Effects of age at first substance use and psychiatric comorbidity on the development of substance use disorders

Minje Sung a, *, Al Erkanli a, Adrian Angold b, E. Jane Costello b

a Department of Biostatistics and Bioinformatics, Duke University Medical School, Box 3454, Durham, NC 27710, USA
b Department of Psychiatry and Behavioral Sciences, Developmental Epidemiology Center, Duke University Medical School, Box 3454, Durham, NC 27710, USA

Received 17 April 2003; received in revised form 17 March 2004; accepted 18 March 2004

Abstract

In this paper, we examine the effects of age at first substance use, and history of psychiatric disorders, on the development of substance use disorder (SUD) by age 16. We use a prospective, longitudinal design to disaggregate the effects of age at first use and time since first use on the development of adolescent SUD. Second, we test the hypothesis that adolescent SUD is an unlikely progression from early substance use unless children also show other early conduct problems. A population sample of 1420 children from the Great Smoky Mountains Study (GSMS) was assessed annually between ages 9 and 16. Logistic regression models were applied within the hierarchical Bayesian framework, where the covariate effects were described by time-varying parameters having a first-order auto-regressive prior distribution. Posterior analyses based on a Gibbs sampling approach revealed that, controlling for years of exposure, the risk of transition to SUD increased with age at onset for onsets before age 13, but began to fall for onset at 14. Among users, use alone, without early conduct problems, led to a 11% prevalence of SUD by age 16. Past conduct disorder (CD) had a strong additive effect at ages 13–15, but at age 16, when substance use and abuse became more normative, the excess risk from prior CD decreased. Boys, but not girls, with a history of depression were at increased risk of SUD. Anxiety increased the risk of SUD in girls at age 16, but not before that. Results only partially support the study hypothesis; early use was a major predictor of adolescent SUD even in the absence of CD.

© 2004 Elsevier Ireland Ltd. All rights reserved.

Keywords: Development; Substance abuse; Conduct disorder; Depression; Substance use disorder; Markov regression; Bayesian estimation; Gibbs sampling

1. Introduction

In this paper, we first examine what Anthony and Petronis called the “uncertainty whether early-onset drug use exerts an independent effect on risk of later drug problems, even though it might presage and serve as a vulnerability marker for later problems” (Anthony and Petronis, 1995). Using retrospective data, they showed that the risk of adult drug problems was linearly related to age at onset of drug use, being twice as high in adults reporting first use before age 13 as in those with first use after age 17. However, the time from first use to problem use was around 4 years, irrespective of age at first use among problem users. Here we look at the situation in adolescence, and using prospective data. Many previous research findings have established that early onset of drug use is associated with later problem use (Andreason et al., 1992; Boyle et al., 1992; DeWit et al., 2000; Kandel et al., 1986; Kaplan et al., 1986; Mills and Noyes, 1984; Robins and Murphy, 1967; Robins and Przybeck, 1985; Yamaguchi and Kandel, 1984). We have found no prospective study testing for a linear relationship between age at onset and probability of later problem drug use in adolescents, but retrospective analysis of the Ontario Mental Health Supplement data (DeWit et al., 2000) found a nonlinear relationship, with the highest risk in those reporting first use at 13–14 for alcohol abuse and 11–12 for alcohol dependence.

Second, we explore the effects of early psychiatric disorders, including conduct disorder (CD), on both the likelihood and the timing of the transition from substance use to substance use disorders (SUD). It is well established that early problem behavior is associated with later drug use and abuse (Anderson et al., 1989; Jessor and Jessor, 1977; Kandel et al., 1978; Kaplan, 1986; Monnelly et al., 1983; Smith and Fogg, 1979; Vaillant and Milofsky, 1982; Wingard...
et al., 1979). However, most of the studies cited start in adolescence. In this paper we examine an earlier developmental period, beginning at age 9 and following children through to age 16. We examine the effects of a range of psychiatric disorders with onset by age 15 on risk for substance use disorders by age 16.

In her analysis of the retrospective Epidemiologic Catchment Area (ECA) data, Robins proposed a more specific hypothesis than one of a general association between conduct problems and SUD.

Abuse is extremely rare for those free of (early) conduct problems, no matter how early substance use began. At every other level of conduct problems, however, the earlier that use begins, the greater is the likelihood of substance abuse. . . . For those first using substances before age 20, the number of conduct problems was an even better predictor of substance abuse than was age of onset (before or after age 15). Among those beginning substance use before age 15 with seven or more conduct problems, more than half developed substance abuse; with only one conduct problem, only 5% did so. When first use occurred between ages 15 and 19, there is still a large effect from number of conduct problems, but the control for age of first use somewhat reduced their impact.” (Robins and McEvoy, 1990) (p. 196).

If it is the case that early CD increases the risk of adult SUD, a logical extension is to expect early CD also to drive the onset of adolescent SUD. Here we test this hypothesis, and also examine the impact of other psychiatric disorders on the transition from use to abuse/dependence.

The present study also adds to the literature in the following ways: (1) It follows a representative population sample of youth prospectively from childhood into adolescence. (2) It uses structured psychiatric interviews with parent and child to generate a range of DSM-IV psychiatric diagnoses. (3) The sample contains both boys and girls, unlike many previous studies. (4) The statistical methodology incorporates the effects of psychiatric comorbidity on transitions from the prior use of substances to SUD. (5) Use of a hierarchical Bayesian modeling approach permits the analysis of a complex, stratified data set that addresses some issues of dynamic age and covariate effects.

2. Methods

2.1. Study population

The Great Smoky Mountains Study (GSMS) is a longitudinal study of the development of psychiatric disorder and substance use and abuse in a predominantly rural area of the southeastern United States. Full details of the study design and instruments used can be found elsewhere (Costello et al., 1996). Briefly, a representative sample of 9, 11, and 13-year olds was recruited from 11 counties in western North Carolina using a household equal probability design. A two-phase sampling procedure was used based on the first-phase sample of 4500 subjects, out of which subsets were retained for the second phase. Youth with behavior problems and American Indian youth were oversampled relative to White youth with low levels of behavior problems. Participants were given a weight inversely proportional to their probability of selection, so that the results presented here are representative of the population from which the sample was drawn. The response rate was 96% at the screening phase and 80% at the interview phase, for a sample of 1420. Between 78 and 95% were reinterviewed each year. Each subject was interviewed annually until age 16. The study was reviewed and approved by the Duke University Institutional Review Board, and parents and children signed informed consent form before each assessment.

2.2. Measures

The Child and Adolescent Psychiatric Assessment (CAPA) (Angold and Costello, 1995; Angold et al., 1995) is an interviewer-based structured psychiatric interview that collects information on the onset dates, duration, frequency, and intensity of symptoms of a wide range of psychiatric disorders. Diagnoses can be made according to DSM-IV, DSM-III-R, or ICD-10 criteria. The interviewer is intensively trained to use the questions and probes provided in the interview schedule to ensure that participants (a) understand the question being asked, (b) provide clear information on behavior or feelings relevant to the symptom, and (c) have the symptom at a clinical level of severity. The areas of interest in this study include substance use, abuse and dependence, and psychiatric symptoms and disorders.

2.3. Substance use, abuse, and dependence

The CAPA asks the child or parent about any use, ever and in the last 3 months, of tobacco (cigarettes, snuff, and chewing tobacco), alcohol (without adult permission), and other substances including cannabis, cocaine, crack, amphetamines, ice, inhalants, heroin, other opioids, LSD, PCP, psilocybin, sedatives, “club drugs” and steroids. If use is reported, information is collected about first use, frequency, quantity, and associated impairment (see below). “Substance use” was defined for these analyses as follows: alcohol use was counted if the child drank any alcohol obtained without parental permission; smoking was counted if the child smoked on average one or more cigarettes a day for 3 months, snuff use if the child used at least half a can per week, chewing tobacco use if half a pouch per week was consumed, and all other drugs if any use was reported over the preceding 3-month period. We report here the proportion of children reporting first use of tobacco in any form.

When any use was reported, questions are asked about maladaptive functioning and functional impairment as a result of the substance use, to determine substance abuse or
Bayesian approach rather than GEE or Generalized Linear

adjusted to obtain properly weighted estimates and variances.

because of the sampling design, analyses needed to be ad-

appropriate modeling strategy was needed to incorporate the

be constant (time-homogeneous) at different ages; thus, an

age effects on the prevalence of substance abuse might not

result in the exclusion of subjects with missing data. Second,

either at random or by study design. Applying GEE would

collection, almost all subjects had at least one missing wave,

three major challenges. First, across the eight waves of data

to address our specific research concerns, we considered

covariates on the outcome. In selecting a modeling strategy

relation structure of the outcome variable over time, and (b)

longitudinal study (a) by describing the within subject cor-

for dealing with missing data and for describing develop-

tion (Zeger et al., 1988 ) provides the most popular set of

those aged 13 at entry) and eight waves (for those aged 9

each subject contributed data from between four waves (for

limited analysis indicated that there was little effect of

cohort on substance abuse. All available data for subjects

up to age 16 was used in these analyses. This means that

each subject contributed data from between four waves (for

months, the month and year of onset is recorded. From

symptom onset dates, we estimated the age at onset

subject had the experience of a diagnosis, by calculating the earliest age at which the

subject had experienced the minimum number of symptoms

required to meet DSM-IV criteria for that disorder.

2.4. Psychiatric disorders

The CAPA focuses on the 3 months immediately preced-

the interview. Three months was chosen because of

concerns about the reliability of memory for longer peri-

ods. For any symptom reported as occurring during the past

3 months, the month and year of onset is recorded. From

these symptom onset dates, we estimated the age at onset

of a diagnosis, by calculating the earliest age at which the

subject had experienced the minimum number of symptoms

required to meet DSM-IV criteria for that disorder.

2.5. Statistical analysis

Data from the three age cohorts were combined, as pre-

liminary analyses indicated that there was little effect of

cohort on substance abuse. All available data for subjects

up to age 16 was used in these analyses. This means that

each subject contributed data from between four waves (for

those aged 13 at entry) and eight waves (for those aged 9

at entry). For data like ours where the response variable is

a repeated binary measure, Generalized Estimating Equa-

tion (Zeger et al., 1988) provides the most popular set of

methods for estimating average or marginal effects of risk

factors. However, the GEE approach has limited capability

for dealing with missing data and for describing develop-

mentally changing effects of risk factors over time. In the

present study, we are trying to maximize the benefit of a

longitudinal study (a) by describing the within subject cor-

relation structure of the outcome variable over time, and (b)

by accounting for the developmental impact of time varying

covariates on the outcome. In selecting a modeling strategy

to address our specific research concerns, we considered

three major challenges. First, across the eight waves of data

collection, almost all subjects had at least one missing wave,

either at random or by study design. Applying GEE would

result in the exclusion of subjects with missing data. Second,

age effects on the prevalence of substance abuse might not

be constant (time-homogeneous) at different ages; thus, an

appropriate modeling strategy was needed to incorporate the

developing prevalence rates over different age levels. Third,

because of the sampling design, analyses needed to be ad-

justed to obtain properly weighted estimates and variances.

To meet these challenges we chose to use a hierarchical

Bayesian approach rather than GEE or Generalized Linear

Mixed Model (GLIMM) (Zeger and Liang, 1992). The ad-

vantage of hierarchical Bayesian methodology is its ability

to deal with missing values in the response variable (under

the assumption of missingness at random or uninformative

drop-out) and to provide flexibility in specifying conditional

independence structures among the parameters in the model.

The essence of the Bayesian approach is to use probability

representation of all unknown parameters in the model and

make inferences about those quantities by posterior probabi-

lity distributions. The missing outcome variables are treated

as unknown quantities, and their posterior distributions are

obtained by simulation.

While GLIMM would use whatever waves of data are

available without deleting whole missing case, the param-

eter estimates would still represent average effects of risk

factors over time, i.e., static effects, which may be inap-

propriate for describing developmental changes over time.

Thus, we employed Bayesian Dynamic Generalized Linear

Models (DGLM) (West et al., 1985) to describe nonlinear,

non-normal longitudinal measures. The DGLM approach

has not used in addiction research, to our best knowledge,

but has been considered in a wide variety of research in fi-

nancial, socio-economic, and engineering studies in recent

years. This modeling approach provides dynamic extensions

of standard generalized linear models. The fundamental idea

of such a dynamic model is that the age specific (dynamic)

regression parameters are estimated by using:

(1) the current information in the data about the association

between the outcome variable and the covariates,

(2) the previous information about the dynamic parameter

described by a dynamic system equation, e.g., \( \beta_t = \beta_{t-1} + \epsilon_t \).

Dynamic multiple logistic regression models were used to

estimate the probability of substance use disorder, as a

function of multiple independent variables. More details on

the model structure are presented in the Appendix A. Even

though the methodology discussed in the Appendix A may

look complex, the observation equation describing the asso-

ciations between the response variable and risk factors is in

a familiar multiple regression format. Also, the practical ad-

vantages of the hierarchical Bayesian models certainly out-

weigh the difficulty with implementation. In fact, the GEE

model, with dummy coded variables for time specific ef-

fects fitted in SAS, failed to estimate the parameters and

the logistic regression implemented in SAS did not produce

comparable results.

In a set of longitudinal data, the observations within sub-

jects are correlated. We use an extension of the transition

model similar to those discussed in Diggle et al. (1994). The

within-subject correlation of the outcome variable over time

is controlled by including the previous outcome variable in

the model as an explanatory variable. Also, the dynamic

modeling approach describes evolving regression para-

meters over various age levels (Cargnioni et al., 1997). Using

the previous outcome values as an explanatory variable, the

dependence. The DSM-IV criteria for dependence were writ-

ten with adults in mind; for example, children are unlikely
to have used long enough to develop major withdrawal syn-
dromes or the physical sequel of longstanding abuse, or to be

at risk for losing jobs or spouses. For these analyses we de-
defined a substance use disorder as abuse, identified as use, ac-
}

panied by significant impairment, i.e., “failure to fulfill

major role obligations at work, school, or home” (American

Psychiatric Association, 1994, p. 182), with or without de-

pendence.


289
missing values of an outcome variable are replaced with posterior samples that are generated by the simulations. Each missing value is multiply imputed, so that uncertainty about them can be quantified.

Our modeling framework integrates two classes of models which can be classified as parameter driven and observation driven Markov models using the terminology of Cox (1981). The observation driven Markov models are the Markov chains where the Markov structure is on the observables which describe the transition from the previous state to current state. In the parameter driven Markov models, the parameters evolve over time according to a first-order Markov chain, which is in the spirit of the dynamic linear models of West et al. (1985).

Our dynamic models were implemented in WinBugs program (Spiegelhalter et al., 1996) using the Gibbs sampling algorithm (Casella and George, 1992). Uniform, non-informative priors are used to describe uncertainty about the regression parameters. Non-informative priors lead to inferences that are comparable to those obtained by non-Bayesian approaches, none of which, to our best knowledge, are available to this kind of data. No apparent convergence problems were encountered in our simulation analysis.

As stated earlier, American Indians ($n = 350$) and youth with behavior problems ($n = 801$) were sampled with a higher probability than non-American Indian youth with low levels of behavior problems ($n = 269$). The latter were selected with a 10% probability from the screened sample. To restore the sample to being representative of the total population of children from the study area, we used the data augmentation approach developed for adjusting for selection bias (Silliman, 1997; Tanner and Wong, 1987). The technique was implemented using the Gibbs sampler, based on the observed data on the 269 non-American Indian youth with low levels of problem behaviors at intake. We adopted this approach because it is important to account for observed data when data are not missing at random. To verify the compatibility of this approach, as a tool for adjusting sampling bias, we compared results from a simple static logistic regression model in WinBugs with those from a weighted analysis conducted using SAS PROC LOGISTIC. The estimated model parameters and variances were almost identical to the results of the model with Bayesian data augmentation, supporting the compatibility of Bayesian data augmentation as an approach to weighted analysis.

Inferences were made using the posterior means and 95% credible intervals obtained from the simulation analysis. Credible intervals are the Bayesian analog of confidence intervals (see, Bernardo and Smith, 1994, p. 259 for a formal definition). Once the posterior distributions are simulated, comparisons between the parameters of interest can be easily made through simple data analytical methods, such as histograms. Although the Bayesian approach does not use a $P$-value, statistical inferences about the effects of risk factors can be made using the 95% credible intervals.

Inferences were made using the posterior means and 95% credible intervals obtained from the simulation analysis. Credible intervals are the Bayesian analog of confidence intervals (see, Bernardo and Smith, 1994, p. 259 for a formal definition). Once the posterior distributions are simulated, comparisons between the parameters of interest can be easily made through simple data analytical methods, such as histograms. Although the Bayesian approach does not use a $P$-value, statistical inferences about the effects of risk factors can be made using the 95% credible intervals.

Results are reported under three headings: (1) onset of substance use, SUD, and psychiatric disorders; (2) relationship between substance use and SUD; (3) impact of psychiatric disorders on the transition to SUD.

### 3. Results

#### 3.1. Rates of onset

Table 1 shows the reported rates of age specific incidence for psychiatric disorders, substance use, and substance use disorder.

<table>
<thead>
<tr>
<th>Age</th>
<th>CD</th>
<th>ANX</th>
<th>DEP</th>
<th>ODD</th>
<th>ADHD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Boy</td>
<td>Girl</td>
<td>Both</td>
<td>Boy</td>
<td>Girl</td>
</tr>
<tr>
<td>Before 10</td>
<td>3.76</td>
<td>0.47</td>
<td>2.13</td>
<td>4.66</td>
<td>5.70</td>
</tr>
<tr>
<td>10</td>
<td>1.38</td>
<td>0.27</td>
<td>0.81</td>
<td>1.45</td>
<td>0.85</td>
</tr>
<tr>
<td>11</td>
<td>3.24</td>
<td>0.49</td>
<td>1.83</td>
<td>0.29</td>
<td>1.61</td>
</tr>
<tr>
<td>12</td>
<td>0.91</td>
<td>0.96</td>
<td>0.94</td>
<td>0.30</td>
<td>1.40</td>
</tr>
<tr>
<td>13</td>
<td>3.34</td>
<td>1.04</td>
<td>2.14</td>
<td>0.40</td>
<td>0.44</td>
</tr>
<tr>
<td>14</td>
<td>1.20</td>
<td>0.06</td>
<td>0.60</td>
<td>0.81</td>
<td>1.66</td>
</tr>
<tr>
<td>15</td>
<td>0.48</td>
<td>0.72</td>
<td>0.61</td>
<td>0.06</td>
<td>0.72</td>
</tr>
<tr>
<td>16</td>
<td>0.15</td>
<td>0.00</td>
<td>0.07</td>
<td>0.00</td>
<td>0.69</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tobacco</th>
<th>Alcohol</th>
<th>Drugs</th>
<th>SUD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boy</td>
<td>Girl</td>
<td>Both</td>
<td>Boy</td>
</tr>
<tr>
<td>Before 10</td>
<td>1.41</td>
<td>0.36</td>
<td>0.91</td>
</tr>
<tr>
<td>10</td>
<td>1.67</td>
<td>1.28</td>
<td>1.48</td>
</tr>
<tr>
<td>11</td>
<td>1.56</td>
<td>0.49</td>
<td>1.02</td>
</tr>
<tr>
<td>12</td>
<td>2.64</td>
<td>1.36</td>
<td>1.99</td>
</tr>
<tr>
<td>13</td>
<td>6.11</td>
<td>4.84</td>
<td>5.46</td>
</tr>
<tr>
<td>16</td>
<td>4.93</td>
<td>4.47</td>
<td>4.69</td>
</tr>
</tbody>
</table>
disorder at each age. The rate of onset at a given age was computed by dividing the weighted number of reported onsets by the weighted total number of subjects interviewed since the age of onset after removing subjects who had onsets previously. There was no SUD before 12, but onsets increased rapidly thereafter, affecting 9.25% of the sample by age 16. Among the substance use onsets, the peak years were 14 (9.24%) and 15 (7.29%) for tobacco, and 13–16 for alcohol (8.68–13.06%) and other drugs (5.50–8.27%). In contrast, psychiatric disorders with the exception of depression had their origins quite early in childhood; new onsets were infrequent for most diagnoses after the middle teens. The onsets of depression were most frequent between ages 13 (1.70%) and 15 (1.37%) while the rate of onset at other age was 0.56% at most. The onset of ADHD was extremely rare after age 10.

Comparisons between boys and girls showed that boys had higher rates of tobacco use onsets and SUD through age 14, alcohol onsets through age 12, and drug onsets through age 16. Thereafter, onsets in girls equaled or exceeded those in boys. In general, boys had higher rates of onset for conduct disorder and oppositional defiant disorder (ODD) at each age level, and girls had higher onset rates for anxiety disorders (ANX) and depression (DEP). Both incidence and prevalence of attention deficit hyperactivity disorder (ADHD) were very rare in adolescence for both boys and girls.

3.2. Timing of onset of substance use and SUD

We tested the suggestion (Anthony and Petronis, 1995) that the time from first use to SUD, among those who develop it, is fairly constant irrespective of age at first use. Fig. 1 presents the results of a survival analysis using Cox regression showing the cumulative probability over time of developing SUD for each age at onset of use. Number of years after onset is on the X-axis. We considered the data as censored survival data as we could not observe all the subjects long enough for all to go through the full period of risk for SUD. While the rate of SUD accelerated with time, regardless of the age at onset of use, the later the age at first reported use, the greater the rate of increase. However, those who began use at 14 or 15 did not have the same risk of SUD by 16 as those who began use before that age; rates of SUD by 16 were 0.7–2% in those beginning use after 13, and 12–28% in those beginning use by age 13. Thus, the results show an interaction between age at SUD diagnosis and age at onset of use.

To further investigate this interaction, Tables 2 and 3 present the relationship between age at onset of use and SUD in two different ways. Table 2 shows the cumulative rates of SUD 2, 4, and 6 years after first use, by age at onset of use. Table 3 shows the cumulative rates by age at onset of SUD. It can be seen from Table 2 that those who were older when they began using substances showed higher rates of SUD 2, 4, and 6 years after first use, by age at onset of use. Table 3 shows that subjects who were younger when they started using were more likely than those with a later onset to have developed SUD by ages 12, 14, and 16—but then they had more years of exposure.

It is apparent that the increasing risk of SUD with time after the onset of use was affected by age at onset. To estimate the effect of age at onset net of age, we first estimated cumulative probabilities of SUD over x years for those with onset at one age and then compared them to cumulative probabilities of SUD for those with onset at an earlier age. We define this quantity as Delta. For example, we compare the effects of onset at 10 and onset at 9 across ‘n’ years after onset, and

![Fig. 1. Estimated cumulative probability of developing SUD for each age at onset of use as a function of time since onset of use.](image-url)
Table 2

Estimated cumulative rates of SUD 2, 4, and 6 years after first use, by age at onset of use

<table>
<thead>
<tr>
<th>Age at onset of use</th>
<th>In 2 years</th>
<th>In 4 years</th>
<th>In 6 years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>95% CI</td>
<td>Mean</td>
</tr>
<tr>
<td>9 or earlier</td>
<td>0.001</td>
<td>0.001, 0.003</td>
<td>0.022</td>
</tr>
<tr>
<td>10</td>
<td>0.003</td>
<td>0.001, 0.006</td>
<td>0.037</td>
</tr>
<tr>
<td>11</td>
<td>0.006</td>
<td>0.004, 0.013</td>
<td>0.094</td>
</tr>
<tr>
<td>12</td>
<td>0.009</td>
<td>0.005, 0.015</td>
<td>0.115</td>
</tr>
<tr>
<td>13</td>
<td>0.024</td>
<td>0.014, 0.037</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>0.017</td>
<td>0.009, 0.027</td>
<td></td>
</tr>
</tbody>
</table>

Table 3

Estimated cumulative rates of SUD by age at onset of use

<table>
<thead>
<tr>
<th>Age at onset of use</th>
<th>Age 12</th>
<th>Age 14</th>
<th>Age 16</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>95% CI</td>
<td>Mean</td>
</tr>
<tr>
<td>9 or earlier</td>
<td>0.010</td>
<td>0.005, 0.017</td>
<td>0.063</td>
</tr>
<tr>
<td>10</td>
<td>0.003</td>
<td>0.001, 0.006</td>
<td>0.037</td>
</tr>
<tr>
<td>11</td>
<td>0.001</td>
<td>0.000, 0.002</td>
<td>0.045</td>
</tr>
<tr>
<td>12</td>
<td>0.009</td>
<td>0.005, 0.015</td>
<td>0.115</td>
</tr>
<tr>
<td>13</td>
<td>0.003</td>
<td>0.001, 0.007</td>
<td>0.138</td>
</tr>
<tr>
<td>14</td>
<td>0.017</td>
<td>0.009, 0.027</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.007</td>
<td>0.002, 0.015</td>
<td></td>
</tr>
</tbody>
</table>

Examine posterior probabilities of this quantity as:

\[ \Delta(10-9) = (\text{cumulative probability of SUD in } n \text{ years for those whose onset is at age } 10) - (\text{cumulative probability of SUD in } n \text{ years for those whose onset is at age } 9). \]

The computed Deltas between any two onset age groups represent how different the cumulative probabilities of SUD are between those two age groups with the same ‘n’ number of years’ exposure after the onset. The posterior distributions of Deltas were simulated in WinBugs. Fig. 2 shows the box plots of the posterior distribution of Deltas computed between two adjacent ages, two (Fig. 2a) and four (Fig. 2b) years after the onset. It is one of the benefits of the Bayesian approach that inferences of such direct comparisons can easily be made with simulated posterior samples. It would have been very difficult to estimate the Deltas and their sampling distributions using a conventional analysis technique such as GEE. In 2 years after the onset (Fig. 2a), the age differences in risk of developing SUD were not large, but the risk of SUD was higher in those with onset at 13 than in those with onset at 10. However, SUD was less likely following onset at 14 than following onset at 13. This suggests that 13 was the onset age which exposed children to the greatest risk of SUD 2 years later.

Only three comparison groups were available to look at SUD 4 years after onset (Fig. 2b). In each case, the
3.3. Relationship between psychiatric disorder and substance use disorder (SUD)

We next tested the hypothesis that, while use is a necessary antecedent of SUD, it is unlikely to lead to SUD except when comorbid with early conduct problems. DSM-IV conduct disorder was the criterion used for conduct problems. We also examined the effect of other psychiatric disorders on risk for SUD in early users, controlling for possible comorbidity among psychiatric disorders and prior SUD. We used hierarchical Bayesian structure in a multiple logistic regression (see Appendix A). The log odds ratio of having SUD at a specific time point was regressed on lifetime history of CD, ANX, DEP and ODD up to the particular time point, as well as prior SUD and age. To simplify the presentation, the effect of ADHD was excluded, as ADHD was very rare in adolescence and the effect of early ADHD on later SUD was non-significant after controlling for comorbidity.

Fig. 3 shows estimated rates of SUD in substance users with and without psychiatric comorbidity. The line with diamond shaped markers shows the estimated probability of SUD at various ages for prior substance users with no psychiatric disorder. The risk of SUD sharply increased after age 14 even in this group, reaching 11% by age 16. Children with a history of ODD showed no increased risk of moving from use to SUD, but each of the other diagnoses was associated with some increase in risk. CD was associated with the greatest increase in risk, which also began earlier than that associated with any other disorder. Users with depression showed a smaller increased risk of SUD, beginning at age 13, while anxiety disorders showed no effect before age 16.

Table 4 displays the posterior estimates of the odds ratios and 95% credible intervals associated with each diagnosis, after controlling for comorbidity among psychiatric disorders and effects of prior SUD and age. The strongest effect was for earlier CD. This effect was strongest at age 14 (OR: 2.0, 95% CI: 1.2, 3.6) and 15 (OR: 2.1, 95% CI: 1.4, 3.2), but at 16 it was weaker (OR: 1.7, 95% CI: 1.1, 2.4), though still statistically significant. Interestingly, anxiety showed little association with current SUD through age 15 but showed a significant association (OR: 1.6, 95% CI: 1.1, 2.5) at age 16. ODD showed weak associations with SUD at any age.

3.3.1. Sex differences in the effects of psychiatric disorder on risk for SUD by 16

To assess the gender differences in the effects of risk factors on SUD, SUD at current time was regressed on main
Table 5

<table>
<thead>
<tr>
<th>Age 13</th>
<th>Age 14</th>
<th>Age 15</th>
<th>Age 16</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR</td>
<td>95% CI</td>
<td>OR</td>
<td>95% CI</td>
</tr>
<tr>
<td>Sex</td>
<td>0.7</td>
<td>0.2, 2.2</td>
<td>0.8</td>
</tr>
<tr>
<td>CD x sex</td>
<td>0.3</td>
<td>0.1, 1.1</td>
<td>0.4*</td>
</tr>
<tr>
<td>DEP x sex</td>
<td>27.4*</td>
<td>1.3, 705.6</td>
<td>0.4</td>
</tr>
<tr>
<td>ANX x sex</td>
<td>1.9</td>
<td>0.2, 23.1</td>
<td>1.6</td>
</tr>
<tr>
<td>ODD x sex</td>
<td>1.1</td>
<td>0.3, 4.3</td>
<td>0.8</td>
</tr>
</tbody>
</table>

* Significant using 95% credible interval.

Table 6

<table>
<thead>
<tr>
<th>Age 12</th>
<th>Age 13</th>
<th>Age 14</th>
<th>Age 15</th>
<th>Age 16</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR</td>
<td>95% CI</td>
<td>OR</td>
<td>95% CI</td>
<td>OR</td>
</tr>
<tr>
<td>CD onset at 15</td>
<td>15.99*</td>
<td>6.17, 45.88</td>
<td>14</td>
<td>14.51*</td>
</tr>
<tr>
<td>CD onset at 14</td>
<td>1.17</td>
<td>0.41, 3.43</td>
<td>13</td>
<td>2.00</td>
</tr>
<tr>
<td>CD onset at 13</td>
<td>2.00</td>
<td>0.72, 5.69</td>
<td>12</td>
<td>2.00</td>
</tr>
<tr>
<td>CD onset at 12</td>
<td>2.00</td>
<td>0.72, 5.69</td>
<td>11</td>
<td>2.00</td>
</tr>
<tr>
<td>CD onset at 11</td>
<td>2.00</td>
<td>0.72, 5.69</td>
<td>10</td>
<td>2.00</td>
</tr>
</tbody>
</table>

* Significant using 95% credible interval.

4. Discussion

4.1. Conclusions

The goal of these analyses was to test some of the ideas about the development of substance use disorders that scholars have derived mainly from the scrutiny of studies using retrospective recall of ages at onset of both substance use and other psychiatric disorders. We adopted a hierarchical Bayesian dynamic modeling approach to assess the precise developmental impact of risk factors on SUD, showing that the effects evolved over time. For example, if a static regression such as GEE had been used, inferences in Table 4 would have shown an averaged effect/trend of risk factors of SUD over time. The advantage of using such a dynamic modeling framework is very well supported by the better model fit to the actual data than a static model. As illustrated in Fig. 4, the static model showed an averaged constant effect of CD on SUD, but the dynamic model showed age specific effects of CD at a given age level.

The first set of analyses addressed the relationship between age at first use and length of time since first use. The belief that early substance use is a risk factor for later...
abuse and dependence has abundant empirical support (e.g., Anthony and Petronis, 1995; Breslau et al., 1993; Brook et al., 2000; Ferguson et al., 1996; Grant and Dawson, 1998; Hanna and Grant, 1999; Schinke et al., 2000). Robins et al. took the next step, arguing from the ECA data that the longer first use was delayed, the lower became the risk of developing a substance use disorder (Robins and Prybylko, 1985; Robins and Prybylko, 1993). This idea has driven prevention research and public policy around the world (e.g., Anthony and Petronis, 1995; Breslau et al., 1993; Brook et al., 2000; De Bernardo et al., 2002; Grant and Harford, 1995; Rohde et al., 1996; Stallings et al., 1997; Swendsen et al., 1998), but evidence for depression as abuse/dependence and depression (e.g., Dansky et al., 2000; De Bernardo et al., 2002; Grant and Harford, 1995; Rohde et al., 1996; Stallings et al., 1997; Swendsen et al., 1998), and have noted an increased risk of depression following ODD and McEvoy, 1990; the present study used data covering only a portion of the time-span available in the ECA data, but has not unequivocal. Our findings showed a nonlinear relationship between age at onset and risk of SUD; for onsets before 13, risk increased with age at onset, but for onset at 14-risk, controlling for period of exposure, began to fall. Because the subjects were only 16 at the last observation we cannot extend the analysis beyond 14 while holding period of risk constant. Thus, our results do not give unequivocal support either to Robins’s conclusion or to Anthony and Petronis’s analysis of the ECA data. However, they may apply only to the particular developmental periods of childhood and adolescence. It is likely that for very young experimenters, the social environment simply does not offer the necessary “affordances” in the way of ready access to illegal substances to permit the development of abuse and dependence.

Challange to the argument that age at first use is a direct risk factor for later abuse also comes from a study of twins (Prescott and Kendler, 2001). Prescott and Kendler argue that the association between early drinking and alcohol dependence is “due primarily to familial characteristics that probably reflect both shared environmental and genetic factors”; that is, early drinking and alcohol dependence are “both manifestations of familial vulnerability to problematic alcohol use” (Prescott and Kendler, 2001). Reflecting other studies, Prescott and Kendler’s sample of adult twins (18–62) reported a fairly linear relationship between age at first drink and lifetime prevalence of alcohol dependence (Prescott and Kendler, 1999). However, they argue that their analysis “provides little support for the ‘mediational’ hypothesis—that early drinking mediates the risk for alcoholism” (Prescott and Kendler, 1999, p. 105). The reason is that the unaffected co-twins of both MZ and DZ twins with alcohol abuse or dependence began drinking earlier than did the co-twins of unaffected twins. Our second goal was to test Robins’s prediction that abuse would be “extremely rare for those free of (early) conduct problems, no matter how early substance use began” (Robins, 1990, p. 203) implies the former, but is not unequivocal.

Anthony and Petronis’s elegant analysis of the retrospective ECA data supported the primacy of age over length of exposure (Anthony and Petronis, 1995). They found that “In the first year after initial illicit drug use, there was no appreciable difference (by age) in the estimated risk of developing a drug problem. However, with the passage of some 5–7 years after initial use, the estimated risk of developing drug problems became clearly greater for youths whose drug taking had started during the pre-adolescent and very early adolescent years. The previously reported relationship between early-onset drug use and subsequent drug problems was not, therefore, an artifact of time” (Anthony and Petronis, 1995, p. 9). The present study used data covering only a portion of the time-span available in the ECA data, but has the advantage of annual assessments and a prospective design. Our findings showed a nonlinear relationship between age at onset and risk of SUD; for onsets before 13, risk increased with age at onset, but for onset at 14-risk, controlling for period of exposure, began to fall. Because the subjects were only 16 at the last observation we cannot extend the analysis beyond 14 while holding period of risk constant. Thus, our results do not give unequivocal support either to Robins’s conclusion or to Anthony and Petronis’s analysis of the ECA data. However, they may apply only to the particular developmental periods of childhood and adolescence. It is likely that for very young experimenters, the social environment simply does not offer the necessary “affordances” in the way of ready access to illegal substances to permit the development of abuse and dependence.

Challange to the argument that age at first use is a direct risk factor for later abuse also comes from a study of twins (Prescott and Kendler, 2001). Prescott and Kendler argue that the association between early drinking and alcohol dependence is “due primarily to familial characteristics that probably reflect both shared environmental and genetic factors”; that is, early drinking and alcohol dependence are “both manifestations of familial vulnerability to problematic alcohol use” (Prescott and Kendler, 2001). Reflecting other studies, Prescott and Kendler’s sample of adult twins (18–62) reported a fairly linear relationship between age at first drink and lifetime prevalence of alcohol dependence (Prescott and Kendler, 1999). However, they argue that their analysis “provides little support for the ‘mediational’ hypothesis—that early drinking mediates the risk for alcoholism” (Prescott and Kendler, 1999, p. 105). The reason is that the unaffected co-twins of both MZ and DZ twins with alcohol abuse or dependence began drinking earlier than did the co-twins of unaffected twins. Our second goal was to test Robins’s prediction that abuse would be “extremely rare for those free of (early) conduct problems, no matter how early substance use began” (Robins, 1990, p. 203) implies the former, but is not unequivocal.

Given the debate about the developmental relationship between ODD and CD, it is notable that ODD was associated with absolutely no increase in risk for SUD, at any age or in either sex. Previous studies have not examined the separate impact of CD and ODD, controlling for comorbidity, so this finding needs replication. The symptoms of ODD focus on temperament (annoying, spiteful, touchy) rather than specific behaviors. The implication is that the risk for SUD, once children have started using substances, has more to do with the kinds of behaviors that they become involved in than the kind of people they are. We found no main effects of emotional disorders once comorbidity had been taken into account. Others have found cross-sectional associations between substance abuse/dependence and depression (e.g., Dansky et al., 2000; De Bernardo et al., 2002; Grant and Harford, 1995; Rohde et al., 1996; Stallings et al., 1997; Swendsen et al., 1998), and have noted an increased risk of depression following SUD (Rohde et al., 2001), but evidence for depression as
Munich longitudinal study of adolescents and young adults as often as depression preceded AUD (Kessler et al., 1997). National Comorbidity study data showed that in adults, alcohol abuse/dependence (AUD) preceded depression about 296


These interactions of sex and comorbidity are examples of the "gender paradox" (Keenan et al., 1999), the thesis that with in a separate paper.

4.2. Limitations

This paper was set out to address the thesis propounded by Robins et al., that the transition from early drug use to adolescent and early-adult drug abuse is rare in those who were not also conduct disordered. There are, however, several differences between our data and the ECA data used by these researchers, apart from age and the number of data waves. Robins focused on drugs other than tobacco or alcohol, while for most of our analyses we have not differentiated among substances, in order to maximize power. In fact cigarettes, alcohol, cannabis, and other drugs were each associated with a 4–20-fold increase in risk of transitioning to SUD by age 16.

The data used in this study differ from those in many other analyses in being right censored at age 16, so we cannot yet say what will happen in adulthood. The International Consortium of Psychiatric Epidemiology (ICPE) collected information on drug use and its onset at seven sites around the world, from adults aged 14 and older (Vega et al., 2002). In the cases they examined (alcohol, cannabis, other drugs) the hazard rate increased across adolescence and young adulthood, peaking at age 18 (alcohol, other drugs) to 20 (cannabis) in the United States, and slightly earlier in Canada. Thus, in these large representative samples onset of use continued to escalate into the 20s, with no sign of a check to the acceleration of onsets that we saw for SUD in this study. However, the ICPE studies were looking at use, not SUD.

DeWit et al. looked more closely at one of the ICPE samples, the Ontario Mental Health Supplement sample of people aged 15 and older, to examine the question of age at first use of alcohol as a predictor of later alcohol abuse and dependence (DeWit et al., 2000). The median ages reported were 15.2 for first drink of alcohol, 17.7 for abuse, and 20.4 for dependence. If we consider the period up to 6 years after first use, which is similar to the period covered in the present study, the group who started drinking at age 13–14 was the one with the fastest acceleration and the highest rate of abuse. In the case of alcohol dependence, it was the group who started drinking at age 11–12 who had the highest rate, followed by those who started drinking at age 13–14. After 20 years, however, the participants who reported the youngest age at first drinking had the highest levels of abuse and dependence, and those who started after age 19, the lowest.

Thus, it is possible that our narrower time-frame has captured the same phenomenon that DeWit et al. observed from retrospective data; i.e., a higher probability of escalation to SUD in those who started using alcohol at age 13–14 (around the time of entry to high school). Later data waves will show whether, as in the Ontario study, the risk to the group who started drug use earlier will accelerate in the late teens and twenties.

Acknowledgements

This research was supported by grants from the National Institute of Drug Abuse (1R01DA11301-01) and the William T. Grant Foundation, and a National Institutes of Health Independent Scientist Award to the last author.

Appendix A

In this Appendix A, we present a brief review of the Dynamic Generalized Linear model (DGLM) framework (West et al., 1985), and describe a detailed modeling structure that is used in data analyses of the present study. In developmental epidemiological research using longitudinal data, the nature of the relationship among variables is highly likely to vary over time; i.e., they are dynamic. The key idea of dynamic models is to relate the model parameters to changes in developmental processes of psychopathology, due to the passage of time. Linear dynamic models consist of an observation equation for a response variable \( Y_{mt} \) in regression form

\[
Y_{mt} = X_{mt} \beta_1 + \epsilon_{mt}, \epsilon_{mt} \sim N(0, \Sigma_t)
\]
where “∼” means “distributed as”, and the system equation that describes the parameters evolving over time, according to a first order Markov dependence process, given by

\[ \beta_t = \beta_{t-1} + w_t, w_t \sim N(0, W_t), \]

where \( X_m \) is a 1 \times p regression components vector, \( \beta_t = (\beta_{t1}, \ldots, \beta_{tp}) \) a p \times 1 fixed-effects regression vector which consists of time varying components. The regression parameter vector, \( \beta_t \), indexed by \( t \) indicates time specific effects, and the system equation reflects the learning process that uses the knowledge from the past in revising the state of knowledge about the current parameter. In the Bayesian paradigm all uncertainty about the unknown quantities are described by probabilities, and inferences about them are made with posterior probabilities, which are computed using Markov Chain Monte Carlo (MCMC) simulation approach due to the computational difficulties in the present study.

We used a modified DGLM structure in the following logistic regression to assess the age specific effect of psychopathology on the probability of SUD:

\[
\logit(\pi_{mt}) = \alpha + \beta_0 SUD_{m,t-1} + X_m \beta_1, \tag{A.1}
\]

where \( \pi_{mt} \) represents the probability of having SUD at time \( t \) for \( m \)th subject, \( SUD_{m,t-1} \) indicates whether an individual had SUD at time \( t-1 \), \( X_m \) is a 1 \times p vector of regressors, \( \beta_0 \) is the associated intercept and \( \beta_1 \) is a 1 \times p fixed-effects regression vector which consists of time varying components. (Note that all the parameters are actually random-effects in a Bayesian analysis.) Thus, \( \alpha \) is the overall mean effect of SUD, \( \beta_0 \) is the association between the current and prior values of the response variable in log-odds-ratio scale, and \( \beta_1 \) represents the time varying effects of other covariates on current SUD. For example, to assess the effects of prior psychiatric disorders on transitions from substance users to SUD at time \( t \), we considered

\[
\logit(\pi_{mt}) = \alpha + \beta_0 SUD_{m,t-1} + \beta_1 ADEW + \beta_2 SUB_{m,t-1} + \beta_3 CD_{m,t-1} + \beta_4 ANX_{m,t-1} + \beta_5 DEP_{m,t-1} + \beta_6 ODD_{m,t-1}, \tag{A.2}
\]

where \( \beta_j \) captures age specific deviations from the baseline, \( SUB_{m,t-1}, \ldots, ODD_{m,t-1} \) indicate past substance use and history of psychiatric disorders (conduct, anxiety, depression, and oppositional defiant disorders respectively) by age \( t-1 \) for subject \( m \). The predictor \( SUD_{m,t-1} \) indicates whether one had SUD at the prior time point. Thus, there may be both onsets and offsets over time.

Fig. 5 is the graphical presentation of the hierarchical Bayesian model with the first order Markov structure on a time varying parameter over different age levels. This illustrates the model structure, and how parameters at a given time point are related to those at other time points. In this figure, a plate represents repeated components for the range, for example (in \( M \)). The arrows show the specific relationships between two nodes. The descriptions of nodes and arrows are as follows:

- Constant node \( \bullet \) describes a quantity fixed by the study design.
- Stochastic node \( \bigcirc \) describes a variable that is given a distribution.
- Deterministic node describes all logical functions of other nodes.
- An arrow ‘’’ connects two nodes by logical functions.
- An arrow ‘’’’’’’ connects two nodes by logical functions.

Fig. 5 illustrates that the probability of having SUD for an individual \( m \) at time \( t \) depends on covariates and whether one had a SUD at the prior time point or not. The covariate and prior SUD effects on the probability of current SUD are represented by regression parameters \( \beta_j \)’s at time \( t \), which are dynamic and evolve over time. In Fig. 5, define \( SUD_{mt} \) as a binary variable representing SUD in subject \( m \) at time \( t \). Thus \( SUD_{mt} \) in a Bernoulli random variable with probability \( \pi_{mt} \) and logit transformation is made on \( \pi_{mt} \). The Bayesian hierarchical modeling framework is as follows.

First Level:

\[
SUD_{mt}(\pi_{mt}) \sim \text{Bernoulli}(\pi_{mt}). \tag{A.2}
\]

for \( m = 1, \ldots, M \), \( t = 1, \ldots, T \).

Second Level:

\[
\logit(\pi_{mt}) = \alpha + \beta_0 SUD_{m,t-1} + X_m \beta_1, \tag{A.3}
\]

where the components are described as in (A.1).

An independent normal distribution is assumed as the prior distribution for \( \alpha \) and \( \beta_1 \)’s as

\[
(\alpha | \mu_\alpha, \tau_\alpha) \sim \text{normal}(\mu_\alpha, \tau_\alpha) \tag{A.3}
\]
\((\beta_j,t,\beta_j,t), (\tau_j), a\alpha, b\alpha) \sim \text{normal}(\beta_j,0), (\tau_j), a\alpha, b\alpha) \sim \text{normal}(0,0), \text{if } t = 1 \quad (A.4)

for \(j = 0,\ldots,p\) with specified mean \(\mu\) and unknown precisions \(\tau_j\) which are inverse of variances, i.e., \(1/\sigma_j^2\).

Third Level:
The unknown precisions (i.e., inverse of variance) in (A.3) and (A.4) are assumed to follow Gamma distribution as

\((\tau_j,a\alpha, b\alpha) \sim \text{Gamma}(a\alpha,b\alpha), (\beta_j,t) \sim \text{normal}(\beta_j,t), (\tau_j) \sim \text{normal}(0,0)\), if \(t = 1\)

References


