Available online at www.sciencedirect.com
ScienceDirect
Drug and Alcohol Dependence 88S (2007) S85–S96
DRUG and ALCOHOL DEPENDENCE
www.elsevier.com/locate/drugalcdep

Adolescent alcohol use and adult alcohol disorders: A two-part random-effects model with diagnostic outcomes

Shelley A. Blozis[a,b,*], Betsy Feldman[a,b], Rand D. Conger[b,c]

[a] Department of Psychology, University of California, Davis, CA 95616, USA
[b] Family Research Group and University of California, Davis, CA 95616, USA
[c] Department of Human and Community Development, University of California, Davis, CA 95616, USA

Received 1 June 2006; accepted 20 December 2006

Abstract

Alcohol use is often analyzed by treating the behavior as a single dimension, such as focusing on frequency of use. Based on data from a longitudinal study, this report considers two distinct aspects of semi-continuous alcohol use data. A two-part random-effects model was used to evaluate change in the log-odds and frequency of use from about age 13 to about age 18 years. Change features were then related to the log-odds of later alcohol disorders. Results suggested differences in the two aspects of use over time and their relationships with later disorders. Most important for the purposes of this study, different methods of analyzing antecedents and consequences of alcohol use trajectories were shown to generate both similar and disparate findings.
© 2007 Elsevier Ireland Ltd. All rights reserved.

Keywords: Adolescents; Adult alcohol use disorders; Trajectories; Growth models; Random effects model; Longitudinal research; Epidemiology; Methodology

1. Introduction

A recent review of the literature on alcohol-related research concluded that drinking behavior by adolescents and adults can have severe negative consequences for individuals and their communities (National Institute on Alcohol Abuse and Alcoholism, 2000). For example, the research demonstrates significant associations between the abuse of alcohol and traffic accidents, major health problems, violent behavior, and disruptions in family life. The estimated annual cost of drinking problems was estimated to be over $180 billion per year. Moreover, developmental trends in drinking behavior appear to follow an age-related sequence, with the frequency and quantity of alcohol use rising during the adolescent and early adult years and diminishing thereafter. This finding poses several major questions for the study of drinking behavior. For example, how can this adolescent and early adult increase in alcohol use be adequately described and explained and what developmental consequences does it have? The present study contributes new information regarding these questions. Specifically, we consider a recently proposed statistical technique expected to deal effectively with the typically observed skewness in alcohol use data and compare it with a more traditional approach.

1.1. Antecedents and consequences of drinking trajectories

Earlier research on trajectories of drinking behavior has tended to focus on the quantity of alcohol use and, in particular, on binge drinking, which is typically defined as four, five, or six drinks in a row on a single occasion (e.g., Hill et al., 2000; Muthén and Muthén, 2000; Tucker et al., 2003). For example, Hill et al. (2000) followed a cohort of children from an average of 10 years of age to an average of 21 years old. For each wave of assessment during this period they asked participants how often during the past 30 days they had consumed five or more drinks in a row on a single occasion. Using contemporary statistical procedures for identifying trajectory classes, they examined both factors that predicted membership in a specific trajectory class as well as several important consequences of class membership. This report by Hill and his colleagues is very similar to other research on developmental trends in binge drinking (e.g., Chassin et al., 2002; Muthén and Muthén, 2000; Oesterle et al.,

* Corresponding author at: Department of Psychology, One Shields Avenue, Davis, CA 95616, USA.
E-mail address: sablozis@ucdavis.edu (S.A. Blozis).

0376-8716/$ – see front matter © 2007 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.drugalcdep.2006.12.008

2004; Tucker et al., 2003). In general, this work has focused on different classes or groups of drinking types.

The research on trajectories in binge drinking provides important insights into the antecedents and consequences of particular classes of binge drinkers. First, all of these studies identify three or more classes of binge drinkers, typically ranging from participants who rarely or never drink to excess to those who persistently and oftentimes increasingly engage in high levels of such behavior. Second, these investigations have found many significant predictors or antecedents of trajectory class membership. Two consistently significant predictors have been participant sex and age of drinking onset. That is, the results show that males compared to females are at greater risk for high and/or increasing trajectories of binge drinking and those participants who demonstrated early onset of binging also are at greater risk than those who start later for entering into the most problematic drinking trajectories. Third, these studies demonstrate agreement in showing that the highest rate trajectories of binge drinking predict alcohol disorders involving abuse and/or dependence at a later point in time.

Although the majority of recent reports on trajectories of alcohol use have focused on binge drinking, some studies have examined drinking frequency (e.g., Flory et al., 2004; Li et al., 2001). For example, in a study of a cohort of youth across the period from 11 to 21 years of age, on each measurement occasion Flory et al. (2004) asked respondents how many drinks they had altogether during the past 30 days. All reported drinks might have been consumed on a single day or on many separate days; thus, quantity of drinking on a single day could not be determined. The important lesson from these studies is that trajectories of drinking frequency are predicted by similar variables as trajectories of binge drinking and that trajectories of frequent drinking predict to alcohol abuse and dependence in the same fashion as trajectories of binge drinking. Indeed, across both types of studies it appears that any drinking trajectory that reflects higher levels of alcohol consumption than the infrequent or non-using trajectory is predicted by similar variables and has similar consequences. Although this observation is likely true for adolescents, we should note that it might not apply equally as well to adult drinking. For example, regular moderate drinking (e.g., a glass of wine with dinner each evening) by adults may actually be health promoting rather than harmful. Regular adolescent drinking, on the other hand, is less likely to occur in this type of normative context.

For adolescents, research suggests that it may be an above average level or increase in the probability of alcohol use that is being predicted or that effectively predicts later alcohol disorders. For this reason, approaches to modeling drinking trajectories that do not rely on the identification of unique trajectory classes may provide meaningful alternatives to this common statistical technique. We consider the positive and problematic aspects of several of the possible methods for analyzing drinking trajectories in the following section. While there are a variety of analytic procedures for assessing drinking behaviors, we limit our discussion to those that concern data that are described as having a high frequency of zeros with the remaining scores being continuously distributed.

Table 1
Alternative growth models for longitudinal, semi-continuous behavioral measures

Method	Common data treatments
Latent class growth model	Treat as continuous or categorical (e.g., use vs. no-use)
Latent curve models	Treat as continuous
Finite mixture growth models	Treat as continuous or categorical (e.g., use vs. no-use)
Latent transition models	Treat as categorical (use vs. no-use)
Two-part random-effects models	Simultaneous treatment of use/no-use and remaining continuous distribution

1.2. Analytic methods for semi-continuous longitudinal drinking data

1.2.1. Latent class growth models. Substance-use data are often described as semi-continuous, that is, non-normal with a high frequency of zeros representing non-use and the remaining values being continuously distributed and often positively skewed. Semi-continuous data present several challenges for the analysis of drinking trajectories and several methods have been proposed to address them (see Table 1). A popular approach for dealing with longitudinal, semi-continuous data is a latent class growth model that assumes individual differences in growth are due to two or more latent class trajectories (Nagin, 1999; Nagin and Tremblay, 2001). As discussed earlier, this is a very common approach to evaluating the antecedents and consequences of drinking trajectories. With this method, individual variation apart from that captured by the class curves is assumed to be due to error. Class membership is unobservable, but the approach yields estimated probabilities of membership for each individual. Associations between hypothesized antecedents and consequences of the different latent classes may then be studied by including in a model such variables as predictors or criterion measures. A benefit of the approach is that it may well describe data that are not normally distributed prior to allowing for the latent classes. Considering substance-use data, the assumption that observations are normal deviations about class curves may not be tenable, however, particularly when data are semi-continuous or change is systematic at the individual level within classes.

1.2.2. Latent growth curve models. An alternative approach is a latent growth curve model that simultaneously characterizes the population curve as well as the curves of individuals (Laird and Ware, 1982; Singer and Willett, 2003). When change follows a linear trend, for example, a model that includes an intercept and linear time effect may well describe change in the response. Random effects (e.g., random intercept and slope) allow each individual to have a trajectory that may be unique from others as well as the population. That is, responses at the individual level are not simply random deviations about a mean curve but rather are assumed to vary systematically as functions of time, with the addition of the random error about the individual's curve. The parameters are combinations of what is common to all individu-

als plus that which is unique to the individual. Given individual differences in change characteristics, it may also be possible to study the moderating effects of person-level (e.g., sex) or contextual (e.g., school size) variables on change features, or the effects of change features on outcome variables. For example, girls and boys may differ on average in their rates of change in alcohol use, or change in alcohol use may be predictive of a later alcohol disorder diagnosis. Several applications of the latent growth curve model to substance-use data have appeared in the literature (Andrews and Duncan, 1998; Chassin et al., 1996; Duncan et al., 1997; Duncan and Duncan, 1996). Under the model, responses are assumed to be normal, however, making the application of the model to semi-continuous substance-use data potentially problematic in some situations.

1.2.3. Finite mixture growth models. A random-effects model that assumes a normal mixture of distributions for the random effects at the second level (Verbeke and Lesaffre, 1996, Section 4.1) represents a combination of features from the latent class growth model and the latent growth curve model. Individual trajectories are assumed to be due in part to latent classes. Within classes, responses at the individual level are assumed to be systematic, such as a function of time and not simply due to random error about a mean curve. A finite growth mixture model may be useful when the distribution of an outcome variable is not normal when considered as a function of time (or other known covariates) but is normal when considered as both a function of time and the latent classes (and possibly known covariates). This model was later extended to allow the latent classes to influence a separate set of outcome variables (Muthén and Shedden, 1999). There have been a few applications of a latent growth mixture model to adolescent substance-use data (e.g., see Colder et al., 2001; Muthén and Muthén, 2000). Similar to the assumption underlying a latent curve model, responses are assumed to be normal, which may also present problems when applying a finite growth mixture model to semi-continuous substance-use data.

1.2.4. Latent transition models. The methods described above concern change in a continuous variable. An alternative approach is to consider behavioral changes as transitions between discrete states, such as the transition from no-use to the onset of substance use. Different methods have been proposed for dealing with data characterized by discrete states. One approach is a latent transition model in which there are presumably multiple states, such as a state of non-use, and where individuals may transition in and out of the different states over time (Langeheine, 1994). This approach allows, for example, one to consider group differences in latent states at different time points, making it possible to study group differences in substance-use onset (Lanza and Collins, 2002). Thus, while potentially useful in characterizing movement through a variety of states, a latent transition model was not developed for the study of the continuous aspect of behavioral data, such as drinking frequency.

This discussion indicates that the distribution of alcohol or other substance use in the population creates analytic problems for many standard techniques of evaluating change over time. For this reason, a few approaches that address the non-normality of substance-use data have been suggested. One approach is to add one to each response and then use a log transformation (e.g., Colder et al., 2001), although this approach may not lead to approximately normal distributions, particularly when there are large numbers of zeros. Another is to assume a censored-normal distribution in which the censoring is assumed to address the high frequency of zeros. This latter approach has been criticized, however, due to the inappropriate treatment of zeros (see Olsen and Schafer, 2001). That is, a censored distribution does not treat the zeros as valid data points, but rather, assumes the values are proxies for negative or missing values.

In many studies of substance use it may be important to include in a statistical model variables that may be related to substance-use behaviors. These may be variables that distinguish different populations under investigation or those that relate to the individuals directly. Common to the methods described above that were intended for continuous data is the assumption that if other variables are introduced into a model, such as to test for the moderating effects of person-level or contextual variables on substance-use behaviors, a covariate is related in the same manner to both the decision to engage in the behavior as well as the amount or frequency of use (Olsen and Schafer, 2001). When semi-continuous substance-use behaviors are analyzed directly, the distinction between the likelihood of engaging in the behavior and the quantity or frequency of consumption, as measured on a continuous scale, when engaged cannot be made. Similarly, when status of use (e.g., onset) alone is studied, this distinction again cannot be made.

1.3. The present study

In the following analyses we evaluate the effectiveness of a two-part random-effects model in describing longitudinal, semi-continuous alcohol use data as compared to a more traditional latent growth curve model. Based on multi-year data from a cohort of adolescents, we consider a two-part random-effects model to address questions regarding the description, antecedents, and consequences of drinking trajectories. The procedure handles both the distributional qualities of semi-continuous substance-use data and the possibility that covariates may differentially predict the probability that one engages in the behavior and the degree of engagement. A related approach is that proposed by Berk and Lachenbruch (2002) that combines a random-effects logistic model and a random-effects model for log-normal repeated measures that also assumes left censoring. Unlike a two-part random-effects model, the procedure does not make the explicit distinction between the two processes that may underlie the response variable.

In the present report, we investigate whether the two-part random-effects model both parsimoniously describes trajectories of drinking frequency and provides sensitive indicators of change that are predictable and consequential. Drawing on the research discussed earlier, we expected that being male and engaging in drinking at a young age would predict higher levels and increases in drinking behavior. We also expected that the growth parameters in the model would predict alcohol disorders

involving abuse and dependence. As a means of comparison, we also fit a latent growth curve model to the alcohol data. Unlike a two-part random-effects model that distinguishes between the two aspects of use, a latent growth curve model treats the response as a continuous measure. Thus, the comparison between these two approaches may be helpful in emphasizing potential benefits of a two-part random-effects model.

1.3.1. Two-part random-effects models. A two-part random-effects model is based on two subcomponents, each addressing a unique aspect of an observed behavior (Olsen and Schafer, 2001; Tooze et al., 2002). Part 1 is a model for an indicator variable that takes on the value of 1 if the individual engaged in the behavior, and 0 if not. The indicator variable is missing otherwise. Part 2 is a model for a continuous variable that is equal to the original observed response if the individual engaged in the behavior, and is missing otherwise. Thus, this part of the model concerns the response conditional on whether the individual engaged in the behavior or not. Considering the two aspects of the behavior longitudinally, change in each is modeled. In the present analysis, the first sub-model of change concerned the log-odds that an individual used alcohol and the second concerned change in the frequency of alcohol consumption when the individual reported some level of use. In a two-part random-effects model, the two model parts that characterize change in each variable are considered simultaneously. Specifically, characteristics of change in each may be allowed to covary, yielding correlations between change features corresponding to one aspect of a behavior with change features relating to the second aspect. In the context of the current study, for example, it was possible to consider the association between the log-odds that one used alcohol at a young age and change in the conditional frequency of use over the study period.

1.3.2. Assumptions concerning no-use. Underlying valid inference of a two-part random-effects model is an assumption that the single discrete value (e.g., zero) is random. For example, when zero represents the absence of a behavior, an observation of zero is assumed to be due to measurement error or an on-and-off engagement in the behavior. This assumption of random zeros is particularly relevant with regard to substance-use data because zero may represent abstinence from a behavior rather than a random observation of no-use as it would for someone who occasionally engages in the behavior. Non-random zeros are referred to as 'structural zeros' (Carlin et al., 2001). In practice, a possible situation is one in which the population contains a combination of individuals characterized by both zero types. One approach to dealing with structural zeros is to exclude cases based on personal, persistent reports of no-use. A potential problem with this approach, however, is that some individuals may be incorrectly identified as true abstainers. One strategy for dealing with this situation in the context of longitudinal data is to consider a finite mixture growth model in which a mixture distribution is assumed for the first part of the model corresponding to use versus no-use to allow for the two populations: one that represents a class of users and the other that represents a class of abstainers. A general description of this type of model is provided in Muthén and Shedden (1999) where a random-effects model is extended to include a mixture distribution for the random-effects at the second level. We applied this procedure to the two-part random-effects model to test the tenability of the assumption of random zeros.

2. Method

2.1. Participants

Data come from the Family Transitions Project, a longitudinal, community epidemiological study of 451 target youth from two-parent families. Data collection began when they were in the seventh (1989) grade and continued on an annual basis through the 1990s with a 90% retention rate through 1999 (see Table 2). The target youth, their parents, and a close-aged sibling participated in the study. Because of a very small minority population in rural Iowa, all participants are of European heritage. Participants in the study were originally recruited to examine the family and developmental effects of the economic downturn in agriculture of the 1980s; for that reason, they were recruited from eight rural counties in Iowa. The original sample of families was primarily lower middle- or middle class. Additional details about the study can be found in Conger and Conger (2002), Conger et al. (1994), and Simons et al. (1996).

2.2. Variables

2.2.1. Adolescent alcohol use data. Alcohol use was first measured when participants were in the seventh grade and averaged 13.2 years old, with approximately annual follow-ups through tenth grade (average age 16.1 years) with an additional assessment during their senior year of high school (average age 18.1 years). Individuals were asked how often they had consumed beer, wine, or hard liquor during the previous year. Their responses were coded 0 for "never," 1 for "less than weekly," 2 for "one to two times per week," and 3 for "three or more times per week." Responses for two categories: (1) beer or wine and (2) hard liquor were summed to form one alcohol consumption frequency score, denoted here as y_{ti}, for the previous year, where t denotes the measurement occasion and i an individual. The resulting alcohol use variable had a possible range of 0–6. Of the total participants, 343 (76.1%) reported use at one or more times over the targeted period. To fit a two-part random-effects model, two variables were

Table 2
Description of study sample (data from 215 male and 236 female participants in the Family Transitions Project, 1989–1999)

Variable	Year						
	1989	1990	1991	1992	1994	1997	1999
Sex (count)							
Male	215	200	189	188	196	191	181
Female	236	224	218	216	228	227	224
Total	451	424	407	404	424	421	405
Age (mean and S.D.)	13(0.42)	14 (0.42)	15 (0.43)	16 (0.43)	18 (0.42)	21 (0.42)	24 (0.43)

Note: For Age, the standard deviation (S.D.) is given in parentheses.

created that related to the observed measures of past-year alcohol use. The first was a dichotomous variable, denoted as u_{ti} that, for time t, was equal to 1 if the individual used alcohol, equal to 0 if the individual did not use alcohol, and missing otherwise. The second variable, denoted as m_{ti}, was, for time t, equal to the original alcohol use variable, y_{ti}, if the individual used alcohol, and was missing otherwise.

2.2.2. Alcohol disorder diagnosis. In addition to reported alcohol use, lifetime and past-year alcohol disorders were assessed using the University of Michigan modification of the Composite International Diagnostic Interview (UM-CIDI, DSM-III-R; Kessler et al., 1994). This fully structured diagnostic interview generates estimates of DSM-III-R (American Psychiatric Association, 1987) psychiatric disorders in terms of both onsets and recurrences for adolescents and adults. World Health Organization field trials show that the CIDI possesses good reliability and validity (Wittchen, 1994). Past-year diagnoses of (a) dependence and (b) abuse without dependence were made twice, first when the target adolescents were on average 21.1 years old, and again approximately 2.5 years later. Two indicator variables were created to represent the alcohol disorder diagnostic categories: dependence ($n = 64$ had positive diagnoses at one or both occasions) and abuse without dependence ($n = 68$ had positive diagnoses at one or both occasions). The majority of participants ($n = 289$) had no diagnosis and $n = 30$ were missing this information. It is worth noting that some individuals may have met criteria for either classification prior to assessment. Table 3 provides descriptive statistics for the alcohol use and disorder variables.

2.2.3. Background variables. Adolescent alcohol use measures were considered functions of age (measured to the nearest month) over a 6-year period. Sex (1 = male; 0 = female) was included to study possible sex differences in alcohol use patterns over time, in addition to sex differences in disorder diagnoses.

2.2.4. Missing data. Common to many longitudinal studies, some alcohol use data were missing. Inference from a random-effects model for longitudinal data is considered valid when missing data are missing at random (Singer and Willett, 2003). For the current study, there were several patterns of missing data. Most of the missing data were due to individuals missing one or more times of assessment or simply having incomplete data for the survey concerning alcohol use for a given occasion, but these individuals remained in the study through 1997 or 1999. Thirty individuals (22 males) were missing data used to assess alcohol disorders, with a small number of these individuals also missing some alcohol use data from previous assessments. Using a pattern-mixture model, we studied the effects of missing data when individuals were missing the data necessary for assessing alcohol disorders by creating an indicator variable that was equal to 1 if an individual had this pattern of missing data and was equal to 0 otherwise (e.g., Hedeker and Gibbons, 1997). This variable was then used to adjust characteristics of change in alcohol use during adolescence by including it as a predictor of the random coefficients in both the two-part random-effects model and the latent growth curve model.

2.3. Data analysis

2.3.1. Two-part random-effects model. A two-part random-effects model was fitted to the repeated measures representing status of use (use versus no-use) and conditional frequency of use over time. The first part of the model was based on a random-effects logistic model for the log-odds of use as a function of an individual's age, Age_i. Assuming linear change in the log-odds, the logit model for a response at time t was

$$\eta_{ti} = \alpha_{0i} + \alpha_{1i}\text{Age}_{ti},$$

where $\eta_{ti} = \log(\pi_{ti}/(1-\pi_{ti}))$ was the log-odds of use for individual i, and π_{ti} was the probability that the individual used (i.e., $P(u_{ti} = 1)$). The coefficients α_{0i} and α_{1i} were each a combination of fixed and random effects (e.g., $\alpha_{0i} = \alpha_0 + a_{0i}$) so that the model could vary across individuals in terms of its parameters. Age_i was centered to 13 years by subtracting 13 from the observed ages at each occasion. Thus, the coefficient α_{0i} denoted the individual log-odds of use at age 13, and α_{1i} denoted the individual-level past-year unit-change in log-odds from about age 13 to about age 18; each coefficient was assumed to randomly deviate about its corresponding fixed effect. The variances of the two random effects (i.e., $\text{var}(a_{0i})$ and $\text{var}(a_{1i})$) characterized the degree of individual differences in the log-odds of use at age 13 and the linear rate of change in the log-odds of use. The two random coefficients were allowed to covary.

The second part of the model was a latent growth curve model for repeated measures of alcohol use frequency when an individual was engaged in the behavior, with frequency centered to 0. The conditional frequency of use at time t, m_{ti}, was considered a linear function of Age_{ti}, plus a random error term, ε_{ti}:

$$m_{ti} = \beta_{0i} + \beta_{1i}\text{Age}_{ti} + \varepsilon_{ti},$$

where for individual i, β_{0i} represented the expected conditional frequency of use at age 13 years. The coefficient β_{1i} represented the expected past-year linear change in frequency. The error, ε_{ti}, represented a random deviation of an observation from an individual's fitted trajectory, and so, denoted the within-subject error.

The individual-level regression coefficients, β_{0i} and β_{1i}, each represented the sum of a fixed coefficient (i.e., β_0 and β_1, respectively) common to all individuals, plus a random component (i.e., b_{0i} and b_{1i}, respectively) unique to

Table 3
Description of study variables (data from 215 male and 236 female participants in the Family Transitions Project, 1989–1996)

Variable	Year					
	1989	1990	1991	1992	1994	1997/1999
Alcohol use (counts)						
Male	64	64	77	101	125	
Female	62	81	96	124	147	
Total	126	145	173	225	272	
Alcohol use frequencies, if using (means, S.D.)						
Male	0.59 (0.20)	0.68 (0.30)	0.84 (0.35)	0.87 (0.41)	1.1 (0.59)	
Female	0.64 (0.39)	0.70 (0.37)	0.88 (0.51)	0.87 (0.44)	0.98 (0.57)	
Overall	0.62 (0.30)	0.69 (0.34)	0.86 (0.44)	0.87 (0.43)	1.1 (0.58)	
Previous year alcohol abuse diagnosis in 1997 or 1999 (counts)						
Male						33
Female						35
Total						68
Previous year alcohol dependence diagnosis in 1997 or 1999 (counts)						
Male						49
Female						15
Total						64

the individual. The coefficients β_{0i} and β_{1i} were assumed to randomly deviate about their respective population values. The variances of the random intercept and slope characterized the degree of individual differences in the conditional frequency of use at age 13 and the rate of change in the frequency over the targeted period, respectively. The two random effects were also allowed to covary with one another.

In a combined model, the two model parts, that for the log-odds of use and that for the conditional frequency of use, were joined by allowing covariances between the random effects at the second level of each model part. Letting the sets of random effects for the two model parts be represented by $\mathbf{a}_i = (a_{0i}, a_{1i})'$ and $\mathbf{b}_i = (b_{0i}, b_{1i})'$, the random-effects at the second level were assumed to have a joint normal distribution:

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Phi = \begin{pmatrix} \Phi_a & \\ \Phi_{ba} & \Phi_b \end{pmatrix}\right),$$

where the expected values of the random-effects were zero, and the matrix Φ was a symmetric block covariance matrix. Block elements of the matrix correspond to the covariance matrix for the random-effects included in the logit model (Φ_a), the continuous response model (Φ_b), and the covariances between the random-effects of the two sub-models, (Φ_{ba}). More specifically, the covariance matrix Φ_a contained the variances of the individual log-odds of use at age 13 (i.e., var(a_{0i})) and the rate of change in log-odds of use over time (var(a_{1i})), in addition to the covariance between the two change characteristics (cov(a_{1i}, a_{0i})). The covariance matrix Φ_b contained the variances of the individual conditional frequency levels of use at age 13 (i.e., var(b_{0i})) and the rate of change in frequency of use over time (var(b_{1i})), in addition to the covariance between these two change characteristics (cov(b_{1i}, b_{0i})). The lower block of the matrix, Φ_{ba}, was not a symmetric covariance matrix but rather contained the covariances between the random effects of the two model parts. This was a particularly interesting component of the model because it provided a means for testing the associations between characteristics of change (e.g., intercept and slope) in both the log-odds of use and the conditional frequency of use.

2.3.1.1. Second-level regressions. At the second level of the first part of the model, the random intercept, α_{0i}, and linear age effect, α_{1i}, were regressed on child's sex and the missing data pattern code, denoted by Miss_i, resulting in two level-2 regressions:

$$\alpha_{0i} = \gamma_{00} + \gamma_{01}\text{Sex}_i + \gamma_{02}\text{Miss}_i + r_{0i}$$

and

$$\alpha_{1i} = \gamma_{10} + \gamma_{11}\text{Sex}_i + \gamma_{12}\text{Miss}_i + r_{1i},$$

where γ_{00} and γ_{10} represented the fixed intercept and slope, respectively, for females who were not missing the alcohol disorder data, γ_{01} and γ_{11} denoted differences in the two effects between males and females holding constant the effects of the missing data pattern, and γ_{02} and γ_{12} denoted differences in the two effects between those missing the alcohol disorder data and those who did not holding constant the effects of child's sex. Specifically, γ_{00} was the log-odds of use at age 13 for females with complete alcohol disorder data, γ_{01} was the sex difference in the log-odds at age 13 between males and females holding constant the effect of the missing data pattern. The coefficient γ_{10} was the annual change in log-odds for females with complete alcohol disorder data, and γ_{11} was the sex difference in the annual change in log-odds between males and females holding constant the effect of the missing data pattern. The errors for these regressions are denoted by r_{0i} and r_{1i}, respectively.

Similarly at the second level of the second part of the model, the intercept, β_{0i}, and linear age effect, β_{1i}, were regressed on child's sex and the missing data pattern, Miss_i, resulting in two level-2 regressions:

$$\beta_{0i} = \eta_{00} + \eta_{01}\text{Sex}_i + \eta_{02}\text{Miss}_i + s_{0i}$$

and

$$\beta_{1i} = \eta_{10} + \eta_{11}\text{Sex}_i + \eta_{12}\text{Miss}_i + s_{1i},$$

where η_{00} and η_{10} represented the intercept and slope, respectively, for females with complete alcohol disorder data, η_{01} and η_{11} denote the differences in the two effects between males and females holding constant the effects of the missing data pattern, and η_{02} and η_{12} denote differences in the two effects between those with incomplete versus complete alcohol disorder data holding constant the effects of child's sex. More specifically, η_{00} is the conditional frequency of use at age 13 for females, and η_{01} is the difference in the conditional frequency of use at age 13 between males and females. The coefficients η_{10} and η_{11} denote the annual change in use for females and the sex difference in the annual change between males and females, respectively. The errors for these regressions are denoted by s_{0i} and s_{1i}, respectively.

With the addition of child's sex and the pattern for missing alcohol disorder data, the random effects at the second level represented individual differences in the intercepts and slopes of each model part adjusted for sex differences and effects due to the missing data. Consequently, this changed the joint distributions of the random effects as follows:

$$\begin{pmatrix} r_i \\ s_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Phi = \begin{pmatrix} \Phi_r & \\ \Phi_{sr} & \Phi_s \end{pmatrix}\right).$$

That is, the matrices in Φ_r and Φ_s represented the variance–covariance matrices for the random effects of the first and second parts of the model, respectively, after adjusting for the effects of child's sex and missing data. The matrix Φ_{sr} was a non-symmetric covariance matrix for the random effects between the two model parts, also after taking into account sex differences and effects due to subject attrition.

2.3.1.2. Using change characteristics of alcohol use to predict later alcohol disorders. With the two aspects of alcohol use each summarized by a model imposing linear trends on the mean and individual trajectories across the targeted period, the individual-level latent measures of log-odds of use and frequency of use at age 13, as well as the rates of change in each, could be considered predictors of later alcohol disorder diagnostic categories, with child's sex as a covariate. Given that both disorder diagnostic measures were dichotomously scored, an appropriate model for their regression was a logistic model in which the log-odds of each diagnosis were taken as a criterion measure. The logit models for the diagnostic measures were

$$\xi_{Di} = D_0 + D_1\text{Sex}_i + D_2\text{Int}_{ui} + D_3\text{Slope}_{ui} + D_4\text{Int}_{mi} + D_5\text{Slope}_{mi}$$

and

$$\xi_{Ai} = A_0 + A_1\text{Sex}_i + A_2\text{Int}_{ui} + A_3\text{Slope}_{ui} + A_4\text{Int}_{mi} + A_5\text{Slope}_{mi},$$

where $\xi_{Di} = \log(\pi_{Di}/(1-\pi_{Di}))$ and $\xi_{Ai} = \log(\pi_{Ai}/(1-\pi_{Ai}))$ were each the log-odds of a dependence and abuse diagnosis, respectively. The log-odds of the two diagnoses were considered functions of the following: child's sex, log-odds of use at age 13 (Int_{ui}), change in the log-odds of use during adolescence (Slope_{ui}), frequency of use at age 13 (Int_{mi}), and change in frequency of use during adolescence (Slope_{mi}). The coefficients D_0 and A_0 represented the mean log-odds of the particular diagnoses for females when the log-odds of use at age 13, change in the log-odds of use during adolescence, frequency of use at age 13, and change in the conditional frequency of use during adolescence were equal to zero. The remaining coefficients (e.g., D_4) represented the effects of the different predictors on the log-odds of a particular diagnosis.

The model with all components considered simultaneously is shown in Fig. 1. As shown in Fig. 1, the logistic random-effects model that specifies the associations between the observed dichotomous responses denoting use and non-use and the latent change features for the log-odds of use is represented in the upper left portion of the figure. The random intercept α_{0i}, labeled in the figure as the "log-odds of use at age 13", and the random linear time effect α_{1i}, labeled in the figure as the "linear change rate in log-odds of use", are shown as predictors of the dichotomous responses, $u_1, \ldots, u_6$, shown in boxes. The paths relating the random effects to the log-odds of use represent fixed and known values. Specifically, the paths from the intercept (i.e., log-odds of use at age 13) to the individual log-odds were fixed to 1 because the intercept represented a constant level of the outcome. The paths from the linear time effect to the log-odds were set equal to the age of the child at each occasion, with adjustments made due to centering Age_i. Thus, the values assigned to these paths could vary between children. The two random coefficients, the random intercept and slope, were allowed to covary, as shown by the double-headed arched arrow between the effects. Child's sex (but not the missing data pattern) is shown as a predictor of the random coefficients. Essentially the same

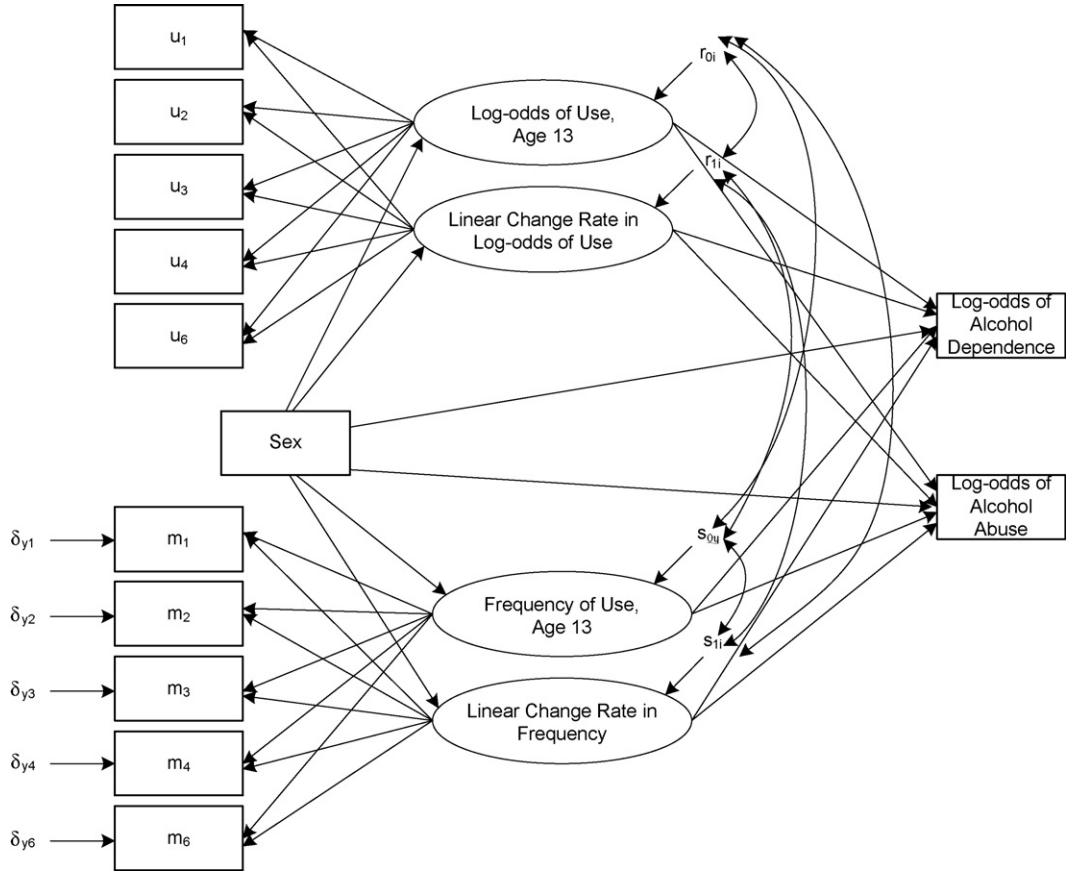

Fig. 1. Structural model relating sex and latent change characteristics of alcohol use based on a two-part random-effects model to diagnostic outcomes.

relationships were specified between the observed conditional drinking frequencies and the corresponding change characteristics, and so they are not described here in detail. Additionally, covariances between the random coefficients of the two model parts are shown by the double-headed arched arrows between these effects.

2.3.2. Latent growth curve model. A latent growth curve model was also fitted to the alcohol use data. For this model, the original alcohol response variable was considered for analysis. Alcohol use frequency was considered a linear function of Age_i, plus a random error term, ε_i:

$$y_{ti} = \beta_{0i} + \beta_{1i}\text{Age}_{ti} + \varepsilon_{ti},$$

where for individual i, β_{0i} represented the expected frequency of use at age 13 years. The coefficient β_{1i} represents the expected past-year linear change in drinking frequency for the individual. The coefficients β_{0i} and β_{1i} were assumed to be sums of fixed effects (i.e., β_0 and β_1, respectively) and random effects (b_{0i} and b_{1i}, respectively) and to deviate at random about the fixed effects, β_{0i} and β_{1i}, respectively. The variances of the random intercept and slope characterized the degree of individual differences in the frequency of use at age 13 and the rate of change in frequency of use over the targeted period, respectively. The two random effects were also allowed to covary with one another. Finally, the error, ε_{ti}, represented a random deviation of an observation from an individual's fitted trajectory at time t. Unlike the second part of the two-part random-effects model, inference from this model was not conditional on an individual's status at a given occasion as one who used or did not use.

At the second level of the model, the individual-level coefficients were considered functions of child's sex and the pattern of missing data relating to the alcohol disorder diagnoses, similar to the two-part random-effects model:

$$\beta_{0i} = \eta_{00} + \eta_{01}\text{Sex}_i + \eta_{02}\text{Miss}_i + s_{0i}$$

and

$$\beta_{1i} = \eta_{10} + \eta_{11}\text{Sex}_i + \eta_{12}\text{Miss}_i + s_{1i},$$

where η_{00} and η_{10} represent the fixed intercept and slope, respectively, for females with complete alcohol disorder data, η_{01} and η_{11} denote sex differences in the two effects holding constant the effects of missing data, and η_{02} and η_{12} denote differences in the two effects for those with complete versus incomplete alcohol disorder data holding constant the effects of child's sex. Given the coding for child's sex and the missing data patterns, η_{00} was the expected frequency of use at age 13 for females with complete alcohol disorder data, and η_{01} was the expected difference in the frequency of use at age 13 between males and females. The coefficients η_{10} and η_{11} denoted the annual change in use for females and the sex difference in the change rate between males and females, respectively, holding constant the effects of the missing data pattern. The errors for these regressions are denoted by s_{0i} and s_{1i}, respectively. Similar to the two-part random-effects model, these errors represent the random change characteristics adjusted for child's sex and the missing data pattern. Thus, their variances and covariance are also adjusted for these effects.

2.3.2.1. Using change characteristics of adolescent alcohol use to predict later alcohol disorders. Similar to the two-part random-effects model, characteristics of change in drinking frequency could be treated as predictors of dichotomous measures of alcohol abuse and dependence diagnoses assessed when individuals were adults, adjusting for sex differences. As was done for the two-part random-effects model, a logistic regression model was used to predict the log-odds of each disorder from child's sex and the latent characteristics of change in alcohol use during adolescence.

The full model is depicted in Fig. 2. As shown in Fig. 2, observed drinking frequencies (shown in boxes), are dependent on the random effects (shown in circles) that relate to the two change characteristics and are labeled in the figure

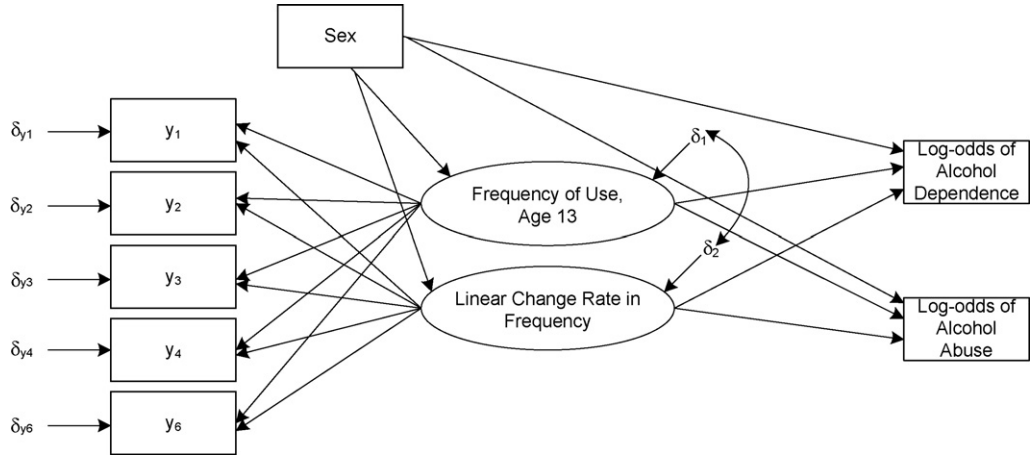

Fig. 2. Structural model relating sex and latent change characteristics of alcohol use based on a latent growth curve model to diagnostic outcomes.

as frequency of use at age 13 and linear change rate in frequency. The paths from the random effects to the observed frequencies represent fixed and known values. Specifically, the paths from the intercept (i.e., frequency at age 13) to the observed frequencies were fixed to 1 because the intercept represented a constant level of the outcome. The paths from the linear time effect to the observed frequencies were set equal to the age of the child at each occasion, with adjustments made due to centering Age_i. Thus, these path coefficients could vary between individuals.

2.3.3. Estimation. The two-part random-effects model that included the regression of the diagnostic outcome categories on alcohol-use change features, with child's sex and the missing data patterns as covariates (as specified earlier), represented a combination of continuous and categorical response variables. Simultaneous estimation of the two model parts was carried out using Mplus version 4 (Muthén and Muthén, 1998–2006) with maximum likelihood estimation and robust standard errors. Estimation of the latent growth curve model that included dichotomous outcomes was also estimated using Mplus.

3. Results

3.1. Two-part random-effects model

In addition to testing a model that assumed linear growth in the two aspects of alcohol use, we also fitted quadratic growth curve models to both outcome variables and found no improvement in model fit. We, therefore, describe results for the linear growth models. Homogeneity of variance was assumed for the variances of the errors at the first level of the model for the continuous responses. Initial results suggested that the variance of the random intercept relating to the continuous part of the model (i.e., that relating to the conditional drinking frequency) was close to or equal to zero. This was reasonable given that there was little variation in the conditional drinking frequencies at the first measurement occasion. Thus, the variance of the random intercept was set equal to zero and the model re-estimated. The effects of the missing data pattern on the difference characteristics of change in both the log-odds and the conditional drinking frequencies were not statistically significant and so were dropped from the model.

For females at age 13, the estimated mean log-odds of use was −1.99 (S.E. = 0.284), corresponding to a predicted probability of 0.12.[1] On average, change in the log-odds of use for females was estimated to be 0.774 (S.E. = 0.104), suggesting an overall increase. Males did not differ, however, from females in the expected log-odds at age 13 (estimated difference = −0.105, S.E. = 0.361) or the corresponding change rate (estimated difference = −0.068, S.E. = 0.124). For females, the estimated conditional frequency of use was 0.181 (S.E. = 0.043), suggesting an overall low mean level of consumption when females were using. The estimated annual increase in the conditional frequency of use over time for females was 0.110 (S.E. = 0.029), suggesting a slow but constant rate of increase from about age 13 to about age 18. Males did not differ from females in the expected change in the conditional frequencies (estimated difference = 0.041, S.E. = 0.024). Thus, based on these results, males and females have comparable mean trajectories for both the log-odds of use as well as the conditional frequencies of use.

With the exception of the variance relating to the random intercept of the growth model for the conditional frequency response, all the variances associated with the random effects, after adjusting for differences due to child's sex, were statistically different from zero, as based on conducting deviance tests, suggesting individual differences in three of the four aspects of change in alcohol use. For these random effects whose variances were statistically different from zero, the estimated covariances were studied. The estimated covariance between the intercept and slope of the model relating to the log-odds of use suggested that higher log-odds of use at age 13 tended to be related to slower rates of increase in the log-odds through adolescence (estimated covariance = −0.697; corresponding correlation, $r = -0.13$). Between the two aspects of alcohol use, there was a positive association between the log-odds of use

[1] For model identification purposes in the logit model, the intercept is fixed to zero and the threshold for the binary variable is estimated. The intercept of the logit model is equal to the negative of the threshold parameter (Muthén and Muthén, 1998–2006, p. 436). The estimated threshold for this model was 2.03, which corresponds to an intercept of −2.03. The predicted probability of use at age 13 was $\exp(-2.03)/(1+\exp(-2.03))$, or more simply, $1/(1+\exp(2.03)) = 0.12$.

at age 13 and the rate of change in the conditional frequency of use (estimated covariance = 0.224; corresponding correlation, $r = 0.45$), indicating faster rates of increase in the frequency of use for those who were most likely to use at age 13. The estimated 95% confidence interval for the covariance between the change in the log-odds of use at age 13 and the change in the conditional frequency of use included zero as an interior point, suggesting no linear association between these aspects of alcohol use.

3.1.1. Evaluating the assumption of random zeros. To evaluate the assumption of random zeros, we first refitted the model described above after removing those individuals who reported no-use during the targeted period. A comparison of the fixed-effects estimates for the complete sample and the subset that excluded those who reported no-use during adolescence suggested a few but minor differences. First, the estimated difference in the log-odds of use for females at age 13 based on the complete sample and the subset was negligible, with an estimate of −1.80 (S.E. = 0.235) based on the complete sample and an estimate of −1.26 (S.E. = 0.271) based on the subset of cases when non-users were excluded. This log-odds of use corresponded to a probability of $1/(1 + \exp(1.26)) = 0.22$. The estimated change in the log-odds of use for females was also negligible, from 0.707 (S.E. = 0.095) for the complete sample to 0.809 (S.E. = 0.109) for the sub-sample. Negligible differences resulted between frequency of use and change in frequency of use between the two samples, as might be expected given that frequency estimates based on the complete sample were conditional, based on observed use at each occasion. The lack of striking differences in model estimates when the two samples were compared suggests that the assumption of random zeros may be tenable.

Olsen and Schafer (2001) also suggested fitting a mixture distribution for the random effects at the second level to evaluate the tenability of the assumption of random zeros. Using a method described in Muthén and Shedden (1999), a mixture distribution for the random effects at the second level was considered that allowed for mean differences in the log-odds of use by assuming two latent classes. Specifically, a two-class mixture distribution for the random effects relating to the log-odds of use was fitted to the complete data set. The revised model did not improve the fit of the model. We therefore proceeded in our analyses assuming that the observed zeros were random across individuals and based the remaining analyses on the complete sample.

3.1.2. Predicting diagnostic outcomes from latent change characteristics of alcohol use during adolescence. As shown in Table 4, in predicting later alcohol disorders the results suggested a sex difference in the log-odds of a dependence diagnosis (estimated difference = 1.79, S.E. = 0.348) but not in the log-odds of an abuse diagnosis (estimated difference = 0.225, S.E. = 0.326), suggesting males were more likely to have a later dependence diagnosis but not an abuse diagnosis. A higher log-odds of drinking at Age 13 was related to a higher log-odds of a later dependence diagnosis (estimated effect = 0.359, S.E. = 0.115) but not to the log-odds of a later abuse diagnosis (estimated effect = 0.019, S.E. = 0.097). Change in the log-odds

Table 4
Parameter estimates for predicting diagnostic outcomes from sex and latent change characteristics of alcohol use, based on a two-part random-effects model

Criterion	Intercept	Predictor			
		Sex	α_{0i}	α_{1i}	β_{1i}
α_{0i}	−1.80 (0.226) [−2.24, −1.35]	−0.133 (0.313) [−0.746, 0.480]			
α_{1i}	0.707 (0.086) [0.538, 0.876]	−0.054 (0.112) [−0.274, 0.166]			
β_{0i}	0.198 (0.046) [0.108, 0.288]				
β_{1i}	0.158 (0.020) [0.119, 0.197]	0.039 (0.024) [−0.008, 0.086]			
Dependence	−3.30 (0.463) [−4.21, −2.39]]	1.79 (0.348) [1.11, 2.47]	0.359 (0.115) [0.134, 0.584]	0.339 (0.401) [−0.447, 1.12]	0.478 (1.29) [−2.05, 3.01]
Abuse without dependence	−3.23 (0.773) [−4.75, −1.71]	0.225 (0.326) [−0.414, .864]	0.019 (0.097) [−0.171, 0.209]	1.45 (0.639) [0.202, 2.71]	1.03 (1.34) [−1.59, 3.64]

Notes: Estimates are maximum likelihood. Standard errors are in parentheses. Estimated 95% confidence intervals for parameters appear in brackets. The coefficients α_{0i}, α_{1i}, β_{0i}, and β_{1i} are the individual log-odds of use at age 13, annual change in log-odds, conditional frequency of use at age 13, and annual change in frequency, respectively. The estimated variance of β_{0i} did not differ statistically from zero so was not considered a predictor of the diagnostic outcomes.

Table 5
Parameter estimates for predicting diagnostic outcomes from sex and latent change characteristics of alcohol use, based on a latent growth curve model

Criterion	Intercept	Predictor		
		Sex	β_{0i}	β_{1i}
Frequency at age 13, β_{0i}	0.347 (0.043) [0.263, 0.431]	−0.112 (0.062) [−0.234, 0.010]		
Change in frequency, β_{1i}	0.192 (0.017) [0.159, 0.225]	0.041 (0.026) [−0.010, 0.092]		
Dependence	−3.77 (0.427) [−4.61, −2.93]	1.75 (0.388) [0.990, 2.51]	1.18 (0.863) [−0.511, 2.87]	2.40 (1.22) [0.009, 4.79]
Abuse without dependence	−2.43 (0.273) [−2.97, −1.90]	−0.008 (0.332) [−0.659, 0.643]	−0.081 (0.818) [−1.68, 1.52]	3.18 (1.36) [0.514, 5.85]

Notes: Estimates are maximum likelihood. Standard errors are in parentheses. Estimated 95% confidence intervals for parameters appear in brackets. Latent change characteristics for alcohol use are adjusted for the effect of subject attrition.

of drinking over time was not predictive of a dependence disorder (estimated effect = 0.339, S.E. = 0.401) but was of an abuse disorder (estimated effect = 1.45, S.E. = 0.639). Change in the conditional drinking frequency was not predictive of either a dependence diagnosis (estimated effect = 0.478, S.E. = 1.29) or an abuse diagnosis (estimated effect = 1.03, S.E. = 1.34).

3.2. Latent curve model

Similar to the two-part random-effects model, a quadratic latent growth model was considered and compared to a linear latent growth model, with no improvement in model fit for the more complex model. We, therefore, describe results for linear growth models. Homogeneity of variance was assumed for the variances of the errors at the first level of the model. Results are presented in Table 5. At 13 years, females had an estimated average drinking frequency of 0.347 (S.E. = 0.043), suggesting on average a low level of use. The estimated change rate for females was 0.192 (S.E. = 0.017), suggesting a slow increase through adolescence. The results suggested no sex differences in the mean drinking frequency at age 13 (estimated effect = −0.112, S.E. = 0.062) or in the change rate (estimated effect = 0.041, S.E. = 0.026), suggesting that on average, males and females had similar behavioral trajectories.

The variances of the random intercept (estimated variance = 0.143, S.E. = 0.047) and slope (estimated variance = 0.019, S.E. = 0.007), adjusted for sex differences, were statistically different from zero, as judged by deviance tests. These results suggested that individuals varied with regard to their frequency levels at age 13 and their rates of change over time. Additionally, a statistically significant and positive covariance between the two conditional random effects (estimated covariance = 0.035; corresponding correlation, $r = 0.67$) suggested that higher levels of use were related to faster rates of increase over time.

Males and females differed in the log-odds of a dependence diagnosis (estimated difference = 1.75, S.E. = 0.388) but not in the log-odds of an abuse diagnosis (estimated difference = −0.008, S.E. = 0.332), with results suggesting that males were more likely than females to have a dependence diagnosis but no difference when considering an abuse diagnosis. Additionally, change in the drinking frequency was predictive of the log-odds of a later dependence diagnosis (estimated effect = 2.40, S.E. = 1.22) as well as the log-odds of a later abuse diagnosis (estimated effect = 3.18, S.E. = 1.36), suggesting a tendency for increases in drinking frequencies to be related to both a later dependence diagnosis and a later abuse diagnosis. However, the 95% confidence interval relating to the effect of change in drinking frequency on the dependence diagnosis had a lower bound value that was quite close to zero.

3.3. Comparison of findings from the two analytic approaches

Results from the two-part random-effects model and the latent growth curve model suggest differences in how drinking behaviors during adolescence develop, as well as how features of change in these behaviors may be related to later alcohol disorder diagnoses. To begin, both approaches were consistent in suggesting no sex difference in drinking frequencies, whether conditional or not conditional of use, at age 13. Both approaches suggested an increase in drinking frequencies through adolescence, with the average change rates being comparable for males and females. The latent growth curve model suggested a tendency for increases in drinking frequencies at age 13 to increase the risk for both a dependence and an abuse diagnosis in adulthood, whereas under the two-part random-effects model, change in the log-odds and the log-odds of early use were predictive of a dependence diagnosis.

4. Discussion

In the present report we evaluated the utility of a two-part random-effects model for estimating developmental trajectories based on alcohol use data. Unlike most approaches to the analysis of longitudinal data that assume a behavior may be represented by a single dimension, a two-part random-effects model assumes a behavior has two components: the first concerning the likelihood that an individual will engage in a behavior, and the second concerning the magnitude or frequency of the behavior when the individual is engaged. This distinction between the two aspects of a behavior may be particularly important when the two differ in their patterns of change or when covariates have different relationships with the two aspects.

Substantively, the results based on a two-part random-effects model for the current sample were generally consistent with earlier research concerning alcohol use. For example, a higher likelihood of early drinking (at ages 13) increased risk for alcohol dependence by about age 23. This finding is consistent with earlier research on trajectories of both drinking frequency (e.g., Flory et al., 2004; Li et al., 2001) and binge drinking (e.g.,

Chassin et al., 2002; Hill et al., 2000). However, because binge drinking represents one aspect of clinical diagnoses for alcohol disorders, we would expect that the findings in terms of antecedents and consequences would likely be more robust if the tested analytic strategy were evaluated using trajectories of binge drinking. Unfortunately that measure is not available across the adolescent years in the present study. There were no differences between males and females in either the log-odds of use or the frequency of use when drinking at age 13, nor were there differences in the rates of change in either aspects of use during adolescence.

As a means of comparison, we also estimated a latent growth curve model in which alcohol use was represented by the original response variable (frequency of use) and corresponding change characteristics of use served as predictors of clinical diagnoses for alcohol disorders, with sex as a covariate. Thus, this analysis considered drinking frequency that was not conditional on whether or not an individual was using. The results suggested no sex differences in the trajectories of frequency of use, similar to the two-part random-effects in which drinking frequencies were considered conditional on whether an individual was using. Unlike the two-part random-effects model, it was the change in the frequency of use that was positively related to both a later dependence and abuse diagnosis.

Despite some of the appealing features of a two-part random-effects model as a method for dealing specifically with semi-continuous data, there may be limitations to the method. The first concerns the assumption of random zeros. This means that any observation equal to zero is assumed to represent a random process, such as an occasion in which a user reports no-use. In studies of adolescent drinking behaviors, it may seem more reasonable to assume that some individuals are true abstainers and thus the model would provide a misfit to the data to the extent that this was the case. Despite this, methods have been proposed to deal with such situations, including a method for incorporating a mixture distribution for the part of the model that relates to whether or not an individuals uses, as described earlier. A second point is that the model does not address the issue of substance-use onset. For example, age at onset may vary across individuals. Although we considered the likelihood of use at an early age of 13 years, some individuals did not report their first use of alcohol until they were older. In adolescent substance use research, onset may be an important feature in characterizing substance use development. A useful addition to the two-part random-effects model for studies where onset is an important feature of development may be an extension of the model to incorporate onset so that that aspect may also be studied and related to other model components. Finally, in the current application of the method, factors that may have been related to either the likelihood of using or the frequency of use were not incorporated into the model. For example, individual trajectories relating to the likelihood of use may be dependent on time-varying environmental factors, such as changes in peer groups or other stressors.

A variety of analytic methods are available for the analysis of substance-use data. Decisions about which ones may be considered for a given study should be dependent on the goals of the study, as well as data characteristics. As was shown here, different treatments of the data lead to both similar and different substantive interpretations, as may have been expected given that the measure of alcohol was treated quite differently by the two approaches considered. In fact, such results suggest that researchers need not be restricted to a single approach, given that the strategies taken are in line with the research questions.

Acknowledgements

The National Institute on Drug Abuse (DA016977) provided primary support for this report. Research on the Iowa study cohort is currently supported by grants from the National Institute of Child Health and Human Development, the National Institute on Drug Abuse, and the National Institute of Mental Health (HD047573, HD051746, and MH051361). Support for earlier years of the study also came from multiple sources, including the National Institute of Mental Health (MH00567, MH19734, MH43270, MH59355, MH62989, and MH48165), the National Institute on Drug Abuse (DA05347), the National Institute of Child Health and Human Development (HD027724), the Bureau of Maternal and Child Health (MCJ-109572), and the MacArthur Foundation Research Network on Successful Adolescent Development Among Youth in High-Risk Settings.

References

American Psychiatric Association, 1987. Diagnostic and Statistical Manual of Mental Disorders: DSM-III-R, 3rd rev. ed. American Psychiatric Association, Washington, DC.

Andrews, J.A., Duncan, S.C., 1998. The effect of attitude on the development of adolescent cigarette use. J. Subst. Abuse 10, 1–7.

Berk, K.N., Lachenbruch, P.A., 2002. Repeated measures with zeros. Stat. Meth. Med. Res. 11, 303–316.

Carlin, J.B., Wolfe, R., Brown, C.H., Gelman, A., 2001. A case study on the choice, interpretation and checking of multilevel models for longitudinal binary outcomes. Biostatistics 2, 397–416.

Chassin, L., Curran, P.J., Hussong, A.M., Colder, C.R., 1996. The relation of parent alcoholism to adolescent substance use: a longitudinal follow-up study. J. Abnormal Psychol. 105, 70–80.

Chassin, L., Pitts, S.C., Prost, J., 2002. Binge drinking trajectories from adolescence to emerging adulthood in a high-risk sample: predictors and substance abuse outcomes. J. Consult. Clin. Psychol. 70, 67–78.

Colder, C.R., Mehta, P., Balanda, K., Campbell, R.T., Mayhew, K.P., Stanton, W.R., Pentz, M.A., Flay, B.R., 2001. Identifying trajectories of adolescent smoking: an application of latent growth mixture modeling. Health Psychol. 20, 127–135.

Conger, R.D., Conger, K.J., 2002. Resilience in midwestern families: selected findings from the first decade of a prospective, longitudinal study. J. Marriage Family 64, 361–373.

Conger, R.D., Elder, G.H.J., Lorenz, F.O., Simons, R.L., Whitbeck, L.B., 1994. Families in Troubled Times: Adapting to Change in Rural America. Aldine de Gruyter, Hawthorne, NY, US.

Duncan, S.C., Alpert, A., Duncan, T.E., Hops, H., 1997. Adolescent alcohol use development and young adult outcomes. Drug Alcohol Depend. 49, 39–48.

Duncan, S.C., Duncan, T.E., 1996. A multivariate latent growth curve analysis of adolescent substance use. Struct. Equation Model. 3, 323–347.

Flory, K., Lynam, D., Milich, R., Leukefeld, C., Clayton, R., 2004. Early adolescent through young adult alcohol and marijuana use trajectories: early predictors, young adult outcomes, and predictive utility. Dev. Psychopathol. 16, 193–213.

Hedeker, D., Gibbons, R.D., 1997. Application of random-effects pattern-mixture models for missing data in longitudinal studies. Psychol. Meth. 2, 64–78.

Hill, K.G., White, H.R., Chung, I., Hawkins, J.D., Catalano, R.F., 2000. Early adult outcomes of adolescent binge drinking: person- and variable-centered analyses of binge drinking trajectories. Alcohol.: Clin. Exp. Res. 24, 892–901.

Kessler, R.C., McGonagle, K.A., Zhao, S., Nelson, C.B., Hughes, M., Eshleman, S., et al., 1994. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity Study. Arch. Gen. Psychiatry 51, 8–19.

Laird, N.M., Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics 38, 963–974.

Langeheine, R., 1994. Latent variable Markov models. In: von Eye, A., Clogg, C.C. (Eds.), Latent Variables Analysis: Applications for Developmental Research. Sage, Thousand Oaks, CA, pp. 373–395.

Lanza, S.T., Collins, L.M., 2002. Pubertal timing and the stages of substance use in females during early adolescence. Prev. Sci. 3, 69–82.

Li, F., Duncan, T.E., Hops, H., 2001. Examining developmental trajectories in adolescent alcohol use using piecewise growth mixture modeling analysis. J. Stud. Alcohol 62, 199–210.

Muthén, B.O., Muthén, L.K., 2000. Integrating person-centered and variable-centered analysis: growth mixture modeling with latent trajectory classes. Alcohol.: Clin. Exp. Res. 24, 882–891.

Muthén, L.K., Muthén, B.O., 1998–2006. Mplus User's Guide, 4th ed. Muthén & Muthén, Los Angeles, CA.

Muthén, B., Shedden, K., 1999. Finite mixture modeling with mixture outcomes using the EM algorithm. Biometrics 55, 463–469.

Nagin, D., 1999. Analyzing developmental trajectories: a semi-parametric group-based approach. Psychol. Meth. 4, 139–157.

Nagin, D., Tremblay, R., 2001. Analyzing developmental trajectories of distinct but related behaviors: a group-based method. Psychol. Meth. 6, 18–34.

National Institute on Alcohol Abuse and Alcoholism, 2000. Tenth Special Report to the U.S. Congress on Alcohol and Health. NIH Publication No. 00-1583.

Oesterle, S., Hill, K.G., Hawkins, J.D., Guo, J., Catalano, R.F., Abbott, R.D., 2004. Adolescent heavy episodic drinking trajectories and health in young adulthood. J. Stud. Alcohol 65, 204–212.

Olsen, M.K., Schafer, J.L., 2001. A two-part random effects model for semicontinuous longitudinal data. J. Am. Stat. Assoc. 96, 730–745.

Simons, R.L., et al. (Eds.), 1996. Understanding Differences Between Divorced and Two Parent Families: Stress, Interaction and Child Outcome. Sage, Thousand Oaks, CA.

Singer, J.D., Willett, J.B., 2003. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford University Press, New York.

Tooze, J.A., Grunwald, G.K., Jones, R.H., 2002. Analysis of repeated measures data with clumping at zero. Stat. Meth. Med. Res. 11, 341–355.

Tucker, J.S., Orlando, M., Ellickson, P.L., 2003. Patterns and correlates of binge drinking trajectories from early adolescence to young adulthood. Health Psychol. 22, 79–87.

Verbeke, G., Lesaffre, E., 1996. A linear random-effects model with heterogeneity in the random effects population. J. Am. Stat. Assoc. 91, 217–221.

Wittchen, H., 1994. Reliability and validity studies of the WHO-composite international diagnostic interview (CIDI): a critical review. J. Psychiatr. Res. 28, 57–84.