Addiction
SSA SOCIETY FOR THE STUDY OF ADDICTION
METHODS AND TECHNIQUES
doi:10.1111/add.12816

Delineating prototypical patterns of substance use initiations over time

Danielle O. Dean, Veronica Cole & Daniel J. Bauer

Department of Psychology, University of North Carolina at Chapel Hill, CB#3270, Davie Hall, Chapel Hill, NC 27599, USA

ABSTRACT

Aims The purpose of this paper is to discover patterns of drug use initiations over time through a multiple event process survival mixture model (MEPSUM model), a novel approach for substance use and prevention research. **Design** The MEPSUM model combines survival analysis and mixture modeling—specifically latent class analysis—to examine individual differences in the timing of initiation and cumulative risk of substance use over time, and is applied to cross-sectional survey data on drug initiations. **Setting** Data are drawn from the 2009 National Survey on Drug Use and Health. **Participants** The survey includes responses from 55 772 individuals (52.05% female). **Measurements** The age of first use of nine different types of substances are examined, including alcohol, tobacco, cocaine and non-medical use of prescription drugs. **Findings** It is argued that six patterns parsimoniously describe the population's risk of initiating different substances over time, described colloquially as general abstainers; early, late and progressive soft drug users; and early and late hard drug users. Both gender and ethnicity significantly predict the patterns, with Caucasians and males having a higher risk for the hard drug-using patterns. The MEPSUM model produced stable results in this application, as the patterns are validated in a split-sample design. **Conclusions** The MEPSUM model provides a statistical framework from which to evaluate patterns of risk for drug initiations over time and predict substance use trajectories relevant to public health interventions. The patterns that result from the model can be used as outcomes for subsequent investigations of etiological and mediating mechanisms.

Keywords Intervention, latent class analysis, substance use initiation, substance use patterns, survival analysis.

Correspondence to: Danielle O. Dean, Department of Psychology, University of North Carolina at Chapel Hill, CB#3270, Davie Hall, Chapel Hill, NC 27599, United States. E-mail: danielledean@unc.edu

Submitted 9 October 2013; initial review completed 12 February 2014; final version accepted 13 November 2014

INTRODUCTION

While considerable research and investment has been made to further our understanding of the process of drug use initiation and persistence, substance abuse and illicit drug use remain widespread. Unfortunately, the use of such substances is associated with other risky health behaviors, as well as adverse impacts on a number of life domains such as employment and educational attainment [1–3]. It is important to understand patterns of substance use initiation and the processes leading to initiation in order to develop targeted efforts to prevent substance use and dependence.

Previous research has found that substance use typically peaks in the developmental period of young or 'emerging' adulthood [4,5]. One author argues that the developmentally distinctive features of this time-period, such as identity explorations and feelings of new freedom, can explain the high substance use initiation and continued use in this age group [6]. The timing of substance use initiation is particularly important, as abuse at a young age is associated with repeated use and a higher probability of dependence [7]. Across the population, substance abuse is also highest during emerging adulthood, making this a particularly important age group to study [8].

Additionally, it is important to consider the relative timing of substance use initiation, as the initiation of one drug is often found to be related to the subsequent use of another type of drug [9,10]. For example, although the number of users of a substance tends to decrease as the severity of the substance increases, a large body of research has documented polydrug use among substance users [11–13]. Some research has found that upwards of three-quarters of individuals in treatment settings were multiple drug users [14].

In evaluating the processes leading to substance initiation and related disorders, many researchers have either focused on a single drug type or have investigated several drugs individually, rather than examining possible interdependencies in the timing of onset across different types of drugs. For example, Kandel & Logan described overall patterns by examining the hazard rate—the conditional probability of initiating, given that one has not yet initiated—of each individual drug type [5]. Similarly, Schwartz *et al*. investigated how tobacco, alcohol, illicit drug and sex initiation were influenced by positive youth development by conducting four separate survival analyses [15]. While analyzing each event separately can be useful, it provides no insight into how the events are related to each other, nor does it reveal individual differences in the patterns of risk over time. Probably contributing to this lack of research surrounding multiple initiations, most multivariate survival analysis models focus on repeatable events or competing risks rather than multiple events [16–19]; for a notable exception in mental health research, see Almansa *et al*. [20].

In an attempt to address these issues, the purpose of this paper is to introduce the multiple event process survival mixture (MEPSUM) model through an application which delineates patterns of drug use initiation over time [21]. The MEPSUM model allows us to examine individual differences in the timing of initiation and cumulative risk of substance use over time across nine different types of substances through the estimation of prototypical, holistic risk patterns. This approach offers two advantages over traditional modeling approaches. First, it provides a fuller and more informative characterization of patterns of onset and potential differences in their prevalence by gender and ethnicity, for example, than approaches which permit the examination of only one or a few drug types or which focus on use at only one point in time. Secondly, one can use this framework to make model-based predictions about patterns of substance abuse for specific individuals—this may be particularly useful for predicting the use of one drug, given the earlier use of another. In addition to presenting a novel statistical framework to drug use data, we intend our results to provide a foundation for generating further hypotheses regarding the mechanisms leading to different patterns of drug use.

METHOD

Participants

The data come from the 2009 National Survey on Drug Use and Health (NSDUH), a cross-sectional study that randomly sampled individuals aged 12 years or older, and is available publicly through the Substance Abuse and Mental Health Data Archive [22]. The 2009 NSDUH recorded data from 55 772 individuals, including 26 744 males (47.95%) and 29 028 females (52.05%). Ethnicity is measured as seven categories, but will be recoded for parsimony into four categories: (1) white (61.93%), (2) African American (12.80%), (3) Hispanic (16.24%) and (4) other (9.03%), as the ethnicity categories Native American, Native Pacific Islands, Asian or more than one ethnicity each comprise fewer than 5% of the sample.

Measures

The age of first use for nine types of substances—alcohol, cocaine, hallucinogens, heroin, inhalants, marijuana, non-medical use of prescription drugs (NMUP), stimulants and tobacco—was the outcome examined. A binary variable was created for each of the nine event processes across ages 10–30 years, indicating whether the event had not yet occurred by that age (coded as 0), occurred at that age (coded as 1) and missing otherwise. The MEPSUM model discussed in more detail below utilizes a survival analysis framework, accounting for individuals who do not experience the event within the time-frame of the study with an assumption that their data are missing at random[1] [23–25]. The sample's observed risk of event occurrence at a given age and cumulative probability of event occurrence are displayed in Fig. 1.

Analysis

MEPSUM models allowing one to seven latent classes—described below as patterns—were fitted in Mplus version 6.12 using the robust maximum likelihood estimator (MLR), accounting for sampling weights [26] with a logit link function to account for the fact the outcomes are binary and corresponding to a hazard model under the assumption of discrete time [27].[2] Example code is provided in the Appendix. Based on examination of the results obtained using unstructured hazard functions—which allows the risk of initiating to take any shape over time—as

[1]This is a mild assumption in this case, as individuals have missing event times due simply to their age at the time of the study.

[2]The intercepts of the hazard functions were constrained to be greater than −15, per the Mplus default, where a logit of −15 implies a functional probability of zero. Because maximum likelihood estimation for latent class and mixture models can be prone to local, suboptimal solutions, all models were run with at least 500 sets of random starting values and the results monitored to ensure the replication of the final log-likelihood [28,29].

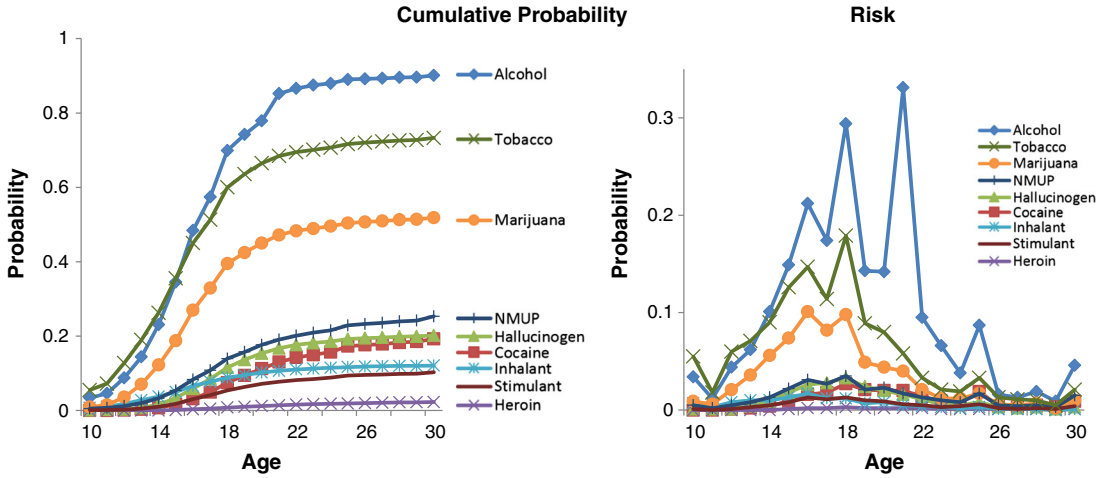


Figure 1 Overall sample estimates, sample estimated, cumulative probability of initiation and unique risk at each time-period

Table 1 Multiple event process survival mixture model (MEPSUM) model with quadratic hazard functions fit information.

Latent classes (no. of patterns)	*'-2LL'*	*Number of free parameters*[a]	*BIC*	*AIC*	*Smallest class size*	*Entropy*
1	−247291.44	27	494859.20	494636.90	NA	NA
2	−226663.50	55	453890.00	453437.10	0.29	0.80
3	−221771.00	71	444268.70	443684.00	0.14	0.73
4	−219857.60	99	440728.50	439913.20	0.09	0.67
5	−218232.90	125	437745.30	436715.80	0.07	0.66
6	**−217353.40**	**151**	**436252.40**	**435008.80**	**0.06**	**0.66**
7	−216563.21	179	434958.65	433484.43	0.03	0.65

Values shown in bold type represent the six-class solution selected. −2LL = −2*log-likelihood; BIC = Bayesian information criterion; AIC = Akaike information criteria; NA = not applicable. Entropy is a measure of classification certainty. [a]To increase the stability of the model, the intercept and quadratic components for a hazard function were constrained to 0 for each event within a latent class where the risk of the event was less than or equal to 0.001 across time-periods. The maximum number of parameters for a given model can be found through this equation, where K represents the number of latent classes: 27*K + (K − 1). For example, the maximum number of parameters before constraints for a three-class model is 83.

well as previous research on the relative risk of substance initiation over time, we chose to implement a more parsimonious parametric form, namely one that is quadratic with age. The use of parametric hazard functions in this investigation represents a novel extension to the originally proposed MEPSUM model [21].

We next examined ethnic and sex differences in the relative prevalence of these patterns of substance use initiation. More specifically, the probability of belonging to each pattern was allowed to differ by sex and ethnicity in a multinomial regression predicting pattern membership (male and Caucasian were selected as reference categories).[3] To evaluate the stability of the results and ensure that the model was not picking up chance patterns in the data, the sample was split randomly in half into an evaluation sample and validation sample (n =27 886 for each) with all initial models fitted to the evaluation sample. The final model was then fitted to the validation sample.

RESULTS

Models with different numbers of patterns and predictors were compared using the Akaike information criterion (AIC) and Bayesian information criterion (BIC), and both continued to decrease when allowing more patterns ([30,31]; Table 1). However, we found that the relative decrease in these criteria became smaller after five or six patterns (Fig. 2). Examining the substantive implications of the model more carefully, there was no single clear optimal number of patterns. This finding is not surprising, as

[3]To avoid potential biases that can arise with a two-step process of first classifying individuals and then predicting class membership, the parameters of the class-specific hazard functions were re-estimated within the MEPSUM model simultaneously with the parameters of the multinomial regression of class membership [32,33]. The size of the classes and the coefficient estimates were monitored for change when these predictors were added to the model to investigate the stability of the results [34,35].

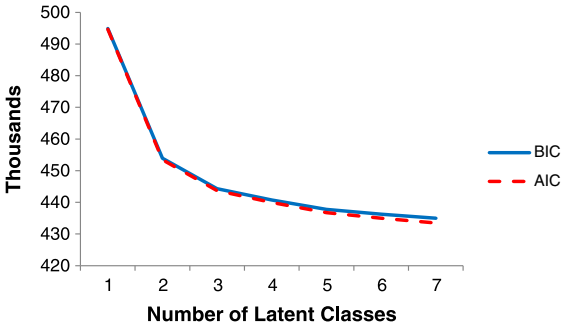


Figure 2 Information criteria: Bayesian information criterion (BIC) and Akaike information criterion (AIC) as a function of the number of patterns[4]

the model is believed to provide a discrete approximation to what is probably a continuous distribution of risk across the population. More patterns result in better approximations with diminishing returns. After careful consideration of the substantive yield provided by competing models, we selected the model allowing for six patterns. This model provided a rich description of prototypical patterns while remaining relatively parsimonious.

In the model without covariates, the six patterns can be described in terms of their risk for initiating different substances over time and the number of individuals within each pattern. The raw parameter estimates in the logit scale for each substance within each pattern are given in Table 2. These estimates imply the risk of initiating different substances shown in Fig. 3. In turn, the cumulative probability of initiating different substances implied by the model are shown in Fig. 4. The patterns may also be described in terms of the median event time, the estimated time at which 50% of individuals have experienced the event (Table 3). The six patterns may be described colloquially as general abstainers (35.7% of the population), later soft drug users (26.7%), early soft drug users (12.1%), progressive soft drug users (8.2%), later hard drug users (10.9%) and early hard drug users (6.3%).

The pattern which may be described as general abstainers had a small probability of initiating any substance use for the first time at any specific age (<0.15 for all events at each age). The cumulative probability of marijuana use as well as NMUP by age 30 is very small at 0.06, the cumulative probability of any tobacco use by age 30 is relatively small at 0.37 and the cumulative probability of initiating alcohol use by age 30 is relatively moderate at 0.74, with the risk of initiating alcohol peaking at age 21 at 0.14. For all other drugs, the cumulative probability of initiation is less than 0.02. This pattern is the only one in which the cumulative risk of initiating alcohol by age 30 is less than 1.00.

The second and third patterns, which may be described as later and early soft drug users, have a high cumulative probability of alcohol use (0.95 by ages 21 and 17, respectively) as well as tobacco use (0.80 by ages 20 and 15, respectively) and moderate to high cumulative probability of marijuana use (0.50 by ages 21 and 17, respectively). The second pattern is characterized by later initiation of these substances than the third latent pattern, the third latent pattern having median event times of 13.5 for tobacco, 14.5 for alcohol and 16.5 for marijuana. The cumulative probability of NMUP for both patterns is moderately low (approximately 0.15 by age 30), and small for all other substances (less than 0.07). Even while the third pattern is characterized by early initiation of alcohol, tobacco and marijuana, the risk of initiating other drugs remains low.

The fourth pattern, described colloquially as 'progressive soft drug users', is relatively unique, as the risk for all drug types is relatively constant rather than having a more peaked shape. Similar to the second and third patterns, which have the highest risk of alcohol, tobacco and marijuana use, the fourth pattern is characterized mainly by soft drug use. However, the risk of other drugs is much higher than the second and third patterns, with the cumulative probability by age 30 of cocaine as well as NMUP being 0.48, hallucinogens being 0.46, inhalants being 0.29, stimulants being 0.23 and heroin being 0.02. While the risk of initiating all substances remains relatively flat over time, the risk of initiating alcohol, tobacco and marijuana use tends to peak at earlier ages (approximately 16–18 years) than the risk of initiating other substances (peaking in mid- to late 20s).

The last two patterns are characterized by a much higher risk of initiation of all substances, with the fifth pattern described colloquially as later hard drug users and the sixth pattern described colloquially as early hard drug users. Both patterns are characterized by high earlier risk of tobacco, alcohol and marijuana use (>0.95 cumulative probability of initiation for all three substances by age 20 for later hard drug-using pattern and by age 15 for early hard drug-using pattern). In both patterns, this is followed by peaking risk of initiation of other substances approximately 2–4 years later. The main distinction between the two patterns is that the peak risk for all substances in the early hard drug-using pattern is earlier than the later hard drug-using pattern, and the overall cumulative risk in the early hard drug-using pattern for all substances is higher (e.g. 0.58 cumulative probability of initiating inhalant use by 24 for early hard drug-using pattern versus 0.24 for later hard drug-using pattern).

Sex and ethnicity were both significant predictors of the probability of an individual belonging to a given pattern. The odds ratios of pattern membership, along with

[4] A lower information criteria is better; however, one must balance model fit with parsimony.

Table 2 Raw parameter estimates of the components of the hazard functions within the patterns.

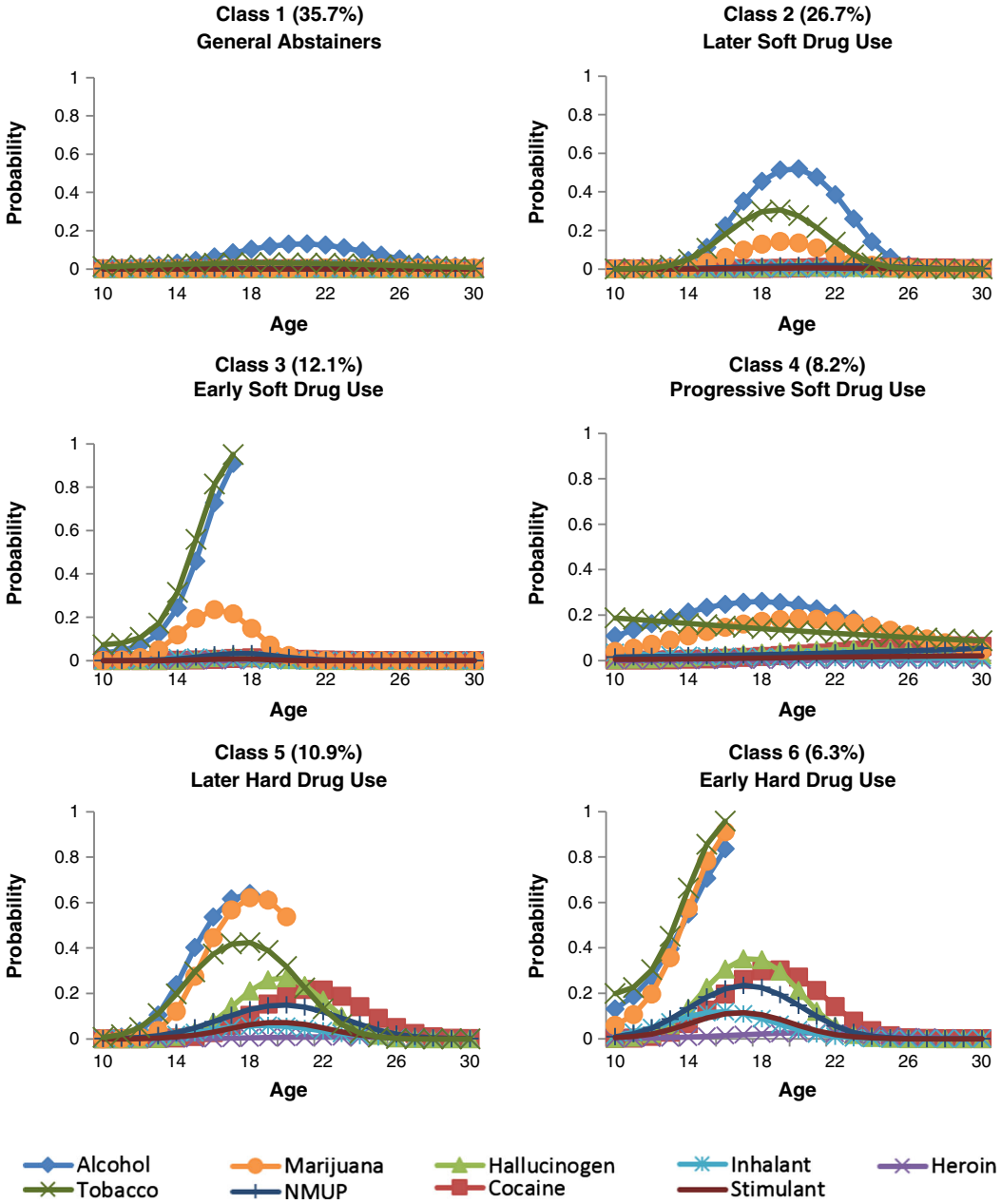
Substance	*Component of function*	*General abstainers*	*Later soft drug users*	*Early soft drug users*	*Progressive soft drug users*	*Later hard drug users*	*Early hard drug users*
Alcohol	Intercept	−6.04	−9.22	−3.36	−2.07	−6.39	−1.90
	Slope	7.71	19.28	2.29	2.58	17.46	3.98
	Quadratic	−3.59	−9.99	8.19	−1.65	−10.98	3.15
Cocaine	Intercept	−11.35	−13.40	−12.98	−6.85	−11.41	−7.21
	Slope	0*	12.74	20.45	4.39	17.75	14.84
	Quadratic	0*	−4.66	−12.09	−1.16	−7.77	−8.62
Hallucinogens	Intercept	−9.69	−15.00	−15.00	−5.62	−11.46	−6.65
	Slope	0*	18.60	30.34	3.92	21.47	16.35
	Quadratic	0*	−8.30	−20.80	−1.48	−11.01	−11.02
Heroin	Intercept	−15.00	−10.40	−15.00	−7.29	−10.28	−7.25
	Slope	0*	0*	0*	0*	8.82	7.00
	Quadratic	0*	0*	0*	0*	−3.48	−3.41
Inhalants	Intercept	−7.45	−6.25	−7.99	−3.32	−6.84	−4.29
	Slope	0*	2.26	15.13	−1.48	9.18	7.91
	Quadratic	0*	−1.77	−14.92	0.45	−5.19	−6.83
Marijuana	Intercept	−8.24	−9.59	−7.86	−3.12	−8.62	−2.79
	Slope	3.32	16.98	21.59	3.16	21.82	6.08
	Quadratic	−0.91	−9.25	−17.43	−1.56	−13.05	4.08
NMUP	Intercept	−6.06	−8.45	−10.52	−4.12	−6.88	−4.75
	Slope	0.41	7.94	19.21	0.52	10.43	10.04
	Quadratic	−0.22	−3.54	−12.89	0.06	−5.30	−7.05
Stimulants	Intercept	−8.38	−11.42	−15.00	−5.50	−9.35	−5.65
	Slope	0*	11.33	28.64	1.39	14.18	10.54
	Quadratic	0*	−5.24	−20.40	−0.29	−7.39	−7.67
Tobacco	Intercept	−4.37	−8.01	−2.56	−1.43	−5.12	−1.41
	Slope	2.14	16.46	−0.09	−0.56	12.57	0.53
	Quadratic	−1.20	−9.41	11.32	0.06	−8.21	11.59

* = Parameter constrained. Regarding constraints, see footnote to Table 1. NMUP = non-medical use of prescription drugs.

95% confidence intervals computed with a Bonferroni correction for multiple comparisons, are given in Table 4. The reference pattern was rotated in order to provide a detailed picture of differences between sex and ethnicities in the occurrence of these patterns. To enhance interpretation, the predicted probabilities of the patterns implied by the model are shown in the top half of Table 5.

Summarizing these results, we find that females are more likely to be general abstainers than males and less likely to belong to any other pattern than males, except that they have the same probability of being in the later soft pattern. African Americans are more likely to be in the later soft pattern as well as the early hard pattern compared to Caucasians, and less likely to be in the early soft pattern as well as the later hard pattern compared to Caucasians. This finding suggests there is a different relationship between the timing of initiations and severity of drugs initiated for African Americans compared to Caucasians. Hispanics and those of other races are less likely to be in the soft drug-using patterns compared to Caucasians, and more likely to be in the general abstainers pattern.

After fitting the model, we explored the MEPSUM model's ability to make person-specific predictions about individuals' trajectories of drug use. These individual trajectories are generated by calculating the individual's predicted probability of membership to each pattern, given their response pattern, and then weighting each pattern's trajectory accordingly, resulting in a unique predicted trajectory [28]. Importantly, these probabilities can be calculated for individuals who are already in the analysis or, as is the case here, for hypothetical respondents. As an example, we calculated the predicted multivariate substance use trajectories for a Caucasian male respondent who initiated alcohol use at age 12 and marijuana use at age 13. The model-predicted cumulative probability of substance use for this individual is shown in Fig. 5. The model predicts that a respondent who tries alcohol by age 12 and marijuana by age 13 is almost certain to try tobacco before the age of 18, and has an overall life-time probability of trying cocaine, hallucinogens and prescription drugs that is greater than 0.7; such a respondent's probability of using stimulants, inhalants or heroin are below 0.5, but considerably above the average probabilities shown in Fig. 1. Mid-to-late adolescence, between ages 14 and 20, appear to be the time of the greatest increases in substance use probability for this subject.



Note: Hazard functions plotted only during ages at which the cumulative probability for initiation was less than 1.00, as the risk of initiating is irrelevant after all individuals have initiated use.

Figure 3 Risk of initiation over time

We then used the model to compare this subject's individual predicted trajectory to that of another, similar subject who differs from the previous subject only in that he did not try marijuana until age 16. We used bootstrapping-based methods to approximate the uncertainty around each subject's predicted cumulative probability trajectory for each drug, by generating 100 alternative predicted trajectories using the estimated sampling covariance matrix of the parameter estimates [36]. Figure 6 shows the resultant bootstrap-predicted uncertainty around the predicted probability of cocaine use for both subjects. This interval facilitates comparisons between the two individual subjects, as the areas of overlap between the two trajectories represents points at which the model predicts the two subjects to have similar or different cumulative risk of trying cocaine.

Finally, the six-pattern model was also fitted in the validation sample to evaluate the stability of the model. The model cross-validated well, as both the risk of initiating the different substances and the overall prevalence rates of the patterns were found to be nearly identical in the two halves of the split sample. Additionally, predicted probabilities of belonging to the patterns by sex and ethnicity were found to be nearly identical between the two samples (Table 5) (r =0.99).

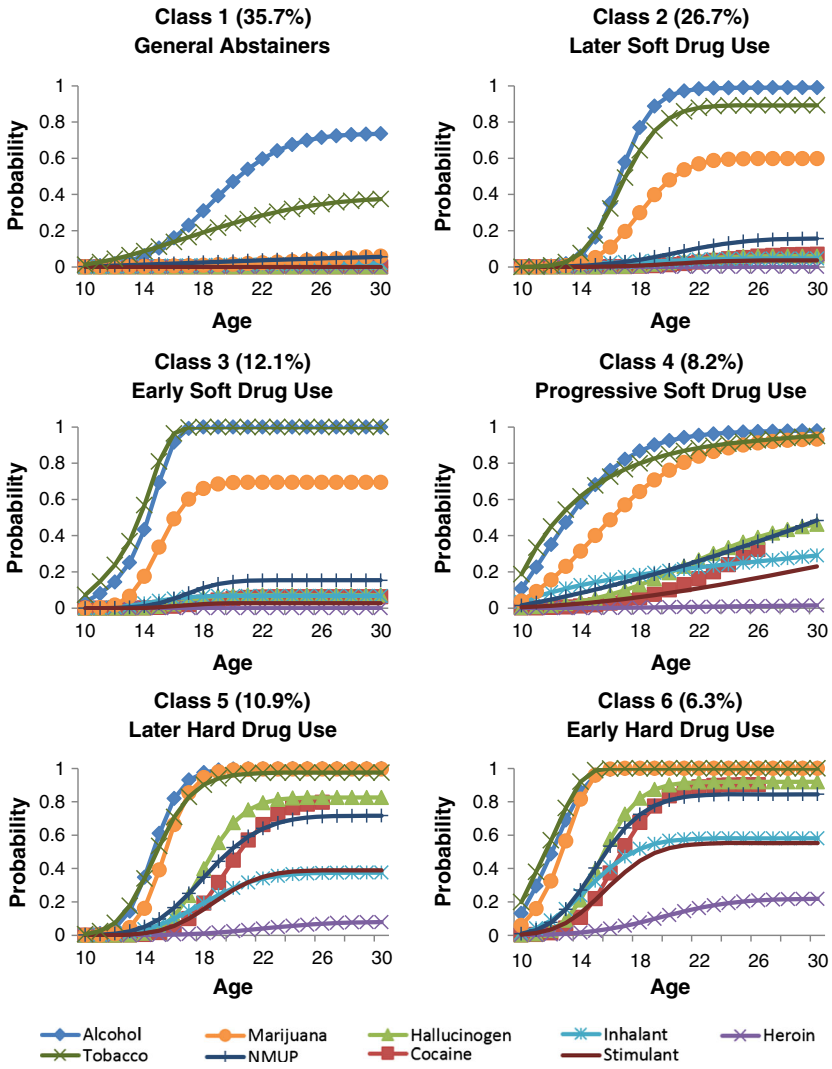


Figure 4 Cumulative probability of initiation over time

Table 3 Median event time: the estimated age at which 50% of individuals have initiated use of the substance.

Event	*General abstainer*	*Later soft drug use*	*Early soft drug use*	*Progressive soft drug use*	*Later hard drug use*	*Early hard drug use*
Alcohol	20.5	16.5	14.5	13.5	14.5	12.5
Cocaine	–	–	–	–	20.5	16.5
Hallucinogens	–	–	–	–	18.5	15.5
Heroin	–	–	–	–	–	–
Inhalant	–	–	–	–	–	17.5
Marijuana	–	20.5	16.5	16.5	15.5	12.5
NMUP	–	–	–	–	19.5	15.5
Stimulant	–	–	–	–	–	19.5
Tobacco	–	17.5	13.5	12.5	14.5	11.5

NMUP = non-medical use of prescription drugs.

DISCUSSION

Application of the MEPSUM model provided a succinct summary of individual differences in the timing of substance use initiation across nine substances in terms of six prototypical patterns. Notably, the differences found between these patterns were mainly differences in the age of peak risk and the overall level, rather than

Table 4 Influence of covariates: odds ratios of pattern membership with 95% confidence intervals.[a]

	Gender (reference male)	*Ethnicity (reference Caucasian)*		
Pattern comparison	*Female*	*African American*	*Hispanic*	*Other Race*
General abstainer versus early hard	**1.92 [1.56,2.35]***	**8.42 [4.74,14.96]***	**2.4 [1.79,3.23]***	**1.54 [1.10,2.15]***
Early soft versus early hard	1.06 [0.81,1.39]	2.19 [0.97,4.95]	0.97 [0.67,1.41]	**0.6 [0.38,0.94]***
Later soft versus early hard	**1.28 [1.04,1.57]***	**5.94 [3.22,10.96]***	1.12 [0.82,1.54]	0.73 [0.51,1.06]
Progressive soft versus early hard	0.89 [0.64,1.24]	**8.8 [4.65,16.65]***	**1.72 [1.13,2.6]***	1.32 [0.82,2.11]
Later hard versus early hard	0.96 [0.74,1.25]	1.57 [0.71,3.48]	0.95 [0.66,1.38]	0.7 [0.45,1.07]
General abstainer versus later hard	**1.99 [1.27,3.13]***	**5.35 [4.35,6.58]***	**2.53 [1.37,4.66]***	**2.2 [1.23,3.93]***
Early soft versus later hard	1.11 [0.76,1.60]	**1.39 [1.00,1.93]***	1.02 [0.54,1.94]	0.86 [0.44,1.68]
Later soft versus later hard	1.33 [0.83,2.13]	**3.77 [2.89,4.93]***	1.18 [0.53,2.61]	1.05 [0.56,1.97]
Progressive soft versus later hard	0.93 [0.60,1.43]	**5.59 [5.35,5.84]***	1.8 [0.03,114.2]	**1.89 [1.14,3.12]***
General abstainer versus progressive soft	**2.15 [1.3,3.55]***	0.96 [0.78,1.18]	1.4 [0.02,96.21]	1.16 [0.85,1.60]
Early soft versus progressive soft	1.19 [0.82,1.74]	**0.25 [0.18,0.35]***	0.57 [0.01,38.11]	**0.46 [0.30,0.69]***
Later soft versus progressive soft	1.43 [0.9,2.28]	**0.68 [0.52,0.88]***	0.65 [0.01,44.39]	**0.55 [0.38,0.81]***
General abstainer versus later soft	1.5 [0.89,2.53]	**1.42 [1.15,1.75]***	**2.14 [1.26,3.64]***	**2.1 [1.54,2.86]***
Early soft versus later soft	0.83 [0.56,1.24]	**0.37 [0.26,0.52]***	0.87 [0.45,1.67]	0.82 [0.55,1.23]
General abstainer versus early soft	**1.8 [1.15,2.83]***	**3.85 [2.92,5.08]***	**2.47 [1.56,3.89]***	**2.56 [1.75,3.74]***

*Bold values indicate the confidence interval is significant at α of 0.05 with a Bonferroni correction. [a]An odds ratio above 1 implies that the target is more likely than the reference to be in the first latent class mentioned in comparison to the second. For example, the odds of females being in the general abstainers class in comparison to the early hard class are 1.92 times higher than males. Confidence intervals which do not contain 1 are significant at the 0.05 level, computed with a Bonferroni correction for multiple comparisons.

Table 5 Cross-validation: predicted probabilities of pattern membership in evaluation and validation sample.

		Pattern					
Sample	*Group*	*General abstainer*	*Later soft*	*Early soft*	*Progressive soft*	*Later hard*	*Early hard*
Evaluation	Male (Caucasian)	0.22	0.29	0.17	0.09	0.15	0.09
	Female (Caucasian)	0.33	0.29	0.14	0.06	0.12	0.07
	African American (male)	0.36	0.34	0.07	0.02	0.05	0.16
	Hispanic (male)	0.37	0.23	0.12	0.11	0.11	0.06
	Other race (male)	0.35	0.22	0.10	0.12	0.11	0.09
Validation	Male (Caucasian)	0.23	0.29	0.17	0.09	0.14	0.08
	Female (Caucasian)	0.34	0.27	0.14	0.07	0.11	0.07
	African American (male)	0.38	0.32	0.09	0.01	0.04	0.16
	Hispanic (male)	0.40	0.23	0.12	0.10	0.09	0.06
	Other race (male)	0.38	0.21	0.11	0.13	0.09	0.07

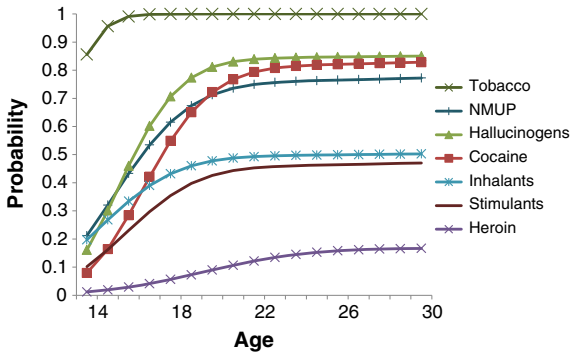


Figure 5 Predicted cumulative probability of initiation for a Caucasian, male subject who uses alcohol at age 12 and marijuana at age 13

large differences in the relative order of onset. When comparing similar patterns which differed in the age of peak risk and overall cumulative probability, the model revealed that the patterns with earlier risk had higher overall cumulative probabilities of initiating drug use. For example, the pattern which could be described colloquially as early hard drug users has higher cumulative probabilities of initiating all substances than the later hard drug users. This finding indicates that the age of initiating drug use is related strongly to the probability of initiating subsequent drugs, consistent with prior research [7]. Adding to prior research, we find this association to be consistent both in patterns related to soft drug use as well as harder drug use.

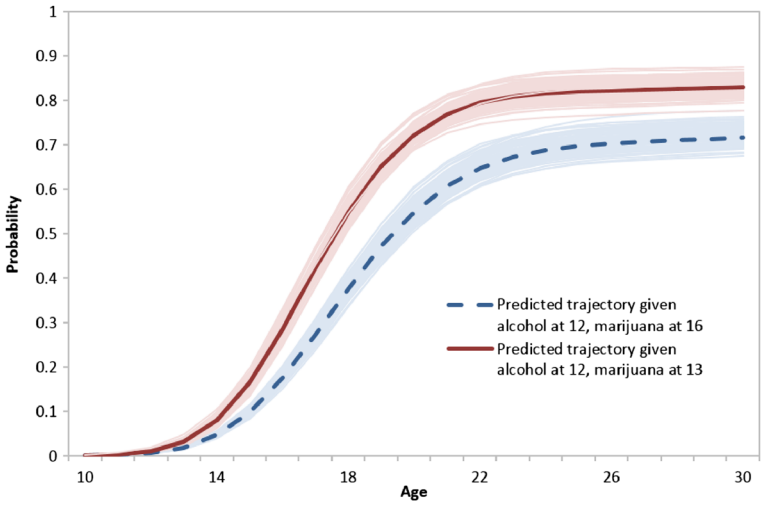


Figure 6 Predicted cumulative probability of initiating cocaine use with bootstrapped predicted error for Caucasian, male subjects with two different ages of marijuana initiation

The model also revealed some interesting insights into individual differences in the prevalence of the patterns depending on sex and ethnicity. Not all of these results might have been found using more traditional methods that focus on a single drug type or examining only use levels across substances at a single point in time. For example, the predicted probabilities of pattern membership suggest that there is a different relationship between the timing of initiations and severity of drugs initiated, depending on ethnicity.

Additionally, the model was used to predict the probability of transition to hard drugs, given early alcohol and marijuana initiation. Importantly, because the MEPSUM can incorporate information about not only the life-time use of a given substance but also the timing of drug use initiation, it can allow more nuanced individual-level predictions. For instance, the two individual subjects who were compared previously—the Caucasian male adolescents who both tried alcohol at age 12 but tried marijuana at different times—might have been recorded as having similar substance use histories in a traditional model (i.e. both had used alcohol and marijuana). However, the MEPSUM was able to use information about the timing of their drug use to make different predictions about further drug-related outcomes. Relevant to coordinating prevention efforts, individual predictions gleaned from the MEPSUM model about the timing of the use of different substances may be particularly helpful in tailoring predictions about substance use to individuals.

A few limitations of the current research must also be noted. First, our analysis did not differentiate between birth cohorts. Thus, follow-up work is needed regarding the stability of the patterns across cohorts. Secondly, substance use initiation is recorded through retrospective reports rather than true longitudinal data, and thus the data are vulnerable to potential retrospective recall bias. Notwithstanding these limitations, however, the MEPSUM model was able to capture and describe individual differences in the initiation of substance use over time for a large number of substances, providing a rich picture of the patterns of substance use over time. In order to understand more clearly the processes leading to substance use initiation as well as how the onset times for different substances are related to each other, it is important to consider individual differences in patterns of initiation of multiple substances rather than dissecting the events in order to apply more traditional methods. In this analysis, the prevalence rates of different patterns of onset were found to differ by sex and ethnicity. Additional research should examine other etiological and mediating mechanisms leading to the observed heterogeneity in the risk for onset of different drug use over time. Thus, in addition to introducing a novel statistical model for drug use data, we hope that the patterns revealed by this model may serve as important outcome variables for future research.

Declaration of interests

None.

Acknowledgements

We would like to thank Dr Andrea Hussong for suggesting that we target this work to an applied audience, and the anonymous reviewers whose feedback led to significant improvements.

References

1. Grella C. E., Anglin M. D., Wugaalter S. E. Cocaine and crack use and HIV risk behaviours among high risk methadone maintenance clients. *Drug Alcohol Depend* 1995; **37**: 15–21.
2. Kosten T., Rounsaville B., Kleber H. Antecedents and consequences of cocaine abuse among opiate addicts: a 25 year follow up. *J Nerv Mental Dis* 1988; **176**: 176–81.
3. Wu L. T., Schlenger W. E., Galvin D. M. The relationship between employment and substance abuse among students aged 12 to 17. *J Adolesc Health* 2003; **32**: 5–15.

4. Chen K., Kandel D. B. The natural history of drug use from adolescence to the mid-thirties in a general population sample. *Am J Public Health* 1995; **85**: 41–7.
5. Kandel D. B., Logan J. A. Patterns of drug use from adolescence to young adulthood: I. Periods of risk for initiation continued use and discontinuation. *Am J Public Health* 1984; **74**: 660–6.
6. Arnett J. J. The developmental context of substance use in emerging adulthood. *J Drug Issues* 2005; **35**: 235–4.
7. Hser Y., Anglin M. D. Addiction treatment and recovery careers. In: Kelly J. F., White W. L., editors. *Addiction Recovery Management: Theory Science and Practice*. New York: Springer Science; 2010, pp. 9–29.
8. Bachman J. G., O'Malley P. M., Schulenberg J. E., Johnston L. D., Bryant A. L., Merline A. C. *The Decline of Substance Use in Young Adulthood: Changes in Social Activities Roles and Beliefs*. Mahwah NJ: Erlbaum; 2002.
9. Yamaguchi K., Kandel D. B. Patterns of drug use from adolescence to young adulthood: II. Sequences of progression. *Am J Public Health* 1984; **74**: 668–72.
10. Hamburg B. A., Kraemer H. C., Jahnke W. A hierarchy of drug use in adolescence: behavioral and attitudinal correlates of substantial drug use. *Am J Psychiatry* 1975; **132**: 1155–63.
11. Booth B., Leukefeld C., Falck R., Wang J., Carlson R. Correlates of rural methamphetamine and cocaine users: results from a multistate community study. *J Stud Alcohol* 2006; **67**: 493–501.
12. Byqvist S. Patterns of drug use among drug misusers in Sweden gender differences. *Subst Use Misuse* 2006; **41**: 1817–35.
13. Flannery B., Morgenstern J., McKay J., Wechsberg W., Litten R. Cooccurring alcohol and cocaine dependence: recent findings from clinical and field studies. *Alcohol Clin Exp Res* 2004; **28**: 976–81.
14. Gossop M., Stewart D., Marsden J. Attendance at Narcotics Anonymous and Alcoholics Anonymous meetings frequency of attendance and substance use outcomes after residential treatment for drug dependence: a 5-year follow-up study. *Addiction* 2003; **103**: 119–25.
15. Schwartz S. J., Phelps E., Lerner J. V., Shi H., Brown C., Lewin-Bizan S. *et al*. Promotion as prevention: positive youth development as protective against tobacco alcohol illicit drug and sex initiation. *Appl Dev Sci* 2010; **14**: 197–211.
16. Hougaard P. *Analysis of Multivariate Survival Data*. New York: Springer; 2000.
17. Willett J. B., Singer J. D. It's *déjà vu* all over again: using multiple-spell discrete time survival analysis. *J Ed Behav Stat* 1995; **20**: 41–67.
18. Hougaard P., Harvald B., Holm N. V. Measuring the similarities between the lifetimes of adult Danish twins born between 1881–1930. *J Am Stat Assoc* 1992; **87**: 17–24.
19. Ventura J., Nuechterlein K. H., Subotnik K. L., Hardesty J., Mintz J. Life events can trigger depressive exacerbation in the early course of schizophrenia. *J Abnorm Psychol* 2000; **109**: 139–44.
20. Almansa J., Vermunt J. K., Forero C. G., Alonso, J. A factor mixture model for multivariate survival data: an application to the analysis of lifetime mental disorders. *Appl Stat* 2014; **63**: 85–102.
21. Dean D. O., Bauer D. J., Shanahan M. J. A discrete-time multiple event process survival mixture (MEPSUM) model. *Psychol Methods* 2014; **19**: 251–264.
22. United States Department of Health and Human Services Substance Abuse and Mental Health Services Administration Office of Applied Studies. *National Survey on Drug Use and Health 2009*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research; 2010.
23. Singer J. D., Willett J. B. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press; 2003.
24. Lee E. T., Wang J. W. *Statistical Methods for Survival Data Analysis*. New York: Wiley; 2003.
25. Muthén B., Masyn K. Discrete-time survival mixture analysis. *J Ed Behav Stat* 2005; **30**: 27–58.
26. Muthén B., Muthén L. *Mplus User's Guide*, 6th edition. Los Angeles CA: Muthén Muthén; 1998–2010.
27. Allison P. D. *Logistic Regression Using the SAS System: Theory and Application*. Cary, NC: SAS Institute; 1999.
28. McLachlan G., Peel D. *Finite Mixture Models*. New York: Wiley; 2000.
29. Hipp J. R., Bauer D. J. Local solutions in the estimation of growth mixture models. *Psychol Methods* 2006; **11**: 36–53.
30. Akaike H. A new look at statistical model identification. *IEEE Trans Autom Control* 1974; **AU-19**: 719–22.
31. Schwarz G. Estimating the dimensions of a model. *Ann Stat* 1978; **6**: 461–4.
32. Clogg C. C. Latent class models: recent developments and prospects for the future. In: Arminger G., Clogg C. C., Sobel M. E., editors. *Handbook of Statistical Modeling for the Social and Behavioral Sciences*. New York: Plenum; 1995, pp. 311–52.
33. Hagenaars J. A. *Loglinear Models With Latent Variables*. London: Sage; 1993.
34. Petras H., Masyn K. General growth mixture analysis with antecedents and consequences of change. In: Piquero A., Weisburd D., editors. *Handbook Quant Criminology*. New York: Springer; 2010, pp. 69–100.
35. Marsh H. W., Lüdtke O., Trautwein U., Morin A. Classical latent profile analysis of academic self-concept dimensions: synergy of person- and variable-centered approaches to theoretical models of self-concept. *Struct Equat Model* 2009; **16**: 191–225.
36. Pek J., Losardo D., Bauer, D. J. Confidence intervals for a semiparametric approach to modeling nonlinear relations among latent variables. *Struct Equat Model* 2011; **18**: 537–53.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web site.

Appendix S1 Example code to fit model in Mplus.