mathbf{x}_{i1}) \\ &= \frac{\exp\left\{\mathbf{x}'_{i1}\boldsymbol{\beta}^{(1)}_{c_1|g}\right\}}{\sum_{c=1}^{C}\exp\left\{\mathbf{x}'_{i1}\boldsymbol{\beta}^{(1)}_{c|g}\right\}}\end{aligned} \tag{2.2}$$

for $c_1 = 1, \ldots, C$, where $\boldsymbol{\beta}^{(1)}_{C|g} = \mathbf{0}$. The transition probability that the i^{th} individual changes their class to $C_{it} = c_t$ from the previous class $C_{i,t-1} = c_{t-1}$ is, for $c_t = 1, \ldots, C$ and $t = 2, \ldots, T$,

$$\begin{aligned}\tau^{(t)}_{c_t|c_{t-1}g}(\mathbf{x}_{it}) &= P(C_{it} = c_t \mid C_{i,t-1} = c_{t-1}, G_i = g, \mathbf{x}_{it}) \\ &= \frac{\exp\left\{\mathbf{x}'_{it}\boldsymbol{\beta}^{(t)}_{c_t|c_{t-1}g}\right\}}{\sum_{c=1}^{C}\exp\left\{\mathbf{x}'_{it}\boldsymbol{\beta}^{(t)}_{c|c_{t-1}g}\right\}},\end{aligned} \tag{2.3}$$

where $\boldsymbol{\beta}^{(t)}_{c_t|c_{t-1}g} = \mathbf{0}$ when $c_t = c_{t-1}$. Note that class C at $t = 1$ serves as a baseline in (2.2), whereas the class in the previous time point is the baseline in (2.3). Therefore, the coefficient $\boldsymbol{\beta}^{(t)}_{c_t|c_{t-1}g}$ in (2.3) can be interpreted as change in the log-odds of transitioning to class c_t at time t from the previous class c_{t-1} versus remaining at the same class as the previous c_{t-1} given a group membership g. The marginal probability of the class membership at time t is not directly estimated in likelihood (2.1) but rather is a function of other parameter:

$$\begin{aligned}\delta_{c_t|g}(\mathbf{x}_{i1}, \ldots, \mathbf{x}_{it}) &= P(C_t = c_t \mid \mathbf{x}_{i1}, \ldots, \mathbf{x}_{it}) \\ &= \sum_{c_1=1}^{C} \cdots \sum_{c_{t-1}=1}^{C} \delta_{c_1|g}(\mathbf{x}_{i1}) \prod_{j=2}^{t} \tau^{(j)}_{c_j|c_{j-1}g}(\mathbf{x}_{ij}).\end{aligned} \tag{2.4}$$

2.2. Expectation-maximization algorithm

Under ordinary circumstances, the ML estimates for the parameters solve the score equation $\sum_{g=1}^{G}\sum_{i\in g}\partial \log L_i/\partial\boldsymbol{\pi} = 0$, where $\boldsymbol{\pi}$ denotes the set of free parameters and $\prod_{i\in g}$ represents the product over the set of individuals in subgroup g. We use the EM algorithm (Dempster *et al.*, 1977) to obtain ML estimates of the model parameters specified in Equation (2.1). The E-step computes the conditional probability that an individual i in subgroup g belongs to class sequence

$\mathbf{c} = (c_1, \ldots, c_T)$ given his or her item responses $\mathbf{y}_i$ and current estimates for $\hat{\boldsymbol{\pi}}$ for the parameters:

$$\begin{aligned} \theta_{i(c_1,\ldots,c_T|g)} &= P(\mathbf{C}_i = \mathbf{c} \mid \mathbf{Y}_i = \mathbf{y}_i, G_i = g) \\ &= \frac{L_i^*}{\sum_{c_1=1}^{C} \cdots \sum_{c_T=1}^{C} L_i^*}, \end{aligned} \tag{2.5}$$

where L_i^* is defined in (2.1). The M-step maximizes the expected complete data likelihood (*i.e.*, the likelihood for the cross-classification by $\mathbf{C}_i$ and $\mathbf{Y}_i$) with respect to the model parameters. This likelihood can be written as

$$\begin{aligned} E\left\{\sum_{g=1}^{G}\sum_{i\in g}\log L_i^*\right\} &= \sum_{g=1}^{G}\sum_{i\in g}\sum_{c_1=1}^{C}\theta_{i(c_1|g)}\log\left\{\delta_{c_1|g}(\mathbf{x}_{i1})\right\} \\ &+\sum_{g=1}^{G}\sum_{i\in g}\sum_{t=2}^{T}\sum_{c_t=1}^{C}\theta_{i(c_{t-1},c_t|g)}\log\left\{\tau^{(t)}_{c_t|c_{t-1}g}(\mathbf{x}_{it})\right\} \\ &+\sum_{g=1}^{G}\sum_{i\in g}\sum_{t=1}^{T}\sum_{c_t=1}^{C}\theta_{i(c_t|g)}\sum_{m=1}^{M}\sum_{k=1}^{r_m}I(y_{imt}=k)\log\left\{\rho_{mkt|c_tg}\right\}, \end{aligned} \tag{2.6}$$

where $\theta_{i(c_{t-1},c_t|g)} = \sum_{c_1}\cdots\sum_{c_{t-2}}\sum_{c_{t+1}}\cdots\sum_{c_T}\theta_{i(c_1,\ldots,c_T|g)}$ and $\theta_{i(c_t|g)} = \sum_{c_{t-1}}\theta_{i(c_{t-1},c_t|g)}$. The first two sums in expression (2.6), which relate to the regression coefficients (*i.e.*, β-parameters), are the log-likelihood functions for the multinomial logistic regression model (Agresti, 2002), except that the unobserved counts for c_t for $t = 1, \ldots, T$ are replaced by the fractional expectations $\sum_{c_1=1}^{C}\theta_{i(c_1|g)}$ and $\sum_{t=2}^{T}\sum_{c_t=1}^{C}\theta_{i(c_{t-1},c_t|g)}$, respectively. Updates estimates for the regression coefficients can be calculated with the standard Newton-Raphson method for multinomial logistic regression, provided that the computational routines allow fractional responses rather than integer counts. In the last sum in (2.6), ρ-parameter can be interpreted as parameter in a multinomial distribution when $\theta_{i(c_t|g)}$ are known. Therefore, ρ-parameter can be updated as

$$\hat{\rho}_{mkt|c_tg} = \frac{\sum_{g=1}^{G}\sum_{i\in g}\theta_{i(c_t|g)}I(y_{imt}=k)}{\sum_{g=1}^{G}\sum_{i\in g}\theta_{i(c_t|g)}}.$$

We can easily extend the EM algorithm to a model with missing observations on measurement items by using the missingness at random assumption (Rubin, 1987). In the E-step, the conditional probability is calculated only with the observed responses of $\mathbf{y}_i$. We denote $\theta^{obs}_{i(c_t|g)}$ to distinguish it from the previous $\theta_{i(c_t|g)}$ that was given in Equation (2.5). The ρ-parameter is then obtained from the provisional estimates $\rho^*_{mkt|c_tg}$ by

$$\hat{\rho}_{mkt|c_tg} = \frac{\sum_{g=1}^{G}\sum_{i\in g,obs_m^{(t)}}\theta^{obs}_{i(c_t|g)}I(y_{imt}=k) + \sum_{g=1}^{G}\sum_{i\in g,mis_m^{(t)}}\theta^{obs}_{i(c_t|g)}\rho^*_{mkt|c_tg}}{\sum_{g=1}^{G}\sum_{i\in g}\theta^{obs}_{i(c_t|g)}}, \tag{2.7}$$

where $\text{obs}_m^{(t)}$ and $\text{mis}_m^{(t)}$ denote the sets of individuals who respond and fail to respond to the m^{th} item at time t, respectively. Because our data set contains missing observations on items, we shall use Equation (2.7) for our calculations.

2.3. Model diagnosis and local identifiability

The log-likelihood ratio statistic(LRT) is the standard way to assess goodness-of-fit by comparing the fit of the estimated model to that of the saturated model. The LRT is asymptotically distributed as chi-square with degrees of freedom equal to (number of possible response patterns − number of parameters estimated − 1). However, the asymptotic assumption for an LRT generally does not hold because LTA tends to involve large contingency tables with many degrees of freedom. Therefore, we empirically determined the distribution of the LRT by generating bootstrap samples of an LRT from the estimated LTA model (Langeheine *et al.*, 1996). Bootstrap samples for an LTA can be constructed in the following way: (a) fit the LTA model to the data set and obtain the observed LRT based on ML estimates, (b) generate hypothetical new data set from the ML estimates, (c) fit the model to the simulated data set by using the EM algorithm, and (d) compute the LRT based on output from step (c). Repeating (b)–(d) many times (200 repetitions were used in our example) produces a bootstrap sample of LRT that does not rely on any known distribution. The area to the right of the observed LRT can be regarded as a bootstrap p-value.

Model identification is required to estimate the parameters of an LTA model correctly. As discussed by McHugh, (1956) and Goodman (1974), a singular Hessian matrix indicates that the model is not locally identifiable for the given data. If the observed data log-likelihood is concave, the inverse of (-1 times) the Hessian matrix of this log-likelihood will consistently estimate the covariance matrix for the ML estimates. However, in many cases we cannot estimate variances due to the boundary solution—any of the estimated parameters is close to 0 or 1. In situations where some estimated parameters lie on the boundary, those values can be fixed 0 or 1 a posteriori to make the remaining parameters identifiable (Formann, 2003).

3. An Analysis of NLSY97

3.1. Data

The National Longitudinal Survey of Youth 1997(NLSY97) was based on a sample of 8984 U.S. adolescents who were between ages of 12 and 18 years during the first survey in 1997. Data for the current analysis were 1407 adolescents aged 12–14 years in 1997 who were identified as the early-onset drinkers (all had started to drink by age 14 years). There are three self-report items measuring adolescent drinking behaviors: how many days they had one or more drinks of an alcoholic beverages during the last 30 days (*Recent Drinking*); how many days they had five or more drinks on the same occasion during the last 30 days (*Binge Drinking*); and how many days they had drinks immediately before or during school or work hours in the last 30 days (*Drinking at School*). The responses for *Recent Drinking* were reduced to a three categories, non-drinker (0 days of drinking), occasional drinker (1–5 days of drinking) and regular drinker (6 or more days of drinking). For *Binge Drinking*, respondents who had consumed five or more drinks on the same occasion at least one time were characterized as binge drinkers. The same binary rule was applied for *Drinking at School*. These three drinking items were tracked over the three survey waves in 1997 (Wave 1), 2000 (Wave 4), and 2003 (Wave 7), corresponding to early adolescence (ages 12–14), middle adolescence (ages 15–17), and late adolescence (ages 18–20), respectively. In addition to these three items, we consider gender as a group variable and race and age as covariates in the multiple-group LTA with logistic regression.

Table 3.1. Goodness-of-fit statistics for a series of multiple-group LTA under various number of classes.

$C^{\dagger}$	Full Model				Restricted Model			
	$k^{\ddagger}$	Log-likelihood	Bootstrap p-value	AIC	$k^{\ddagger}$	Log-likelihood	Bootstrap p-value	AIC
2	26	−5939.75	0.000	11931.51	18	−5980.00	0.000	11996.01
3	52	−5762.20	0.004	11628.40	40	−5780.78	0.004	11641.56
4	86	−5725.91	0.076	11623.81	70	−5736.32	0.208	11612.63
5	128	−5695.99	0.160	11647.98	108	−5710.32	0.128	11636.64
6	178	−5669.56	0.112	11695.12	154	−5682.64	0.108	11673.29

† is the number of classes; ‡ is the number of parameters.

3.2. Model selection

As shown in Bandeen-Roche *et al.* (1997), we do not need to consider covariates when selecting the number of classes due to the marginalization property in LTA. To keep the interpretability of classes stable over time, we fitted LTA models in which ρ-parameters were constrained to be equal over time (*i.e.*, $\rho_{mk|cg} = \rho_{mk1|cg} = \cdots = \rho_{mkT|cg}$). However, since we have pooled samples over two different subgroups (*i.e.*, male and female), there was strong possibility that their ρ-parameters could vary across gender. To select the number of classes, we began by fitting a two-class LTA, where ρ-parameters were allowed to vary across gender (referred to *Full Model*). We then increased the number of classes and fitted a three-class LTA. The procedure was repeated until we reached a six-class LTA. To examine whether the ρ-parameters vary across gender, we compared another series of LTA models, where the ρ-parameters were constrained to be equal across gender (referred to *Restricted Model*). Table 3.1 shows a series of LTA models with evaluations based on the bootstrap p-values and AIC for goodness-of-fit . We used 100 different sets of starting values and selected the solution with the best fit in order to avoid local maxima. Among the full models, the four-class LTA is the most parsimonious model with the bootstrap p-value of 0.076, which indicates an adequate model fit. In addition, the four-class LTA provides the smallest AIC value (11623.81) as shown in Table 3.1. Among the restricted models, the four-class LTA also shows an appropriate model fit based on the bootstrap p-value (0.208) and a smallest AIC value (11612.63). Based on AIC values of these two four-class LTA models combined with the results from the bootstrap p-value, we select the restricted four-class LTA model to analyze the data set. Thus, for the remainder of this paper, we will focus on the four-class LTA model where the ρ-parameters are invariant across gender.

To investigate local identifiability for the selected model, we fixed some parameters to make the remaining parameters identifiable, as proposed by Formann (2003). Under the restricted four-class LTA, the EM algorithm reached the final solution with 20 boundary estimates. The estimates of six ρ-parameters and 14 τ-parameters were less than 0.001. Hence, we set those parameters to zero a posteriori, reducing the number of parameters to be estimated to 50. The Hessian matrix in the ML solution was non-singular under this condition, that the four-class LTA under consideration was locally identifiable. Finally, we added race (Caucasian American, African American, and other) and age as covariates to investigate variations in the relative occurrence of class membership.

3.3. Parameter estimates

Under the restricted four-class LTA model, the estimated ρ-parameters are presented in Table 3.2. The values under the first column of *Recent Drinking* are the estimated probabilities of having reported one to five days of drinking in the last 30 days (*i.e.*, occasional drinking) given a class

Table 3.2. Estimated probabilities of responding 'any use' to the drinking items for each class (ρ-parameter).

Class	Drinking items			
	Recent Drinking		*Binge*	*Drinking*
	Occasional	Regular	*Drinking*	*at School*
1. Not current drinkers	0.067	0.000	0.000	0.000
2. Light drinkers	0.942	0.058	0.392	0.094
3. Regular binge drinkers	0.267	0.733	0.956	0.189
4. Problematic binge drinkers	0.810	0.190	0.624	1.000

Table 3.3. Estimated class prevalence over time (δ-parameter).

Class	Male: Year			Female: Year		
	1997	2000	2003	1997	2000	2003
1. Not current drinkers	0.710	0.472	0.351	0.647	0.454	0.393
2. Light drinkers	0.190	0.202	0.190	0.282	0.361	0.351
3. Regular binge drinkers	0.023	0.271	0.448	0.033	0.174	0.249
4. Problematic binge drinkers	0.077	0.055	0.011	0.037	0.011	0.007

membership. The second column of *Recent Drinking* provides the estimated probabilities of having reported six or more days of drinking in the last 30 days (*i.e.*, regular drinking) for a given class membership. For example, the majority of Class 2 responded to 'occasional' category in *Recent Drinking* (0.942); however, most of adolescents in Class 3 reported themselves as regular drinkers (0.733). This indicates that Class 3 is composed of adolescents who have consumed alcohol on a regular basis during the past 30 days. The last two columns show the probabilities of having consumed five or more drinks on the same occasion at least one time in the last 30 days for *Binge Drinking* and consumed an alcoholic beverage right before or during school or work hours at least once in the last 30 days for *Drinking at School* given a class membership, respectively. An inspection of the estimated ρ-parameters leads to the adoption of the class names in Table 3.2. The combination of all three items support a meaningful interpretation for each class as follow: early-onset adolescents in Class 1 are 'not current drinkers' who have not been involved in any drinking in the previous 30 days; those in Class 2 are 'light drinkers' who drink occasionally but have no history of binge drinking or drinking at work or school; those in Class 3 are 'regular binge drinkers' who both drink regularly and engage in binge drinking; and those in Class 4 are 'problematic binge drinkers' who do not drink regularly; however, both engage in binge drinking and drink right before or during school or work hours in the last 30 days.

Table 3.3 presents the estimated marginal probability of belonging to a particular class over time for male and female. To obtain marginal probabilities 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,

$$\bar{\delta}_{c_t|g} = \frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{c_1=1}^{C}\cdots\sum_{c_{t-1}=1}^{C}\hat{\delta}_{c_1|g}(\mathbf{x}_{i1})\prod_{j=2}^{t}\hat{\tau}^{(j)}_{c_j|c_{j-1}g}(\mathbf{x}_{ij})\right\}, \tag{3.1}$$

where $\hat{\delta}_{c_1|g}(\mathbf{x}_{i1})$ and $\hat{\tau}^{(j)}_{c_j|c_{j-1}g}(\mathbf{x}_{ij})$ are the estimated δ and τ values evaluated at the ML estimates for $\boldsymbol{\beta}^{(1)}_{c_1|g}$ and $\boldsymbol{\beta}^{(t)}_{c_t|c_{t-1}g}$, respectively. During early adolescence (*i.e.*, year of 1997), about 71.0% of male early-onset adolescents belonged to Class 1 (not current drinkers). However, the prevalence of Class 1 decreased to 35.1% by the time they were reached 18–20 years in 2003. For early-onset females during early adolescence, about 64.7% belonged to Class 1, less prevalent than male; however, this

Table 3.4. Estimated regression coefficients of age and race for the class prevalence during early adolescence in 1997.

Covariate	Male: Class†			Female: Class†			d.f.	chi-square	p-value
	2	3	4	2	3	4			
Age	0.403	0.881	0.306	0.332	1.237	0.468	6	24.820	0.000
Race‡									
African American	−0.147	−0.224	0.265	−0.797	−1.202	0.174	12	14.771	0.254
Other race	0.225	0.826	−0.210	−0.067	−24.746	0.511			

† Class 1 (not current drinkers) is the baseline class; ‡ Caucasian American is the baseline race.

class was more prevalent when they were reached 18–20 years in 2003, comparing to male early-onset adolescents. Interestingly, the prevalence of Class 2 (light drinkers) were stable over time both for male and female adolescents; however, female adolescents were more likely belong to 'light drinkers' than their male counterpart across all time points. During early adolescence, not many adolescents belonged to Class 3 (regular binge drinkers) or Class 4 (problematic binge drinkers). However, Class 3 emerged both for male and female during middle adolescence but became more prevalent for male early-onset adolescents. The difference in the prevalence of Class 3 between male and female increased from 9.7% during middle adolescence to 19.9% during late adolescence.

Table 3.4 provides the regression coefficients of age and race for the prevalence of each class during early adolescence in 1997. The exponentiated coefficients may be interpreted as change in estimated odds ratios for one-unit increase in a covariate. For instance, the estimated odds of belonging to Class 3 (regular binge drinkers) versus Class 1 (not current drinkers) are $\exp(0.881) = 2.413$ and $\exp(1.237) = 3.445$ times higher as an unit-year increase in age among male and female adolescents, respectively. Speaking broadly, older adolescents are more likely to belong to the classes that are characterized by intensified drinking behaviors, and the likelihood ratio statistic shows that age makes a statistically significant contribution to the model. However, the three-level of race does not explain the difference in the prevalence of class membership during early adolescence. We generated a single binary indicator for race (Caucasian American/Other race), and fitted a four-class LTA with logistic regression. The likelihood ratio test shows that race with two-level binary indicator are not statistically significant, confirming our suspicion that there is no race effect on the class prevalence during early adolescence.

Table 3.5 gives transitional probabilities of moving from one class in 1997 to another in 2000 and from one in 2000 to another in 2003 conditional on gender. To obtain transition probabilities in Table 3.5, we averaged the ML estimates of the subject-specific transition probabilities over the sample,

$$\bar{\tau}^{(t)}_{c_t|c_{t-1}g} = \frac{1}{n}\sum_{i=1}^{n} \frac{\exp\left\{\mathbf{x}'_{it}\hat{\boldsymbol{\beta}}^{(t)}_{c_t|c_{t-1}g}\right\}}{\sum_{c=1}^{C}\exp\left\{\mathbf{x}'_{it}\hat{\boldsymbol{\beta}}^{(t)}_{c|c_{t-1}g}\right\}}.$$

The diagonal values in Table 3.5 represent the probability of remaining the same class as previous time point. For example, during early-mid adolescence (*i.e.*, from 1997 to 2000), the probability of remaining in Class 1 (not current drinkers) for males (0.510) is higher than the probability for females (0.477), implying that not current drinkers among females are more likely to forward to the advanced alcohol classes during early-mid adolescence. However, from mid to late adolescence the probability of remaining in Class 1 is higher for females, becoming female not current drinkers more stable than their males counterpart. The probabilities of remaining in Class 2 (light drinkers) among females (0.342 from early to mid adolescence and 0.542 from mid to late adolescence) are higher

Table 3.5. Estimated transition probabilities (τ-parameter).

Class in 1997	Class in 2000: Male				Class in 2000: Female			
	1	2	3	4	1	2	3	4
1	0.510	0.195	0.253	0.041	0.477	0.385	0.124	0.015
2	0.344	0.316	0.253	0.087	0.369	0.342	0.289	0.000
3	0.412	0.000	0.588	0.000	0.384	0.184	0.433	0.000
4	0.405	0.017	0.450	0.128	0.562	0.438	0.000	0.000
Class in 2000	Class in 2003: Male				Class in 2003: Female			
	1	2	3	4	1	2	3	4
1	0.474	0.240	0.286	0.000	0.595	0.313	0.093	0.000
2	0.317	0.338	0.311	0.034	0.233	0.542	0.224	0.000
3	0.117	0.036	0.846	0.000	0.146	0.123	0.731	0.000
4	0.466	0.000	0.459	0.075	0.337	0.000	0.000	0.663

Table 3.6. Estimated regression coefficients for the transitional probability of 'not current drinkers' ('not current drinkers' class is the baseline class).

Covariate	Class in 2000†: Male			Class in 2000†: Female			d.f.	chi-square	p-value
	2	3	4	2	3	4			
Age	−0.017	0.632	0.498	0.241	−0.239	0.498	6	18.612	0.005
Race‡									
African American	−0.460	−1.949	0.958	−0.871	−2.224	33.692	12	49.717	0.000
Other race	−0.145	−0.026	1.317	−0.378	−1.172	33.352			
Covariate	Class in 2003†: Male			Class in 2003†: Female			d.f.	chi-square	p-value
	2	3	4	2	3	4			
Age	0.268	0.317	.	0.221	0.259	.	4	4.822	0.306
Race‡									
African American	−0.411	−0.840	.	−1.227	−1.228	.	8	22.353	0.004
Other race	−0.250	−0.487	.	−1.045	0.495	.			

† Class 1 (not current drinkers) is the baseline class; ‡ Caucasian American is the baseline race.

than those for males; however, the probabilities of remaining in Class 3 (regular binge drinkers) among females are lower than those for males consistently over time. Interestingly, the diagonal probabilities of males are higher for most of latent classes during early-mid adolescence. However, during mid-late adolescence the diagonal probabilities of females are higher for all latent classes, implying that females are more likely to remain in the same class as their middle adolescence in late adolescence.

Table 3.6 shows the regression coefficients of age and race for the transitions of Class 1 (not current drinkers). Note that these coefficients reflect the transitional probabilities provided in the two first rows of Table 3.5 (*i.e.*, the membership of Class 1 in 1997 and 2000). As mentioned before, we set τ-parameters less than 0.001 to zero a posteriori; however, the regression coefficients regarding the fixed τ-parameters are not estimated. You can see that there are two zero transition probabilities of Class 1 from 2000 to 2003 in Table 3.5. The τ-parameters are fixed as zero, and the corresponding regression coefficients are presented as "·" in Table 3.6 (not estimated). All covariates, except age from 2000 to 2003, are significantly related with the transition of Class 1.

4. Discussion

This research has applied multiple-group LTA model to compare the sequential patterns of drinking behaviors between male and female early-onset adolescents. In our analysis using the data set from

NLSY97, we identified plausible latent classes from drinking items. In addition, the multiple-group LTA differentiated sequential patterns of drinking behaviors across gender. Based on our selected LTA model, early-onset adolescents, regardless of their gender, showed four common latent classes: 'not current drinker', 'light drinkers', 'regular binge drinkers', and 'problematic binge drinkers'. We estimated the effects of age and race on the class membership during early adolescence. For both males and females, older adolescents were more likely to belong to the classes characterized by intensified drinking behaviors during early adolescence. In addition, these covariates were used to predict the transition of early-onset drinkers who belonged to Class 1 (not current drinkers) during their adolescence. We found that white early-onset adolescents that belong to Class 1 were more likely than their counterpart to move forward more advanced drinking classes.

We implemented the ML algorithm to estimate the unknown parameters via EM iterations. As an alternative to this ML estimation, Bayesian inference via Markov chain Monte Carlo(MCMC) methods may be an attractive method of fitting an LTA model. Bayesian methods open up new possibilities for model checking and assessment of fit via the posterior check distribution, and they provide interval estimates without difficulty when hypothesis tests involving combinations of parameters are necessary to address specific research questions. Note that, in our example, the δ-parameter (see in Table 3.3) and τ-parameter (see in Table 3.5) are not directly estimated but rather are functions of other parameters. The standard errors of those estimated parameters cannot be easily calculated using the ML method. Although using a Bayesian approach via MCMC may overcome some of the shortcomings of ML, the ML methods are an efficient way to analyze LTA models in many studies. They have good finite sample properties and provide reliable selection tools. The results of simulation study (Chung *et al.*, 2011) suggested that the ML method via the EM algorithm have good finite sample properties in the latent class profile analysis(LCPA), especially when measurement parameter are strong. Future work should explore the performance of a Bayesian approach for LTA models and compare it to the performance of an ML method via simulation study. We expect that each method will display its own unique strengths and weaknesses under different conditions.

References

Agresti, A. (2002). *Categorical Data Analysis*, (Second ed.). Hoboken, Wiley, New Jersey.

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.

Centers for Disease Control and Prevention (2010). Youth risk behavior surveillance - United States, 2009, *Surveillance Summaries, MMWR*, **59** (SS-5).

Chung, H., Anthony, J. C. and Schafer, J. L. (2011). Latent class profile analysis: an application to stage-sequential process of under-age drinking behaviours, *Journal of the Royal Statistical Society, Series A*, **174**.

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.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via em algorithm (with discussion), *Journal of the Royal Statistical Society, Series B*, **39**, 1–38.

Formann, A. K. (2003). Latent class model diagnosis from a frequentist point of view, *Biometrics*, **59**, 189–196.

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

Hingson, R., Heeren, T., Jamanka, A. and Howard, J. (2000). Age of drinking onset and unintentional injury involvement after drinking, *Journal of the American Medical Association*, **284**, 1527–1533.

Hingson, R., Heeren, T. and Zakocs, R. (2001). Age of drinking onset and involvement in physical fights after drinking, *Pediatrics*, **108**, 872–877.

Langeheine, R., Pannekoek, J. and Van de Pol, F. (1996). Bootstrapping goodness-of-fit measures in categorical data analysis, *Sociological Methods and Research*, **24**, 492–516.

Lynskey, M. T., Bucholz, K. K., Madden, P. A. F. and Health, A. C. (2007). Early-onset alcohol-use behaviors and subsequent alcohol-related driving risks in young women: A twin study, *Journal of Studies on Alcohol and Drugs*, **68**, 798–804.

McHugh, R. B. (1956). Efficient estimation and local identification in latent class analysis, *Psychometrika*, **21**, 331–347.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*, Wiley, New York.