



Published in final edited form as:

J R Stat Soc Ser A Stat Soc. 2010 January 1; 173(1): 145–164. doi:10.1111/j.1467-985X.2009.00607.x.

Latent transition models with latent class predictors: attention deficit hyperactivity disorder subtypes and high school marijuana use

Beth A. Reboussin and

Wake Forest University School of Medicine, Winston-Salem, USA

Nicholas S. Ialongo

Johns Hopkins University Bloomberg School of Public Health, Baltimore, USA

Summary

Attention deficit hyperactivity disorder (ADHD) is a neurodevelopmental disorder which is most often diagnosed in childhood with symptoms often persisting into adulthood. Elevated rates of substance use disorders have been evidenced among those with ADHD, but recent research focusing on the relationship between subtypes of ADHD and specific drugs is inconsistent. We propose a latent transition model (LTM) to guide our understanding of how drug use progresses, in particular marijuana use, while accounting for the measurement error that is often found in self-reported substance use data. We extend the LTM to include a latent class predictor to represent empirically derived ADHD subtypes that do not rely on meeting specific diagnostic criteria. We begin by fitting two separate latent class analysis (LCA) models by using second-order estimating equations: a longitudinal LCA model to define stages of marijuana use, and a cross-sectional LCA model to define ADHD subtypes. The LTM model parameters describing the probability of transitioning between the LCA-defined stages of marijuana use and the influence of the LCA-defined ADHD subtypes on these transition rates are then estimated by using a set of first-order estimating equations given the LCA parameter estimates. A robust estimate of the LTM parameter variance that accounts for the variation due to the estimation of the two sets of LCA parameters is proposed. Solving three sets of estimating equations enables us to determine the underlying latent class structures independently of the model for the transition rates and simplifying assumptions about the correlation structure at each stage reduces the computational complexity.

Keywords

Attention deficit hyperactivity disorder; Estimating equations; Latent class; Latent transition; Marijuana

1. Introduction

Attention deficit hyperactivity disorder (ADHD) is a neurodevelopmental disorder which is most often diagnosed in childhood and characterized by symptoms of inattention, hyperactivity and impulsiveness. It is estimated that 3–7% of school-aged youth in the USA have ADHD (American Psychiatric Association, 2000) and 1–5% in Europe (Swanson *et al.*, 1998; Polanczyk *et al.*, 2007). Many children with ADHD experience behavioural and

psychosocial difficulties. In particular, high rates of substance use disorders have been evidenced in studies of adolescents who were diagnosed with ADHD in childhood (Barkley *et al.*, 1990; Levin and Kleber, 1995). Whereas early studies focused on global assessments of substance use, more recent work suggests that individuals with ADHD might be at risk for using specific drugs; however, the results are not consistent. Molina and Pelham (2003) found that, although ADHD adolescents were more likely than controls to report alcohol-related problems, they were not more likely to use alcohol, cigarettes or marijuana. In contrast, others have found increased rates of nicotine dependence, regular cannabis smoking and daily smoking but not alcohol use or dependence among adolescents with ADHD (Galera *et al.*, 2005, 2008; Biederman *et al.*, 2006). Just as there has been an increased focus on specific drugs and ADHD, researchers are beginning to look at the relationship between specific symptoms or subtypes of ADHD and substance use. Currently three ADHD subtypes are recognized in American Psychiatric Association (2000); predominantly inattentive, predominantly hyperactive-impulsive and a combined subtype. Inattentiveness symptoms have been shown to be associated with marijuana and nicotine dependence (Abrantes *et al.*, 2005), tobacco use (Burke, *et al.*, 2001) and frequency of alcohol, marijuana and tobacco use (Molina and Pelham, 2003). Symptoms of hyperactivity and impulsiveness have been shown to be associated with alcohol use (Burke *et al.*, 2001), illicit drug use (Molina and Pelham, 2003), earlier age of initiation of substance use (Molina and Pelham, 2003; Elkins *et al.*, 2007) and marijuana and nicotine dependence (Elkins *et al.*, 2007).

In this paper, we propose to use a latent transition model (LTM) to study stage sequential transitions in use of marijuana through the high school years and to examine the influence of ADHD subtypes on these transitions. The transition concept, which was introduced in the 1990s, depicts drug involvement as a sequence of transitions from earliest opportunities to use a drug, to first use and consequences, followed by the drug dependence syndrome (Stenbacka *et al.*, 1993; Anthony and Helzer, 1995). Within this framework it has been suggested that determinants of transitions across stages of drug involvement may be different (Clayton, 1992; Anthony and Helzer, 1995). This may explain the inconsistent findings regarding the relationship between drug use and ADHD. For example, ADHD may influence drug use initiation differently from how it does for more serious drug involvement. In addition to positing drug use as a stage sequential process with the potential for different influences at each stage, LTMs offer two additional advantages. First, LTMs are an empirically based procedure that allows the data to guide our understanding of how drug use progresses by creating homogeneous groups of individuals with similar drug use profiles. Secondly, LTMs account for the measurement error that is often found in self-reported substance use data (Harrison and Hughes, 1997; Golub *et al.*, 2000) by assuming that each response is an imperfect measure of drug use.

The current study extends the LTM of Reboussin *et al.* (1999) by incorporating a latent class predictor to represent ADHD subtypes. This is appealing because data that are based on clinic-based samples have demonstrated higher rates of the combined subtype (Lahey *et al.*, 1994; Faraone *et al.*, 1998) whereas population-based studies have had mixed results with some suggesting that the inattentive subtype is the most prevalent (Gaub and Carlson, 1997; Froehlich *et al.*, 2007) and others the combined type (Angold *et al.*, 2002; Ford *et al.*, 2003). In addition, we know that using different diagnostic criteria (e.g. those of American Psychiatric Association (2000) and version 10 of the international classification of diseases) results in different ADHD prevalence estimates (Swanson *et al.*, 1997; Polanczyk *et al.*, 2007). Therefore, rather than rely on specific diagnostic criteria to define ADHD subtypes, we use latent class analysis (LCA) to identify empirically subgroups of individuals in an urban sample of high school students with similar profiles of inattentive, hyperactive and impulsive behaviours. Parameter estimation involves an estimating equations analogue of

the pseudolikelihood method for estimation of the parameters of interest, namely the transition model parameters. We begin by fitting two separate LCA models by using second-order estimating equations:

- a. a longitudinal LCA model to define stages of marijuana use, which is hereafter referred to as the latent stage model, and
- b. a cross-sectional LCA model to define the ADHD subtypes.

The LTM model parameters describing the probability of transitioning between the LCA-defined stages of marijuana use and the influence of the LCA-defined ADHD subtypes on these transition rates are then estimated by using a set of first-order estimating equations given the LCA model parameter estimates. A robust estimate of the LTM parameter variance that accounts for the variation due to the estimation of the two sets of LCA parameters is proposed. By solving three separate estimating equations that only require specification of first and second moments and that allow us to make simplifying assumptions about the correlation structure at each stage, computational complexity is significantly reduced.

2. Latent transition model with latent class predictor

2.1. The latent stage model of drug use

Let $\mathbf{y}_{it} = (y_{i1t}, \dots, y_{ipt})$, $t = 1, \dots, T$, be a vector of binary responses regarding drug use behaviours for individual i at time t where $y_{ijt} = 1$ if the response to behaviour j at time t is 'yes' and $y_{ijt} = 0$ otherwise. We refer to \mathbf{y}_{it} as the drug use behaviour profile at time t . We assume that the co-occurrence of the behaviours comprising the drug use profile \mathbf{y}_{it} can be explained by an underlying classification of individuals into subgroups (classes) with similar drug use profiles representing stages of drug use. In a statistical sense this means that, within a stage of drug use, behaviours are independent. This is the axiom of local independence that forms the basis for LCA. Since latent class membership is not observed without error, this assumption is not verifiable; however, its adequacy under various class assumptions will be discussed later. For estimation purposes, we assume there are D latent stages indexed by $1, \dots, D$. The latent stage of drug use for individual i at time t is denoted by D_{it} .

Information about the latent stages of drug use are conveyed through two sets of parameters; the probability of reporting behaviour j at time t within stage m of drug use,

$$p_{jmt} = P(y_{ijt} = 1 | D_{it} = m),$$

and the proportion of individuals in stage of m of drug use at time t , or the latent class prevalence,

$$\pi_{mt} = P(D_{it} = m).$$

We shall refer to $\boldsymbol{\theta}_{1t} = (\mathbf{p}_t, \boldsymbol{\pi}_t)$ as the latent stage parameters for time t where $\boldsymbol{\theta}_1 = (\boldsymbol{\theta}_{11}, \dots, \boldsymbol{\theta}_{1T})$. The conditional response probabilities p_{jmt} aid in the interpretation of the stages of drug use by characterizing the behaviour of individuals within a particular stage. Although in principle it is possible to allow the conditional response probabilities to vary over time, this implies that the definition of drug use stages is changing, which would substantially complicate interpretation of the longitudinal model. Therefore, we constrain the conditional response probabilities to be constant over time, i.e. $p_{jmt} = p_{jm} \forall t = 1, \dots, T$. This is

analogous to constraining the factor loadings to be equal over time in a longitudinal factor analysis model (which is sometimes referred to as factor invariance). To restrict the conditional response probabilities to the admissible interval, a logistic representation was used, $p_{jm} = \exp(\zeta_{jm}) / \{1 + \exp(\zeta_{jm})\}$. Consistent with the longitudinal framework, the proportion of individuals in each drug use stage π_{mt} is allowed to vary over time.

2.2. The latent class attention deficit hyperactivity disorder predictor model

Let $\mathbf{z}_i = (z_{i1}, \dots, z_{ir})$ be a vector of responses regarding ADHD symptoms measured before the drug use behaviours for individual i where $z_{iu} = 1$ if ADHD symptom u is present and $z_{iu} = 0$ otherwise. We refer to \mathbf{z}_i as the ADHD symptom profile for individual i . We assume that the co-occurrence of the ADHD symptoms can be explained by an underlying classification of individuals into subgroups (classes) with similar symptom profiles representing ADHD subtypes. For estimation purposes, we assume that there are A subtypes indexed by $1, \dots, A$. The latent ADHD subtype for individual i is denoted by A_i . Similarly to the latent stage model, information about ADHD subtypes is conveyed through two sets of parameters; the probability of reporting ADHD symptom u for individuals with ADHD subtype l ,

$$p_{ul}^* = P(z_{iu} = 1 | A_i = l),$$

and the proportion of individuals with ADHD subtype l ,

$$\pi_l^* = P(A_i = l).$$

We refer to $\theta_2 = (\mathbf{p}^*, \boldsymbol{\pi}^*)$ as the ADHD measurement parameters. Similarly to the latent stage model in Section 2.1, to restrict the conditional response probabilities to the admissible interval, a logistic representation was used, $p_{ul}^* = \exp(\zeta_{ul}^*) / \{1 + \exp(\zeta_{ul}^*)\}$.

2.3. The latent transition model

Scientific interest focuses on changes in stages of drug use over time and the influence of ADHD subtypes on these transitions. This makes the modelling of transition probabilities $\tau_{it}(k, m) = P(D_{it} = m | D_{i,t-1} = k)$ as a function of predictors natural. We begin by considering the first-order transition model of Reboussin *et al.* (1999) where the transition probabilities are modelled as a function of observed covariates \mathbf{x}_{it} , discrete or continuous and possibly time dependent:

$$\log \left\{ \frac{\tau_{it}(k, m; \mathbf{x}_{it})}{\tau_{it}(k, 1; \mathbf{x}_{it})} \right\} = \alpha_m + \beta_{km} + \Upsilon_m \mathbf{x}_{it} + \Psi_{km} \mathbf{x}_{it} \quad (1)$$

where $t = 2, \dots, T$, $m = 2, \dots, D$, $k = 1, \dots, D$, $\beta_{1m} = 0$ and $\Psi_{1m} = 0$. Model (1) is a multinomial logistic regression model. The probability that an individual transitions from stage $D_{i,t-1} = k$ at time $t - 1$ to stage $D_{it} = m$ at time t given covariates \mathbf{x}_{it} is represented by the logistic function:

$$\tau_{it}(k, m; \mathbf{x}_{it}) = \frac{\exp(\alpha_m + \beta_{km} + \Upsilon_m \mathbf{x}_{it} + \Psi_{km} \mathbf{x}_{it})}{1 + \sum_{l=2}^D \exp(\alpha_l + \beta_{kl} + \Upsilon_l \mathbf{x}_{it} + \Psi_{kl} \mathbf{x}_{it})} \quad (2)$$

The parameter α_m in model (1) is the log-odds of transitioning from stage $D_{i,t-1} = 1$ (e.g. no drug use) at time $t - 1$ to stage $D_{it} = m$ at time t relative to remaining in stage 1 at time t given $\mathbf{x}_{it} = 0$. The log-odds of transitioning from stage $D_{i,t-1} = k$ at time $t - 1$ to stage $D_{it} = m$ at time t relative to remaining in stage k at time t when $k \neq 1$ and $m > k$ is given by $\alpha_m - \alpha_k + \beta_{km} - \beta_{kk}$ when $\mathbf{x}_{it} = 0$. If we assume for illustration that there is a single binary variable x_{it} , then v_m is the log-odds-ratio comparing the odds of transitioning from stage 1 to stage m for $x_{it} = 1$ versus $x_{it} = 0$. Similarly, $v_{lm} - v_{lk} + \psi_{km} - \psi_{kk}$ is the log-odds-ratio comparing the odds of transitioning from stage k to stage m for $x_{it} = 1$ versus $x_{it} = 0$ when $k \neq 1$ and $k > m$.

We extend model (1) to examine how transition probabilities depend on latent ADHD subtypes by including an interaction between an individual's prior latent stage of drug use $D_{i,t-1}$ and latent ADHD subtype A_i as shown below:

$$\log \left\{ \frac{\tau_{it}(k, m; \mathbf{x}_{it}, A_i=l)}{\tau_{it}(k, 1; \mathbf{x}_{it}, A_i=l)} \right\} = \alpha_m + \beta_{km} + \Upsilon_m \mathbf{x}_{it} + \Psi_{km} \mathbf{x}_{it} + \delta_{lm} + \Gamma_{lkm} \tag{3}$$

where $m = 2, \dots, D, k = 1, \dots, D, l = 1, \dots, A, \delta_{1m} = 0$ and $\Gamma_{lkm} = 0$ if $k = 1$ or $l = 1$. We shall refer to $\theta_3 = (\alpha, \beta, \mathbf{T}, \mathbf{\Psi}, \delta, \mathbf{\Gamma})$ as the latent transition parameters. The parameter δ_{lm} is the log-odds-ratio comparing the odds of transitioning from stage 1 to stage m for a youth of ADHD subtype l compared with a youth of ADHD subtype 1. The log-odds-ratio that compares the odds of transitioning from stage k to stage m where $k \neq 1$ and $m > k$ for a youth of ADHD subtype l compared with a youth of ADHD subtype 1 is given by $\delta_{lm} - \delta_{lk} + \gamma_{lkm} - \gamma_{lkk}$. This model can be extended further to allow the effect of ADHD subtype on the transition probabilities to depend on observed covariates \mathbf{x}_{it} , e.g. age and gender, by including an interaction between ADHD subtype and \mathbf{x}_{it} and a three-way interaction between prior stage of drug use, ADHD subtype and \mathbf{x}_{it} .

3. Estimation

3.1. Estimation of θ_1

Similarly to Reboussin *et al.* (1999), we propose first to solve a set of estimating equations $U_{1t}(\theta_{1t})$ for estimating the latent stage parameters θ_{1t} at time t that incorporate information from both the first- and the second-order moments of the observed drug use profile \mathbf{y}_{it} . This is unlike estimating equations for generalized linear models which only use information in the first-order moments (Zeger and Liang, 1986). Information in the second-order moments is necessary for identification of the latent stage model parameters in which the covariance between responses is of scientific interest.

The estimating equations are formed by equating the observed responses \mathbf{y}_{it} and $\mathbf{w}_{it} = \{(y_{ijt} - \mu_{ijt})(y_{iht} - \mu_{iht}); j < h = 1, \dots, p\}$ to their expected values $\boldsymbol{\mu}_{it} = E[\mathbf{y}_{it}]$ and $\boldsymbol{\sigma}_{it} = E[\mathbf{w}_{it}]$. The first-order moments are given by $\mu_{ijt} = E[y_{ijt}] = \sum_{m=1}^D p_{jmt} \pi_{mt}$. Under the assumption of conditional independence, the second-order moments are now

$$\sigma_{ijht} = E[(y_{ijt} - \mu_{ijt})(y_{iht} - \mu_{iht})] = \sum_{m=1}^D p_{jmt} p_{hmt} \pi_{mt} - \mu_{ijt} \mu_{iht}.$$

The estimating equations are weighted by the matrix \mathbf{C}_{it} of first-order derivatives of the first two moments with respect to the set of parameters θ_{1t} and a working $p(p + 1)/2 \times p(p + 1)/2$ covariance matrix \mathbf{R}_{it} of \mathbf{y}_{it} and \mathbf{w}_{it} . The covariance matrix is referred to as working because,

as demonstrated by Liang *et al.* (1992), parameter estimates and standard errors remain consistent even if the covariance is misspecified.

The proposed second-order estimating equations are then

$$U_{1t}(\theta_{1t}) = \sum_{i=1}^N \mathbf{C}'_{it} \mathbf{R}_{it}(\theta_{1t})^{-1} \begin{pmatrix} \mathbf{y}_{it} - \mu_{it}(\theta_{1t}) \\ \mathbf{w}_{it}(\theta_{1t}) - \sigma_{it}(\theta_{1t}) \end{pmatrix} = 0. \tag{4}$$

The solution of the estimating equations (4) can occur simultaneously for all $t, t = 1, \dots, T$, assuming independence between time periods. The cross-products of indicators at adjacent time points, e.g. $y_{ijt} y_{ih,t-1}$, are not included since they contain the same information as the first moments in the estimating equations for θ_3 . Cross-products of indicators that are more than 1 unit apart in time are also not incorporated owing to the computational burden. It is reasonable to assume that the indicators at time t provide most of the information about the latent stage variable D_{it} . Failure to incorporate cross-products of indicators that are more than 1 unit apart in time results in potential loss of efficiency in parameter estimation but no risk of bias. To avoid higher order moment specifications, we assume that $\text{cov}(\mathbf{y}_{it} \mathbf{w}_{it}) = 0$ and $\text{cov}(\mathbf{w}_{it})$ is diagonal. These estimating equations are solved simultaneously for $\theta_{1t}, t = 1, \dots, T$, by using a Newton–Raphson iterative procedure. A robust variance estimator of $\hat{\theta}_1$ that is consistent even when the working covariance matrix of \mathbf{y}_i and \mathbf{w}_i is misspecified is given in Appendix A.

3.2. Estimation of θ_2

The ADHD measurement parameters θ_2 are estimated by solving a set of estimating equations $U_2(\theta_2)$ similar to the estimating equations (4) that incorporate information from both the first- and the second-order moments of the observed ADHD symptom profile \mathbf{z}_i . Unlike equations (4), we consider the ADHD symptom profile at a single time point. The estimating equations are formed by equating the observed responses \mathbf{z}_i and

$\mathbf{w}_i^* = \{(z_{ij} - \mu_{ij}^*)(z_{ih} - \mu_{ih}^*); j < h = 1, \dots, r\}$ to their expected values $\mu_i^* = E[\mathbf{z}_i]$ and $\sigma_i^* = E[\mathbf{w}_i^*]$. The

first-order moments are given by $\mu_{ij}^* = E[z_{ij}] = \sum_{l=1}^A p_{jl}^* \pi_l^*$. Under the assumption of conditional independence, the second-order moments are given by

$$\sigma_{ijh}^* = E[(Z_{ij} - \mu_{ij}^*)(Z_{ih} - \mu_{ih}^*)] = \sum_{l=1}^A p_{jl}^* p_{hl}^* \pi_l^* - \mu_{ij}^* \mu_{ih}^*.$$

The estimating equations are weighted by the matrix \mathbf{E}_i of first-order derivatives of the first two moments with respect to the set of parameters θ_2 and a working $r(r+1)/2 \times r(r+1)/2$ covariance matrix \mathbf{R}_i^* of \mathbf{z}_i and \mathbf{w}_i^* . The proposed second-order estimating equations

$$U_2(\theta_2) = \sum_{i=1}^N \mathbf{E}'_i \mathbf{R}_i^*(\theta_2)^{-1} \begin{pmatrix} \mathbf{z}_i - \mu_i^*(\theta_2) \\ \mathbf{w}_i^*(\theta_2) - \sigma_i^*(\theta_2) \end{pmatrix} = 0 \tag{5}$$

are solved for θ_2 by using a Newton–Raphson iterative procedure. Similarly to the estimation of θ_1 , we avoid higher order moment specifications by assuming that $\text{cov}(\mathbf{z}_i \mathbf{w}_i^*) = 0$

and $\text{cov}(\mathbf{w}_i^*)$ is diagonal. A robust variance estimator of θ_2 that is consistent even when the working covariance matrix of \mathbf{z}_i and \mathbf{w}_i^* is misspecified is given in Appendix B.

3.3. Estimation of θ_3

We propose to solve the following set of first-order estimating equations (Liang and Zeger, 1986; Zeger and Liang, 1986) for the latent transition parameters θ_3 similarly to Reboussin *et al.* (1999),

$$U_3(\theta_3, \widehat{\theta}_1, \widehat{\theta}_2) = \sum_{i=1}^N \mathbf{B}_i V_i^{-1}(\theta_3, \widehat{\theta}_1, \widehat{\theta}_2) \{\mathbf{y}_i^* - \boldsymbol{\eta}_i(\theta_3, \widehat{\theta}_1, \widehat{\theta}_2)\} = 0 \tag{6}$$

where $\mathbf{y}_i^* = (\mathbf{y}_{i2}, \dots, \mathbf{y}_{iT})$, $\boldsymbol{\eta}_i = (\boldsymbol{\eta}_{i2}, \dots, \boldsymbol{\eta}_{iT})$, $\boldsymbol{\eta}_{it} = E[\mathbf{y}_{it} | \mathbf{y}_{i,t-1}, \mathbf{x}_{it}, \mathbf{z}_i; \theta_3, \widehat{\theta}_1, \widehat{\theta}_2]$, $\widehat{\theta}_1$ is a \sqrt{N} -consistent estimator of the latent stage parameters θ_1 discussed in Section 3.1 and $\widehat{\theta}_2$ is a \sqrt{N} -consistent estimator of the ADHD measurement parameters θ_2 discussed in Section 3.2. The matrix \mathbf{B}_i is a matrix of first-order derivatives of the first conditional moments with respect to θ_3 .

Assuming that the covariates \mathbf{x}_{it} and ADHD symptom profile \mathbf{z}_i are only included because of their expected association with the transition probabilities, the first conditional moment of y_{ijt} given $\mathbf{y}_{i,t-1}$, \mathbf{x}_{it} and \mathbf{z}_i can be expressed as

$$\eta_{ijt} = \sum_{m=1}^D \sum_{k=1}^D \sum_{l=1}^A p_{jmt} \tau_{it}(k, m; \mathbf{x}_{it}, A_i=l) P(D_{i,t-1}=k | \mathbf{y}_{i,t-1}) P(A_i=l | \mathbf{z}_i)$$

where

$$P(D_{i,t-1}=k | \mathbf{y}_{i,t-1}) = \frac{\prod_{s=1}^p p_{sk,t-1}^{y_{is,t-1}} (1 - p_{sk,t-1})^{1-y_{is,t-1}} \pi_{k,t-1}}{\sum_{f=1}^D \prod_{q=1}^p p_{qf,t-1}^{y_{iq,t-1}} (1 - p_{qf,t-1})^{1-y_{iq,t-1}} \pi_{f,t-1}}$$

and

$$P(A_i=l | \mathbf{z}_i) = \frac{\prod_{s=1}^r p_{sl}^{*z_{is}} (1 - p_{sl}^*)^{1-z_{is}} \pi_l^*}{\sum_{f=1}^A \prod_{q=1}^r p_{qf}^{*z_{iq}} (1 - p_{qf}^*)^{1-z_{iq}} \pi_f^*}$$

We refer to the matrix V_i as a working covariance matrix. Let $V_{it} = \text{cov}(\mathbf{y}_{it} | \mathbf{y}_{i,t-1}, \mathbf{x}_{it}, \mathbf{z}_i)$ be the working covariance matrix for the indicators at the t th time point. To simplify calculations, we assume that $V_i = \text{diag}(V_{i2}, \dots, V_{iT})$ so that $\text{cov}(\mathbf{y}_{it}, \mathbf{y}_{it'} | \mathbf{y}_{i,t-1}, \mathbf{y}_{i,t'-1}, \mathbf{x}_{it}, \mathbf{x}_{it'}, \mathbf{z}_i) = 0 \forall t \neq t'$. If the random process $\{\mathbf{y}_{it}, t = 1, \dots, T\}$ is a first-order Markov chain this

assumption is met. The (j, h) -element of V_{it} is given by $V_{it}(j, h) = E[y_{ijt} y_{iht} | \mathbf{y}_{i,t-1}, \mathbf{x}_{it}, \mathbf{z}_i] - \eta_{ijt} \eta_{iht}$ where

$$E[y_{ijt} y_{iht} | \mathbf{y}_{i,t-1}, \mathbf{x}_{it}, \mathbf{z}_i] = \sum_{m=1}^D \sum_{k=1}^D \sum_{l=1}^A p_{jmt} p_{hmt} \tau_{it}(k, m; \mathbf{x}_{it}, A_i=l) P(D_{i,t-1}=k | \mathbf{y}_{i,t-1}) P(A_i=l | \mathbf{z}_i).$$

The generalized estimating equations (GEEs) (6) are not explicitly solvable for θ_3 so an iterative approach is necessary for parameter estimation. Starting with an initial estimate $\widehat{\theta}_3^0$, equation (6) can be solved by using the following Newton–Raphson iterative scheme:

$$\widehat{\theta}_3^f = \widehat{\theta}_3^{f-1} + \left(\sum_{i=1}^N \widehat{B}_i \widehat{V}_i^{-1} \widehat{B}_i \right)^{-1} \sum_{i=1}^N \widehat{B}_i \widehat{V}_i^{-1} (\mathbf{y}_i^* - \widehat{\eta}_i), \quad f=1, 2, \dots,$$

where $\widehat{B}_i = \widehat{B}_i(\widehat{\theta}_3^{f-1}, \widehat{\theta}_1, \widehat{\theta}_2)$, $\widehat{V}_i = V_i(\widehat{\theta}_3^{f-1}, \widehat{\theta}_1, \widehat{\theta}_2)$, $\widehat{\eta}_i = \eta_i(\widehat{\theta}_3^{f-1}, \widehat{\theta}_1, \widehat{\theta}_2)$ and f denotes the estimate at the f th iteration.

Following the work of Zeger and Liang (1986) and Liang and Zeger (1986), the GEEs (6) yield consistent and asymptotically multivariate normal estimates of θ_3 as long as the first conditional moment η_i is correctly specified. The covariance of y_i^* need not be correctly specified. If $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are \sqrt{N} -consistent estimators of θ_1 and θ_2 respectively, this approach can be considered an estimating function analogue of the pseudolikelihood method for the parameters of interest θ_3 (Gong and Samaniego, 1981). If the covariance of y_i^* is correctly specified and θ_1 and θ_2 are known, a model-based consistent estimator of the asymptotic covariance of $\widehat{\theta}_3$ is

$$\widehat{\text{var}}(\widehat{\theta}_3)_{\text{mod}} = \left(\sum_{i=1}^N \widehat{B}_i' \widehat{V}_i^{-1} \widehat{B}_i \right)^{-1} / N \tag{7}$$

where θ_3 is replaced by $\widehat{\theta}_3$. A robust estimator of the variance of $\widehat{\theta}_3$ which accounts for the variation due to the estimation of θ_1 and θ_2 and which remains consistent even when the covariance of y_i^* is misspecified is given in Appendix C.

4. Use of marijuana and attention deficit hyperactivity disorder subtype example

Our application concerns a sample of 495 high school students who participated in a randomized school-based, preventive intervention trial which targeted early learning and aggressive or disruptive behaviour in the first grade in nine schools in Baltimore, Maryland, USA (Ialongo *et al.*, 1999). In 1993, 799 urban first-graders were recruited from 27 classrooms in nine Baltimore city public elementary schools. Students and their families were interviewed annually, although no assessments were conducted in the fourth and fifth grades. Beginning in the sixth grade, parental consent was obtained to participate in middle and high school assessments in which youths would be asked about their experiences with drugs. Of the original 799 adolescents who were recruited in the first grade, 495 (85%) completed face-to-face interviews in the eighth grade that included questions about

inattentive, hyperactive and impulsive behaviours and questions about use of marijuana in the ninth and 10th grades. These 495 youths comprised the sample of interest for studying transitions in high school use of marijuana. Approximately 55% of the sample were male and 87% were African American. Because data on use of marijuana were not collected until the sixth grade, we could not compare the 495 youths participating in this study and the 304 youths from the original sample who were not included in this study on the response variables. However, *t*-tests revealed no differences between these groups in terms of the first-grade behavioural measures of self-reported anxious and depressive symptoms or teacher ratings of concentration problems, hyperactivity and impulsiveness. The 304 youths who were not participating in this study had slightly higher scores on the teacher-rated aggressive–disruptive behaviour scale. The sample of 495 youths with data available in the ninth and 10th grade decreased to 432 youths in the 11th grade and 415 youths in the 12th grade. χ^2 -tests revealed no differences in terms of measures of marijuana use in the ninth and 10th grades between the 415 youths with complete data and the 80 youths with missing assessments. These youths also did not differ on first-grade behavioural measures.

We characterized use of marijuana in high school by considering responses to five questions that were asked in the spring of the ninth, 10th, 11th and 12th grades.

- a. Have you used marijuana since this time last year?
- b. Did you use marijuana in the last month?
- c. How many times have you used marijuana in the past month?
- d. Have you ever used marijuana every day or almost every day for 2 or more weeks?
- e. During the past 12 months, have you gotten into trouble at home, at school or with the police because you used marijuana?

For analysis, we dichotomized the third question as three or more times in the past month *versus* two or fewer to represent someone who has used marijuana more than just a couple of times in the past month. The last question combined responses to three individual questions that were asked during the assessments into a single indicator of social problems. This was done because of the low individual prevalences and their lack of discriminatory power.

We first applied the latent stage model of drug use that was described in Section 2.1 to examine the structure underlying the five behaviours comprising the marijuana use profile. We started with the most parsimonious one-stage model ('all marijuana use the same') with progression to a less parsimonious model with three stages of marijuana use. Because models with different numbers of stages (classes) are not nested, precluding the use of a difference likelihood ratio test, we must rely on measures of fit such as Akaike's information criterion (AIC), which is a global fit index which combines goodness of fit and parsimony. In comparing different sets of models with the same set of data, models with lower values are preferred. Because the AIC requires a likelihood for model comparison and estimating equations approaches are non-likelihood based, we used a modified version of the AIC (Pan, 2001; Reboussin *et al.*, 2006, 2008). The AIC suggested a best fitting model based on two stages of use of marijuana (AIC1=111811; AIC2 =110810; AIC3 =175981). However, rather than rely solely on global indices of fit like the AIC, we also examined the validity of the latent stage model assumption of local independence more directly. Specifically, we use a modified version of Garrett and Zeger's (2000) log-odds-ratio check that was suggested by Uebersax (2000). This method involved calculating the log-odds-ratio in both the observed and the expected two-way tables for all pairs of marijuana use behaviours. The observed data log-odds-ratio is then expressed as a *z*-score relative to the expected data log-odds-ratio. This *z*-value is then used as a guide to detect items that are locally dependent. Uebersax

(2000) suggested that the p -values should not be interpreted literally but rather that the focus should be on the relative magnitude of the z -values. A threshold of ± 1.5 was conservatively chosen as suggestive of local dependence. Under the two-stage model of use of marijuana, z -values exceeding the threshold provided evidence for violation of the local independence assumption. The addition of a third stage removed all local dependences.

Although we can always achieve local independence by increasing the number of latent classes (Suppes and Zanotti, 1981), doing so may yield spurious classes that are not immediately interpretable to experts in the field. For this reason, we examined the resultant latent structures to evaluate their interpretability. As displayed in Fig. 1, the two-stage model divided use of marijuana into 'no use' (stage 1) and 'use' (stage 2). However, the second stage is predominantly characterized by past-month users with moderate levels of use (approximately 80% report past-month use and 60% report using three or more times in the past month). In contrast, the three-stage model (Fig. 2) creates a stage of use of marijuana that is characterized by less frequent use. Under the two-stage model infrequent use is difficult to classify with 15% of those in the 'no-use' stage reporting past year use and 20% of those in the 'current use' stage not reporting past-month use. With the three-stage model, infrequent use becomes an intermediary stage of use of marijuana between no use (stage 1) and current, more frequent and problematic marijuana use (stage 3). The three-stage model is consistent with Anthony and Helzer's (1995) transition concept where an individual might experiment with a drug once or twice and then over time proceed to a more frequent and problematic pattern of drug use as opposed to going directly from no use to frequent use. Since the two-stage model provided the best overall fit on the basis of the AIC but did not fully explain the heterogeneity in the marijuana use profiles as evidenced by residual local dependences and because the three-stage model provided a substantively meaningful description of the progression of use of marijuana, we present results for both the two-stage and the three-stage models when examining associations with ADHD subtypes. We note that our stepwise approach has the advantage that the latent structures of marijuana use and ADHD are determined separately. Hence, the acceptance of a two- or three-stage model of marijuana use does not influence the choice of the appropriate number of ADHD classes.

Before examining associations with ADHD, we were interested in estimating the overall probability of transitioning between stages of use of marijuana during the high school years. Given the estimates of the latent stage parameters from the two- and three-stage models just described, we fit the LTM (1) where the observed covariate vector $x_{it} = (x_{it1}, x_{it2})$ contained two binary indicator variables representing the 10th–11th- and 11th–12th-grade transitions relative to the 9th–10th-grade transition respectively. Presented in Tables 1 and 2 are the estimated probabilities of transitioning from one stage of use of marijuana at the current grade to another stage of use of marijuana in the following grade for the two- and three-stage models respectively. Both models suggest that there is a greater risk of transitioning out of the no-use stage during the first 2 years of high school, i.e. 9th–10th grade and 10th–11th grade compared with 11th–12th grade. The rate of advance based on the two-stage model was 7.9% from ninth to 10th grade, 6.8% from 10th to 11th grade and 3.3% from 11th to 12th grade. On the basis of the estimated two-stage LTM, youths were significantly more likely to transition from no use to use during the 9th–10th-grade transition compared with the 11th–12th-grade transition (odds ratio OR = 4.14; 95% confidence interval CI = (2.15, 7.96)). Similarly, under the three-stage model, the rate of advance was 16.7% from ninth to 10th grade, 13.6% from 10th to 11th grade and 10.5% from 11th to 12th grade. The likelihood of transitioning from no use to infrequent use (stage 2) was 54% greater from ninth to 10th grade compared with from 11th to 12th grade (OR = 1.54; 95% CI = (0.98, 2.44)). This difference was marginally significant ($p = 0.058$). Although much less common, the risk of transitioning from no use to frequent use (stage 3) was almost five times greater from 10th to 11th grade compared with 11th to 12th grade (OR = 4.72; 95% CI = (1.08,

20.64)). The 9th–10th-grade transition posed the greatest risk for movement from infrequent (stage 2) to frequent (stage 3) use of marijuana but these differences were not statistically significant.

Next, we were interested in the influence of LCA-derived ADHD subtypes on these transitions in use of marijuana. We considered 11 items from three subscales of the ‘Teacher observation for classroom adaptation’ interview measuring symptoms of inattention, hyperactivity and impulsivity in the eighth grade. This is a structured interview with the teacher administered by a trained assessor (Werthamer-Larsson *et al.*, 1991). Items were rated on a six-point frequency scale where 1 is almost never, 2 rarely, 3 sometimes, 4 often, 5 very often and 6 always. We dichotomized items so that an individual symptom was considered present if the teacher reported that it was present often, very often or always and absent if it was observed sometimes, rarely or almost never. We characterized ADHD symptom profiles in the eighth grade by considering teachers’ responses to the following items where items 1–5 are associated with inattention problems, items 6–8 with hyperactivity and items 9–11 with impulsiveness:

1. trouble completing assignments;
2. difficulty concentrating;
3. easily distracted;
4. cannot stay on task;
5. has difficulty organizing tasks;
6. cannot sit still;
7. is always on the go or acts as if driven by a motor;
8. fidgeted and/or squirmed a lot;
9. cannot wait for turn;
10. interrupts or intrudes on others;
11. blurts out answers before the question is complete.

We applied the latent class ADHD predictor model in Section 2.2 to examine the structure underlying the ADHD symptom profile that was just described. We started with the most parsimonious one-class model with progression to a four-class model of ADHD. The AIC suggested a best-fitting model based on the three classes of ADHD (AIC1 = 63962; AIC2 = 60154; AIC3 = 58902; AIC4 = 59124). A check of the local independence assumption via the log-odds-ratio residuals for the three-class model indicated that there were no residual dependences. As stated previously, because statistical analysis may yield models that are not substantively meaningful despite their optimality based on goodness-of-fit statistics, we also examined the resultant three-class latent structure of ADHD to evaluate its interpretability. Under the three-class model that is displayed in Fig. 3, 65% of youths do not have ADHD on the basis of teacher reports. Approximately a quarter of youths were reported by teachers to exhibit only symptoms of inattention. The third class of youths, representing 12% of the sample, were reported to exhibit symptoms of inattention, hyperactivity and impulsiveness. We refer to this as the combined subtype. Although there was no evidence for a hyperactive–impulsive subtype in our sample of youth, the inattentive and combined subtypes were consistent with the diagnostic subtypes that are defined in American Psychiatric Association (2000).

On the basis of the three-class model of ADHD, we then examined the influence of the ADHD subtypes on the probability of transitioning between stages of use of marijuana. We

did this by fitting model (3) but included an interaction between grade and the ADHD subtype and a three-way interaction between prior latent stage, grade and ADHD subtype. This allowed us to estimate grade-specific influences of ADHD subtypes on transitions. On the basis of the two-stage model of marijuana use, youths with the inattentive and combined ADHD subtypes were more likely than youths without ADHD to transition from no use of marijuana (stage 1) to use of marijuana (stage 2) during high school. These differences were statistically significant for the inattentive subtype during the 9th–10th- (OR = 2.50; 95% CI = (1.17, 5.36)) and 10th–11th- (OR = 1.89; 95% CI = (1.07, 3.33)) grade transitions and for the combined subtype during the 10th–11th- (OR = 2.26; 95% CI = (1.11, 4.57)) and 11th–12th- (OR = 4.72; 95% CI = (1.35, 16.52)) grade transitions. As indicated by the three-stage model, this increased risk of transitioning for the ADHD subtypes relative to youths without ADHD was stage specific. Youths with the combined ADHD subtype were four times more likely to transition from no use of marijuana to infrequent use during the 10th–11th-grade transition relative to youths without ADHD (OR = 4.01; 95% CI = (1.57, 10.26)) as seen in Fig. 4(a). Youths with the inattentive subtype were five times more likely to transition from no use of marijuana to frequent use (Fig. 4(b)) from 9th to 10th grade compared with youths without ADHD (OR = 4.90; 95% CI = (1.17, 20.50)). Youths with the inattentive subtype were also marginally more likely to transition from infrequent to frequent use of marijuana (Fig. 4(c)) during this same time period compared with youths without ADHD (OR = 8.24; 95% CI = (0.86, 79.06)).

5. Discussion

We presented both an alternative model and a method for examining the relationship between ADHD and drug use. First, the LTM model of Reboussin *et al.* (1999) was extended to incorporate a latent class predictor. Using a multinomial logistic regression model, the odds to be in the current latent stage of drug use were modelled as a function of the prior stage, the latent class predictor and their interaction. This allowed us to examine the effect of empirically derived ADHD subtypes on stage sequential transitions in use of marijuana over time. The flexibility of this modelling approach also afforded us the ability to include observed covariates like grade in the model so that transition probabilities could be different over the course of high school. In addition to the innovative modelling of the transition probabilities, model development of the underlying latent structures and estimation occurred in a stepwise fashion. In the first step, the latent stage model of marijuana use that was described in Section 2.1 was used to create homogeneous groups of individuals with similar marijuana use profiles. This approach allowed us to examine empirically the progression of use of marijuana while accounting for the measurement error that occurs in self-reported substance use data. Although the two-stage model provided the best overall fit on the basis of global indices of fit, the three-stage model provided a better explanation of the heterogeneity in the profiles by delineating marijuana use further into infrequent and frequent use. On the basis of both the two- and the three-stage models of marijuana use, we then examined the rates of transitioning between stages over time as a function of grade by using the multinomial logistic regression model (1). Not only could we estimate the transition probabilities by using this model, but we could also perform inference on the log-odds-ratios by comparing the odds of transitioning between stages of use of marijuana relative to remaining in the same stage for different high school transitions (e.g. ninth to 10th *versus* 10th to 11th). Interestingly, we found that the movement from ninth to 10th grade and 10th to 11th grade posed the greatest risk of transitioning out of no use of marijuana compared with 11th to 12th grade.

Next, using the cross-sectional LCA model in Section 2.2, we could derive the ADHD subtypes independently from the derivation of the marijuana use stages. Rather than rely on diagnostic criteria for ADHD derived from clinic-based samples, we could explore profiles

of inattentive, hyperactive and impulsive behaviours in our community-based sample of youth that might be associated with use of marijuana regardless of whether diagnostic criteria were met. Using this approach, we found evidence for two subtypes of ADHD that corresponded well to the diagnostic subgroups: both an inattentive and a combined subtype. There was not evidence in our sample for a strictly hyperactive–impulsive subtype. Given the estimates of the latent stage parameters described in Section 2.1 and the ADHD measurement parameters in Section 2.2, we could estimate the influence of the latent ADHD subtypes on the rates of transitioning by using the LTM model (3). Although the two-stage model of use of marijuana found that the two ADHD subtypes had a greater likelihood of transitioning to use of marijuana compared with youth without ADHD, it could not delineate between the type of transition, i.e. no use to infrequent use (initiation) and no use or infrequent use to frequent use (escalation or progression). The richer three-class model provided evidence that different subtypes had influences at different stages of use of marijuana. Using this approach, we found that the combined ADHD subtype had a significantly greater risk of transitioning from no use of marijuana to infrequent (initiation) use during the 10th–11th-grade transition relative to youths without ADHD. Although our combined subtype includes symptoms of inattention in addition to symptoms of hyperactivity and impulsiveness, this finding is generally consistent with Elkins *et al.* (2007) who found that the hyperactive–impulsive subtype was associated with initiation of use of marijuana. Youths with the inattentive subtype were significantly more likely to transition directly from no use of marijuana to frequent use and marginally more likely to transition from infrequent to frequent use during the 9th–10th-grade transition relative to those without ADHD (escalation or progression). These findings were similar to Abrantes *et al.* (2005), who found that the inattentive subtype was associated with an increased risk of dependence on marijuana.

This particular example of high school use of marijuana collected from a sample of students participating in a randomized school-based intervention trial demonstrated the usefulness of the LTM approach for modelling stage sequential transitions in behaviour while accounting for the measurement error that often occurs in self-reported substance use data. The flexibility of the multinomial logistic regression model for the latent transition parameters allowed us to incorporate both observed and latent class predictors easily. By including both grade and latent-class-derived ADHD subtypes, we could fit a combined and more sensitive model of progression of drug use that did not rely on clinical criteria and was consistent with the transition concept of Anthony and Helzer (1995) allowing for different influences at each stage of drug use. We should note, however, that the small sample size and relatively low prevalence of use of marijuana in our example resulted in some insignificant findings despite the magnitude of the odds ratios relating the transition probabilities. The results of simulation studies for the LTM of Reboussin *et al.* (1999) indicate that the estimating equations approach has good finite sample properties even with samples as small as 400. This suggests that the low prevalence of the underlying stages of use of marijuana in our example in combination with the small sample size may have resulted in an underpowered study.

Limitations of the study should be noted. As in most longitudinal studies, some of the original study population was lost to follow-up, i.e. participated in the first-grade assessments but not the middle and high school assessments. GEE-type estimation approaches were developed for the analysis of multivariate categorical data when non-response is classified as missing completely at random. When the missingness depends only on the observed covariates (which is termed covariate-dependent missingness), GEEs provide consistent regression parameter estimates (Lipsitz *et al.*, 2000; Preisser *et al.*, 2002). However, when missingness depends on the observed outcomes as well as covariates (which is termed missingness at random), GEEs may provide biased results. Although those who

were lost to follow-up for the middle and high school assessments scored slightly higher on teacher-rated aggressive–disruptive behaviour, they did not differ on first-grade behavioural measures of anxiety, depression, concentration problems, hyperactivity or impulsiveness. In addition, attrition in our sample of 495 in the 11th and 12th grades was not related to use of marijuana in the ninth and 10th grades or the first-grade behavioural measures.

We also recognize that identifiability is a well-known problem in latent class modelling. For simple cases, the conditions for latent class parameters to be identified are known. In the absence of any restrictions, three classes become identifiable when there are four dichotomous variables (Lazarsfeld and Henry, 1968; McHugh, 1956). Conditions for model identifiability become more complex with covariate effects on underlying and measured variables (see Huang and Bandeen-Roche (2004)). However, an advantage of our three-stage estimating equations approach is that the latent class structures are determined without covariates in the first two stages. The latent transition regression model is then fitted given the latent class model parameters. For the latent stage model of use of marijuana, we have five binary indicators and fit two- and three-stage models. Although we have longitudinal data, we constrain the conditional response probabilities to be the same over time. Hence, within each time point, the latent class model is theoretically identifiable. The cross-sectional latent class predictor model for ADHD with three classes and 13 indicators is also theoretically identifiable. Despite the theoretical identifiability of our models, the model may not be empirically identifiable given our data. A drawback of the semiparametric GEE approach is that we cannot check the empirical identifiability of our fitted model by examining the matrix of second-order partial derivatives of the log-likelihood. We did check the condition of the matrix of the second-order derivatives of the quasi-score function and it was non-singular for the three-class models. However, methods for diagnosing model identifiability for latent class models estimated by using GEEs are needed and should be the focus of future research.

Perhaps the most important strength of our approach is that it enabled us to determine the underlying stages of use of marijuana and ADHD subtypes independently of the multinomial logistic regression model for the transition rates. Separate estimation of the model parameters by using three sets of estimating equations reduced the complexity that was introduced by the longitudinal nature of the data and the extension to the incorporation of a latent class predictor. Compared with a full likelihood approach, estimation of model parameters and standard errors was made less problematic by virtue of simplifying assumptions to the working covariance matrix at each stage of estimation. Although a possible loss of efficiency associated with making these simplifying assumptions (and possibly an explanation for some of the insignificant findings) warrants further exploration, a robust estimate of the LTM parameter variance that accounted for the variation due to the estimation of the latent stage and latent class parameter estimates was easily obtained. This made inference on the latent transition odds ratios possible under the stagewise estimating equation approach whereas estimation of standard errors in likelihood-based approaches is problematic, so standard errors are often not reported.

Finally, latent transition analysis is an attractive approach for modelling the evolution of stage sequential developmental processes. Not only does it allow the data to guide our understanding of how these processes unfold but also it accounts for the measurement error that is often found in self-report data. Although much of the application of latent transition modelling has occurred in the context of substance use, methods for modelling transitions between health states are important more broadly as are methods allowing for multiple indicators of health. It is increasingly difficult to quantify health with a single measure leading to widespread use in public health studies of questionnaires and surveys involving a series of self-report questions. Although each question on its own is an imperfect measure,

together the questions may describe variation in a health profile. The extension of these multiple indicator methods to the predictor variable in the form of a latent class variable is also important for identifying subgroups of individuals with homogeneous risk profiles. These combined models have the potential to inform our understanding of the aetiology of stage sequential developmental processes and to aid in the development of targeted intervention and prevention programmes that are stage and subgroup specific.

Acknowledgments

This work was supported by mentored research scientist development awards K01 DA-016279 and R01 DA-11796 from the National Institute on Drug Abuse and R01 MH-57005 and T32 MH-18834 from the National Institute of Mental Health.

References

- Abrantes AM, Strong DR, Ramsey SE, Lewisohn PM, Brown RA. Substance use disorder characteristics and externalizing problems among inpatient adolescent smokers. *J Psychact Drugs*. 2005; 37:391–399.
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4. Washington DC: American Psychiatric Association; 2000.
- Angold A, Erkanli A, Farmer EM, Fairbank JA, Burns BJ, Keeler G, Costello EJ. Psychiatric disorder, impairment, and service use in rural African American and white youth. *Arch Gen Psychiatr*. 2002; 59:893–901. [PubMed: 12365876]
- Anthony, JC.; Helzer, JE. Epidemiology of drug dependence. In: Tsuang, M.; Tohen, M.; Zahner, G., editors. *Textbook of Psychiatric Epidemiology*. New York: Wiley; 1995.
- Barkley RA, Fischer M, Edelbrock CS, Smallish L. The adolescent outcome of hyperactive children diagnosed by research criteria: an 8-year prospective follow-up study. *J Am Acad Chld Adolesc Psychiatr*. 1990; 29:546–557.
- Biederman J, Monuteaux MC, Mick E, Spencer T, Wilens TE, Silva JM, Snyder LE, Faraone SV. Young adult outcome of attention deficit hyperactivity disorder: a controlled 10-year follow-up study. *Psychol Med*. 2006; 36:167–179. [PubMed: 16420713]
- Burke JD, Loeber R, Lahey BB. Which aspects of ADHD are associated with tobacco use in early adolescence? *J Chld Psychol Psychiatr*. 2001; 42:493–502.
- Clayton, RR. Transitions in drug use: risk and protective factors. In: Glantz, MD.; Pickens, RW., editors. *Vulnerability to Drug Abuse*. Washington DC: American Psychological Association; 1992.
- Elkins IJ, McGue M, Iacono WG. Prospective effects of attention-deficit/hyperactivity disorder, conduct disorder and sex on adolescent substance use and abuse. *Arch Gen Psychiatr*. 2007; 64:1145–1152. [PubMed: 17909126]
- Faraone SV, Biederman J, Weber W, Russell RL. Psychiatric, neuropsychological, and psychosocial features of DSM-IV subtypes of attention-deficit/hyperactivity disorder: results from a clinically referred sample. *J Am Acad Chld Adolesc Psychiatr*. 1998; 37:185–193.
- Ford T, Goodman R, Meltzer H. The British Child and Adolescent Mental Health Survey 1999: the prevalence of DSM disorders. *J Am Acad Chld Adolesc Psychiatr*. 2003; 38:716–722.
- Froehlich TE, Lanphear BP, Epstein JN, Barbaresi WJ, Katusic SK, Kahn RS. Prevalence, recognition, and treatment of attention-deficit/hyperactivity disorder in a national sample of US children. *Arch Ped Adolesc Med*. 2007; 161:857–864.
- Galera C, Bouvard MP, Messiah A, Fombonne E. Hyperactivity-inattention symptoms in childhood and substance use in adolescence: the youth gazel cohort. *Drug Alc Depend*. 2008; 94:30–37.
- Galera C, Fombonne E, Chastang JF, Bouvard M. Childhood hyperactivity-inattention symptoms and smoking in adolescence. *Drug Alc Depend*. 2005; 78:101–108.
- Garrett ES, Zeger SL. Latent class model diagnosis. *Biometrics*. 2000; 56:1055–1067. [PubMed: 11129461]
- Gaub M, Carlson CL. Behavioural characteristics of DSM-IV ADHD subtypes in a school-based population. *J Abnorm Chld Psychol*. 1997; 25:103–111.

- Golub A, Labouvie E, Johnson BD. Response reliability and the study of adolescent substance use progression. *J Drug Iss.* 2000; 30:103–108.
- Gong G, Samaniego FJ. Pseudo-maximum likelihood estimation: theory and applications. *Ann Statist.* 1981; 9:861–869.
- Harrison, L.; Hughes, A., editors. *Research Monograph*. Department of Health and Human Services, Division of Epidemiology and Prevention Research, National Institute on Drug Abuse; Rockville: 1997. The validity of self-reported drug use: improving the accuracy of survey estimates; p. 167
- Huang GH, Bandeen-Roche K. Latent class regression with covariate effects on underlying and measured variables. *Psychometrika.* 2004; 69:5–32.
- Ialongo NS, Werthamer L, Kellam SG, Brown CH, Wang S, Lin Y. Proximal impact of two first-grade preventive interventions on the early risk behaviors for later substance abuse, depression, and antisocial behavior. *Am J Commtty Psychol.* 1999; 27:599–641.
- Lahey BB, Applegate B, McBurnett K, Biederman J, Greenhill L, Hynd GW, Barkley RA, Newcorn J, Jensen P, Richters J. DSM-IV field trials for attention deficit hyperactivity disorder in children and adolescents. *Am J Psychiatr.* 1994; 151:1673–1685. [PubMed: 7943460]
- Lazarsfeld, PF.; Henry, NW. *Latent Structure Analysis*. Boston: Houghton Mifflin; 1968.
- Levin FR, Kleber HD. Attention deficit hyperactivity disorder and substance use: relationships and implications for treatment. *Harv Rev Psychiatr.* 1995; 2:246–258.
- Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika.* 1986; 73:13–22.
- Liang KY, Zeger SL, Qaqish B. Multivariate regression analyses for categorical data (with discussion). *J R Statist Soc B.* 1992; 54:3–40.
- Lipsitz SR, Molenberghs G, Fitzmaurice GM, Ibrahim J. GEE with Gaussian estimation of the correlations when data are incomplete. *Biometrics.* 2000; 56:528–536. [PubMed: 10877313]
- McHugh RB. Efficient estimation and local identifiability in latent class analysis. *Psychometrika.* 1956; 21:331–347.
- Molina BSG, Pelham WE. Childhood predictors of adolescent substance use in a longitudinal study of children with ADHD. *J Abnorm Psychol.* 2003; 112:497–507. [PubMed: 12943028]
- Pan W. Akaike's information criterion in generalized estimating equations. *Biometrics.* 2001; 57:120–125. [PubMed: 11252586]
- Polanczyk G, de Lima MS, Horta BL, Biederman J, Rohde LA. The world-wide prevalence of ADHD: a systematic review and meta-regression analysis. *Am J Psychiatr.* 2007; 164:942–948. [PubMed: 17541055]
- Preisser JS, Lohman KK, Rathouz PJ. Performance of weighted estimating equations for longitudinal binary data with drop-outs missing at random. *Statist Med.* 2002; 21:3035–3054.
- Reboussin BA, Ip EH, Wolfson M. Locally dependent latent class models with covariates: an application to under-age drinking in the USA. *J R Statist Soc A.* 2008; 171:877–897.
- Reboussin BA, Liang KY, Reboussin DM. Estimation equations for a latent transition model with multiple discrete indicators. *Biometrics.* 1999; 55:839–845. [PubMed: 11315015]
- Reboussin BA, Lohman KK, Wolfson M. Modeling adolescent drug use patterns in cluster-unit trials with multiple sources of correlation using robust latent class regressions. *Ann Epidem.* 2006; 16:850–859.
- Stenbacka M, Allebeck P, Romlesjo A. Initiation into drug abuse: the pathway from being offered drugs to trying cannabis and progression to intravenous drug abuse. *Scand J Socl Med.* 1993; 21:31–39.
- Suppes P, Zanotti M. When are probabilistic explanations possible? *Synthese.* 1981; 48:191–199.
- Swanson JM, Sergeant J, Taylor E, Sonuga-Burke E, Cantwell D, Jensen P. Attention deficit hyperactivity disorder and hyperkinetic disorder. *Lancet.* 1998; 351:429–433. [PubMed: 9482319]
- Uebersax, JS. A practical guide to local dependence in latent class models. 2000. (Available from <http://ourworld.compuserve.com/homepages/jsuebersax>.)
- Werthamer-Larsson L, Kellam S, Wheller L. Effects of first-grade classroom environment on shy behaviour, aggressive/disruptive behaviour and concentration problems. *Am J Commtty Psychol.* 1991; 19:585–602.

Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*. 1986; 42:121–130. [PubMed: 3719049]

Appendix A

Consider a first-order Taylor series approximation of $U_1(\theta_1)$ about $\theta_1 = \hat{\theta}_1$ which results in

$$\sqrt{N}(\hat{\theta}_1 - \theta_1) \approx -\frac{1}{\sqrt{N}} \text{diag}(K_1, \dots, K_T)_{|\theta_1 = \hat{\theta}_1}^{-1} U_1(\theta_1)$$

where

$$K_t = \left(\sum_{i=1}^N \tilde{C}_{it} \tilde{R}_{it}^{-1} \tilde{C}_{it}' \right)^{-1}, \quad t=1, \dots, T,$$

and $U_1(\theta_1)$ is defined in Section 3.1. Then

$$\sqrt{N}(\hat{\theta}_1 - \theta_1) \rightarrow N(0, V_{11} = M^{-1} S M^{-1'})$$

where

- a. $M = \text{diag}(K_1, \dots, K_T)$,
- b. $S = \sum_{i=1}^N C_i R_i^{-1} F_i F_i' R_i^{-1} C_i'$ where C_i and R_i are block diagonal matrices with elements C_{it} and R_{it} , $t = 1, \dots, T$, and
- c. F_i is a vector with elements

$$F_{it} = \begin{pmatrix} y_{it} - \mu_{it}(\theta_{1t}) \\ y_{it} - \sigma_{it}(\theta_{1t}) \end{pmatrix}, \quad t=1, \dots, T.$$

A consistent estimator of the asymptotic covariance of $\hat{\theta}_1$ that is robust to misspecification of the covariance of y_i and w_i is obtained by replacing θ_1 by $\hat{\theta}_1$, and the covariance of y_i and w_i by its empirical estimate $F_i F_i'$.

Appendix B

Consider a first-order Taylor series approximation of $U_2(\theta_2)$ about $\theta_2 = \hat{\theta}_2$ which results in

$$\sqrt{N}(\hat{\theta}_2 - \theta_2) \approx -\frac{1}{\sqrt{N}} \left(\sum_{i=1}^N \tilde{E}_i \tilde{R}_i^{-1} \tilde{E}_i' \right)_{|\theta_2 = \hat{\theta}_2}^{-1} U_2(\theta_2)$$

where $U_2(\theta_2)$ is defined in Section 3.2. Then,

$$\sqrt{N}(\widehat{\theta}_2 - \theta_2) \rightarrow N(0, V_{22} = M^{*-1} S^* M^{*-1'})$$

where

- a. $M^* = (\sum_{i=1}^N \widehat{E}_i \widehat{R}_i^{*-1} \widehat{E}_i)^{-1}$,
- b. $S = \sum_{i=1}^N E_i R_i^{*-1} F_i^* F_i^{*'} R_i^{*-1} E_i$ and
- c.

$$F_i^* = \begin{pmatrix} z_i - \mu_i^*(\theta_2) \\ w_i^* - \sigma_i^*(\theta_2) \end{pmatrix}.$$

A consistent estimator of the asymptotic covariance of $\widehat{\theta}_2$ that is robust to misspecification of the covariance of z_i and w_i^* is obtained by replacing θ_2 by $\widehat{\theta}_2$, and the covariance of z_i and w_i^* by its empirical estimate $F_i^* F_i^{*'}$.

Appendix C

The variance of the estimate of θ_3 is constructed by using two first-order Taylor series approximations of $U_3(\theta_3, \widehat{\theta}_1, \widehat{\theta}_2)/\sqrt{N}$ about

- a. $\widehat{\theta}_1 = \theta_1$ and $\widehat{\theta}_2 = \theta_2$, and
- b. $\theta_3 = \widehat{\theta}_3$

given by

$$\frac{U_3(\theta_3, \widehat{\theta}_1, \widehat{\theta}_2)}{\sqrt{N}} + \frac{\partial U_3(\theta_3, \widehat{\theta}_1, \widehat{\theta}_2) / \partial \widehat{\theta}_1 |_{\widehat{\theta}_1 = \theta_1, \widehat{\theta}_2 = \theta_2}}{N} L_1 + \frac{\partial U_3(\theta_3, \widehat{\theta}_1, \widehat{\theta}_2) / \partial \widehat{\theta}_2 |_{\widehat{\theta}_1 = \theta_1, \widehat{\theta}_2 = \theta_2}}{N} L_2, \tag{8}$$

$$\frac{\partial U_3(\theta_3, \widehat{\theta}_1, \widehat{\theta}_2) / \partial \theta_3 |_{\theta_3 = \widehat{\theta}_3}}{N} \sqrt{N}(\widehat{\theta}_3 - \theta_3) \tag{9}$$

where $L_1 = \sqrt{N}(\widehat{\theta}_1 - \theta_1)$ and $L_2 = \sqrt{N}(\widehat{\theta}_2 - \theta_2)$. On the basis of the approximation in model (1) and applying the weak law of large numbers, we have that $U_3(\theta_3, \widehat{\theta}_1, \widehat{\theta}_2)/\sqrt{N}$ is asymptotically normally distributed with mean 0 and variance equal to

$$V = V_{33} + G V_{11} G' + H V_{22} H' + 2G \text{cov}(L_1, L_3) + 2G \text{cov}(L_1, L_2) H + 2\text{cov}(L_3, L_2) H'$$

where

$$G = \sum_{i=1}^N \frac{E[\partial U_{3i}(\theta_3, \theta_1, \theta_2) / \partial \theta_1']}{N},$$

$$H = \sum_{i=1}^N \frac{E[\partial U_{3i}(\theta_3, \theta_1, \theta_2) / \partial \theta_2']}{N},$$

$$V_{33} = \sum_{i=1}^N \frac{E[U_{3i}(\theta_3, \theta_1, \theta_2) U_{3i}(\theta_3, \theta_2, \theta_1)']}{N}$$

and $U_{3i}(\theta_3, \theta_1, \theta_2)$ is the contribution of individual i to the estimating equation for θ_3 . The covariances of L_1, L_2 and L_3 with each other are based on their first-order Taylor series approximations so that

$$\text{cov}(L_1, L_3) = \left\{ \sum_{i=1}^N \frac{E[\partial U_{1i}(\theta_1) / \partial \theta_1]}{N} \right\}^{-1} \sum_{i=1}^N E[U_{1i}(\theta_1) U_{3i}(\theta_3, \theta_1, \theta_2)'],$$

$$\text{cov}(L_3, L_2) = \sum_{i=1}^N E[U_{3i}(\theta_3, \theta_1, \theta_2) U_{2i}(\theta_2)'] \left\{ \sum_{i=1}^N \frac{E[\partial U_{2i}(\theta_2) / \partial \theta_2]}{N} \right\}^{-1}$$

where U_{1i} is the contribution of individual i to the estimating equation for θ_1 and U_{2i} is the contribution of individual i to the estimating equation for θ_2 . Because the estimating equations for θ_1 and θ_2 are orthogonal, $\text{cov}(L_1, L_2) = 0$. Rearranging equation (2), $\sqrt{N}(\hat{\theta}_3 - \theta_3)$ is asymptotically normally distributed with mean 0 and variance equal to

$$V_{33}^* = A^{-1} V_{33} A^{-1'}$$

where $A = -E[\partial U_3(\theta_3, \theta_1, \theta_2) / \partial \theta_3] / N$. If we replace the expected by the observed information and θ_3, θ_1 and θ_2 by their \sqrt{N} -consistent estimates we obtain a consistent estimate of V_{33}^* after adjusting for the estimation of θ_1 and θ_2 .

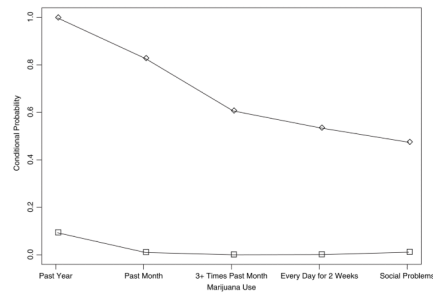


Fig. 1. Estimated stages of use of marijuana among 9th–12th-grade students on the basis of a latent two-stage model of use: □, no use, grade-specific latent stage prevalences (92%, 86%, 82%, 86%); *, use, grade-specific latent stage prevalences (8%, 14%, 18%, 14%)

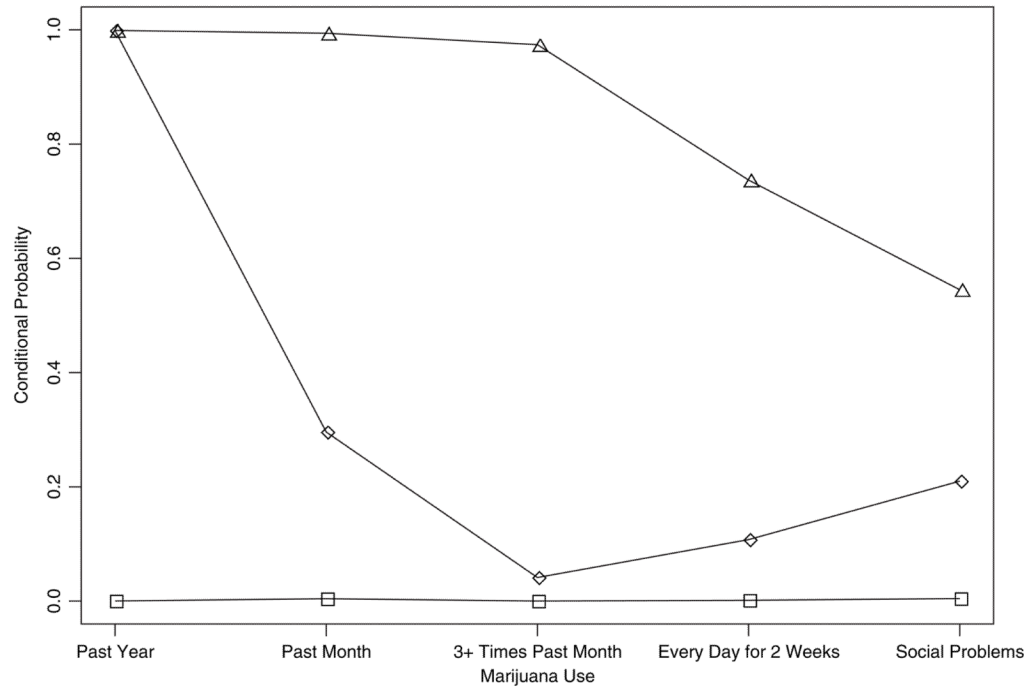


Fig. 2. Estimated stages of use of marijuana among 9th–12th-grade students on the basis of a latent three-stage model of use: □, no use, grade-specific latent stage prevalences (88%, 77%, 75%, 75%); ◇, infrequent use, grade-specific latent stage prevalences (7%, 16%, 15%, 17%); △, frequent use, grade-specific latent stage prevalences (4%, 7%, 10%, 7%)

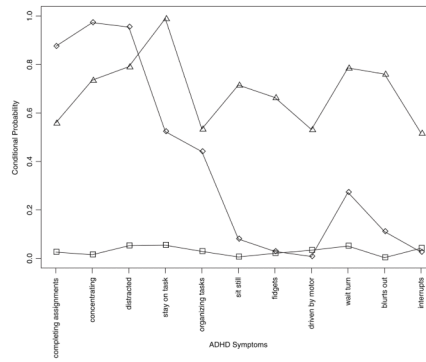


Fig. 3. Estimated subtypes of ADHD among eighth-grade students on the basis of a latent three-class model of ADHD: □, no ADHD (65%); ♦, inattentive subtype (23%); ▲, combined subtype

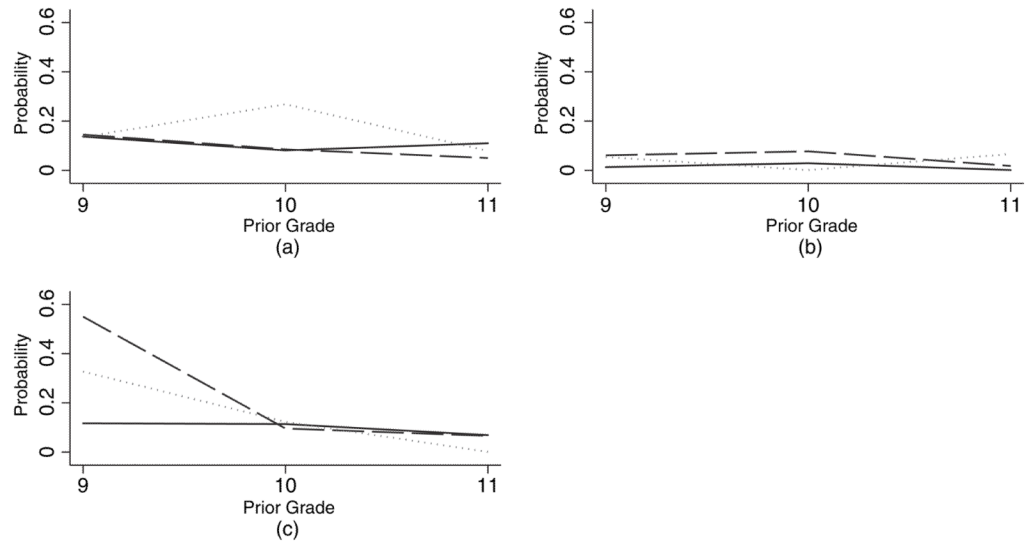


Fig. 4. Estimated probability of transitioning from (a) no use of marijuana to infrequent use, (b) no use of marijuana to frequent use and (c) infrequent use of marijuana to frequent use: —, no ADHD; — —, inattentive subtype; ·····, combined subtype

Table 1

Estimated transition rates between stages of use of marijuana from ninth to 12th grade based on a latent two-stage model of use

Prior stage	Transition rates for the following transitions:					
	9th–10th grade		10th–11th grade		11th–12th grade	
	No use	Use	No use	Use	No use	Use
No use	0.921	0.079	0.932	0.068	0.967	0.033
Use	0.149	0.851	0.339	0.661	0.258	0.742

Table 2

Estimated transition rates between stages of use of marijuana from ninth to 12th grade based on a latent three-stage model of use

Prior stage	Transition rates for the following transitions:								
	9th–10th grade			10th–11th grade			11th–12th grade		
	No use	Infrequent use	Frequent use	No use	Infrequent use	Frequent use	No use	Infrequent use	Frequent use
No use	0.832	0.138	0.029	0.864	0.101	0.035	0.895	0.097	0.008
Infrequent use	0.331	0.414	0.255	0.476	0.412	0.111	0.401	0.544	0.054
Frequent use	0.119	0.215	0.666	0.234	0.095	0.671	0.143	0.248	0.609