Genetic risk for schizophrenia influences substance use in emerging adulthood: an event-level polygenic prediction model

Travis T. Mallard, B.S.; K. Paige Harden, Ph.D.; & Kim Fromme, Ph.D.

University of Texas at Austin

Correspondence to: Travis T. Mallard, B.S., Department of Psychology, The University of Texas at Austin, 108 E. Dean Keeton Stop A8000, Austin, TX 78712, USA. E-mail: travis.mallard@utexas.edu

Abstract

Emerging adulthood is a peak period of risk for alcohol and illicit drug use. As genetic risk for schizophrenia has been previously associated with substance use disorders, we examine how schizophrenia-associated genetic variants are related to trajectories of five substance use behaviors as they occurred in daily life across emerging adulthood. Non-Hispanic European participants provided DNA samples and completed daily reports of alcohol and drug use for one month per year across four years (*N*=28,372 individual observations of *N*=318 participants). The present study yields two major insights. First, results indicated that genetic risk for schizophrenia predicted emerging adults' overall likelihood to engage in illicit drug use and polysubstance use (concurrent illicit drug use and alcohol use or binge drinking), but did not predict alcohol use-only phenotypes. Second, the present findings suggest that genetic variants related to schizophrenia predict the rate of age-related change in substance use.

Introduction

Emerging adulthood, which spans the ages 18 to 25 years, is a peak developmental period for the initiation and escalation of alcohol and drug use[1,2]. Approximately 75% of lifetime cases of substance use disorders develop by the mid- to late-20s[3–5]. Moreover, problematic substance use often co-occurs with other forms of psychopathology[6,7], which further increases risk for negative health outcomes. Alcohol and drug use, therefore, continue to be two of the greatest contributors to preventable morbidity and mortality in the United States[8,9].

Substance use behaviors are heritable and highly polygenic[10,11], and several twin studies have reported that multiple genetic factors contribute to a rather heterogeneous genetic etiology of substance use[12,13]. Furthermore, advances in statistical genetics suggest that the co-occurrence of substance use and other mental health problems is due, in part, to a shared genetic etiology. That is, while a portion of the underlying genetic etiology may specifically increase liability for alcohol and/or drug use, other genetic risk factors for substance use may also be related to psychopathology more broadly[14–16].

Recently, genome-wide polygenic scores (GPS) have been used to examine shared, cross-trait genetic influences on several psychiatric phenotypes[17]. Notably, GPS provide individual-specific estimates of genetic liability for a given trait by aggregating the effects of thousands of single nucleotide polymorphisms (SNPs) identified in large genome-wide association studies (GWASs). Because GPS leverage the results from well-powered GWASs, GPS approaches are well-suited to investigate aggregate genetic effects in smaller samples[18]. Here, we examine the extent to which trajectories of alcohol and illicit drug use in a university sample of emerging adults are influenced by a GPS for schizophrenia.

Our focus on the schizophrenia GPS is motivated by evidence suggesting that schizophrenia and substance use disorders (SUDs) share a portion of their underlying genetic architecture[19]. For instance, a recent study identified a significant genetic correlation (r_g=.22) between schizophrenia and lifetime cannabis use by using bivariate linkage disequilibrium (LD) score regression[20]. Similarly, numerous cross-sectional GPS studies have reported that genetic liability for schizophrenia predicted alcohol, amphetamine, cannabis, cocaine, opioid, and sedative use disorders[20–25]. While previous studies have related genetic risk for schizophrenia to substance use disorders, no study has considered how the influence of genetic risks function in the context of development—when do genetic risks manifest during emerging adulthood?

In the present manuscript, we sought to extend this research through a "deep phenotyping" approach to investigate the effect of GPS for schizophrenia (henceforth referred to as GPS_{SCZ}) on substance use as it occurred in daily life within a longitudinal study of emerging adults. To accomplish this aim, we first collected daily self-report data related to substance use from across a four-year period (*N*=28,372 observations, *M*=89.22 observations). We then extended polygenic prediction methods to event-level phenotypes, which increase measurement precision of behavior in the natural environment and reduce recall bias common to cross-sectional research. Finally, we used a longitudinal design to elucidate whether genetic risk shapes how substance use *changes* across emerging adulthood.

To calculate GPS_{SCZ}, we leveraged results from the Psychiatric Genomic Consortium's (PGC) most recent GWAS of schizophrenia[26]. We then investigated the effect of GPS_{SCZ} on five event-level phenotypes: any alcohol use, binge drinking, illicit drug use, any concurrent alcohol and drug use, and concurrent binge drinking and drug use. Specifically, we tested: (1) whether GPS_{SCZ} predicted an individual's overall likelihood to engage in substance use on a given day,

and (2) whether GPS_{SCZ} predicted the rapidity of longitudinal, age-related change in substance use. As such, this paper lends critical insight into the heterogeneous genetic etiology of substance use behaviors and how they manifest in daily life during emerging adulthood.

Materials and method

Participants

The sample was recruited from a larger cohort of participants who underwent a longitudinal investigation of college students. Briefly, matriculating students at a large public university were recruited and surveyed at 10 waves over 6 years for a longitudinal study of substance use and related behaviors. Recruitment procedures for the full study have been described in previously published articles[27,28], and additional information (e.g., recruitment flow chart) is provided in Supplementary Note sections 1 and 2.

A subset of the full sample completed a daily monitoring protocol and provided DNA for genotyping procedures (N=517, 64% non-Hispanic European, 67% female). To avoid potential effects associated with population stratification, the analyses detailed below were limited to the non-Hispanic European portion of the sample (N=330, 65% female). Twelve participants were excluded from analyses following the quality control procedures described below (Final N=318, 65% female, M_{age}=18.44 years, SD_{age}=0.32 years). The university's Institutional Review Board approved all study procedures.

Genotyping protocol and quality control

Participants provided 2 mL of saliva in Oragene-Discover (Oragene™, DNAgenotek, Ottawa, Ontario, Canada) collection kits that were distributed and returned via mail. DNA for participants was extracted, purified, and diluted from saliva samples at the Institute for Behavior Genetics at the University of Colorado, Boulder. Purified and diluted samples were sent to the

Neuroscience Genomics Core at the University of California, Los Angeles, for chip-based genotyping. Samples were assayed on an Illumina BeadLab platform using an Illumina Infinium PsychArray BeadChip array (San Diego, CA), which assays ~265,000 SNPs across the genome.

Genotypic data were subjected to quality control procedures recommended for chip-based genomic data[29,30]. Samples were excluded from statistical analyses because of poor call rate (<98%), non-European ancestry (sigma>6.0), inconsistent self-reported sex and biological sex, and relatedness (p□>.125). SNPs were excluded from analyses if the minor allele frequency (MAF) was < 0.05, if more than 2% of genotype data was missing, or if the SNP was not in Hardy-Weinberg equilibrium (HWE P<0.00005). Finally, although the present analyses are limited to participants of non-Hispanic European descent, EIGENSTRAT[31] was used to extract the top ten genomic principal components of ancestry. Additional information on genotyping procedures is reported in Supplementary Note section 3.

Imputation

Unknown genotypes were imputed on the *Michigan Imputation Server* (https://imputationserver.sph.umich.edu). Variants were phased with Eagle v2.3[32] and imputed with Minimac3 1.0.13[33], using Phase 3 v5 of the 1000 Genomes Project[34] as a reference panel. To ensure all markers were of high quality, several post-imputation quality control thresholds were applied. Imputed SNPs with a MAF < .01, INFO score < .50, or HWE P-value < .00005 were excluded from all statistical analyses. As a result, the final set of 6,945,034 genotyped and imputed variants was of high quality.

Genome-wide polygenic scores

GPS were calculated for 318 unrelated non-Hispanic European participants using β-weights and P-values from summary statistics provided by the PGC. Summary statistics for

102,636 LD-pruned SNPs were derived from the PGC's 2014 GWAS of schizophrenia, which consisted of up to 36,989 cases and 113,075 controls[26]. All SNPs had a MAF > .01, INFO > .90, and pairwise LD < .25. PLINK 1.9[35] was then used to calculate a GPS for each participant in the sample by summing trait-associated SNPs weighted by the beta coefficients and *P*-values derived from the PGC GWAS. To aid interpretation of results, GPS were then z-standardized, establishing a mean (*M*) of 0 and a standard deviation (*SD*) of 1. This procedure was repeated to generate GPS at five discrete *P*-value thresholds (P_T=1.00, .50, .30, .10, and .01; see Supplementary Note section 4 for descriptive statistics), as well as a GPS using only the 128 independent markers identified by the PGC as associated with schizophrenia (i.e., P_T=5e-8). As each substance use phenotype was tested at six P_T, an α level of *P*=.0017 (α=.05/30=.0017) was established to correct for multiple testing.

Longitudinal event-level design

Participants completed up to 30 consecutive days of online self-monitoring in each of their first four years of college. A random selection of 40-43 students was invited to participate in the study each week to ensure sufficient monitoring across the entire calendar year. During their annual reporting period, participants were instructed to use the self-monitoring website (maintained by DatStat, Seattle, WA) to answer questions about the previous day. A further description of the longitudinal event-level design is presented in Supplementary Note section 5.

Phenotypic measures and quality control

Each day, participants answered questions about the previous day related to time-varying demographics (e.g., weight), alcohol consumption (*"How many drinks did you consume yesterday?" and "Of the times that you drank this day, how long was your heaviest drinking episode?"*), and illicit drug use (*"Did you use illicit drugs yesterday?"*). If participants endorsed

illicit drug use on any given day, they were asked to specify whether the drug use occurred while sober or during a drinking episode. Five event-level substance use phenotypes were assessed: any alcohol use, binge drinking, illicit drug use, any concurrent alcohol and drug use, and concurrent binge drinking and drug use. Operant definitions and descriptive statistics for all substance use phenotypes are presented in Supplementary Note section 5.

The self-monitoring website recorded the time and date of each daily report, which was subsequently used to determine the participant's age (rounded to two decimal points) on a given day. To reduce potential bias attributable to over-exclusion or inclusion of noncompliant participants, eight participants who did not provide at least 14 days of monitoring data were excluded from statistical analyses. The final sample included 318 participants with 28,372 event-level observations.

Analytic approach

A two-level hierarchical linear model with robust standard errors (HLM)[36] was used to analyze the relationships between GPS_{SCZ}, participant age (AGE), and the five substance use phenotypes. Events were nested within participants for all statistical analyses. Notably, the HLM included a random intercept and random slope to simultaneously assess the person-centered general and developmentally-specific effects of GPS_{SCZ}. Genomic principal components of ancestry (GPC_1 … GPC_{10}), biological sex (SEX), and age at beginning of college (AGE_{W1}) were included as trait-level covariates in all analyses. The full model is described below.

LEVEL 1 MODEL

$$\mathrm{Prob}(\mathrm{OUTCOME} = 1|\pi) = \varphi$$

$$\mathrm{Log}\left[\frac{\varphi}{(1-\varphi)}\right] = \eta$$

$$\eta = \pi_0 + \pi_1\,(\mathrm{AGE})$$

LEVEL 2 MODEL

$$\pi_0 = \beta_{00} + \beta_{01}\,(\mathrm{GPS_{SCZ}}) + \beta_{02\ldots011}\,(\mathrm{GPC_1} \ldots \mathrm{GPC_{10}}) + \beta_{012}\,(\mathrm{SEX}) + \beta_{013}\,(\mathrm{AGE_{W1}}) + r_0$$

$$\pi_1 = \beta_{10} + \beta_{11}\,(\mathrm{GPS_{SCZ}}) + \beta_{12\ldots111}\,(\mathrm{GPC_1} \ldots \mathrm{GPC_{10}}) + \beta_{112}\,(\mathrm{SEX}) + r_1$$

All substance use phenotypes were analyzed using a logit model. The Level 1 (event level) equation modeled the likelihood of a participant engaging in substance use on a given day as a function of a person-specific random intercept (π_0) and a person-centered random slope describing within-person variability in the likelihood of using substances as a function of event-level age (π_1). Importantly, event-level age was centered on the person mean and thus reflects within-person, age-related change in substance use over time[36]. Overall, the Level 1 equation tested whether an individual's age on a given reporting day predicted whether they were more likely to use illicit drugs on some days relative to others.

The Level 2 (person level) equation modeled between-person variability in the likelihood to use substances when aggregating across all occasions. Here, the intercept for substance use phenotypes (π_0), which represents the person-average likelihood to engage in substance use across all events, was modeled as a function of the effect of $\mathrm{GPS_{SCZ}}$ (β_{01}), as well as the effects of ancestry ($\beta_{02} \ldots \beta_{11}$), sex ($\beta_{12}$), and age at first wave of data collection (β_{13}). We additionally modeled the random slopes for event-level age as a function of the effects of $\mathrm{GPS_{SCZ}}$ (β_{11}), ancestry ($\beta_{12} \ldots \beta_{111}$), and sex ($\beta_{112}$). The top ten principal components of ancestry and age at first wave were centered on the grand mean, while $\mathrm{GPS_{SCZ}}$ and sex were uncentered. Between-person residuals were included for all event-level slopes (r_0 and r_1) to allow for heterogeneity in the magnitude of within-person effects. Overall, the Level 2 model tested whether $\mathrm{GPS_{SCZ}}$, sex,

and age predicted (1) participants' overall likelihood to use substances when aggregating across all events and (2) age-related changes in the likelihood to use substances as participants grew older.

Results

The general and developmentally-specific effects of GPS_{SCZ} on all five substance use phenotypes are presented in Tables 1 and 2, respectively. The unique effects of GPS_{SCZ} on the intercept and slope of substance use phenotypes are also illustrated in Figures 1A and 1B. Here, we represent the effects of GPS_{SCZ} as odds ratios, which reflect change in the odds of an outcome given a one unit increase in the predictor. Moreover, as we standardized all GPS prior to analysis, we can characterize the effect of a 1 *SD* increase in genetic risk for schizophrenia on the odds of substance use in our event-level data.

After correcting for multiple testing, we observed significant effects of GPS_{SCZ} on illicit drug use, any concurrent alcohol use and drug use, and concurrent binge drinking and drug use. GPS_{SCZ} did not predict the likelihood to engage in any alcohol use or binge drinking. Additionally, we did not observe any consistently significant effect of age at the beginning of college (range=17.10 to 19.26) or biological sex on the likelihood to engage in any substance use. However, event-level age (that is, getting older over the course of college) was positively associated with a greater likelihood to engage in all forms of substance use (all $P<.001$). The specific results for each substance use phenotype are detailed below.

Illicit drug use

Our results indicated that GPS_{SCZ} predicted a greater overall likelihood to use illicit drugs at $P_{T=}1.00$ and 5e-8. We additionally observed nominally significant effects of GPS_{SCZ} at $P_{T=}.50$, .10, and .01 (all $P<.05$). We detected the most robust effect at $P_T=1.00$ ($B=.120$, OR=1.128,

P<.001). Here, a 1 *SD* increase in GPS_{SCZ} was associated with a relative 12.8% increase in the likelihood to engage in illicit drug use on any given day.

Furthermore, the results of our HLM indicated that age-related changes in illicit drug use varied as a function of GPS_{SCZ} at $P_{T=}$5e-8. Specifically, we found that GPS_{SCZ} was negatively associated with the event-level slope between age and illicit drug use (*B*=-.090, OR=.914, *P*<.001). So, although participants with higher GPS_{SCZ} at $P_{T=}$5e-8 were more likely to use illicit substances overall, they also experienced a less substantial increase in the likelihood of using drugs as they grew older. This effect is illustrated in Figure 2.

Concurrent alcohol and drug use

GPS_{SCZ} predicted a greater overall likelihood to engage in concurrent alcohol and drug use at P_T=1.00, .50, .10, and 5e-8. We also observed a nominally significant association at P_T=.30 (*P*<.05). The most robust effect of GPS_{SCZ} on contemporaneous alcohol and drug use was observed at P_T=5e-8 (*B*=.209, OR=1.233, *P*<.001). At P_T=5e-8, a 1 *SD* increase in GPS_{SCZ} was associated with a relative 23.3% increase in the likelihood to use alcohol and drugs at the same time on any given reporting day.

Moreover, age-related change in concurrent alcohol and drug use varied as a function of GPS_{SCZ} at P_T=.10 and 5e-8. These age × genotype interaction effects differed by P_T. Whereas GPS_{SCZ} at P_T=.10 was positively associated with the slope between each participant's age and concurrent alcohol and drug use (*B*=.090, OR=1.095, *P*<.001), GPS_{SCZ} at P_T=5e-8 was negatively associated with the slope of age-related change in concurrent alcohol and drug use (*B*=-.078, OR=0.925, *P*<.001). These effects are illustrated in Figures 3A and 3B.

Concurrent binge drinking and drug use

We observed a similarly robust association between genetic risk for schizophrenia and concurrent binge drinking and drug use. Specifically, GPS_{SCZ} predicted participants' overall likelihood to simultaneously binge drink and use drugs at all P_T except for P_T=.30. Once more, we observed the most robust signal at P_T=5e-8 (*B*=.184, OR=1.203, *P*<.001), where a 1 *SD* increase in GPS_{SCZ} was associated with a relative 20.3% increase in the likelihood to engage in concurrent binge drinking and drug use on any given day. Genetic risk for schizophrenia was also positively associated with age-related change in concurrent binge drinking and drug use at P_T=.50 and .10. We observed the most robust association of GPS_{SCZ} on the slope at P_T=.10 (*B*=.093, OR=1.097, *P*<.001).

Discussion

Here, we described the first longitudinal event-level examination of genetic risk for schizophrenia and its effects on daily substance use in a sample of university undergraduates. Specifically, we tested whether (1) GPS_{SCZ} predicted a greater overall likelihood to engage in substance use on a given day and (2) whether within-person age-related changes in substance use varied as a function of GPS_{SCZ}. We report two major findings. First, we found that GPS_{SCZ} predicted an individual's overall likelihood to engage in illicit drug use and polysubstance use (concurrent illicit drug use and alcohol use or binge drinking), but it did not predict alcohol use-only phenotypes. Second, we found preliminary support for the hypothesis that genetic variants related to schizophrenia predict the rate of age-related change in substance use.

Overall, a greater polygenic loading for schizophrenia consistently predicted more recurrent substance use. We observed robust associations between GPS_{SCZ} and the event-level likelihood to engage in illicit drug use, concurrent alcohol and drug use, and concurrent binge drinking and drug use at several P_T, ranging from the most liberal threshold (P_T=1.00) to the

most conservative threshold (P_T=5e-8). Whereas many prior studies have only examined the effect of GPS_{SCZ} on substance use with diagnostic phenotypes, we identified genetic influences on trajectories of substance use in daily life. As a result, our findings both corroborate and build upon recent studies that have reported GPS_{SCZ} predicted problematic alcohol and drug use[20–24]. Taken together, our results align with growing evidence of a shared genetic architecture between schizophrenia and substance use.

Moreover, the effect of GPS_{SCZ} on substance use also predicted age-related change in substance use over four years; however, these effects on intraindividual change varied depending on the P_T used to construct the GPS. Generally, the most conservative P_T threshold (P_T=5e-8) resulted in a negative effect of GPS_{SCZ} on the slope between participant age and substance use. This indicated that the SNPs most strongly associated with schizophrenia conferred greater risk for substance use during late adolescence and early emerging adulthood, but became less impactful over time, as individuals with low polygenic risk for schizophrenia "caught up" in terms of their substance use behaviors. Conversely, when GPS_{SCZ} was calculated with a more liberal P_T threshold, we observed a positive effect of GPS_{SCZ} on the slope between participant age and substance use, which indicated that the effect of broad genetic risk for schizophrenia on substance use grows stronger over time. We postulate that these nuances could be explained by the inclusion and/or exclusion of certain causal variants in schizophrenia, but further investigation is beyond the scope of this paper.

Interestingly, GPS_{SCZ} did not predict phenotypes that involved only alcohol consumption (alcohol use and binge drinking). In contrast, a small genetic correlation between schizophrenia and alcohol consumption was recently reported in a sample of older adults[37]. Given the relatively small genetic correlation between schizophrenia and alcohol consumption (r_g=.13), it is possible

that we were not powered to detect cross-trait effects. Alternatively, different genetic factors may influence alcohol consumption at different stages of development[38]. In the present sample of emerging adults, alcohol use and binge drinking are relatively normative behaviors and, as such, they may be less influenced by genetic factors during this developmental period. Indeed, research has demonstrated that genetic influences on alcohol consumption typically increase across the lifespan[2,39].

Our findings should be interpreted in light of several limitations. First, our measure of illicit drug use did not identify the specific substance that was consumed, so we have limited insight into substance-specific patterns of drug use. However, as previous investigations have identified cannabis as the most commonly used illicit drug in this cohort[27], we would hypothesize cannabis to be the most commonly used drug in the present study. Second, we restricted the analyses reported here to non-Hispanic European participants to reduce the risk of spurious findings caused by population stratification. Consequently, the findings of our study may not generalize to other ancestral populations. A third potential limitation is our relatively moderate sample size. However, concerns about statistical power in the present study are partially attenuated by the fact that (1) we were well-powered for our within-person approach (see Supplementary Note section 6 for power simulations), (2) we leveraged *a priori* effect size estimates from a well-powered GWAS of schizophrenia[26], and (3) we examined aggregate genomic variation rather than individual SNPs of small effect.

Despite these limitations, this study exhibited notable strengths in its novel approach. As the first longitudinal, event-level investigation of genetic risk for schizophrenia and its effects on substance use in daily life, this work provides ecologically valid evidence that these two psychiatric conditions are influenced, in part, by shared genetic factors. This finding is consistent

with previous research that has identified shared genetic influences between schizophrenia and substance use[20–24]. Notably, this study demonstrated that genetic variants associated with schizophrenia predict important behavioral phenotypes in a sample of healthy, university students, where schizophrenia prevalence is expected to be minimal. In doing so, we present a critical extension of previous work, which has primarily examined the genetic underpinnings of substance use in clinical samples. Moreover, the present study contributes to the broader literature by providing a framework to characterize broad and developmentally-specific effects of genetic variation. Here, we used a novel combination of GPS and repeated phenotyping to characterize the influence of genetic risk for schizophrenia on patterns of age-related change in substance use across emerging adulthood. Future studies that employ similar methods will be uniquely poised to identify novel targets for prevention and intervention by elucidating mechanisms that contribute to the development and maintenance of psychiatric traits and disorders.

Acknowledgments

The authors would like to acknowledge the National Institute for Alcohol Abuse and Alcoholism (R01AA013967, R01AA020637) for their support of this research. Additionally, the authors thank James Ashenhurst, Emily Wilhite, Elise Marino, and the rest of the SAHARA Lab for their assistance in data collection.

Conflict of Interest

The authors have no conflicts of interest to report.

Table 1. *The effect of genetic risk for schizophrenia on the intercept of substance use outcomes*

	P_T=1.00			P_T=.50			P_T=.30		
Outcome	***B***	**OR**	**95% CI**	***B***	**OR**	**95% CI**	***B***	**OR**	**95% CI**
AU	-.039	0.961	(0.869,1.064)	-.010	0.990	(0.896,1.094)	-.030	0.970	(0.872,1.079)
BD	-.040	0.961	(0.870,1.062)	-.001	0.999	(0.905,1.102)	-.058	0.944	(0.860,1.036)
DU	.153	1.165	(1.057,1.285)	.111	1.117	(1.013,1.233)	.081	1.085	(0.953,1.234)
PSU1	.133*	1.142	(1.070,1.219)	.121*	1.129	(1.062,1.199)	.095	1.100	(1.025,1.181)
PSU2	.086*	1.090	(1.035,1.148)	.086*	1.090	(1.034,1.148)	.037	1.037	(0.982,1.096)
	P_T=.10			P_T=.01			P_T=5e-8		
Outcome	***B***	**OR**	**95% CI**	***B***	**OR**	**95% CI**	***B***	**OR**	**95% CI**
AU	.046	1.047	(0.955,1.148)	.011	1.011	(0.925,1.105)	.118	1.125	(1.024,1.237)
BD	.010	1.010	(0.927,1.101)	-.014	0.986	(0.904,1.076)	.072	1.075	(0.977,1.183)
DU	.107	1.113	(1.008,1.229)	.082	1.085	(1.002,1.176)	.120*	1.128	(1.057,1.202)
PSU1	.133*	1.143	(1.076,1.214)	.062	1.064	(0.997,1.134)	.209*	1.233	(1.175,1.294)
PSU2	.085*	1.089	(1.038,1.142)	.046	1.047	(1.000,1.097)	.184*	1.203	(1.157,1.250)

Note: * = significant after correction for multiple comparisons (*P* = .0017). AU=alcohol use, BD=binge drinking, DU=drug use, PSU1=concurrent alcohol and drug use, PSU2=concurrent binge drinking and drug use.

Table 2. *The effect of genetic risk for schizophrenia on the slope between age and substance use outcomes*

	P_T=1.00			P_T=.50			P_T=.30		
Outcome	***B***	**OR**	**95% CI**	***B***	**OR**	**95% CI**	***B***	**OR**	**95% CI**
AU	-.006	0.994	(0.943,1.047)	-.009	0.991	(0.945,1.040)	-.003	0.997	(0.950,1.046
BD	-.004	0.996	(0.955,1.039)	-.017	0.984	(0.944,1.025)	-.027	0.974	(0.937,1.012)
DU	-.030	0.970	(0.891,1.056)	-.025	0.975	(0.895,1.062)	.001	1.001	(0.899,1.116)
PSU1	-.009	0.991	(0.942,1.042)	.029	1.030	(0.981,1.080)	.046	1.048	(0.994,1.104)
PSU2	-.002	0.998	(0.970,1.026)	.054*	1.056	(1.024,1.089)	.041	1.042	(1.009,1.076)
	P_T=.10			P_T=.01			P_T=5e-8		
Outcome	***B***	**OR**	**95% CI**	***B***	**OR**	**95% CI**	***B***	**OR**	**95% CI**
AU	.015	1.015	(0.967,1.065)	.026	1.026	(0.979,1.076)	-.002	0.998	(0.950,1.049)
BD	.018	1.019	(0.980,1.059)	.026	1.027	(0.986,1.070)	-.033	0.968	(0.932,1.004)
DU	.016	1.016	(0.939,1.099)	-.016	0.984	(0.937,1.033)	-.090	0.914	(0.876,0.954)
PSU1	.090*	1.095	(1.046,1.145)	.015	1.015	(0.972,1.059)	-.078*	0.925	(0.894,0.957)
PSU2	.093*	1.097	(1.063,1.133)	.002	1.002	(0.969,1.037)	-.020	0.980	(0.955,1.005)

Note: * = significant after correction for multiple comparisons (*P* = .0017). AU=alcohol use, BD=binge drinking, DU=drug use, PSU1=concurrent alcohol and drug use, PSU2=concurrent binge drinking and drug use.

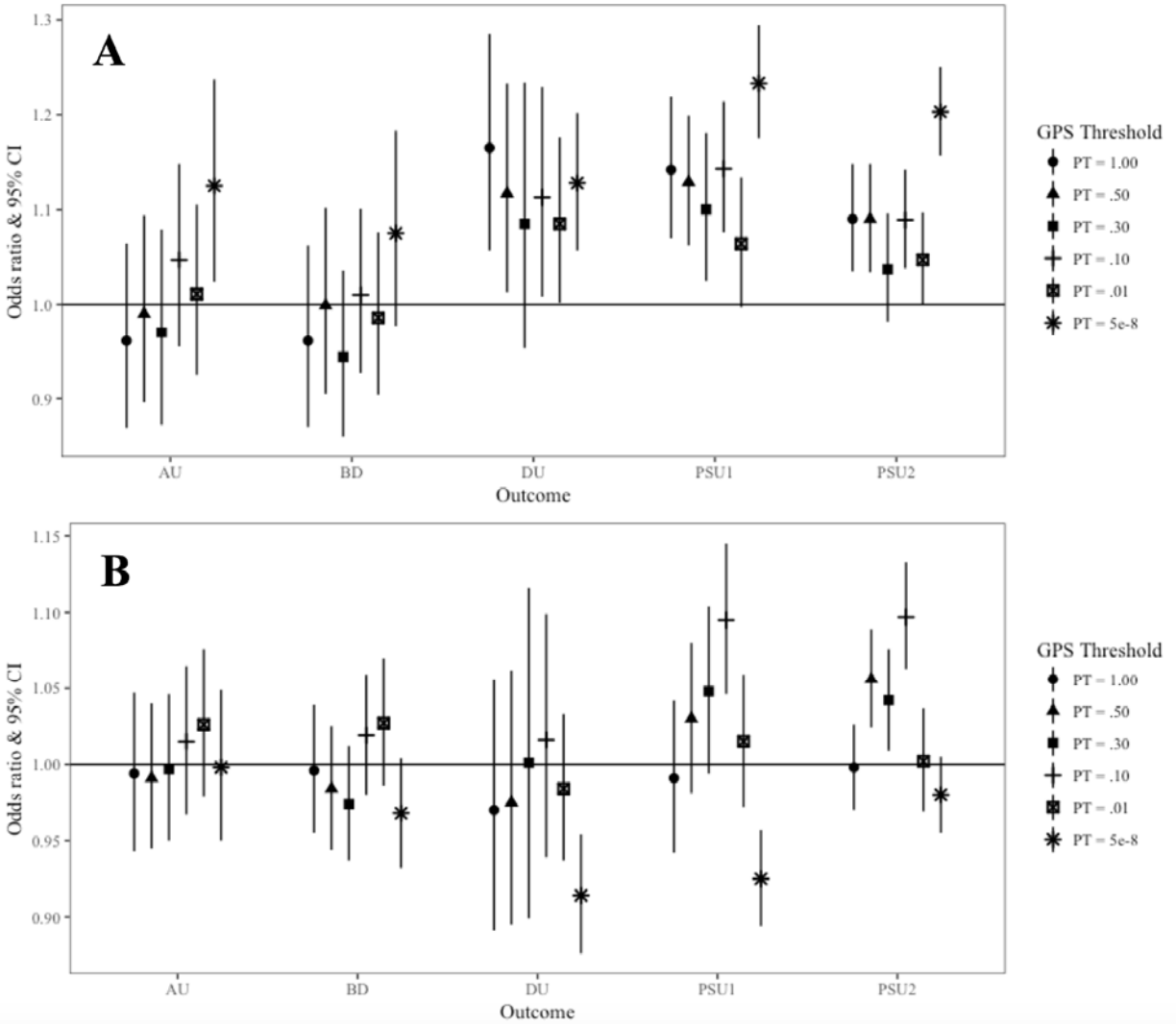

Figure 1. Odds ratios illustrating the effects of genetic risk for schizophrenia on the (A) intercept and (B) age-related slope of the five substance use phenotypes. Note: CI=confidence interval, GPS=genome-wide polygenic score, PT=*P*-value threshold, AU=alcohol use, BD=binge drinking, DU=drug use, PSU1=concurrent alcohol and drug use, PSU2=concurrent binge drinking and drug use.

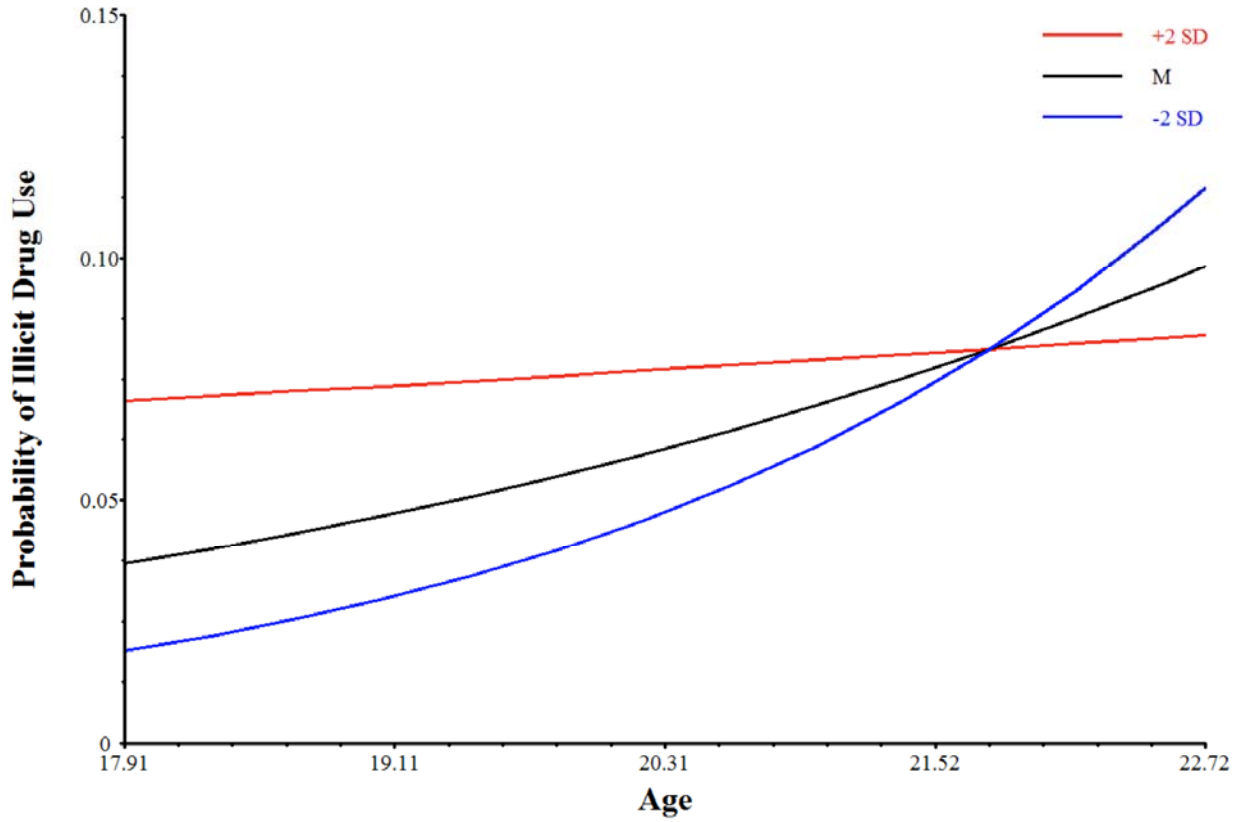

Figure 2. The effects of genetic risk for schizophrenia (GPS_{SCZ}) on the likelihood to use illicit drugs as a function of increasing age at P_T=5e-8. Results show that GPS_{SCZ} predicted a greater overall likelihood to engage in illicit drug use, but individuals with lower genetic liability for schizophrenia experienced a more substantial age-related increase in drug use.

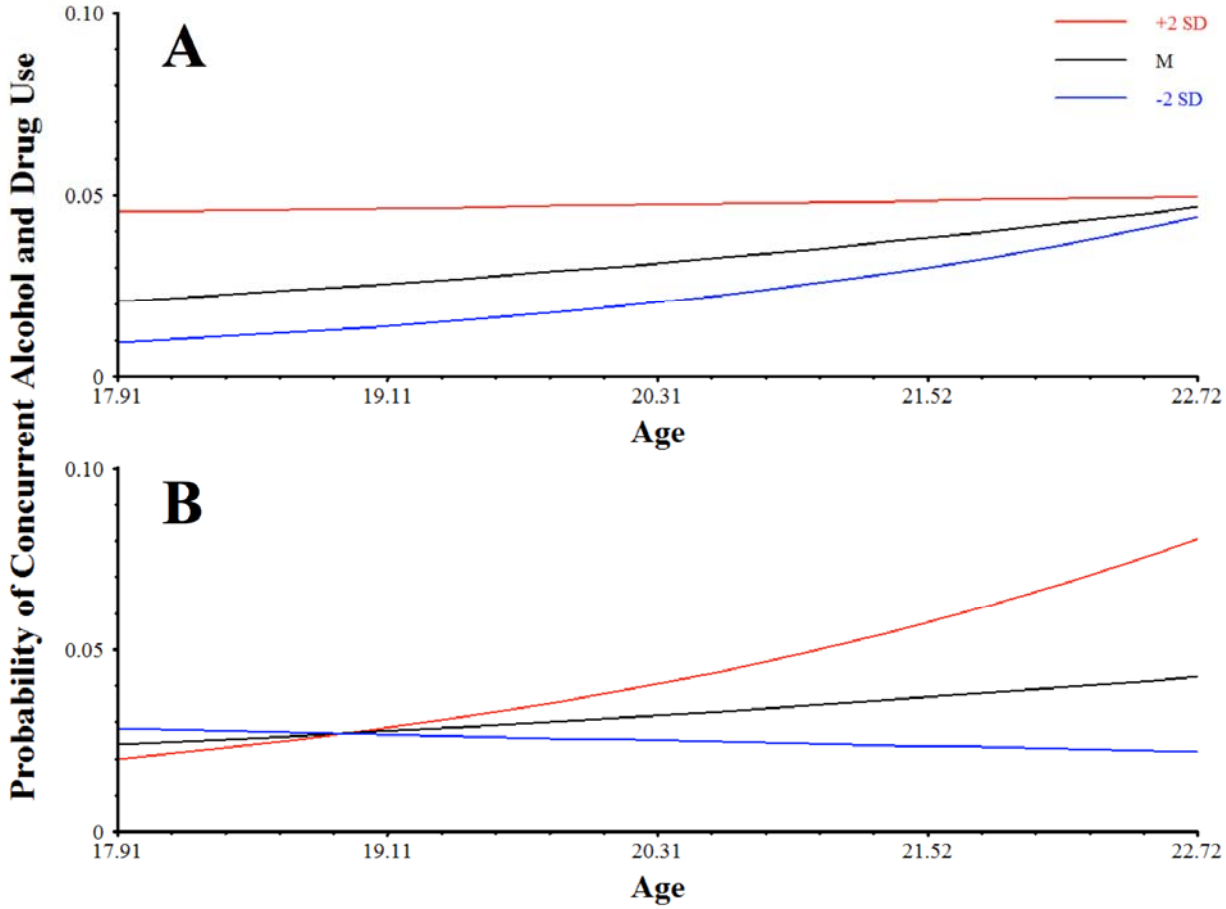

Figure 3. The effects of genetic risk for schizophrenia (GPS_{SCZ}) on the likelihood to engage in concurrent alcohol and drug use as a function of increasing age. Figure 3A illustrates the effect of GPS_{SCZ} at P_T=5e-8, whereas Figure 3B illustrates the effect of GPS_{SCZ} at P_T=.10. Here, we see that GPS_{SCZ} predicts a greater overall likelihood to use illicit drugs in both cases, but the developmentally-specific effects of genetic liability for schizophrenia depend on the variants that are included.

References

1. Johnston LD, O'malley PM, Bachman JG, Schulenberg JE. Monitoring the Future National Survey Results on Drug Use, 1975-2010. Volume II, College Students & Adults Ages 19-50. *Inst Soc Res*. 2011.

2. Kendler KS, Schmitt E, Aggen SH, Prescott CA. Genetic and Environmental Influences on Alcohol, Caffeine, Cannabis, and Nicotine Use From Early Adolescence to Middle Adulthood. *Arch Gen Psychiatry*. 2008;65(6):674. doi:10.1001/archpsyc.65.6.674.

3. Kessler RC, Berglund P, Demler O, et al. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005;62(6):593. doi:10.1001/archpsyc.62.6.593.

4. Kessler RC, Amminger GP, Aguilar-Gaxiola S, Alonso J, Lee S, Ustün TB. Age of onset of mental disorders: a review of recent literature. *Curr Opin Psychiatry*. 2007;20(4):359-364. doi:10.1097/YCO.0b013e32816ebc8c.

5. Christie KA, Burke JD, Regier DA, Rae DS, Boyd JH, Locke BZ. Epidemiologic evidence for early onset of mental disorders and higher risk of drug abuse in young adults. *Am J Psychiatry*. 1988;145(8):971-975. doi:10.1176/ajp.145.8.971.

6. Grant BF, Goldstein RB, Saha TD, et al. Epidemiology of DSM-5 Alcohol Use Disorder: results from the National Epidemiologic Survey on Alcohol and Related Conditions-III. *JAMA Psychiatry*. 2015;72(8):757. doi:10.1001/jamapsychiatry.2015.0584.

7. Grant BF, Saha TD, Ruan WJ, et al. Epidemiology of DSM-5 Drug Use Disorder: results from the National Epidemiologic Survey on Alcohol and Related Conditions-III. *JAMA Psychiatry*. 2016;73(1):39. doi:10.1001/jamapsychiatry.2015.2132.

8. Johnson NB, Hayes LD, Brown K, Hoo EC, Ethier KA. CDC National Health Report:

leading causes of morbidity and mortality and associated behavioral risk and protective factors--United States, 2005-2013. *MMWR Surveill Summ*. 2014;63 Suppl 4(1545-8636 (Electronic)):3-27. doi:su6304a2 [pii].

9. Whiteford HA, Degenhardt L, Rehm J, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet*. 2013;382(9904):1575-1586. doi:10.1016/S0140-6736(13)61611-6.

10. Polderman TJC, Benyamin B, de Leeuw CA, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet*. 2015;47(7):702-709. doi:10.1038/ng.3285.

11. Gratten J, Wray NR, Keller MC, Visscher PM. Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nat Neurosci*. 2014;17(6):782-790. doi:10.1038/nn.3708.

12. Kendler KS, Prescott CA, Myers J, Neale MC. The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Arch Gen Psychiatry*. 2003;60(9):929. doi:10.1001/archpsyc.60.9.929.

13. Kendler KS, Aggen SH, Prescott CA, Crabbe J, Neale MC. Evidence for multiple genetic factors underlying the DSM-IV criteria for alcohol dependence. *Mol Psychiatry*. 2012;17(12):1306-1315. doi:10.1038/mp.2011.153.

14. Johnson W, Hicks BM, McGue M, Iacono WG. How intelligence and education contribute to substance use: hints from the Minnesota Twin family study. *Intelligence*. 2009;37(6):613-624. doi:10.1016/j.intell.2008.12.003.

15. Caspi A, Houts RM, Belsky DW, et al. The p factor: one general psychopathology factor in the structure of psychiatric disorders? *Clin Psychol Sci*. 2014;2(2):119-137.

doi:10.1177/2167702613497473.

16. Pettersson E, Larsson H, Lichtenstein P. Common psychiatric disorders share the same genetic origin: a multivariate sibling study of the Swedish population. *Mol Psychiatry*. 2016;21(5):717-721. doi:10.1038/mp.2015.116.

17. Krapohl E, Euesden J, Zabaneh D, et al. Phenome-wide analysis of genome-wide polygenic scores. *Mol Psychiatry*. 2016;(May):1-6. doi:10.1038/mp.2015.126.

18. Belsky DW, Israel S. Integrating genetics and social science: genetic risk scores. *Biodemography Soc Biol*. 2014;60(2):137-155. doi:10.1080/19485565.2014.946591.

19. Polimanti R, Agrawal A, Gelernter J. Schizophrenia and substance use comorbidity: a genome-wide perspective. *Genome Med*. 2017;9(1):25. doi:10.1186/s13073-017-0423-3.

20. Verweij KJH, Abdellaoui A, Nivard MG, et al. Short communication: genetic association between schizophrenia and cannabis use. *Drug Alcohol Depend*. 2017;171:117-121. doi:10.1016/j.drugalcdep.2016.09.022.

21. Reginsson GW, Ingason A, Euesden J, et al. Polygenic risk scores for schizophrenia and bipolar disorder associate with addiction. *Addict Biol*. 2017. doi:10.1111/adb.12496.

22. Power RA, Verweij KJH, Zuhair M, et al. Genetic predisposition to schizophrenia associated with increased use of cannabis. *Mol Psychiatry*. 2014;19(11):1201-1204. doi:10.1038/mp.2014.51.

23. Carey CE, Agrawal A, Bucholz KK, et al. Associations between polygenic risk for psychiatric disorders and substance involvement. *Front Genet*. 2016;7(AUG):1-10. doi:10.3389/fgene.2016.00149.

24. Kalsi G, Euesden J, Coleman JRI, et al. Genome-wide association of heroin dependence in Han Chinese. *PLoS Genet*. 2016:1-18. doi:10.1371/journal.pone.0167388.

25. Hartz SM, Horton A, Oehlert M, et al. Association between substance use disorder and polygenic liability to schizophrenia. *Biol Psychiatry*. 2017. doi:10.1016/j.biopsych.2017.04.020.

26. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511:421-427. doi:10.1038/nature13595.

27. Fromme K, Corbin WR, Kruse MI. Behavioral Risks during the Transition from High School to College. *Dev Psychol*. 2008;44(5):1497-1504. doi:10.1016/j.surg.2006.10.010.Use.

28. Ashenhurst JR, Harden KP, Corbin WR, Fromme K. Trajectories of Binge Drinking and Personality Change Across Emerging Adulthood. *Psychol Addict Behav*. 2015;29(4):978-991. doi:10.14440/jbm.2015.54.A.

29. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat Protoc*. 2010;5(9):1564-1573. doi:10.1038/nprot.2010.116.

30. Turner S, Armstrong LL, Bradford Y, et al. Quality control procedures for genome wide association studies. *Curr Proc Hum Genet*. 2011;68(1):1-24. doi:10.1002/0471142905.hg0119s68.Quality.

31. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904-909. doi:10.1038/ng1847.

32. Loh P-R, Palamara PF, Price AL. Fast and accurate long-range phasing in a UK Biobank cohort. *Nat Genet*. 2016;48(7):811-816. doi:10.1038/ng.3571.

33. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48(10):1284-1287. doi:10.1038/ng.3656.

34. 1000 Genomes Project Consortium T 1000 GP, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393.

35. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4(1):7. doi:10.1186/s13742-015-0047-8.

36. Raudenbush SW, Bryk AS. *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2nd ed. Sage; 2002.

37. Clarke T-K, Adams MJ, Davies G, et al. Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112,117). *bioRxiv*. 2017.

38. Edwards AC, Kendler KS. Alcohol consumption in men is influenced by qualitatively different genetic factors in adolescence and adulthood. *Psychol Med*. 2013;43(9):1857-1868. doi:10.1017/S0033291712002917.

39. van Beek JHDA, Kendler KS, de Moor MHM, et al. Stable Genetic Effects on Symptoms of Alcohol Abuse and Dependence from Adolescence into Early Adulthood. *Behav Genet*. 2012;42(1):40-56. doi:10.1007/s10519-011-9488-8.