Psychological Medicine

cambridge.org/psm

Prediction of drug abuse recurrence: a Swedish National Study

K. S. Kendler[1,2,3], H. Ohlsson[4], K. Sundquist[4] and J. Sundquist[4]

[1]Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA; [2]Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA; [3]Department of Human and Molecular Genetics, Virginia Commonwealth University, Richmond, VA, USA and [4]Center for Primary Health Care Research, Lund University, Malmö, Sweden

Original Article

Cite this article: Kendler KS, Ohlsson H, Sundquist K, Sundquist J (2017). Prediction of drug abuse recurrence: a Swedish National Study. *Psychological Medicine* **48**, 1367–1374. https://doi.org/10.1017/S0033291717002938

Received: 22 February 2017
Revised: 1 September 2017
Accepted: 7 September 2017
First published online: 10 October 2017

Key words:
Co-relative analysis; drug abuse; prognosis; recurrence; relapse

Author for correspondence:
K. S. Kendler, E-mail: kenneth.kendler@vcuhealth.org

Abstract

Background. Relapse from drug abuse (DA) is common, but has rarely been studied in general population samples using a wide range of objective predictors.

Method. Using nationwide registries, we ascertained 44 523 subjects first registered for DA between the ages of 15 and 40 in 1998 to 2004 and followed for 8 years. We predicted relapse in subjects defined as a second DA registration. We also predicted DA relapse in relative pairs concordant for DA but discordant for relapse.

Results. In multivariate regression analyses, the strongest predictors for relapse were prior criminal behavior, male sex, being on social welfare, low school achievement, prior alcoholism, and a high-risk father. A risk index trained from these analyses on random split-halves demonstrated a risk ratio of 1.11 [95% confidence intervals (CIs) 1.10–1.11] per decile and an ROC value of 0.70 (0.69–0.71). Co-relative analyses indicated that a modest proportion of this association was causal, with the remainder arising from familial confounders. A developmental structural equation model revealed a complex interviewing of risk pathways to DA with three key mediational hubs: low educational attainment, early age at first registration, and being on social welfare.

Conclusions. In a general population sample, using objective registry information, DA relapse is substantially predictable. However, the identified risk factors may not be valid targets for interventions because many index familial risk and may not impact causally on probability of relapse. Risk for DA relapse may reflect an inter-weaving, over developmental time, of genetic–temperamental vulnerability, indices of externalizing behaviors and social factors reflecting deprivation.

After initial onset, the course of substance use disorders can be highly variable ranging from a life-long remission to many repeated episodes of recurrence and stable addictions lasting decades (Brownell *et al.* 1986; Hser *et al.* 2007*b*; Calabria *et al.* 2010; Sarvet & Hasin, 2016). However, most population-based studies on the course of substance use disorders have examined alcohol dependence (Sarvet & Hasin, 2016) and there remains considerable uncertainty about the nature and the predictors of the course of illicit drug abuse (DA) (Calabria *et al.* 2010). A wide range of predictors of a poor course for DA have been proposed, including male sex, childhood adversity, prior attention-deficit/hyperactivity disorder (ADHD), other substance use disorders, prior criminality, family history of substance problems, low educational attainment, comorbid mood disorders, low social support, recent stressors, low levels of social attachment, low SES, and an early age at onset (McLellan *et al.* 1983; Biederman *et al.* 1998; White *et al.* 2004; Hser *et al.* 2007*a*, *b*; Lopez-Quintero *et al.* 2011; van der Pol *et al.* 2015; Sarvet & Hasin, 2016). However, many of these prior studies had methodological limitations including modest sample sizes, short-term follow-ups, unrepresentative sampling (e.g. attenders at a single treatment facility), examination of a limited set of predictors, and lack of attention paid to the problems of causal inference and the delineation of mediational pathways of risk.

In this report, we seek to address several of these limitations. We examine a national sample of first-time drug abusers in Sweden ($n = 44\,523$) identified from medical and criminal registries and follow their course in these registers for 8 years. Our analyses have the following aims:

1. Develop univariate and multivariate models predicting DA relapse from a diverse set of risk factors obtained from various Swedish Registries.
2. From the multivariate analyses, develop a risk prediction model trained in one random half of the sample and then test its performance in the second half.
3. To gain insight into the causal nature of our risk prediction score, examine the model performance within pairs of relatives concordant for DA but discordant for relapse.
4. Develop a structural equation model of relapse prediction to illuminate mediational processes and developmental pathways to relapse.

CAMBRIDGE UNIVERSITY PRESS

Methods

We used Swedish population-based registers with national coverage, linking them using each person's unique identification number which, to preserve confidentiality, was replaced by a serial number. We secured ethical approval for this study from the Regional Ethical Review Board of Lund University (No. 2008/409). From the population-based registers, we selected all individuals who had their first lifetime registration for DA during the period 01-01-1998 to 12-31-2004. Furthermore, we required that individuals were between 15 and 40 years of age at the time of this registration and that their father and mother were included in the Swedish Multigenerational Register. DA relapse was defined as a new DA registration within the ensuing 8 years. Registrations within 180 days from the initial registration, which could have resulted from the same criminal events or medical disorders associated with initial registration, were not considered a relapse and such data were censored from all analyses. The outcome variable DA was identified in the Swedish medical registries by ICD codes [ICD9: Drug psychoses (292) and Drug dependence (304), Non-dependent abuse of drugs (305; excluding 305.0); ICD10: Mental and behavioral disorders due to psychoactive substance use (F10–F19), except those due to alcohol (F10) or tobacco (F17)]; in the Suspicion Register by codes 3070, 5010, 5011, and 5012, that reflect crimes related to DA; and in the Crime Register by references to laws covering narcotics (law 1968:64, paragraph 1, point 6) and drug-related driving offences (law 1951:649, paragraph 4, subsection 2 and paragraph 4A, subsection 2). A definition of all our predictor variables can be found in online Supplementary Table A1.

Of the 44 523 individuals 26 072 had no missing values among the included predictors; 11 984 had below 5% and 6467 had more than 5% missingness. To impute values in the regression models, we used the Predicted Regression imputation method within specific groups of questions: that is using regression models to predict missing values based on similar covariates. As the outcome (DA relapse) in our population is high (58%), we report risk ratios (RR) instead of odds ratios (OR). When events are common, the RR and OR deviate to a higher extent compared to when the outcome is rare, and the OR overestimates the more easily interpretable RR (Altman *et al.* 1998). We therefore used a modified Poisson approach to estimate the RR and corresponding 95% confidence interval (CI) (Zou, 2004).

We first developed univariate and multivariate models to predict relapse. To assess the relative importance of individual predictors in the models, we used the t-statistic for each model parameter using the Caret Package in R (R Development Core Team, 2014). We thereafter fitted the multivariate regression model to a random half as a training-sample. Results from that model were then applied to the second random half and were evaluated using an ROC model. We created a risk-score based on the deciles of the predicted probabilities and used it as a predictor variable in the model.

Next, we sought to assess the degree to which the association between the risk-score and relapse reflect confounding by familial risk factors (genetic and/or environmental) using a co-relative design. From the Swedish Multi-Generation Register we identified all first-cousin and full-sibling pairs. Using a stratified model with a separate stratum for each cousin and sibling pair, we refitted the models. Within each stratum, the RR for the risk score was adjusted for unmeasured shared genetic and environmental factors among siblings. The regression models were performed using SAS 9.3 (SAS Institute, 2011). We then combined the population, full-sibling, and cousin datasets, and performed two co-relative analyses. The first allowed all coefficients for each sample to be independent. In the second, we modeled the genetic resemblance assuming that it equaled 0 for the population, +0.125 for cousins and +0.5 for full-siblings. We compared this model, using the AIC (Akaike, 1987), with the previous model. If the second model fitted the data well, we obtained improved estimation of the risk index-DA relapse association among all types of relatives. In this model we were also able to extrapolate an OR for monozygotic twins (there were two monozygotic twin registered who were concordant for DA but discordant for relapse).

For the structural equation model, we organized the predictor variables into 5 tiers that reflected roughly five developmental periods: (1) birth and familial factors (Father High Risk Behavior, Mother High Risk Behavior, Low parental education, and sex); (2) childhood and early adolescence (Full-sibling Drug Abuse, Half Sibling Drug Abuse, ADHD); (3) late adolescence [School Achievement, Criminal Behavior (CB), Psychiatric Illness, Alcohol Use Disorder]; (4) early adulthood (Low educational attainment, Unemployed, Marital Status, On Social Welfare); (5) at initial DA registration (Age at Registration, Source of Registration). The structural equation model of relapse prediction consisted of path and correlation coefficients connecting the 18 observed variables in the model. We began with a fully saturated model and used a combination of three approaches to produce a model with an optimal balance of explanatory power and parsimony. DA relapse was treated as a dichotomous variable with an assumed underlying normal liability distribution. Note also that variables in the first tier are interconnected by correlations, depicted by two-headed arrows in the figures, rather than partial regression coefficients, depicted by one-headed arrows. In the first step, observing the significance levels of individual paths, we fixed sets of paths to zero when the associated z value was <1.96. Second, some paths remained significant that were too small to be meaningful. Therefore, the second step was to set all paths to zero with a value of <0.05, regardless of z value. Third, we added and subtracted a number of paths that were marginal by significance and/or magnitude to see if we could arrive at a better overall fit and, indeed, produced a modest improvement in fit and explanatory power. We utilized three fit indices that reflect the success of the model in balancing explanatory power and parsimony: the Tucker–Lewis index (TLI), the comparative fit index (CFI) and the root-mean-square error of approximation (RMSEA). For the TLI and the CFI, values between 0.90 and 0.95 are considered acceptable and values $\geqslant$0.95 as good. For the RMSEA, good models have values $\leqslant$0.05. The fit function was weighted least squares. The structural model was fitted in Mplus, version 7.31 (Muthén & Muthén, 2015). Mplus uses all data that is available to estimate the model using full information maximum likelihood.

Results

Sample

We identified 44 523 subjects first registered for DA between the ages of 15 and 40 from 1998 to 2004. We examined the registry data for all these subjects to determine if they had a relapse in the 8 years following their initial registration. Their mean (S.D.) age of first registration was 24.7 (6.8) years and the 15th, 50th

https://doi.org/10.1017/S0033291717002938 Published online by Cambridge University Press

and 75th percentiles equaled 19, 23 and 29, respectively. Over the course of follow-up, 58.0% met our definition of relapse. Of those who relapsed, 44.5 and 75.5% did so within 1 and 3 years, respectively.

As seen in Table 1, 73.4% of the samples were male. At the start of follow-up, 15.3% had a psychiatric diagnosis, 54.3% a prior criminal registration, 11.7% a prior registration for alcohol use disorder (AUD), 95.7% were unmarried, 38.7% unemployed, and 39.0% on social welfare.

Regression analyses

Examining first the univariate regression analyses, and using the multivariate *t*-statistic as an effect size estimate, the six strongest predictors of relapse were, in order, prior CB, low school achievement, being on social welfare, being male, having had a low education, and having a high-risk father (Table 1). Given the expected high inter-correlations among a number of the risk factors, effect sizes of many of them declined substantially in the multivariate analysis. However, the order was relatively similar with the top six predictors being prior CB, male sex, social welfare, low school achievement, prior alcohol use disorder, and having a high-risk father.

Aggregate risk prediction

We fitted our multivariate regression model to a random half of our cohort as a training-sample. Results from that model were then applied to the second random half as the test sample. We divided the test sample into 10 risk groups and found the RR for relapse per decile was 1.11 (1.10–1.11).

Figure 1 and online Supplementary Table A2 display the RRs for these deciles compared with the lowest risk group. Those in the 9th and 10th deciles of risk had, in our test sample, relapse rates of 77.8 and 82.0%, with RRs of, respectively, 2.28 (2.21–2.35) and 2.52 (2.44–2.62). A formal ROC analysis (online Supplementary Fig. A1) showed an area under the curve estimate of 0.70 (0.69–0.71).

To gain insight into the causal relationship of our risk score and rates of relapse, we examined our risk prediction model

Table 1. Univariate and multivariate regression analysis of relapse of drug abuse from twenty-five putative risk factors

	%	Missing	T-statistics (univariate)	T-statistics (multivariate)	Risk ratio* (univariate)	Risk ratio* (multivariate)
Male *v.* Female	73.4% (M)		34.49	23.37	**1.41 (1.38; 1.45)**	**1.29 (1.26; 1.32)**
High Risk Father	44.1%		24.40	9.25	**1.22 (1.20; 1.24)**	**1.08 (1.06; 1.10)**
High Risk Mother	23.7%		21.23	8.16	**1.21 (1.19; 1.23)**	**1.07 (1.05; 1.09)**
Low Parental Education	–	271	19.31	0.47	**1.08 (1.08; 1.09)**	1.00 (0.99; 1.01)
Psychiatric Illness	15.3%		3.52	5.11	**1.04 (1.02; 1.06)**	**1.06 (1.04; 1.08)**
ADHD	0.5%		3.82	1.57	**1.22 (1.12; 1.33)**	1.07 (0.99; 1.17)
Low School Achievement	–	15 144	38.08	13.76	**1.20 (1.19; 1.21)**	**1.06 (1.05; 1.07)**
Criminal Behavior	54.3%		53.14	28.52	**1.57 (1.54; 1.60)**	**1.32 (1.29; 1.34)**
Alcohol Use Disorder	11.7%		22.55	11.29	**1.30 (1.27; 1.32)**	**1.12 (1.10; 1.14)**
Not Married *v.* Married	95.7%	2496	1.62	1.44	1.02 (0.99; 1.05)	**1.05 (1.00; 1.09)**
Spouse with *v.* without DA	0.5%	2496	4.46	1.64	**1.11 (1.00; 1.23)**	**1.12 (1.10; 1.14)**
Low Education	–	2659	30.67	8.42	**1.17 (1.16; 1.19)**	**1.05 (1.04; 1.06)**
Unemployed	38.7%	2659	17.08	2.07	**1.16 (1.14; 1.18)**	**1.02 (1.00; 1.04)**
On Social Welfare	39.0%	2659	36.55	15.35	**1.36 (1.34; 1.38)**	**1.14 (1.12; 1.15)**
Full-Sibling Drug Abuse	7.8%		12.72	5.90	**1.20 (1.17; 1.23)**	**1.08 (1.05; 1.10)**
Age at Registration	–		0.43	6.63	**1.00 (1.00; 1.00)**	**1.00 (1.00; 1.01)**
First registration medical *v.* criminal	29.1% (M)		15.09	0.33	0.87 (0.85; 0.89)	0.99 (0.97; 1.01)
Half-Sibling Drug Abuse	4.1%		9.72	4.02	**1.20 (1.16; 1.24)**	**1.07 (1.04; 1.10)**
Full-Cousin Drug Abuse	7.4%		4.61	0.40	**1.07 (1.04; 1.10)**	1.01 (0.98; 1.03)
Neighborhood SES	–	1631	17.32	0.05	**1.04 (1.03; 1.04)**	1.00 (0.99; 1.00)
% DA in Neighborhood	–	1631	15.45	3.06	**1.06 (1.06; 1.07)**	**1.02 (1.01; 1.02)**
Parental Death	5.6%		4.33	0.91	**1.08 (1.04; 1.11)**	1.02 (0.99; 1.05)
Disposable Income	–	2659	24.22	7.93	**1.17 (1.15; 1.19)**	**1.06 (1.03; 1.07)**
Child – not living with	10.5%		7.62	1.44	**1.04 (1.02; 1.07)**	0.98 (0.95; 1.01)
Child – living with	12.4%		16.14	2.63	**1.27 (1.23; 1.31)**	**1.04 (1.01; 1.07)**

*Significant risk ratios ($p < 0.05$) are **bolded**.

https://doi.org/10.1017/S0033291717002938 Published online by Cambridge University Press

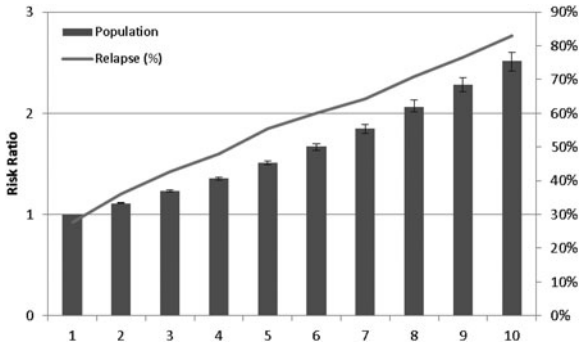

Fig. 1. The risk-ratio for Drug Abuse relapse as a function of the Decile of our Risk Index Created from our Multiple Regression Analyses. The figure also shows the predicted relapse rates by decile of risk.

within cousin and full-sibling pairs concordant for DA but discordant for relapse ($n = 621$ and 212, respectively) (Table 2). The effect sizes of our model were moderately attenuated to a RR per decile of 1.10 and 1.07 in the cousin and sibling pairs, respectively. We then fitted our co-relative model to these data along with a very small number of DA concordant but relapse discordant MZ twin pairs. This model fitted the data well (model AIC 259189.0 *v.* 259191.0 for the observed data) and predicted the RR per decile in discordant MZ pairs to equal 1.02.

Structural equation analysis

To examine mediational pathways, we developed a structural model predicting DA relapse. To avoid undue complexity, we trimmed several variables which contributed very modestly to the predictive effect of the model: Parental Death, Disposable Income, Neighborhood Socioeconomic status, Neighborhood Drug Abuse, Child, and DA in cousin. The trimmed model for the remaining 17 variables fitted the data well (CFI = 0.993, TLI = 0.984; RMSEA = 0.011) and explained 24.8% of the variance in liability to DA relapse.

Table 3 summarizes the results of our developmental SEM model. All the correlational and path coefficients are provided in online Supplementary Table A3. The variables with the highest aggregate effect on relapse were, in order, sex, CB, school achievement, high-risk mother, being on social welfare, prior AUD, and young age at first registration. Table 3 also shows the proportion of the total effect of each variable that was indirect. In interpreting these values, it is important to note that more distal (earlier) variables in SEM models have more opportunities for indirect effects, which decline as the variables get closer to the dependent variable – here relapse. Of the variables 'early' in the model, sex, high-risk mother, and DA in a full-sibling had substantial direct effects. In the middle of the model, school achievement, CB, and AUD also have substantial direct effects. Table 3 also summarizes the major mediating variables in the model. For most predictors early in the model, the three key mediators were low school achievement, CB and being on social welfare. The effect of many of the predictors later in the model were mediated through AUD, age at registration, and being on social welfare.

Table 2. The observed and predicted risk ratio for drug abuse relapse per decile of risk score in the general population, and cousins, full siblings and MZ twins concordant for drug abuse but discordant for relapse

Risk for drug abuse relapse per decile		
	Observed	Predicted
Population	1.11 (1.10; 1.11)	1.11 (1.10; 1.11)
Cousins	1.10 (1.07; 1.12)	1.10 (1.09; 1.11)
Full siblings	1.07 (1.04; 1.10)	1.06 (1.05; 1.08)
MZ twins	2 discordant pairs	1.02 (0.99; 1.05)
AIC	259190.98	259189.00

The results of our best-fit path model are depicted in Fig. 2, which is color coded. The color of the paths reflects the tier of the dependent variable, while the color underlining the path coefficient indexes the tier of the predictor variable. Such figures are rich visual summaries. We can only comment on a few prominent results. First, the three variables that originate the most paths are 'male *v.* female' with 10, school achievement and CB with 8, and ADHD with 7. Second, the variables receiving the most paths – which might be considered hubs in the causal network – are age at registration with 11, social welfare with 10 and low education with 6. Third, relapse – the ultimate dependent variable – received 9 paths from all five of the antecedent levels of our developmental model. The strongest direct paths to relapse are from male sex (+0.19), prior CB (+0.19), social welfare (+0.14), low school achievement (+0.11), and prior AUD (+0.11). Fourth, the model shows modest and indirect effects on relapse risk for two measures of low social integration: being unmarried and unemployed. Fifth, the figure illustrates the complex pathways from a history of externalizing behaviors in close relatives to relapse risk involving direct paths from high-risk mother and a drug-abusing sibling, and mediation through variables such as poor school achievement, prior criminality, low educational attainment, and early age at first registration. Finally, the figure well illustrates the close inter-relationship in the prediction of DA relapse between social factors reflecting deprivation (low parental education, poor school achievement, unemployment, being on social welfare), those reflecting genetic–temperamental factors (ADHD, psychiatric illness) and those which index externalizing behaviors (criminality and AUD).

Discussion

We investigated predictors of relapse over an 8-year period after a first DA registration in a large general population Swedish sample. We review our four aims in turn. Our first was to develop univariate and multivariate regression models including 25 diverse risk factors obtained from several Swedish Registries. These models were prospective; all risk factors were assessed prior to or at the time of the first DA registration. DA relapse was predicted by a diverse set of prospective risk factors. We could find support in the prior literature for a negative impact on the course of DA for many but not all of our risk factors: (i) externalizing behaviors in relatives (Hser *et al.* 2007*a*; van der Pol *et al.* 2015), (ii) male gender (Lopez-Quintero *et al.* 2011; Farmer *et al.* 2015), (iii) low SES home of origin (Hser *et al.* 2007*a*), (iv) poor educational attainment (Hser *et al.* 2007*a*), (v) childhood adversity (van der Pol *et al.* 2015), (vi) prior ADHD (Biederman *et al.* 1998; White *et al.* 2004) or other psychiatric diagnoses (McLellan *et al.* 1983; White *et al.* 2004; Hser *et al.* 2007*b*; Florez-Salamanca *et al.* 2013; Farmer *et al.* 2016), and (vii) prior externalizing psychopathology (crime and AUD)

https://doi.org/10.1017/S0033291717002938 Published online by Cambridge University Press

Table 3. The total and indirect effect of predictor variables from our structural equation model on prediction of relapse of drug abuse and the major mediating variables

Predictor variable	Total effect	Indirect effect	Major mediating variables
Male *v.* female	0.278 (0.263; 0.292)	0.087 (0.070; 0.105)	Criminal behavior, on social welfare
High risk father	0.073 (0.066; 0.080)	0.073 (0.066; 0.080)	School achievement, criminal behavior
High risk mother	0.152 (0.135; 0.168)	0.071 (0.064; 0.078)	On social welfare, criminal behavior
Low parental education	0.074 (0.068; 0.079)	0.074 (0.068; 0.079)	School achievement, criminal behavior
Full-sibling drug abuse	0.090 (0.068; 0.112)	0.032 (0.026; 0.038)	Criminal behavior, on social welfare
Half-sibling drug abuse	0.028 (0.019; 0.038)	0.028 (0.019; 0.038)	Criminal behavior, school achievement
ADHD	0.080 (0.067; 0.092)	0.080 (0.067; 0.092)	Age at registration, school achievement
Low school achievement	0.227 (0.213; 0.241)	0.100 (0.093; 0.108)	Criminal behavior, on social welfare
Criminal behavior	0.248 (0.229; 0.268)	0.055 (0.041; 0.068)	Alcohol use disorder, on social welfare
Psychiatric illness	0.039 (0.028; 0.049)	0.039 (0.028; 0.049)	Alcohol use disorder, age at registration
Alcohol use disorder	0.096 (0.073; 0.119)	−0.011 (−0.017; −0.005)	Age at registration, on social welfare
Marital status	0.039 (0.032; 0.046)	0.039 (0.032; 0.046)	Age at registration, on social welfare
Low education	0.024 (0.020; 0.028)	0.024 (0.020; 0.028)	On social welfare, age at registration
Unemployed	0.057 (0.049; 0.065)	0.057 (0.049; 0.065)	On social welfare
On social welfare	0.137 (0.118; 0.155)	0	
Age at registration	0.092 (0.077; 0.108)	0	
Source of registration	0.041 (0.020; 0.063)	0	

(Lopez-Quintero *et al.* 2011; Feingold *et al.* 2015; Sarvet & Hasin, 2016). DA relapse was also predicted by a range of risk factors assessed at the time of the initial DA registration including: (i) age at that first registration (Hser *et al.* 2007*a*), (ii) marital status (Hser *et al.* 2007*b*; van der Pol *et al.* 2015), history of DA in the spouse, and presence of children, (iii) employment status (Hser *et al.* 2007*a*; van der Pol *et al.* 2015), (iv) SES [disposable income and being on welfare (Hser *et al.* 2007*b*)], and (v) neighborhood environment (mean SES and rates of DA in the immediate neighborhood). Our findings are particularly consistent with the emphasis of Hser *et al.* that measures of social involvement and success (e.g. level of education, presence and quality of employment, presence and quality of social supports including marriage) substantially predicts risk of DA relapse (Hser *et al.* 2007*a*, *b*).

Our multivariate model is particularly useful for comparing the unique predictive power of our individual risk factors. Of note, the six strongest predictors represented diverse kinds of predisposing variables including prior externalizing behaviors (CB and AUD prior to first DA registration), familial factors (a high risk father), poor educational success (low school achievement), economic hardship (receiving social welfare), and sex (being male).

Our second aim was to develop, from our multiple regression results, a risk prediction model. To obtain a valid measure of its performance, we trained the model in a random half of the sample and then tested its performance in the second random half. The model's performance was relatively robust with an ROC area under the curve value of 0.70. Each decile of risk increased the probability of relapse by 11% with the observed relapse rate equaling 82% in the highest risk decile.

Such risk prediction models are empirical and can identify those at high probability of relapse without knowledge of the causal relationship between the risk index and the outcome. Therefore, in our third aim, using a co-relative design, we sought to clarify the degree to which our risk index was likely directly causally impacting on relapse rates *v.* indexing familial confounders. This is a relevant concern because of prior work both in Swedish and other twin and adoption samples that familial factors impact strongly on risk for DA (Tsuang *et al.* 1996; Kendler *et al.* 2003, 2012, 2015*b*). Examining the risk for relapse in pairs of cousins, full-siblings and MZ twins concordant for DA but discordant for relapse, we could estimate that our risk index was much less potent at predicting relapse in MZ twins than in the general population. This finding suggests that familial factors (which could be genetic and/or environmental) both predispose to many of our risk measures and to relapse in DA so that many of our risk factors and relapse are correlated but not strongly causally related.

These results illustrate the difference, in epidemiological follow-up studies, between the goals of prediction and informed intervention. Our ROC analyses suggest that our risk index would have potential utility in clinical settings for the prediction of individuals at high risk for relapse. However, our co-relative analyses suggest that caution should be indicated on planned interventions based on these analyses. While it would be tempting to recommend, on the basis of our findings, interventions to provide employment or improve educational attainment as a way to reduce risk for DA relapse, our results suggest that such approaches might be ineffective. The lack of efficacy would arise because these risk factors might index familial characteristics that in turn predict relapse risk rather than they themselves directly impacting on the probability of recurrence of the abuse.

Our fourth aim was to construct a developmental structural equation model to predict relapse so that we could illustrate mediational processes and clarify pathways to relapse. The resulting model indeed indicated the complexity of the mediational

https://doi.org/10.1017/S0033291717002938 Published online by Cambridge University Press

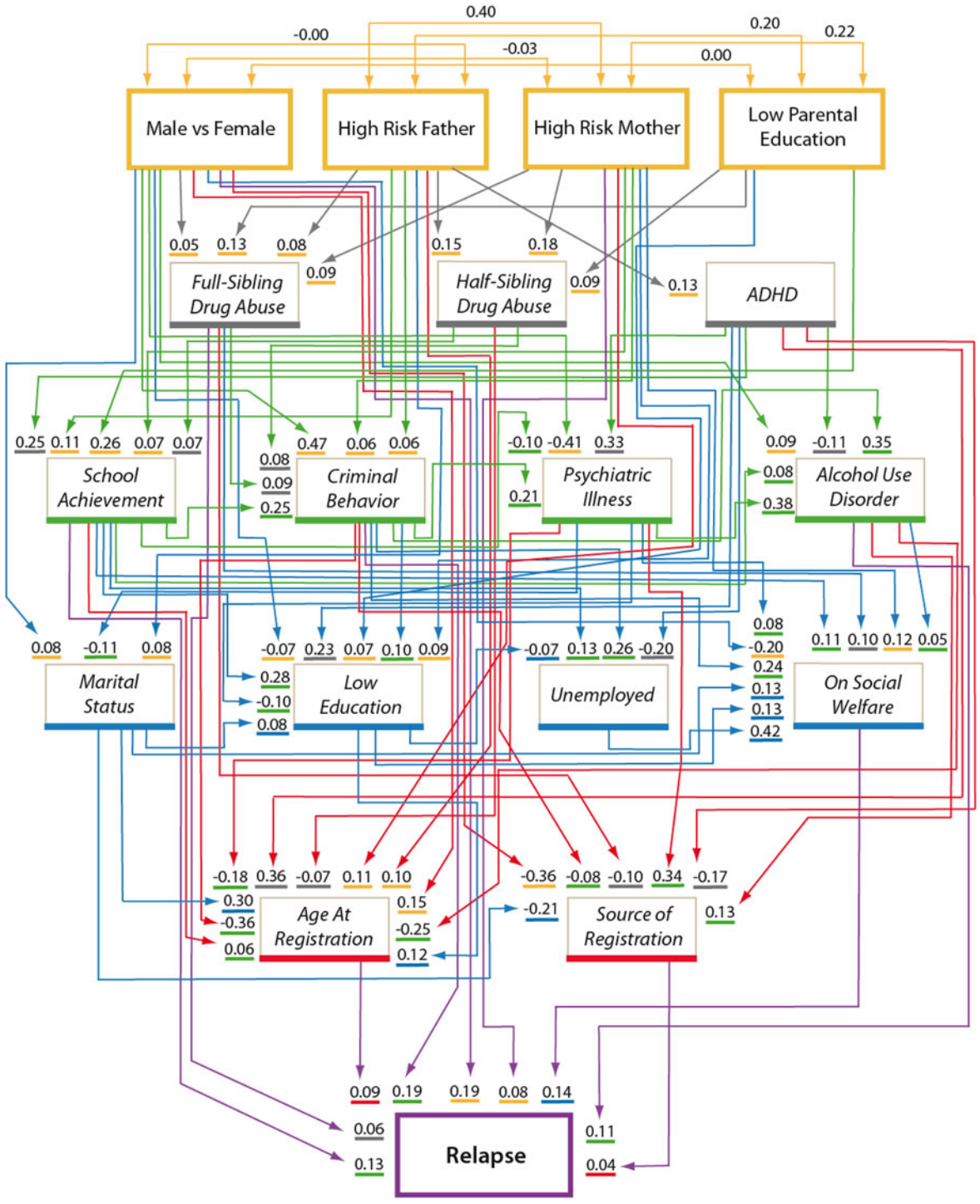

Fig. 2. A Structural Equation Path Model for the Prediction of Drug Abuse Relapse – Two-headed arrows represent correlation coefficients while one-headed arrows represent path coefficients or standardized partial regression coefficients. The variables were ordered to approximate a developmental process within 5 color-coded tiers that approximated five developmental periods: (1) birth and familial factors – yellow, (2) childhood and early adolescence – brown, (3) late adolescence – green), (4) early adulthood – blue, (5) at initial DA registration – red. The color of the paths reflects the tier of the dependent variable while the color underlining the path coefficient indexes the tier of the predictor variable. See the online Supplementary Table A1 for a detailed description of each variable in the model.

pathways to DA relapse. A surprising number of risk factors from early developmental phases (e.g. male sex, high-risk mother, and drug-abusing sibling) had direct effects on risk of relapse. The key mediational 'hubs' in the entire model were low education, age at registration, and being on social welfare. These integrated models are perhaps most useful for showing how often pathways of risk cross 'levels' or 'domains' of scientific inquiry. Our model clearly shows the inter-weaving in the prediction of DA relapse of social factors reflecting deprivation, genetic–temperamental vulnerability and indices of externalizing behaviors.

Our suggested path model can also provide some insight into the high degree of familial confounding we observed for the

https://doi.org/10.1017/S0033291717002938 Published online by Cambridge University Press

association between risk prediction score and the probability of DA relapse. Many of the early and influential risk factors in our structural model either directly reflect family risk (high-risk parents, DA in siblings) or are themselves highly familial [school achievement (Krapohl *et al.* 2014), criminal history (Kendler *et al.* 2015*a*) or poor educational attainment (Branigan *et al.* 2013)]. Given that these risk factors themselves and those variables which they influence constitute a substantial proportion of our risk score, it is understandable that our score would be considerably less effective at predicting the probability of relapse within closely related individuals than it would be in the general population.

Our approach to the prediction of DA relapse is empirically rather than theoretically oriented and is congruent with those described by Brandon *et al.* In an overview of theories of relapse in substance use disorders they write: 'Recent integrative theories of relapse have shied away from highly specific causal models toward broader models that categorize classes of risk variables' (Brandon *et al.* 2007, p. 269). Our findings are not easily integrated into the individual theoretical relapse models outlined by Brandon *et al.* (e.g. instrumental conditioning, behavioral economic, cognitive/social learning, and negative affectivity) that drive drug relapse. Rather our results support an empirical multifactorial model for the prediction of drug relapse incorporating a diverse array of risk factors from a range of familial, psychological, social, and economic domains.

Limitations

These results should be interpreted in the context of seven potential methodological limitations. First, our results are limited to the Swedish population and may not extrapolate to other countries. Patterns of illicit drug use and abuse in Sweden are broadly representative of those found in Northern Europe (Hibell *et al.* 2000; Kraus *et al.* 2003). Second, DA was ascertained using medical and criminal records which are not dependent on subject cooperation or accurate recall, which can be problematic for illegal behaviors. Compared with interviews, these methods likely generate false-positive and false-negative diagnoses. Overall, it is likely that we are assessing more severe cases of DA, or at least those that have clear medical and legal consequences. While large interview-based general population studies of DA prevalence do not exist in Sweden, lifetime prevalence of DA/dependence in near-by Norway is only slightly higher than the estimates we obtain using our methods in Sweden (Kringlen *et al.* 2001).

Third, our method of assessing relapse differed from nearly all prior studies, which may reduce the comparability of our findings. However, the overall rate and timing of relapse (58.0% with over three-quarters of the relapses within 3 years) in our sample were broadly in line with those reported in previous studies (Calabria *et al.* 2010; Farmer *et al.* 2015; Feingold *et al.* 2015; Sarvet & Hasin, 2016). However, we cannot, with the data available, confidently discriminate episode reoccurrence *v.* continuance as we have no definitive way to assess remission. However, we did censor any registrations that occurred within 180 days of the original registration because we considered it likely to reflect the same 'episode' of DA as did the first registration.

Fourth, because of concerns that our methods of ascertainment of DA could alter our predictors of relapse, we explored, using our multivariate model, whether predictors of future medical DA registrations differed from those found for criminal DA registration (see online Supplementary Table A4). Assuming p values <0.02 to be of primary interest given the 25 tests performed, eight noteworthy differences were found. Male sex, prior CB, low education, early age at registration, low income, and living with a child more strongly predicted relapse through the criminal registry. By contrast, psychiatric illness and having a medical first DA registration predicted more strongly relapse detected in the medical registry. So, not surprisingly, eight of our 25 risk factors differentially predicted relapse in a criminal *v.* a medical context.

Fifth, an advantage of our study was our comprehensive sampling of a large population-based cohort. But our analyses therefore assumed that risk prediction for DA relapse was broadly similar in major population subgroups. To test this, we examined, again using our multivariate model, the variability of our predictors across sex and age. Online Supplementary Fig. A2 and Table A5 depict the multivariate risk ratios for all of our predictors in our multivariate model in the entire sample and then in males/females and the younger and older halves of the sample. Given the 50 tests performed, we regarded only those with p values of <0.001 as of interest of which there were nine. Being male, having a high risk father, a lower school achievement, and having an early age at registration were more potent predictors of relapse in the younger half of the sample. Prior CB and being on social welfare were more potent in the older half. Having prior psychiatric illness and having an initial criminal registration were stronger predictors of relapse in females while early age at registration was stronger in males. So, a small number of predictors performed differently in men *v.* women and a slightly larger number by age, with most of these predicting relapse more strongly in the younger half of the sample.

Sixth, while there are important advantages in using registry data for the prediction of DA relapse, including large and representative samples and objective predictors, an important disadvantage is our inability to include a number of important variables that would need to be assessed by self-report such as levels of anxiety and depression, motivational constructs, physiological symptoms, and degree of social support (Brownell *et al.* 1986).

Conclusions

We examined the predictors of relapse in the 8 years after a first registration for DA in a general population sample of over 44 000 affected Swedish individuals. Univariate and multivariate regression analyses showed a wide diversity of risk factors predicted relapse which when combined into a risk index in a test random split-half sample, achieved an under the curve ROC value of 0.70. Co-relative analyses suggested that only a modest proportion of the association of the index to relapse risk was causal with the majority due to familial confounding factors. These findings suggest that prevention efforts for DA may need to focus on modifiable aspects of the familial environment rather than later putative risk factors that arise in adolescence and early adulthood. To explore mediational pathways, we then presented a structural equation model for the prediction of DA relapse. This model clearly demonstrated the inter-weaving, over developmental time, of risk factors for DA especially social factors reflecting deprivation, genetic–temperamental vulnerability and indices of externalizing behaviors.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033291717002938.

https://doi.org/10.1017/S0033291717002938 Published online by Cambridge University Press

Acknowledgements. This project was supported by grant R01DA030005 from the National Institutes of Health, the Swedish Research Council (K2012-70X-15428-08-3), the Swedish Research Council for Health, Working Life and Welfare (In Swedish: Forte; Reg.nr: 2013-1836), the Swedish Research Council (2012-2378; 2014-10134), and FORTE (2014-0804) as well as ALF funding from Region Skåne awarded.

Declaration of Interest. The authors have no conflicts of interest to declare.

Ethical Standards. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. We secured ethical approval for this study from the Regional Ethical Review Board of Lund University (Grant no. 2008/409).

References

Akaike H (1987) Factor analysis and AIC. *Psychometrika* **52**, 317–332.

Altman DG, Deeks JJ and Sackett DL (1998) Odds ratios should be avoided when events are common. *British Medical Journal* **317**, 1318.

Biederman J, *et al.* (1998) Does attention-deficit hyperactivity disorder impact the developmental course of drug and alcohol abuse and dependence? *Biological Psychiatry* **44**, 269–273.

Brandon TH, Vidrine JI and Litvin EB (2007) Relapse and relapse prevention. *Annual Reviews of Clinical Psychology* **3**, 257–284.

Branigan AR, McCallum KJ and Freese J (2013) Variation in the heritability of educational attainment: an international meta-analysis. *Social Forces* **92**, 109–140.

Brownell KD, *et al.* (1986) Understanding and preventing relapse. *American Psychologist* **41**, 765–782.

Calabria B, *et al.* (2010) Systematic review of prospective studies investigating "remission" from amphetamine, cannabis, cocaine or opioid dependence. *Addiction Behavior* **35**, 741–749.

Farmer RF, *et al.* (2015) Natural course of cannabis use disorders. *Psychological Medicine* **45**, 63–72.

Farmer RF, *et al.* (2016) Association of comorbid psychopathology with the duration of cannabis use disorders. *Psychology of Addiction Behavior* **30**, 82–92.

Feingold D, *et al.* (2015) Natural outcome of cannabis use disorder: a 3-year longitudinal follow-up. *Addiction* **110**, 1963–1974.

Florez-Salamanca L, *et al.* (2013) Probability and predictors of cannabis use disorders relapse: results of the National Epidemiologic Survey on alcohol and related conditions (NESARC). *Drug Alcohol and Dependence* **132**, 127–133.

Hibell B, *et al.* (2000) *The 1999 ESPAD Report: Alcohol and Other Drug Use Among Students in 30 European Countries*. The Swedish Council for Information on Alcohol and Other Drugs (Pompidou Group), Modin Tryck AB: Stockholm, Sweden.

Hser YI, *et al.* (2007*a*). Trajectories of heroin addiction: growth mixture modeling results based on a 33-year follow-up study. *Evaluation Review* **31**, 548–563.

Hser YI, Longshore D and Anglin MD (2007*b*). The life course perspective on drug use: a conceptual framework for understanding drug use trajectories. *Evaluation Review* **31**, 515–547.

Kendler KS, *et al.* (2003) Specificity of genetic and environmental risk factors for use and abuse/dependence of cannabis, cocaine, hallucinogens, sedatives, stimulants, and opiates in male twins. *American Journal of Psychiatry* **160**, 687–695.

Kendler KS, *et al.* (2015*a*). A Swedish national twin study of criminal behavior and its violent, white-collar and property subtypes. *Psychological Medicine* **45**, 2253–2262.

Kendler KS, *et al.* (2015*b*). A population-based Swedish twin and sibling study of cannabis, stimulant and sedative abuse in men. *Drug Alcohol and Dependence* **149**, 49–54.

Kendler KS, *et al.* (2012) Genetic and familial environmental influences on the risk for drug abuse: a national Swedish adoption study. *Archives of General Psychiatry* **69**, 690–697.

Krapohl E, *et al.* (2014) The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 15273–15278.

Kraus L, *et al.* (2003) Estimating prevalence of problem drug use at national level in countries of the European Union and Norway. *Addiction* **98**, 471–485.

Kringlen E, Torgersen S and Cramer V (2001) A Norwegian psychiatric epidemiological study. *American Journal of Psychiatry* **158**, 1091–1098.

Lopez-Quintero C, *et al.* (2011) Probability and predictors of remission from life-time nicotine, alcohol, cannabis or cocaine dependence: results from the National Epidemiologic Survey on alcohol and related conditions. *Addiction* **106**, 657–669.

McLellan AT, *et al.* (1983) Predicting response to alcohol and drug abuse treatments. Role of psychiatric severity. *Archives of General Psychiatry* **40**, 620–625.

Muthén LK and Muthén BO (2015) *Mplus User's Guide: 1998–2015*, 7th edn. Muthén & Muthén: Los Angeles, CA.

R Development Core Team (2014) *The R Project for Statistical Computing; R: A Language and Environment for Statistical Computing. [3.1.2]*. R Foundation for Statistical Computing: Vienna, Austria.

Sarvet AL and Hasin D (2016) The natural history of substance use disorders. *Current Opinions in Psychiatry* **29**, 250–257.

SAS Institute I (2011) *SAS/STAT User's Guide, Version 9.3*. SAS Institute Inc.: Cary, NC.

Tsuang MT, *et al.* (1996) Genetic influences on abuse of illicit drugs: a study of 3,297 twin pairs. *American Journal of Medical Genetics* **67**, 473–477.

van der Pol P, *et al.* (2015) Three-year course of cannabis dependence and prediction of persistence. *European Addiction Research* **21**, 279–290.

White AM, *et al.* (2004) Predictors of relapse during treatment and treatment completion among marijuana-dependent adolescents in an intensive outpatient substance abuse program. *Substance Abuse* **25**, 53–59.

Zou G (2004) A modified poisson regression approach to prospective studies with binary data. *American Journal of Epidemiology* **159**, 702–706.