

## Original Article

**Cite this article:** Mallard TT, Harden KP, Fromme K (2019). Genetic risk for schizophrenia is associated with substance use in emerging adulthood: an event-level polygenic prediction model. *Psychological Medicine* 49, 2027–2035. <https://doi.org/10.1017/S0033291718002817>

Received: 20 February 2018

Revised: 14 August 2018

Accepted: 6 September 2018

First published online: 12 October 2018

**Key words:**

Emerging adulthood; longitudinal; multilevel modeling; polygenic prediction; schizophrenia; substance use

**Author for correspondence:**

Travis T. Mallard, E-mail: [travis.mallard@utexas.edu](mailto:travis.mallard@utexas.edu)

# Genetic risk for schizophrenia is associated with substance use in emerging adulthood: an event-level polygenic prediction model

Travis T. Mallard, K. Paige Harden and Kim Fromme

Department of Psychology, University of Texas at Austin, 108 E. Dean Keeton Stop A8000, Austin, TX 78712, USA

**Abstract**

**Background.** Emerging adulthood is a peak period of risk for alcohol and illicit drug use. Recent advances in psychiatric genetics suggest that the co-occurrence of substance use and psychopathology arises, in part, from a shared genetic etiology. We sought to extend this research by investigating the influence of genetic risk for schizophrenia on trajectories of four substance use behaviors as they occurred across emerging adulthood.

**Method.** Young adult participants of non-Hispanic European descent provided DNA samples and completed daily reports of substance use for 1 month per year across 4 years ( $N = 30\,085$  observations of  $N = 342$  participants). A schizophrenia polygenic score was included in two-level hierarchical linear models designed to test associations between genetic risk for schizophrenia, participant age, and four substance use phenotypes.

**Results.** Participants with a greater schizophrenia polygenic score experienced greater age-related increases in the likelihood of using substances across emerging adulthood ( $p < 0.005$ ). Additionally, our results suggest that the polygenic score was positively associated with participants' overall likelihood to engage in illicit drug use but not alcohol-related substance use.

**Conclusions.** This study used a novel combination of polygenic prediction and intensive longitudinal methods to characterize the influence of genetic risk for schizophrenia on patterns of age-related change in substance use across emerging adulthood. Results suggest that genetic risk for schizophrenia has developmentally specific effects on substance use behaviors in a non-clinical population of young adults.

**Introduction**

Emerging adulthood, which spans the ages of 18–25 years, is a peak developmental period for the initiation and escalation of alcohol and drug use (Kendler *et al.*, 2008; Johnston *et al.*, 2011). Approximately 75% of lifetime cases of substance use disorders develop by the mid-to late-20s (Christie *et al.*, 1988; Kessler *et al.*, 2005, 2007), and problematic substance use in this period often co-occurs with other forms of psychopathology (Grant *et al.*, 2015, 2016). Psychiatric comorbidity increases risk for negative health outcomes, contributing significantly to the morbidity and mortality associated with alcohol and drug use (Whiteford *et al.*, 2013; Johnson *et al.*, 2014). Whereas early research hypothesized that high rates of comorbidity between psychopathology and substance use reflected self-medication behaviors (i.e. efforts to alleviate distress engendered by schizophrenia symptoms), recent research has generated additional theories of comorbidity.

Advances in psychiatric genetics suggest that the co-occurrence of substance use and other mental health problems is due, in part, to a shared genetic etiology (Polimanti *et al.*, 2017). While a portion of the underlying genetic etiology of substance use may specifically increase liability for alcohol and/or drug use *per se*, other genetic risk factors for substance use may also be related to psychopathology more broadly (Johnson *et al.*, 2009; Caspi *et al.*, 2014; Pettersson *et al.*, 2016). Given the substantial heritability and polygenicity of substance use behaviors (Gratten *et al.*, 2014; Polderman *et al.*, 2015), it has been posited that some genetic variants dually confer risk for substance use and psychopathology, perhaps influencing biological pathways common to multiple psychiatric conditions (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013; Network and Pathway Analysis Subgroup of the Psychiatric Genomics Consortium, 2015). Indeed, twin and family studies have reported that substance use behaviors arise from a heterogeneous etiology comprised of multiple genetic factors (Kendler *et al.*, 2003, 2012).

Recently, polygenic scores have been used to examine shared, cross-trait genetic influences on several psychiatric phenotypes (Krapohl *et al.*, 2016). Polygenic scores 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 this approach leverages the results from well-powered GWASs, it is well-suited to the investigation of aggregate genetic effects with modest sample sizes (Belsky and Israel, 2014). Here, we apply this method in a university sample of emerging adults, where we examine the extent to which a schizophrenia polygenic score influences trajectories of alcohol and illicit drug use.

Our focus on genetic risk for schizophrenia is motivated by evidence suggesting that schizophrenia and substance use share a portion of their underlying genetic architecture (Polimanti *et al.*, 2017). For instance, recent studies have found that schizophrenia has modest but significant genetic correlations with cannabis use (Pasman *et al.*, 2018), alcohol use (Clarke *et al.*, 2017), and risk preferences (Karlsson Linnér *et al.*, 2018). Similarly, several cross-sectional studies have reported that schizophrenia polygenic scores predict alcohol, amphetamine, cannabis, cocaine, opioid, and sedative use disorders (Power *et al.*, 2014; Carey *et al.*, 2016; Kalsi *et al.*, 2016; Hartz *et al.*, 2017; Reginsson *et al.*, 2017; Verweij *et al.*, 2017; Gurriarán *et al.*, 2018). However, while previous studies have related genetic risk for schizophrenia to diagnosed substance use disorders or lifetime substance use, no study has considered how this genetic risk functions in the context of development: *when* does genetic risk for schizophrenia influence substance use?

The growing support for a shared genetic architecture between schizophrenia and substance use spans numerous studies employing various methodologies. However, these studies were cross-sectional and used either lifetime history or diagnostic phenotypes. In the present paper, we sought to extend this research through a person-centered, high-resolution, longitudinal approach that investigates the effect of genetic risk for schizophrenia on substance use as it occurred in the daily lives of emerging adults. To accomplish this aim, we collected daily self-report data related to substance use across a 4-year period ( $N = 30\,085$  observations,  $M = 87.97$  observations per person). We then extended polygenic prediction methods to event-level phenotypes, which increase measurement precision of behavior in the natural environment and can characterize within-person patterns of variation (Molenaar and Campbell, 2009). Finally, we constructed a hierarchical linear model (HLM) to test whether genetic risk shapes how substance use *changes* across emerging adulthood, a developmental period in which genetic risks associated with schizophrenia and substance use often manifest (Kessler *et al.*, 2007).

To calculate a schizophrenia polygenic score, we used results from the Psychiatric Genomic Consortium's (PGC) most recent GWAS of the disorder (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). We then investigated the effect of genetic risk for schizophrenia on four event-level phenotypes: daily alcohol use, binge drinking, illicit drug use, and concurrent alcohol and drug use. Specifically, we tested: (i) whether the schizophrenia polygenic score predicted an individual's overall likelihood to engage in substance use on a given day, and (ii) whether the schizophrenia polygenic score predicted the magnitude of longitudinal, age-related change in substance use. In accordance with previous research, we hypothesized that genetic risk for schizophrenia would be positively associated with all forms of substance use. Furthermore, given that schizophrenia often onsets between late adolescence and early adulthood, we hypothesized that genetic risk for schizophrenia would be associated with a greater likelihood to use substances as participants grew older. In testing these hypotheses, we hope to lend insight into the heterogeneous genetic etiology of substance use behaviors and when they manifest in development.

## Materials and methods

### Participants

The present sample was recruited from a larger cohort of subjects who participated in a longitudinal investigation of alcohol abuse and behavioral risks among college students. Recruitment procedures for the full study have been described in previously published papers (Fromme *et al.*, 2008; Ashenhurst *et al.*, 2015; Mallard *et al.*, 2018). A subset of the full sample completed a daily monitoring protocol and provided DNA for genotyping procedures ( $n = 541$ , 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 ( $n = 354$ , 66% female). Twelve participants were excluded from analyses following the quality control procedures described below (final  $N = 342$ , 66% female,  $M_{\text{age}} = 18.44$  years,  $s.d._{\text{age}} = 0.32$  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 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 (Anderson *et al.*, 2010; Turner *et al.*, 2011). Samples were excluded from statistical analyses because of poor call rate ( $<98\%$ ), inconsistent self-reported sex and biological sex, and relatedness ( $\hat{p} > 0.125$ ). SNPs were excluded from analyses if more than 2% of genotype data was missing. Thresholds for minor allele frequency (MAF) and Hardy-Weinberg equilibrium (HWE) were applied after phasing and imputation (described below), as variant-level filtering has been shown to have a detrimental effect on imputation quality (Roshyara *et al.*, 2014).

Finally, although the present analyses were limited to participants who self-reported non-Hispanic European descent, flashPCA2 was used to (i) extract the top ten genomic principal components of ancestry and (ii) identify ancestral outliers (i.e. participants with a greater level of admixture than reported). First, principal components of ancestry were estimated using the European samples from Phase 3 v5 of the 1000 Genomes Project (1000 Genomes Project Consortium *et al.*, 2015) as a reference sample. Outliers were then defined as any participant with a score greater than or equal to four standard deviations from the mean on the first and/or second principal component of ancestry (i.e. the range present in European samples from Phase 3 v5 of the 1000 Genomes Project); five participants met this exclusion criterion. Scatterplots of the principal component scores were then examined to confirm that no ancestral outliers remained in the sample.

### Imputation

Unknown genotypes were imputed on the *Michigan Imputation Server* (<https://imputationserver.sph.umich.edu>). Variants were phased with Eagle v2.3 (Loh *et al.*, 2016) and imputed with Minimac3 1.0.13 (Das *et al.*, 2016), using Phase 3 v5 of the 1000 Genomes Project (1000 Genomes Project Consortium *et al.*, 2015) as a reference panel. To ensure all markers were of

high quality, several post-imputation quality control thresholds were applied. After phasing and imputation, SNPs with a MAF <0.01, INFO score <0.90, or HWE  $p$  value <0.00001 were excluded from all statistical analyses. These procedures yielded a final set of 5 250 123 high-quality genotyped and imputed variants.

### Schizophrenia polygenic score

A schizophrenia polygenic score was calculated for 342 unrelated participants of non-Hispanic European ancestry by using summary statistics from the PGCs 2014 GWAS of schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). Specifically, summary statistics were obtained for the 15 358 497 variants analyzed in PGC cohorts of European ancestry, which consisted of 32 405 cases, 42 221 controls, and 1235 trios. These variants were restricted to 4 509 191 bi-allelic SNPs that were present in both datasets after the quality control procedures described above. Linkage disequilibrium (LD)-based clumping was then used to identify a set of 121 702 independent SNPs ( $r^2 < 0.1$  in the present sample) with the lowest  $p$  value in a given 1 Mb window. An additional LD threshold was imposed to ensure that these independent SNPs were not in long-range LD with each other ( $r^2 > 0.1$  within a 10 Mb window). This process identified a final set of 118 719 independent SNPs to be used for the polygenic score.

Before calculating the schizophrenia polygenic score, the odds ratios (ORs) reported by the PGC were log-transformed to identify the beta coefficients associated with the effect allele for each SNP. PLINK 1.9 (Chang *et al.*, 2015) was then used to calculate a polygenic score for each participant by multiplying the number of effect alleles (0, 1, or 2) at a given SNP by its associated beta coefficient and summing across all included SNPs. Finally, the schizophrenia polygenic score was  $z$ -standardized to aid interpretation of results, establishing a mean of 0 and a standard deviation (S.D.) of 1.

### Longitudinal event-level design and phenotyping

Participants completed up to 30 consecutive days of online self-monitoring in each of their first 4 years of college. At the beginning of the study, a random sample of 200 students was invited to participate in a daily monitoring study. A random selection of 40–43 students thereafter 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.

Each day, participants answered questions about the previous day related to time-varying characteristics (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. Four event-level substance use phenotypes were assessed: any alcohol use, binge drinking, illicit drug use, and concurrent alcohol and drug use. Operant definitions for these substance use phenotypes are presented below.

- Alcohol use was defined as consuming at least one standard drink during the reporting day.

- Binge drinking was defined as consuming alcohol at a rate of 2 or 2.5 standard drinks per hour for at least 2 h (i.e. equivalent to the NIAAA definition of four or five drinks within a 2-h period, depending on sex).
- Illicit drug use was defined as consuming any illicit drugs during the reporting day.
- Concurrent alcohol and drug use was defined as simultaneously consuming alcohol and drugs during the reporting day.

Additionally, the self-monitoring website recorded the time and date of each daily report, which was used to determine the participant's age (rounded to two decimal points) on a given day. This approach allowed us to model age as a continuous event-level predictor that varied within a 30-day reporting period (e.g. increasing 17.92 to 18.05 during the first reporting period), as well as between reporting periods (e.g. increasing from 18.05 to 18.92 between the first and second reporting periods). 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 30 085 event-level observations from 342 participants.

### Analytic approach

A two-level HLM (Raudenbush and Bryk, 2002) with robust standard errors was used to analyze the relationships between the schizophrenia polygenic score (PGS<sub>SCZ</sub>), participant age (AGE), and the four substance use phenotypes. Events were nested within participants for all statistical analyses. As between-person and within-person relationships are not necessarily synonymous (Molenaar and Campbell, 2009), the HLM included a random intercept and random slope to account for individual differences in the overall level of substance use and rate of age-related change in substance use, respectively. Principal components of ancestry (PC<sub>1</sub> ... PC<sub>10</sub>), biological sex (SEX), and age at beginning of college (AGE<sub>W1</sub>) were included as trait-level covariates in all analyses. The full model is described below.

#### Level 1 model

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

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

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

#### Level 2 model

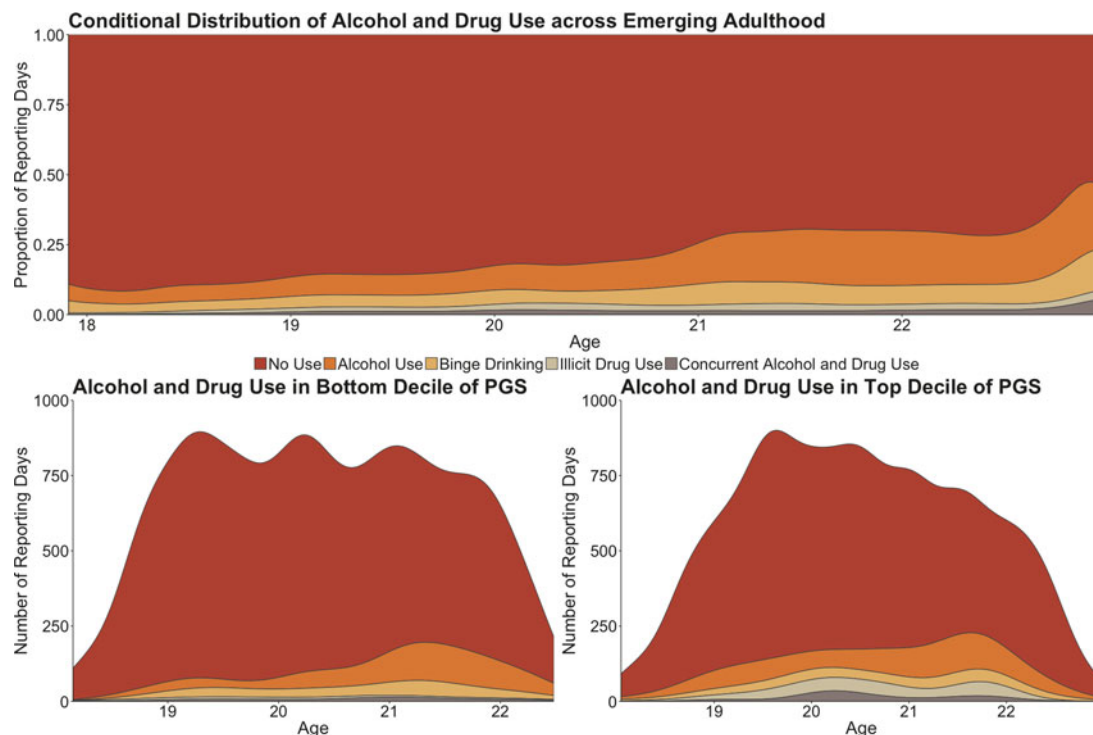
$$\pi_0 = \beta_{00} + \beta_{01}(\text{PGS}_{\text{SCZ}}) + \beta_{02...011}(\text{PC}_1 \dots \text{PC}_{10}) + \beta_{012}(\text{SEX}) + \beta_{013}(\text{AGE}_{\text{W1}}) + r_0$$

$$\pi_1 = \beta_{10} + \beta_{11}(\text{PGS}_{\text{SCZ}}) + \beta_{12...111}(\text{PC}_1 \dots \text{PC}_{10}) + \beta_{112}(\text{SEX}) + r_1$$

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 ( $\pi_0$ ) and a person-centered random slope describing within-person variability in the likelihood of using substances as a function of event-level age ( $\pi_1$ ). Importantly, event-level age was centered on the person mean

**Table 1.** Descriptive statistics for each substance use phenotype stratified by reporting year

	Year 1 ( <i>n</i> = 317)		Year 2 ( <i>n</i> = 321)		Year 3 ( <i>n</i> = 309)		Year 4 ( <i>n</i> = 236)		Total ( <i>N</i> = 342)	
	Non-abstainers	Events	Non-abstainers	Events	Non-abstainers	Events	Non-abstainers	Events	Non-abstainers	Events
Alcohol use	198	998	220	1326	246	1824	205	1522	316	5670
Binge drinking	120	371	145	504	163	591	142	434	254	1900
Illicit drug use	51	163	57	329	48	292	33	192	103	976
Concurrent alcohol and drug use	28	68	43	132	35	112	29	85	79	397

**Fig. 1.** In the top panel, a conditional density plot illustrates how patterns of substance use changed across development in the entire sample. The X-axis of the plot represents the age of participants at each event-level observation, while the Y-axis represents the proportion of reporting days the substance use occurred for that age interval. In the bottom panels, two density plots illustrate how patterns of substance use vary by genetic risk for schizophrenia. Here, the X-axis again represents the age of participants at each event-level observation, but the Y-axis now represents the actual number of reporting days that substance use occurred for that age interval. PGS, polygenic score.

and thus reflects within-person, age-related change in substance use over time (Raudenbush and Bryk, 2002). Overall, the Level 1 equation tested the extent to which a person showed systematic age-related change in substance use.

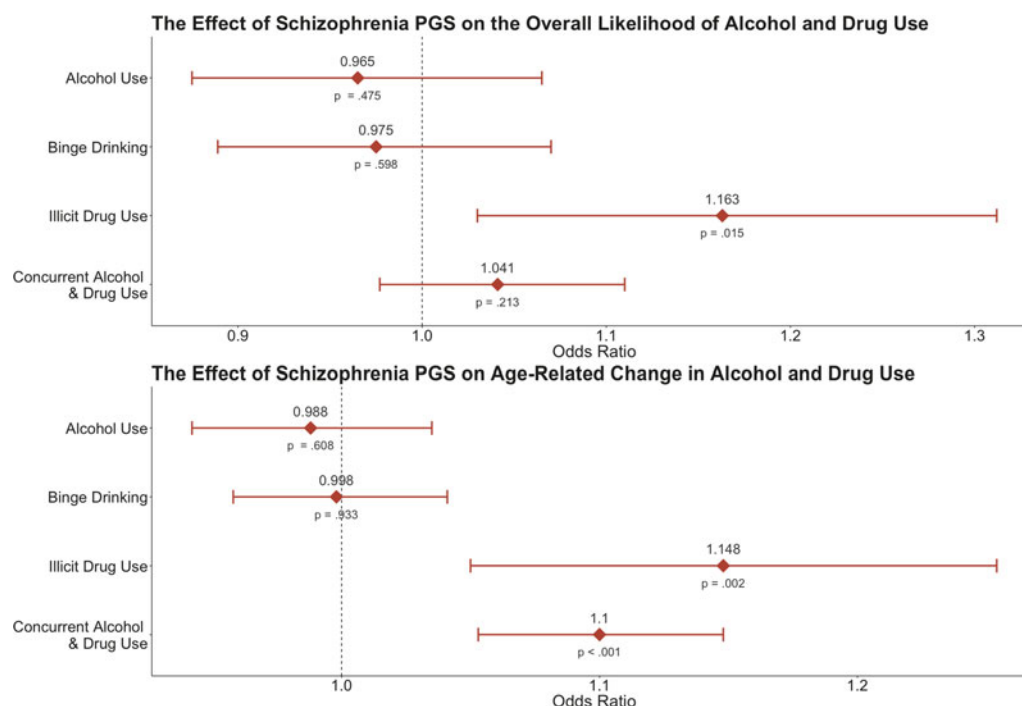
The Level 2 (person level) equation modeled between-person variability in the likelihood to use substances when aggregating across all occasions. Here, the intercept for substance use phenotypes ( $\pi_0$ ), which represents the person-average likelihood to engage in substance use across all events, was modeled as a function of the effect of the schizophrenia polygenic score ( $\beta_{01}$ ), as well as the effects of ancestry ( $\beta_{02} \dots \beta_{11}$ ), sex ( $\beta_{12}$ ), and age at first wave of data collection ( $\beta_{13}$ ). We additionally modeled the random slopes for event-level age as a function of the effects of the schizophrenia polygenic score ( $\beta_{11}$ ), ancestry ( $\beta_{12} \dots \beta_{111}$ ), and sex ( $\beta_{112}$ ). The first 10 principal components of ancestry and age at first wave were centered on the grand mean, while the polygenic score 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 the schizophrenia polygenic score, sex, and mean age predicted (i) participants' overall likelihood to use substances when aggregating across all events and (ii) age-related changes in the likelihood to use substances as participants grew older.

## Results

Descriptive statistics for each substance use phenotype are presented stratified by year in Table 1, while the longitudinal distributions for each phenotype are illustrated in Fig. 1. Throughout the course of the study, alcohol consumption and binge drinking were reported at least once by 92.4% and 74.3% of the sample, respectively, while illicit drug use and concurrent alcohol and drug use were reported by 30.1% and 23.1% of the sample,





**Fig. 2.** ORs illustrating the effect of the schizophrenia polygenic score on the intercept (i.e. overall likelihood) and age-related slope (i.e. age-related change) of the four substance use phenotypes. The bars for each estimate reflect the 95% confidence interval, while the corresponding *p* value is listed below each point. PGS, polygenic score.

respectively. The number of reporting days that each participant completed was not associated with age or the polygenic scores for schizophrenia, but males completed fewer daily monitoring reports ( $r = -0.183$ ,  $p = 0.001$ ). Per recent reflections on statistical power and reproducibility (Benjamin *et al.*, 2018), and because we were effectively conducting a series of eight regression models (intercepts and slopes as outcomes for four phenotypes), we interpret results at  $p \leq 0.05$  as suggestive and results at  $p \leq 0.005$  as significant. This significance threshold is slightly more conservative than a Bonferroni-corrected threshold for eight tests (which would be  $p \leq 0.00625$ ).

Effects of the schizophrenia polygenic score on the intercept and slope of all four substance use phenotypes are presented in Fig. 2 and Table 2. Here, we represent the effects of the polygenic score as ORs, which reflect change in the odds of an outcome given a one unit increase in the predictor. Moreover, as we standardized the schizophrenia polygenic score prior to analysis, we characterize how a 1 S.D. increase in genetic risk influences the likelihood to use substances on any given reporting day.

To briefly summarize the results reported in Table 2, we observed significant associations between the schizophrenia polygenic score and both illicit drug use and concurrent alcohol use and drug use; however, the schizophrenia polygenic score was not associated with likelihood to engage in alcohol use or binge drinking alone. Additionally, biological sex was not associated with any form of substance use, but age at the beginning of college (i.e. between-person differences in age) showed a positive suggestive association with all types of use (all  $p < 0.05$ ; online Supplementary Table S1). Getting older over the course of the study (i.e. event-level age) was also positively associated with a greater likelihood to engage in all forms of substance use (all  $p < 0.001$ ; Fig. 1; online Supplementary Tables S1–S3). The specific results for the drug use and concurrent alcohol and drug use phenotypes are described in detail below.

**Table 2.** Effect of the schizophrenia polygenic score on the intercept and age-related slope of each substance use phenotype

	<i>B</i>	OR (95% CI)	<i>p</i> Value
Intercept			
Alcohol use	−0.036	0.965 (0.875–1.065)	0.475
Binge drinking	−0.025	0.975 (0.889–1.070)	0.598
Illicit drug use	0.151	1.163 (1.030–1.312)	0.015 <sup>†</sup>
Concurrent alcohol and drug use	0.040	1.041 (0.977–1.110)	0.213
Slope			
Alcohol use	−0.012	0.988 (0.942–1.035)	0.608
Binge drinking	−0.002	0.998 (0.958–1.041)	0.933
Illicit drug use	0.138	1.148 (1.050–1.254)	0.002*
Concurrent alcohol and drug use	0.095	1.100 (1.053–1.148)	<0.001*

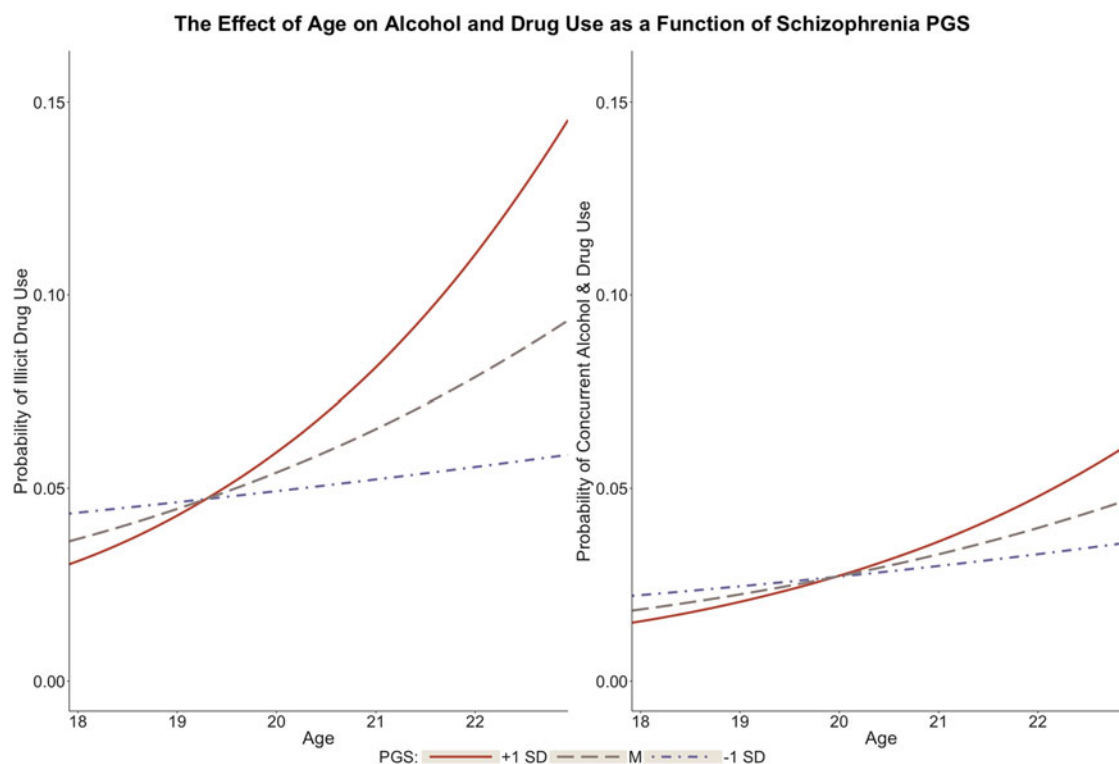
OR, odds ratio; CI, confidence interval.

<sup>†</sup>Suggestive of statistical significance at  $p \leq 0.05$ .

\*Statistically significant at  $p \leq 0.005$ .

### Illicit drug use

Results suggested that the schizophrenia polygenic score was associated with a greater overall likelihood to use illicit drugs ( $B = 0.151$ , OR = 1.163,  $p = 0.015$ ). Here, a 1 S.D. increase in genetic risk for schizophrenia was associated with a relative 16.3% increase in the likelihood to engage in illicit drug use on any given day across the entire study. Furthermore, results indicated that age-related changes in illicit drug use varied as a function of the polygenic score: higher genetic risk was significantly associated with the event-level slope between age and



**Fig. 3.** The effects of event-level age on the likelihood to engage in illicit drug use and concurrent alcohol and drug use as a function of the schizophrenia polygenic score, as measured in the current study. In both cases, we see that greater genetic risk for schizophrenia predicts a more substantial increase in age-related substance use. PGS, polygenic score; M, mean, s.d., standard deviation.

illicit drug use ( $B = 0.138$ ,  $OR = 1.148$ ,  $p = 0.002$ ). So, participants with higher schizophrenia-associated genetic risk were more likely to use illicit substances overall *and* experienced a more substantial increase in the likelihood of using drugs as they grew older. This effect is illustrated in Fig. 3.

#### Concurrent alcohol and drug use

The schizophrenia polygenic score was not associated with a greater overall likelihood to engage in concurrent alcohol and drug use ( $B = 0.040$ ,  $OR = 1.041$ ,  $p > 0.05$ ), but it was associated with age-related change in concurrent alcohol and drug use (the event-level slope for age;  $B = 0.095$ ,  $OR = 1.010$ ,  $p < 0.001$ ). Participants with higher schizophrenia-associated genetic risk experienced a more substantial increase in the likelihood of concurrent alcohol and drug use as they grew older. This effect is illustrated in Fig. 3.

#### Discussion

The current paper describes the first longitudinal, event-level examination of genetic risks discovered in large-scale GWAS of schizophrenia and their relationship with daily substance use in a sample of university students. Specifically, we tested whether a genome-wide polygenic score measuring schizophrenia-associated genetic risk predicted a greater overall likelihood to engage in substance use on a given day and within-person age-related changes in substance use. We report two major findings. First, we found suggestive evidence that genetic risk for schizophrenia predicted an individual's overall likelihood to engage in illicit drug use, but it did not predict the likelihood that participants would engage in

any form of alcohol-related substance use. Second, we found that genetic risk for schizophrenia significantly predicted the rate of age-related change in illicit drug use and concurrent alcohol and drug use. Whereas many prior studies have only examined the effect of a schizophrenia polygenic score on substance use disorders or lifetime use outcomes, we identified genetic influences on substance use in the daily lives of emerging adults from a non-clinical sample. As a result, our findings corroborate and build upon recent studies that have reported associations between polygenic scores for schizophrenia and problematic substance use (Power *et al.*, 2014; Carey *et al.*, 2016; Kalsi *et al.*, 2016; Hartz *et al.*, 2017; Reginsson *et al.*, 2017; Verweij *et al.*, 2017; Gurriarán *et al.*, 2018).

Although our daily measure of illicit drug use did not identify the specific substance that was consumed, related investigations of this cohort have identified cannabis as the most commonly used illicit drug (Fromme *et al.*, 2008). Schizophrenia and cannabis use share a modest but significant genetic correlation ( $r_g = 0.24$ ; Pasman *et al.*, 2018). Researchers have recently begun to interrogate the relationship between these two phenotypes, finding that genetic risk for schizophrenia exerts a causal influence on the liability to use cannabis (Pasman *et al.*, 2018). One possibility is that genetic variants that confer risk for schizophrenia also influence cannabis use by impacting some shared pathophysiology (Chambers *et al.*, 2001; Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013; Network and Pathway Analysis Subgroup of the Psychiatric Genomics Consortium, 2015). Alternatively, individuals with a higher polygenic loading for schizophrenia may experience prodromal symptoms or neurocognitive impairment that leads them to use cannabis (i.e. a graded iteration of the 'self-medication' hypothesis). Our results indicate

that genetic risk for schizophrenia begins to influence substance use during emerging adulthood, suggesting that studying this developmental period may be critical to disentangling the complex relationship between these two phenotypes.

Interestingly, the schizophrenia polygenic score did not predict phenotypes that only involved 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 (Clarke *et al.*, 2017). Given the relatively weak genetic correlation between schizophrenia and alcohol consumption ( $r_g = 0.13$ ), 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 (Edwards and Kendler, 2013). 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 (Kendler *et al.*, 2008; van Beek *et al.*, 2012).

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, the monthly rates of alcohol and drug use observed in this study are quite similar to those reported by college students in the *Monitoring the Future* study (Johnston *et al.*, 2011), so we are still able to generate insight into general patterns of substance use. Second, our analyses were restricted 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 current study are partially attenuated by the fact that (i) we were well-powered for our within-person approach, (ii) we leveraged *a priori* effect size estimates from a well-powered GWAS of schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), and (iii) we examined aggregate genomic variation rather than individual SNPs of small effect. The fourth limitation is that we cannot test whether observed associations operate through the experience of psychiatric symptoms that may precede or co-occur with substance use, or whether age-related escalation of illicit drug use would be apparent even among those with zero psychiatric symptoms. Future research with longitudinal measurements of the co-occurrence between substance use behaviors and psychiatric symptoms could help clarify more precisely how genetic risk for schizophrenia influences substance use behaviors and their change over time.

Despite these limitations, the strength of this study is its novel combination of genome-wide data with high-resolution phenotyping. As the first longitudinal, event-level investigation of a schizophrenia polygenic score and its association with substance use in daily life, this work provides ecologically valid evidence that these two psychiatric conditions are influenced, in part, by shared genetic factors. Notably, this study demonstrated that genetic risk for schizophrenia can predict important behavioral phenotypes in a sample of non-clinical 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 current study contributes to the broader literature by illustrating how combining polygenic prediction and intensive longitudinal methods can be used to characterize broad and developmentally specific effects of genetic variation. As repeated event-level measurement can be used to examine concurrent dynamic processes, person-centered approaches may facilitate greater insight into the specific temporal dynamics or causal relationships of co-occurring phenomena as they unfold in the life of an individual person. For instance, future studies that employ similar methods may be uniquely poised to elucidate the schizophrenia-cannabis use association during emerging adulthood by simultaneously assessing daily experiences with prodromal symptoms and cannabis use. In doing so, future studies using polygenic prediction methods will be better suited to investigate *when* and *how* genotypic differences contribute to complex human behavior.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/S0033291718002817>

## References

- 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA and Abecasis GR (2015). A global reference for human genetic variation. *Nature* 526, 68–74.
- Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP and Zondervan KT (2010). Data quality control in genetic case-control association studies. *Nature Protocols* 5, 1564–1573.
- Ashenhurst JR, Harden KP, Corbin WR and Fromme K (2015). Trajectories of binge drinking and personality change across emerging adulthood. *Psychology Of Addictive Behaviors* 29, 978–991.
- Belsky DW and Israel S (2014). Integrating genetics and social science: genetic risk scores. *Biodemography and Social Biology* 60, 137–155.
- Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers E-J, Berk R, Bollen KA, Brembs B, Brown L, Camerer C, Cesarini D, Chambers CD, Clyde M, Cook TD, De Boeck P, Dienes Z, Dreber A, Easwaran K, Efferson C, Fehr E, Fidler F, Field AP, Forster M, George EI, Gonzalez R, Goodman S, Green E, Green DP, Greenwald AG, Hadfield JD, Hedges LV, Held L, Hua Ho T, Hoijtink H, Hruschka DJ, Imai K, Imbens G, Ioannidis JPA, Jeon M, Jones JH, Kirchner M, Laibson D, List J, Little R, Lupia A, Machery E, Maxwell SE, McCarthy M, Moore DA, Morgan SL, Munafó M, Nakagawa S, Nyhan B, Parker TH, Pericchi L, Perugini M, Rouder J, Rousseau J, Savalei V, Schönbrodt FD, Sellke T, Sinclair B, Tingley D, Van Zandt T, Vazire S, Watts DJ, Winship C, Wolpert RL, Xie Y, Young C, Zinman J and Johnson VE (2018). Redefine statistical significance. *Nature Human Behaviour* 2, 6–10.
- Carey CE, Agrawal A, Bucholz KK, Hartz SM, Lynskey MT, Nelson EC, Bierut LJ and Bogdan R (2016). Associations between polygenic risk for psychiatric disorders and substance involvement. *Frontiers in Genetics* 7, 1–10.
- Caspi A, Houts RM, Belsky DW, Goldman-Mellor SJ, Harrington H, Israel S, Meier MH, Ramrakha S, Shalev I, Poulton R and Moffitt TE (2014). The *p* factor: one general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science* 2, 119–137.
- Chambers RA, Krystal JH and Self DW (2001). A neurobiological basis for substance abuse comorbidity in schizophrenia. *Biological Psychiatry* 50, 71.
- Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM and Lee JJ (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4, 1–16.
- Christie KA, Burke JD, Regier DA, Rae DS, Boyd JH and Locke BZ (1988). Epidemiologic evidence for early onset of mental disorders and higher risk of drug abuse in young adults. *American Journal of Psychiatry* 145, 971–975.
- Clarke T-K, Adams MJ, Davies G, Howard DM, Hall LS, Padmanabhan S, Murray AD, Smith BH, Campbell A, Hayward C, Porteous DJ, Deary IJ

- and McIntosh AM (2017). Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank ( $N = 112117$ ). *Molecular Psychiatry* 22, 1376–1384.
- Cross-Disorder Group of the Psychiatric Genomics Consortium (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *The Lancet* 381, 1371–1379.
- Das S, Forer L, Schönerr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, Schlessinger D, Stambolian D, Loh PR, Iacono WG, Swaroop A, Scott LJ, Cucca F, Kronenberg F, Boehnke M, Abecasis GR and Fuchsberger C (2016). Next-generation genotype imputation service and methods. *Nature Genetics* 48, 1284–1287.
- Edwards AC and Kendler KS (2013). Alcohol consumption in men is influenced by qualitatively different genetic factors in adolescence and adulthood. *Psychological Medicine* 43, 1857–1868.
- Fromme K, Corbin WR and Kruse MI (2008). Behavioral risks during the transition from high school to college. *Developmental Psychology* 44, 1497–1504.
- Grant BF, Goldstein RB, Saha TD, Chou SP, Jung J, Zhang H, Pickering RP, Ruan WJ, Smith SM, Huang B and Hasin DS (2015). Epidemiology of DSM-5 alcohol use disorder: results from the national epidemiologic survey on alcohol and related conditions-III. *JAMA Psychiatry* 72, 757.
- Grant BF, Saha TD, Ruan WJ, Goldstein RB, Chou SP, Jung J, Zhang H, Smith SM, Pickering RP, Huang B and Hasin DS (2016). Epidemiology of DSM-5 drug use disorder: results from the national epidemiologic survey on alcohol and related conditions-III. *JAMA Psychiatry* 73, 39.
- Gratten J, Wray NR, Keller MC and Visscher PM (2014). Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nature Neuroscience* 17, 782–790.
- Gurriarán X, Rodríguez-López J, Flórez G, Pereiro C, Fernández JM, Fariñas E, Estévez V, Group GS, Arrojo M and Costas J (2018). Relationships between substance abuse/dependence and psychiatric disorders based on polygenic scores. *Genes, Brain and Behavior*, e12504.
- Hartz SM, Horton A, Oehlert M, Carey CE, Agrawal A, Bogdan R, Chen L-S, Hancock DB, Johnson EO, Pato C, Pato M, Rice JP and Bierut LJ (2017). Association between substance use disorder and polygenic liability to schizophrenia. *Biological Psychiatry* 82, 709–715.
- Johnson W, Hicks BM, McGue M and Iacono WG (2009). How intelligence and education contribute to substance use: hints from the Minnesota Twin family study. *Intelligence* 37, 613–624.
- Johnson NB, Hayes LD, Brown K, Hoo EC and Ethier KA (2014). CDC national health report: leading causes of morbidity and mortality and associated behavioral risk and protective factors – United States, 2005–2013. *MMWR. Surveillance Summaries* 63(Suppl 4), 3–27.
- Johnston LD, O'malley PM, Bachman JG and Schulenberg JE (2011). Monitoring the future national survey results on drug use, 1975–2010. Volume II, college students & adults ages 19–50. *Institute for Social Research*.
- Kalsi E, Euesden J, Coleman JRI, Ducci F, Aliev F, Newhouse SJ, Liu X, Ma X, Wang Y, Collier DA, Asherson P, Li T and Breen G (2016). Genome-wide association of heroin dependence in Han Chinese. *PLoS Genetics* 11, e0167388.
- Karlsson Linnér R, Biroli P, Kong E, Meddens SFW, Wedow R, Fontana MA, Lebreton M, Abdellaoui A, Hammerschlag AR, Nivard MG, Okbay A, Rietveld CA, Timshel PN, Tino SP, Trzaskowski M, de Vlaming R, Zünd CL, Bao Y, Buzdugan L, Caplin AH, Chen C-Y, Eibich P, Fontanillas P, Gonzalez JR, Joshi PK, Karhunen V, Kleinman A, Levin RZ, Lill CM, Meddens GA, Muntané G, Sanchez-Roige S, van Rooij FJ, Taskesen E, Wu Y, Zhang F, Team 23andMe Research, eQTLgen Consortium, International Cannabis Consortium, Psychiatric Genomics Consortium, Social Science Genetic Association Consortium, Auton A, Boardman JD, Clark DW, Conlin A, Dolan CC, Fischbacher U, Groenen PJ, Harris KM, Hasler G, Hofman A, Ikram MA, Jain S, Karlsson R, Kessler RC, Kooyman M, MacKillop J, Männikkö M, Morcillo-Suarez C, McQueen MB, Schmidt KM, Smart MC, Sutter M, Thurik AR, Uitterlinden AG, White J, de Wit H, Yang J, Bertram L, Boomsma D, Esko T, Fehr E, Hinds DA, Johannesson M, Kumari M, Laibson D, Magnusson PK, Meyer MN, Navarro A, Palmer AA, Pers TH, Posthuma D, Schunk D, Stein MB, Svento R, Tiemeier H, Timmers PR, Turley P, Ursano RJ, Wagner GG, Wilson JF, Gratten J, Lee JJ, Cesarini D, Benjamin DJ, Koellinger P and Beauchamp JP (2018). Genome-wide study identifies 611 loci associated with risk tolerance and risky behaviors. *bioRxiv*. <https://doi.org/10.1101/261081>
- Kendler KS, Prescott CA, Myers J and Neale MC (2003). The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Archives of General Psychiatry* 60, 929.
- Kendler KS, Schmitt E, Aggen SH and Prescott CA (2008). Genetic and environmental influences on alcohol, caffeine, cannabis, and nicotine use from early adolescence to middle adulthood. *Archives of General Psychiatry* 65, 674.
- Kendler KS, Aggen SH, Prescott CA, Crabbe J and Neale MC (2012). Evidence for multiple genetic factors underlying the DSM-IV criteria for alcohol dependence. *Molecular Psychiatry* 17, 1306–1315.
- Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR and Walters EE (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry* 62, 593.
- Kessler RC, Amminger GP, Aguilar-Gaxiola S, Alonso J, Lee S and Ustün TB (2007). Age of onset of mental disorders: a review of recent literature. *Current Opinion in Psychiatry* 20, 359–364.
- Krapohl E, Euesden J, Zabaneh D, Pingault J-B, Rimfeld K, von Stumm S, Dale PS, Breen G, O'Reilly PF and Plomin R (2016). Phenome-wide analysis of genome-wide polygenic scores. *Molecular Psychiatry* 21, 1188–1193.
- Loh P-R, Palamara PF and Price AL (2016). Fast and accurate long-range phasing in a UK Biobank cohort. *Nature Genetics* 48, 811–816.
- Mallard TT, Ashenhurst JR, Harden KP and Fromme K (2018). GABRA2, alcohol, and illicit drug use: an event-level model of genetic risk for polysubstance use. *Journal of Abnormal Psychology* 127, 190–201.
- Molenaar PCM and Campbell CG (2009). The new person-specific paradigm in psychology. *Psychological Science* 18, 112–117.
- Network and Pathway Analysis Subgroup of the Psychiatric Genomics Consortium (2015). Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nature Neuroscience* 18, 199–209.
- Pasman JA, Verweij KJH, Gerring Z, Stringer S, Sanchez-Roige S, Treur JL, Abdellaoui A, Nivard MG, Baselmans BML, Ong J-S, Ip HF, van der Zee MD, Bartels M, Day FR, Fontanillas P, Elson SL, Team the 23andMe Research, de Wit H, Davis LK, MacKillop J, Consortium IC, Derringer JL, Branje SJT, Hartman CA, Heath AC, van Lier PAC, Madden PAF, Magi R, Meuw W, Montgomery GW, Oldehinkel AJ, Pausova Z, Ramos-Quiroga JA, Paus T, Ribases M, Kaprio J, Boks MPM, Bell JT, Spector TD, Gelernter J, Boomsma DI, Martin NG, MacGregor S, Perry JRB, Palmer AA, Posthuma D, Munafò MR, Gillespie NA, Derks EM and Vink JM (2018). Genome-wide association analysis of lifetime cannabis use ( $N = 184\,765$ ) identifies new risk loci, genetic overlap with mental health, and a causal influence of schizophrenia on cannabis use. *bioRxiv*. <https://doi.org/10.1101/234294>
- Pettersson E, Larsson H and Lichtenstein P (2016). Common psychiatric disorders share the same genetic origin: a multivariate sibling study of the Swedish population. *Molecular Psychiatry* 21, 717–721.
- Polderman TJC, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM and Posthuma D (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics* 47, 702–709.
- Polimanti R, Agrawal A and Gelernter J (2017). Schizophrenia and substance use comorbidity: a genome-wide perspective. *Genome Medicine* 9, 25.
- Power RA, Verweij KJH, Zuhair M, Montgomery GW, Henders AK, Heath AC, Madden PAF, Medland SE, Wray NR and Martin NG (2014). Genetic predisposition to schizophrenia associated with increased use of cannabis. *Molecular Psychiatry* 19, 1201–1204.
- Raudenbush SW and Bryk AS (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd Edn. London: Sage.
- Reginsson GW, Ingason A, Euesden J, Bjornsdottir G, Olafsson S, Sigurdsson E, Oskarsson H, Tyrfinsson T, Runarsdottir V, Hansdottir I, Steinberg S, Stefansson H, Gudbjartsson DF, Thorgerisson TE and Stefansson K



- (2017). Polygenic risk scores for schizophrenia and bipolar disorder associate with addiction. *Addiction Biology* **23**, 485–492.
- Roshyara NR, Kirsten H, Horn K, Ahnert P and Scholz M (2014). Impact of pre-imputation SNP-filtering on genotype imputation results. *BMC Genetics* **15**, 88.
- Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427.
- Turner S, Armstrong LL, Bradford Y, Carlson CS, Dana C, Crenshaw AT, De Andrade M, Doheny KE, Jonathan L, Hayes G, Jarvik G, Jiang L, Kullo IJ, Li R, Manolio TA, Matsumoto M, Mccarty CA, Andrew N, Mirel DB, Paschall JE, Pugh EW, Luke V, Wilke RA, Zuvich RL and Ritchie MD (2011). Quality control procedures for genome wide association studies. *Current Proceedings in Human Genetics* **68**, 1–24.
- van Beek JHDA, Kendler KS, de Moor MHM, Geels LM, Bartels M, Vink JM, van den Berg SM, Willemsen G and Boomsma DI (2012). Stable genetic effects on symptoms of alcohol abuse and dependence from adolescence into early adulthood. *Behavior Genetics* **42**, 40–56.
- Verweij KJH, Abdellaoui A, Nivard MG, Sainz Cort A, Ligthart L, Draisma HHM and Minică CC (2017). Short communication: genetic association between schizophrenia and cannabis use. *Drug and Alcohol Dependence* **171**, 117–121.
- Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, Charlson FJ, Norman RE, Flaxman AD, Johns N, Burstein R, Murray CJ and Vos T (2013). Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *The Lancet* **382**, 1575–1586.