Published in final edited form as:

Addiction. 2020 February ; 115(2): 337–346. doi:10.1111/add.14815.

Sibling Comparisons Elucidate the Associations between Educational Attainment Polygenic Scores and Alcohol, Nicotine, and Cannabis

Jessica E. Salvatore[1,2], **Peter B. Barr**[1], **Mallory Stephenson**[1], **Fazil Aliev**[1,3], **Sally I-Chun Kuo**[1], **Jinni Su**[4], **Arpana Agrawal**[5], **Laura Almasy**[6,7], **Laura Bierut**[5], **Kathleen Bucholz**[5], **Grace Chan**[8], **Howard J. Edenberg**[9], **Emma C. Johnson**[5], **Vivia V. McCutcheon**[5], **Jacquelyn L. Meyers**[10], **Marc Schuckit**[11], **Jay Tischfield**[12], **Leah Wetherill**[13], **Danielle M. Dick**[1,14,15]

[1]Department of Psychology, Virginia Commonwealth University, Box 842018, Richmond, VA 23284-2018.

[2]Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth, University, Box 980126, Richmond, VA 23298

[3]Department of Business Administration, Karabuk University, 78050 Karabuk, Turkey

[4]Department of Psychology, Arizona State University, Box 871104, Tempe, AZ 85287-1104

[5]Department of Psychiatry, Washington University in St. Louis, 660 S. Euclid, CB 8134, St., Louis, MO 63110

[6]Department of Genetics, University of Pennsylvania, 415 Curie Boulevard Philadelphia, PA, 19104-6145

[7]Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, 3615, Civic Center Blvd, ARC 1016-C, Philadelphia, PA 19104

[8]Department of Psychiatry, University of Connecticut School of Medicine, 263 Farmington, Avenue, Farmington, CT 06030-2103

[9]Department of Biochemistry and Molecular Biology, Indiana University, 635 Barnhill Dr.,, Indianapolis, IN 46202

[10]Department of Psychiatry, SUNY Downstate Medical Center, 450 Clarkson Avenue Brooklyn, NY 11203

[11]Department of Psychiatry, University of California-San Diego, 9500 Gilman Drive La Jolla,, CA 92093

[12]Department of Genetics and the Human Genetics Institute of New Jersey, 145 Bevier Road, Piscataway, NJ 08854-8082.

Correspondence concerning this article should be addressed to: Jessica E. Salvatore, Department of Psychology and the Virginia Institute for Psychiatric and Behavioral Genetics, VCU PO Box 842018, 806 West Franklin Street, Richmond, VA 23284-2018, USA. jesalvatore@vcu.edu. Telephone: +1 804-828-8132.

Conflicts of interest: None

[13]Department of Medical and Molecular Genetics, Indiana University, 410 W. 10th Street, Indianapolis, IN 46202

[14]Department of Human & Molecular Genetics, Virginia Commonwealth University, Box, 980033, Richmond, VA, USA 23298

[15]College Behavioral and Emotional Health Institute, Virginia Commonwealth University, Box, 842018 Richmond, VA, 23284

Abstract

Background and Aims.—The associations between low educational attainment and substance use disorders (SUDs) may be related to a common genetic vulnerability. We aimed to elucidate the associations between polygenic scores for educational attainment and clinical criterion counts for three SUDs (alcohol, nicotine, and cannabis).

Design.—Polygenic association and sibling comparison methods. The latter strengthens inferences in observational research by controlling for confounding factors that differ between families.

Setting.—Six sites in the United States.

Participants.—European ancestry participants 25 years of age and older from the Collaborative Study on the Genetics of Alcoholism (COGA). Polygenic association analyses included 5582 (54% female) participants. Sibling comparisons included 3098 (52% female) participants from 1226 sibling groups nested within the overall sample.

Measurements.—Outcomes included criterion counts for DSM-5 Alcohol Use Disorder (AUDSX), Fagerström Nicotine Dependence (NDSX), and DSM-5 Cannabis Use Disorder (CUDSX). We derived polygenic scores for educational attainment (*EduYears-GPS*) using summary statistics from a large (>1 million) genome-wide association study of educational attainment.

Findings.—In polygenic association analyses, higher *EduYears-GPS* predicted lower AUDSX, NDSX, and CUDSX ($p<0.01$, effect sizes (R^2) ranging from 0.30%−1.84%). These effects were robust in sibling comparisons, where sibling differences in *EduYears-GPS* predicted all three SUDs ($p<0.05$, R^2 0.13%−0.20%).

Conclusions.—Individuals who carry more alleles associated with educational attainment tend to meet fewer clinical criteria for alcohol, nicotine, and cannabis use disorders, and these effects are robust to rigorous controls for potentially confounding factors that differ between families (e.g., socioeconomic status, urban-rural residency, and parental education).

Keywords

alcohol; nicotine; cannabis; polygenic risk score; sibling comparisons; Collaborative Study on the Genetics of Alcoholism

---

Researchers have studied the associations between educational attainment and substance use disorders (SUDs) for more than a century [1, 2]. Cross-sectional studies consistently link use of tobacco, alcohol, and cannabis with high school dropout [2], and greater educational

attainment with lower rates of SUD diagnoses [3–5]. There is a substantial body of work exploring the hypotheses that SUDs influence early termination of education, and that early termination of education influences SUDs [6], with evidence supporting both temporal orderings [5, 7–10]. A third hypothesis is also plausible: that the associations between low educational attainment and SUDs are attributable at least in part to a common general vulnerability. Genetic factors represent one type of general vulnerability. Consistent with this possibility, genetic epidemiologic data indicate that there is a set of genetic factors that influence both low educational attainment and a higher likelihood of developing SUDs [11–14]. There is also evidence that familial factors confound the associations between educational attainment and multiple forms of substance use and dependence [6, 15], although the specific source of this familial confounding (i.e., genes or the rearing environment) was not specified in those studies.

Recent advances in characterizing the molecular genetic basis of complex traits and behaviors have stimulated interest in translating findings from genetic epidemiological studies, which use patterns of resemblance among individuals of known genetic relatedness to make inferences about latent genetic influences on traits and behaviors, into a molecular genetic framework [16, 17]. This is typically accomplished using a polygenic scoring approach, where researchers leverage genome-wide association results from large, well-powered discovery samples to calculate personalized indices of the weighted number of trait-associated alleles carried by each participant in an independent sample [18, 19]. In polygenic analyses, one examines the associations between these polygenic scores and other traits and behaviors to examine their shared genetic etiology.

In this study, we combined polygenic association and sibling comparison methods to elucidate the associations between polygenic scores for educational attainment [20] and clinical criterion counts for three common SUDs (alcohol, nicotine, and cannabis) in a sample of adults of European ancestry. Sibling comparisons [21–24] provide a complementary tool to clarify the nature of associations observed in polygenic analyses. Biological full siblings reared together share the same home environment and a substantial portion of their genetic variation (50% on average), allowing for control of measured and unmeasured familial factors such as socioeconomic status, religious upbringing, urban-rural residency, parental education, and familial polygenic load, that are also known to influence SUD outcomes. Controlling for these potential confounders shared by siblings is important because too often polygenic associations are over-interpreted as evidence that a particular set of alleles has pleiotropic effects across traits or disorders. For this reason, testing the alternative explanation that polygenic associations are attributable to familial confounding is important for understanding the molecular genetic basis underlying the links between low educational attainment and SUDs. This is particularly critical in view of the enthusiasm to incorporate polygenic scores as part of precision medicine efforts to identify and intervene with individuals deemed genetically at risk.

Significant associations between an educational attainment polygenic score and SUD criterion counts within a sibling comparison design would be consistent with the interpretation that carrying more alleles associated with educational attainment is associated with a lower likelihood of developing SUD problems. In contrast, if sibling differences in

educational attainment polygenic scores do not predict SUD criterion counts it suggests that polygenic associations are confounded by other shared familial factors. This difference is important, considering that social advantage is related to both educational attainment polygenic scores [25–27] and rates of SUDs [28].

Materials and Methods

Participants

Participants came from the Collaborative Study on the Genetics of Alcoholism (COGA) [29–31], whose objective is to identify genes involved in alcohol dependence and related disorders. Probands (i.e., index individuals) were identified through alcohol treatment programs at six U.S. sites. Probands and their families were invited to participate if the family was sufficiently large (usually sibships > 3 with parents available) with two or more members in the COGA catchment area. Comparison families were recruited from the same communities. The Institutional Review Boards at all data gathering sites approved this study and written consent was obtained from all participants. COGA data are available via dbGaP (phs000763.v1.p1, phs000125.v1.p1) or through the National Institute on Alcohol Abuse and Alcoholism.

We defined two study samples within COGA. The first sample included all participants of European ancestry 25 years of age or older with both genome-wide association data and relevant SUD phenotypic information (n = 5582 individuals from 1093 extended families; 3009 (54%) female; M_{age} = 42.29 years, age range = 25 – 91 years). We limited the sample to those of European ancestry to avoid population stratification [32] because the educational attainment genome-wide association study (GWAS) weights come from a European ancestry discovery sample. SNPrelate [33] was implemented to estimate principal components from GWAS data and subsequently used to determine European ancestry. We implemented the age minimum to balance the needs to ensure that the majority of participants had passed through the period of highest risk for onset of the SUDs without unduly limiting sample size. Epidemiological data regarding age of onset for SUDs [34–36] guided our decision to select age 25 as the cutoff, which also mirrors the cutoff used in analyses of educational attainment in US Census data [37].

The second sample was a subset of the first sample, limited to groups of European ancestry biological full siblings (confirmed by genotyping) nested within the larger COGA sample. This process identified 4733 individuals nested within 1655 sibling groups (2–12 siblings per group). As detailed below, the n = 4733 sibling GWAS samples were used to calculate the educational attainment polygenic scores used for the sibling comparison analyses. The sample was subsequently filtered by age at phenotypic assessment for the linear mixed model analyses. In total, the sibling comparison analyses included 3098 individuals [1616 (52%) female] who were 25 years of age or older (M_{age} = 37.89 years) from 1226 sibling groups nested within 773 extended families.

Measures

Genotyping.—Genotyping for the COGA European ancestry participants was performed using the Illumina 1M, Illumina OmniExpress (Illumina, San Diego, CA), and Smokescreen (BioRealm, Walnut, CA) arrays. Quality control and imputation procedures are described in Lai et al. [31] and in the Supporting Information section 1.

Substance Use Disorder Clinical Criterion Counts.—Clinical criterion counts for alcohol (AUDSX), nicotine (NDSX), and cannabis (CUDSX) were obtained from the reliable and valid Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) [38, 39]. Criterion counts for alcohol and cannabis use disorder were made according to DSM-5 [40] and thus each had a possible range of 0–11. The criterion count for NDSX came from the Fagerström Test for Nicotine Dependence [41] and had a possible range of 0–10. The criterion count distributions showed right skews; to address this in inferential analyses, we applied a logarithmic transformation (left anchored at 1).

Covariates and Measures for Robustness and Sensitivity Analyses.—We included sex, age at last interview, cohort (indexed using three dummy-coded variables derived from participant year of birth: [1896–1930) set as reference; [1930–1950); [1950–1970); [1970–2010)) and the first two principal components for genetic ancestry in all analyses.

We conducted a series of robustness and sensitivity analyses to probe and interpret the effects from our primary analyses. For robustness analyses, we used participants' educational attainment, assessed as highest level of education completed. Potential responses ranged from 0–17 years (primary or secondary school = actual year; technical school/1 year college = 13 years; 2 years college = 14 years; 3 years college = 15 years; 4 years college = 16 years; any graduate degree = 17 years). In sensitivity analyses of the sibling data, we used participants' reports of their living arrangements while growing up from a set of thirteen options (see Supporting Information section 2) to evaluate whether the pattern of effects changed when the sample was limited to siblings who reported the same living arrangements (and thus likely shared the same rearing environment). An early version of the SSAGA did not query living arrangements; accordingly, we were only able to confirm that siblings grew up together for a subset of the sample (see Supporting Information section 2).

Statistical Methods

Educational Attainment Genome-Wide Polygenic Scores (*EduYears-GPS*).—We used results from the Social Science Genetic Association Consortium (SSGAC) GWAS of educational attainment [20], to construct educational attainment genome-wide polygenic scores (*EduYears-GPS*) in the COGA sample. Although polygenic scores are often described as polygenic *risk* scores, we prefer the term genome-wide polygenic score for this study. This is because 'risk' connotes a negative outcome, whereas educational attainment is typically valued. After removing palindromic SNPs (which can be ambiguous with respect to the reference allele in different samples), we used the *clump* and *score* procedures in PLINK [42] to sum each individual's total number of minor alleles from the score SNPs, with each SNP weighted by the negative log of the GWAS association *p* value and sign of

the association (beta) statistic. Clumping was done with respect to the linkage disequilibrium (LD) pattern in the COGA EA sample (founders only) using a 500 kb physical distance and an LD threshold of $r^2 \geq 0.25$. Following conventions for polygenic scoring using the pruning-and-thresholding approach [18], we calculated a series of GPS in COGA that included SNPs meeting increasingly stringent *p*-value thresholds in the discovery GWAS ($P<0.50$, $P<0.40$, $P<0.30$, $P<0.20$, $P<0.10$, $P<0.01$, $P<0.001$, $P<.0001$).

Association of *EduYears-GPS* and SUDs.—We examined associations between *EduYears-GPS* and the SUD criterion counts in separate linear mixed models using the nlme package version 3.1–128 [43] for R version 3.2.3 [44]. We conducted preliminary analyses to identify the *EduYears-GPS* most strongly associated with criterion counts for each SUD (see Supporting Information Table 1), and present results using the threshold with the strongest association. We conducted these preliminary analyses separately for polygenic scores meeting increasingly stringent p-value thresholds using linear mixed models, which allowed us to account for the nested structure of the COGA family-based data; other methods for optimizing the p-value threshold [e.g., PRSice; 45] do not allow for nested data. In addition to the covariates described above, we also included a count measure of the number of SNPs available for scoring for each participant. Marginal effect sizes for fixed effects were calculated using the MuMIn package version 1.15.6 [46].

Sibling Comparisons of *EduYears-GPS* and SUDs.—We used the $n = 4733$ sibling GWAS sample to calculate the *EduYears-GPS*-mean (for each sibling group) and *EduYears-GPS*-deviation scores (for each individual within that sibling group). We then filtered the sample based on participants' age at last interview to retain those who were 25 years of age or older (age cutoff selected to ensure that participants had passed through the period of highest risk for onset of the SUDs) for our primary sibling comparisons sample; additional information regarding this process can be found in Supporting Information section 3. Using all available GWAS data from a sibling group to calculate the *EduYears-GPS*-mean and *EduYears-GPS*-deviation scores has the advantage of providing a more precise estimate for these variables (since genotype does not change with age), versus limiting calculation of *EduYears-GPS*-mean to those siblings who also met the phenotypic age threshold. In separate linear mixed models, we then examined whether *EduYears-GPS*-deviation predicted SUDs after controlling for *EduYears-GPS*-mean. The sibling comparison is captured by the *EduYears-GPS*-deviation parameter, and indicates whether sibling differences in *EduYears-GPS* predict SUDs; this parameter captures the within-family effect. The *EduYears-GPS*-mean parameter captures whether family-level differences in *EduYears-GPS* predict SUDs, reflecting the between-family effect.

Robustness and Sensitivity Analyses.—We conducted robustness analyses to examine whether findings changed when statistically controlling for educational attainment in both the association and sibling comparison analyses. Sibling differences and family means for phenotypic educational attainment (i.e., EduYears-deviation and EduYears-mean) were calculated using the same procedure described above for *EduYears-GPS*-deviation and *EduYears-GPS*-mean.

We conducted sensitivity analyses to see whether effects changed when using a more conservatively defined subsample of siblings who were known to have the same living arrangements while growing up or who were born within 3 years of the eldest. These more conservative definitions assume that siblings who report the same living arrangements growing up and who are born in closer proximity to one another are likely to share more features of their home environment than siblings who report different living arrangements or who are born further apart. In total, 1702 individuals (54% female) from 739 sibling groups were available for this analysis. We also examined whether the effects were robust when sibships that included monozygotic twins (8 sibling groups) were removed from the analysis. Monozygotic twins share 100% of their genetic variation, and we wanted to ensure that our results were not driven by genotyping errors or PLINK's handling of SNPs set to missing (as part of cleaning for Mendelian errors) during polygenic score calculation. Sample size as a function of the filters employed for these sensitivity analyses are shown in Supporting Information Figure 1.

Results

Descriptive statistics

Descriptive statistics for the SUD criterion counts and educational attainment for the full sample (n = 5582) and the sibling subsample (n = 3098) are summarized in Table 1. Representativeness analyses of the sibling subsample are summarized in Supporting Information section 4.

Polygenic Association for *EduYears-GPS* and SUDs

We identified the $P<0.30$ threshold for AUDSX, $P<0.20$ for NDSX, and $P<0.01$ for CUDSX as the *EduYears-GPS* thresholds most strongly associated with each SUD criterion count. As shown in Table 2, higher *EduYears-GPS* was associated with lower SUD criteria. The *EduYears-GPS* accounted for 0.79%, 1.84%, and 0.30% of the variance in AUDSX, NDSX, and CUDSX, respectively.

Sibling Comparisons of *EduYears-GPS* and SUDs

We carried forward the substance-specific thresholds that were most strongly associated with each criterion count from above into the sibling comparisons to examine whether *EduYears-GPS*-deviation predicted each SUD criterion count after controlling for *EduYears-GPS*-mean.

The results of the sibling comparisons are shown in Table 3. Individuals with higher *EduYears-GPS* compared to their siblings had lower alcohol, nicotine, and cannabis criterion counts. Sibling differences in *EduYears-GPS* accounted for 0.17%, 0.20%, and 0.13% of the variance in AUDSX, NDSX, and CUDSX, respectively. There were also family-level effects, whereby those in sibling groups with higher *EduYears-GPS*-mean had lower alcohol, nicotine, and cannabis criterion counts. These family-level effects accounted for 0.29%, 1.89%, and 0.22% of the variance in AUDSX, NDSX, and CUDSX, respectively.

Robustness Analyses

After controlling for participants' measured (phenotypic) educational attainment in the polygenic analyses, *EduYears-GPS* continued to be associated with AUDSX and NDSX (but not CUDSX) (Supporting Information Table 2). After controlling for sibling and family differences in educational attainment in the sibling comparison analyses, the effects of sibling differences in *EduYears-GPS* on SUD criterion counts were attenuated for NDSX and CUDSX ($P = 0.09$ to 0.13) but remained significant for AUDSX ($P = 0.01$) (Supporting Information Table 3). Sibling and family differences in educational attainment were also significantly associated with SUD criterion counts. Individuals with higher educational attainment compared to their siblings, and individuals from sibling groups with higher educational attainment had lower AUDSX, NDSX, and CUDSX.

Sensitivity Analyses

In the first set of sensitivity analyses, we examined whether effects held when the sample was limited to the groups of siblings who were known to have grown up together (N = 739 sibling groups). In the second set of sensitivity analyses, we examined whether the effects held when the sample was limited to those who were born within 3 years of the first born in a sibling group. In the third set of sensitivity analyses, we examined whether the effects were also robust when sibships that included monozygotic twins (8 sibling groups) were removed from the analysis. Across all three sets of sensitivity analyses in smaller, more conservative test samples, we continued to find that individuals with higher *EduYears-GPS* than their siblings had lower SUD criterion counts (Supporting Information Tables 4–6). The only exception to this was that the effect of sibling differences in *EduYears-GPS* on CUDSX was attenuated ($P = 0.08$) in the sensitivity analyses limited to those born within 3 years of the first born in a sibling group.

Discussion

The present study illustrates how sibling comparisons can improve our understanding of the shared genetic etiology underlying educational attainment and substance use problems. Consistent with previous findings that educational attainment has a negative genetic correlation with alcohol problems [11, 13], cannabis use disorder [14], and smoking [12], we found that individuals met fewer SUD criteria when they carried more alleles associated with educational attainment. We replicated these effects within a sibling comparisons design, where we found that individuals met fewer clinically significant substance use criteria when they carried more alleles associated with higher educational attainment *than their siblings*. Sibling comparisons are uniquely powerful because they control for unmeasured confounding factors shared by siblings that could otherwise explain the association between educational attainment polygenic scores and substance use disorder criteria: factors such as socioeconomic status, urban-rural residency, and parental education. Thus, our findings suggest that the association between educational attainment polygenic scores and SUDs is not completely explained by confounders that differ between families.

These findings add important nuance to discussions regarding the nature of associations between educational attainment and problematic substance use. First, our findings are

consistent with previous findings that educational outcomes reflect many genetically influenced traits and behaviors, including SUD-associated factors like behavior problems, attention-deficit hyperactivity disorder, and personality [25, 26, 47–50], not simply intelligence or cognitive ability. Interestingly, in our robustness analyses, the educational attainment polygenic scores predicted alcohol use disorder and nicotine dependence criterion counts above and beyond participants' observed (phenotypic) educational attainment. This highlights that these polygenic scores index factors linked to educational persistence and SUDs that are not fully captured by educational attainment itself. In contrast, for cannabis, the educational attainment polygenic score did not have unique predictive power above and beyond the educational attainment phenotype.

Second, our sibling comparison analyses demonstrated that polygenic scores were significant predictors of SUD criteria even within families. For outcomes like SUDs, which have considerable influences that vary among families, ruling out familial confounding is particularly important. In addition to significant sibling differences, we also found that between-family differences in *EduYears-GPS* predicted SUDs. This suggests that both the overall polygenic loading of one's family and one's relative polygenic loading within that family are important predictors of risk for SUDs. The associations between sibling differences in polygenic scores and SUDs were attenuated somewhat after controlling for sibling differences in phenotypic educational attainment. This attenuation may reflect the relative statistical power of polygenic scores compared to the phenotypes from which they are derived, as well the likelihood that some of the effect of sibling differences in educational attainment polygenic scores is likely to be mediated through sibling differences in educational persistence, as has been documented previously [26].

These results should be considered in the context of several limitations. First, the COGA sample is enriched with individuals with SUDs, and the results may not generalize to lower-risk samples. Second, the sibling comparison design assumes that siblings are reared together. Not all COGA participants were asked about their living arrangements while growing up, and so we could not test whether this assumption was met for all sibling groups. However, to address this concern, we restricted the analyses to the sibling groups where it was possible to determine that they grew up together, and to siblings who were born close together in time (and thus more likely to share aspects of their rearing environment compared to siblings born further apart). The pattern of effects remained significant and in the same direction in these sensitivity analyses, suggesting that the effects observed in our sibling comparisons of polygenic scores were not driven by differences in siblings' rearing environments.

Third, because genetic associations can differ across ancestral groups, we focused here on the European ancestry subset of COGA because the discovery GWAS for educational attainment used a European ancestry sample. It is unknown whether the same pattern of effects would be observed in samples of non-European ancestry.

Fourth, polygenic scores by design only capture common genetic variation. Fifth, despite evidence for polygenic association even after controlling for family-level confounders, the polygenic scores accounted for a relatively small amount of variance. This limited predictive

power cautions against incorporating polygenic scores into clinical screening or intervention efforts for substance use disorders.

As efforts to characterize how polygenic predispositions influence complex behavioral outcomes increase in popularity [16], we believe that environmentally-informed designs such as sibling comparisons will become a particularly useful tool to illuminate the “chains of risk” from genotype to phenotype. For example, sibling differences can be elaborated upon to include examination of how subtle differences in polygenic loading between siblings impact individual differences or selection into particular environments. In turn, these mediating phenotypes may be particularly actionable targets for prevention and intervention efforts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The Collaborative Study on the Genetics of Alcoholism (COGA), Principal Investigators B. Porjesz, V. Hesselbrock, H. Edenberg, L. Bierut, includes eleven different centers: University of Connecticut (V. Hesselbrock); Indiana University (H.J. Edenberg, J. Nurnberger Jr., T. Foroud); University of Iowa (S. Kuperman, J. Kramer); SUNY Downstate (B. Porjesz); Washington University in St. Louis (L. Bierut, J. Rice, K. Bucholz, A. Agrawal); University of California at San Diego (M. Schuckit); Rutgers University (J. Tischfield, A. Brooks); Department of Biomedical and Health Informatics, The Children’s Hospital of Philadelphia; Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA (L. Almasy), Virginia Commonwealth University (D. Dick), Icahn School of Medicine at Mount Sinai (A. Goate), and Howard University (R. Taylor). Other COGA collaborators include: L. Bauer (University of Connecticut); J. McClintick, L. Wetherill, X. Xuei, Y. Liu, D. Lai, S. O’Connor, M. Plawecki, S. Lourens (Indiana University); G. Chan (University of Iowa; University of Connecticut); J. Meyers, D. Chorlian, C. Kamarajan, A. Pandey, J. Zhang (SUNY Downstate); J.-C. Wang, M. Kapoor, S. Bertelsen (Icahn School of Medicine at Mount Sinai); A. Anokhin, V. McCutcheon, S. Saccone (Washington University); J. Salvatore, F. Aliev, B. Cho (Virginia Commonwealth University); and Mark Kos (University of Texas Rio Grande Valley). A. Parsian and M. Reilly are the NIAAA Staff Collaborators.

We continue to be inspired by our memories of Henri Begleiter and Theodore Reich, founding PI and Co-PI of COGA, and also owe a debt of gratitude to other past organizers of COGA, including Ting-Kai Li, P. Michael Conneally, Raymond Crowe, and Wendy Reich, for their critical contributions. This national collaborative study is supported by NIH Grant U10AA008401 from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and the National Institute on Drug Abuse (NIDA). Additional support for this project comes from K01AA024152 (JES); K02DA032573 (AA); and R01DA040411 (ECJ).

Funding support for GWAS genotyping performed at the Johns Hopkins University Center for Inherited Disease Research was provided by the National Institute on Alcohol Abuse and Alcoholism, the NIH GEI (U01HG004438), and the NIH contract “High throughput genotyping for studying the genetic contributions to human disease” (HHSN268200782096C). GWAS genotyping was also performed at the Genome Technology Access Center in the Department of Genetics at Washington University School of Medicine which is partially supported by NCI Cancer Center Support Grant #P30 CA91842 to the Siteman Cancer Center and by ICTS/CTSA Grant# UL1RR024992 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research.

References

1. Esch P, Bocquet V, Pull C, Couffignal S, Lehnert T, Graas M et al. The downward spiral of mental disorders and educational attinment: a systematic review on early school leaving, BMC psychiatry 2014: 14.
2. Townsend L, Flisher AJ, King G A systematic review of the relationship between high school dropout and substance use, Clin Child Fam Psychol Rev 2007: 10: 295–317. [PubMed: 17636403]

3. Kessler RC, Foster CL, Saunders WB, Stang PE Social consequences of psychiatric disorders, I: Educational attainment, Am J Psychiat 1995: 152: 1026–32. [PubMed: 7793438]
4. Stinson FS, Ruan WJ, Pickering R, Grant BF Cannabis use disorders in the USA: prevalence, correlates and co-morbidity, Psychol Med 2006: 36: 1447–60. [PubMed: 16854249]
5. Breslau J, Lane M, Sampson N, Kessler RC Mental disorders and subsequent educational attainment in a US national sample, J Psychiatr Res 2008: 42: 708–16. [PubMed: 18331741]
6. Verweij KJ, Huizink AC, Agrawal A, Martin NG, Lynskey MT Is the relationship between early-onset cannabis use and educational attainment causal or due to common liability?, Drug Alcohol Depend 2013: 133: 580–6. [PubMed: 23972999]
7. Breslau J, Miller E, Chung WJ, Schweitzer JB Childhood and adolescent onset psychiatric disorders, substance use, and failure to graduate high school on time, J Psychiatr Res 2011: 45: 295–301. [PubMed: 20638079]
8. Martin MJ, Conger RD, Sitnick SL, Masarik AS, Forbes EE, Shaw DS Reducing Risk for Substance Use by Economically Disadvantaged Young Men: Positive Family Environments and Pathways to Educational Attainment, Child Dev 2015: 86: 1719–37. [PubMed: 26307026]
9. Fothergill KE, Ensminger ME, Green KM, Crum RM, Robertson J, Juon HS The impact of early school behavior and educational achievement on adult drug use disorders: A prospective study, Drug Alcohol Depend 2008: 91: 191–99.
10. Green KM, Zebrak KA, Fothergill KE, Robertson JA, Ensminger ME Childhood and adolescent risk factors for comorbid depression and substance use disorders in adulthood, Addict Behav 2012: 37: 1240–47. [PubMed: 22762959]
11. Latvala A, Dick DM, Tuulio-Henriksson A, Suvisaari J, Viken RJ, Rose RJ et al. Genetic correlation and gene-environment interaction between alcohol problems and educational level in young adulthood, J Stud Alcohol Drugs 2011: 72: 210–20. [PubMed: 21388594]
12. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR et al. An atlas of genetic correlations across human diseases and traits, Nat Genet 2015: 47: 1236–41. [PubMed: 26414676]
13. Walters RK, Polimanti R, Johnson EC, Mcclintick JN, Adams MJ, Adkins AE et al. Trans-ancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders, Nature Neuroscience 2018: 21: 1656–69. [PubMed: 30482948]
14. Bergen SE, Gardner CO, Aggen SH, Kendler KS Socioeconomic status and social support following illicit drug use: Causal pathways or common liability?, Twin Res Hum Genet 2008: 11: 266–74. [PubMed: 18498205]
15. Grant JD, Scherrer JF, Lynskey MT, Agrawal A, Duncan AE, Haber JR et al. Associations of alcohol, nicotine, cannabis, and drug use/dependence with educational attainment: evidence from cotwin-control analyses, Alcohol Clin Exp Res 2012: 36: 1412–20. [PubMed: 22587016]
16. Martin AR, Daly MJ, Robinson EB, Hyman SE, Neale BM Predicting Polygenic Risk of Psychiatric Disorders, Biol Psychiatry 2018.
17. Maier RM, Visscher PM, Robinson MR, Wray NR Embracing polygenicity: a review of methods and tools for psychiatric genetics research, Psychol Med 2017: 48: 1055–67. [PubMed: 28847336]
18. Bogdan R, Baranger D. a. A., Agrawal A Polygenic risk scores in clinical psychology: bridging genomic risk to individual differences, Annu Rev Clin Psychol 2018: 14: 119–57. [PubMed: 29579395]
19. Dudbridge F Polygenic epidemiology, Genet Epidemiol 2016: 40: 268–72. [PubMed: 27061411]
20. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals, Nat Genet 2018: 50: 1112–21. [PubMed: 30038396]
21. D'onofrio BM, Lahey BB, Turkheimer E, Lichtenstein P Critical need for family-based, quasi-experimental designs in integrating genetic and social science research, Am J Public Health 2013: 103: 46–55.
22. Lahey BB, D'onofrio. All in the Family: Comparing Siblings to Test Causal Hypotheses Regarding Environmental Influences on Behavior, Curr Dir Psychol Sci 2010: 19: 319–23. [PubMed: 23645975]
23. Rutter M Proceeding from observed correlation to causal inference: the use of natural experiments, Perspect Psychol Sci 2007: 2: 377–95. [PubMed: 26151974]

24. Donovan SJ, Susser E Commentary: Advent of sibling designs, Int J Epidemiol 2011: 40: 345–9. [PubMed: 21450688]
25. Belsky DW, Moffitt TE, Corcoran DL, Domingue B, Harrington H, Hogan S et al. The genetics of success: how single-nucleotide polymorphisms associated with educational attainment relate to life-course development, Psychol Sci 2016: 27: 957–72. [PubMed: 27251486]
26. Domingue BW, Belsky D, Conley D, Harris KM, Boardman JD Polygenic Influence on Educational Attainment: New evidence from The National Longitudinal Study of Adolescent to Adult Health, AERA Open 2015: 1: 1–13. [PubMed: 28164148]
27. Selzam S, Krapohl E, Von Stumm S, O'reilly PF, Rimfeld K, Kovas Y et al. Predicting educational achievement from DNA, Mol Psychiatry 2017: 22: 267–72. [PubMed: 27431296]
28. Galea S, Nandi A, Vlahov D The social epidemiology of substance use, Epidemiologic Reviews 2004: 26: 36–52. [PubMed: 15234946]
29. Begleiter H, Reich T, Hesselbrock V, Porjesz B, Li TK, Schuckit MA et al. The Collaborative Study on the Genetics of Alcoholism, Alcohol Health Res W 1995: 19: 228–36.
30. Bucholz KK, Mccutcheon VV, Agrawal A, Dick DM, Hesselbrock VM, Kramer JR et al. Comparison of parent, peer, psychiatric, and cannabis use influences across stages of offspring alcohol involvement: Evidence from the COGA Prospective Study, Alcohol Clin Exp Res 2017: 41: 359–68. [PubMed: 28073157]
31. Lai D, Wetherill L, Bertelsen S, Carey CE, Kamarajan C, Kapoor M et al. Genome-wide association studies of alcohol dependence, DSM-IV criterion count and individual criteria., Genes Brain Behav 2019: 18: e12579.
32. Cardon LR, Palmer LJ Population stratification and spurious allelic association, Lancet 2003: 361: 598–604. [PubMed: 12598158]
33. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS A high-performance computing toolset for relatedness and principal component analysis of SNP data, Bioinformatics 2012: 28: 3326–8. [PubMed: 23060615]
34. Grant BF, Goldstein RB, Saha TD, Chou SP, Jung J, Zhang H 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: 757–66. [PubMed: 26039070]
35. Hasin DS, Kerridge BT, Saha TD, Huang B, Pickering R, Smith SM et al. Prevalence and correlates of DSM-5 cannabis use disorder, 2012–2013: findings from the National Epidemiologic Survey on Alcohol and Related Conditions-III, Am J Psychiat 2016: 173: 588–99. [PubMed: 26940807]
36. Breslau N, Johnson EO, Hiripi E, Kessler R Nicotine dependence in the United States: Prevalence, trends and smoking persistence, Arch Gen Psychiat 2001: 58: 810–16. [PubMed: 11545662]
37. Census Bureau U.S.. Current Population Survey, 2018 Annual Social and Economic Supplement; 2018.
38. Hesselbrock M, Easton C, Bucholz KK, Schuckit M, Hesselbrock V A validity study of the SSAGA--a comparison with the SCAN, Addiction 1999: 94: 1361–70. [PubMed: 10615721]
39. Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JL et al. A new, semi-structured psychiatric interview for use in genetic linkage studies: A report on the reliability of the SSAGA, J Stud Alcohol 1994: 55: 149–58. [PubMed: 8189735]
40. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, fifth edition Arlington, VA: American Psychiatric Publishing; 2013.
41. Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO The Fagerstrom Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire, Br J Addict 1991: 86: 1119–27. [PubMed: 1932883]
42. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D et al. PLINK: a tool set for whole-genome association and population-based linkage analyses, Am J Hum Genet 2007: 81: 559–75. [PubMed: 17701901]
43. Pinheiro J, Eispack Authors, Heisterkamp S, Van Willigen B, R-Core. Linear and nonlinear mixed effects models; 2018.
44. R Development Core Team. R: A language and environment for statistical computing Vienna, Austria; 2014.

45. Euesden J, Lewis CM, O'reilly PF PRSice: Polygenic Risk Score software, Bioinformatics 2015: 31: 1466–8. [PubMed: 25550326]

46. Barton K MuMIn: Multi-Model Inference. R package version 1.15.6; 2016.

47. Krapohl E, Rimfeld K, Shakeshaft NG, Trzaskowski M, Mcmillan A, Pingault JB et al. The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence, Proc Natl Acad Sci USA 2014: 111: 15273–8.

48. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA et al. Genome-wide association study identifies 74 loci associated with educational attainment, Nature 2016: 533: 539–42. [PubMed: 27225129]

49. Hagenaars SP, Harris SE, Davies G, Hill WD, Liewald DC, Ritchie SJ et al. Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N=112 151) and 24 GWAS consortia, Mol Psychiatry 2016: 21: 1624–32. [PubMed: 26809841]

50. De Zeeuw EL, Van Beijsterveldt CE, Glasner TJ, Bartels M, Ehli EA, Davies GE et al. Polygenic scores associated with educational attainment in adults predict educational achievement and ADHD symptoms in children, Am J Med Genet B Neuropsychiatr Genet 2014: 165B: 510–20. [PubMed: 25044548]

Table 1.

Descriptive Statistics

Full sample (n = 5582; 54% Female)					
Measure	*N*	Mean	SD	Min	Max
Age	5582	42.29	13.26	25	91
AUDSX	5582	3.77	3.83	0	11
NDSX	4754	2.61	3.00	0	10
CUDSX	5578	1.56	2.81	0	11
Educational Attainment (years)	5578	13.43	2.33	2	17
Sibling subsample (n = 3098; 52% Female)					
Measure	*N*	Mean	SD	Min	Max
Age	3098	37.89	10.85	25	81
AUDSX	3098	4.39	3.90	0	11
NDSX	2752	2.58	3.00	0	10
CUDSX	3097	1.94	3.04	0	11
Educational Attainment (years)	3095	13.55	2.29	5	17

Table 2.

Associations between *EduYears-GPS* genome-wide polygenic scores and substance use disorder criterion counts in the full sample.

	Alcohol Use Disorder Criterion Count (n = 5582)		Nicotine Dependence Criterion Count (n = 4754)		Cannabis Use Disorder Criterion Count (n = 5578)	
Parameter	***b***	**95% CI**	***b***	**95% CI**	***b***	**95% CI**
Intercept	**−2.79**	**[−3.78, −1.79]**	**−1.47**	**[−2.59, −0.34]**	**−1.01**	**[−1.74, −0.28]**
Sex (female)	**−0.58**	**[−0.62, −0.54]**	**−0.22**	**[−0.27, −0.18]**	**−0.34**	**[−0.37, −0.30]**
Age	**−0.01**	**[−0.01, −5.12E-03]**	**5.00E-03**	**[−3.84E-05, 9.76E-03]**	**−8.60E-03**	**[−0.01, −5.01E-03]**
PC1	**83.39**	**[9.18, 157.60]**	−23.63	[−107.81, 60.54]	−42.38	[−108.61, 23.86]
PC2	26.39	[−11.30, 64.07]	29.71	[−12.40, 71.83]	15.99	[−17.76, 49.73]
EduYears-GPS count	**1.01E-05**	**[8.35E-06, 1.19E-05]**	**9.29E-06**	**[6.78E-06, 1.18E-05]**	**3.70E-05**	**[2.77E-05, 4.62E-05]**
Cohort 2	**0.22**	**[0.10, 0.33]**	0.15	[−0.01, 0.31]	−0.03	[−0.14, 0.07]
Cohort 3	**0.41**	**[0.25, 0.58]**	0.14	[−0.08, 0.35]	**0.44**	**[0.29, 0.58]**
Cohort 4	0.16	[−0.05, 0.36]	−0.02	[−0.28, 0.24]	**0.31**	**[0.12, 0.49]**
EduYears-GPS	**−19,360.64**†	**[−25,072.31, −13,648.97]**	**−24,663.58**†	**[−29,961.23, −19,365.93]**	**−2,551.10**	**[−3,781.71, −1,320.49]**
EduYears-GPS ΔR^2	**0.79%**		**1.84%**		**0.30%**	

Notes. **Boldface** indicates estimate P ≤ 0.05.

†denotes that the *EduYears-GPS* effect was robust after controlling for phenotypic educational attainment (see Supporting Information Table 2). The *EduYears-GPS* thresholds for each substance were: alcohol (P<0.30); nicotine (P<0.20); cannabis (P<0.01). Abbreviations: PC = principal component for genetic ancestry; Cohort = dummy-coded variables indexing year of birth, defined as [1930–1950); [1950–1970); and [1970–2010); *EduYears-GPS* count = number of single nucleotide polymorphisms available for polygenic scoring; *EduYears-GPS* = educational attainment genome-wide polygenic score; ΔR^2 = change in r-squared.

Table 3.

Sibling comparisons of substance use disorder criterion counts as a function of *EduYears-GPS* genome-wide polygenic scores.

	Alcohol Use Disorder Criterion Count (n = 3098)		Nicotine Dependence Criterion Count (n = 2752)		Cannabis Use Disorder Criterion Count (n = 3097)	
Parameter	***b***	**95% CI**	***b***	**95% CI**	***b***	**95% CI**
Intercept	**−6.88**	**[−8.26, −5.50]**	**−2.75**	**[−4.29, −1.21]**	**−1.87**	**[−3.00, −0.74]**
Sex (female)	**−0.51**	**[−0.57, −0.46]**	**−0.19**	**[−0.25, −0.13]**	**−0.40**	**[−0.46, −0.35]**
Age	2.00E-03	[−3.83E-03, 7.45E-03]	**0.02**	**[0.01, 0.02]**	−4.00E-03	[−9.90E-03, 1.45E-03]
PC1	−10.02	[−117.21, 97.16]	**−186.36**	**[−305.61, −67.10]**	**−126.75**	**[−235.49, −18.00]**
PC2	−0.50	[−53.35, 52.36]	41.30	[−17.28, 99.88]	−4.57	[−58.33, 49.18]
EduYears-GPS count	**1.78E-05**	**[1.54E-05, 2.01E-05]**	**1.37E-05**	**[1.05E-05, 1.70E-05]**	**5.47E-05**	**[4.09E-05, 6.85E-05]**
Cohort 2	**0.34**	**[0.12, 0.56]**	0.27	[−0.03, 0.57]	0.03	[−0.20, 0.26]
Cohort 3	**0.53**	**[0.27, 0.79]**	0.32	[−0.02, 0.66]	**0.63**	**[0.37, 0.90]**
Cohort 4	0.28	[−0.03, 0.59]	0.22	[−0.18, 0.61]	**0.48**	**[0.17, 0.79]**
EduYears-GPS-deviation	**−19,694.89**†	**[−31,041.68, −8,348.11]**	**−16,676.08**	**[−27,013.63, −6,338.52]**	**−3,829.85**	**[−6,486.43, −1,173.26]**
EduYears-GPS-mean	**−13,147.15**	**[−23,328.50, −2,965.80]**	**−29,327.98**	**[−38,688.84, −19,967.11]**	**−2,870.93**	**[−5,437.72, −304.13]**
EduYears-GPS-deviation ΔR^2		**0.17%**		**0.20%**		**0.13%**
EduYears-GPS-mean ΔR^2		**0.29%**		**1.89%**		**0.22%**

Notes. **Boldface** indicates estimate P ≤ 0.05.

†denotes that the *EduYears-GPS*-deviation effect was robust after controlling for sibling and family differences in phenotypic educational attainment (see Supporting Information Table 3).The *EduYears-GPS* thresholds for each substance were: alcohol (P<0.30); nicotine (P<0.20); cannabis (P<0.01). Abbreviations: PC = principal component for genetic ancestry; Cohort = dummy-coded variables indexing year of birth, defined as [1930–1950); [1950–1970); and [1970–2010); *EduYears-GPS* count = number of single nucleotide polymorphisms available for polygenic scoring; *EduYears-GPS*-deviation = the difference of an individual's *EduYears-GPS* from the sibling group mean; *EduYears-GPS*-mean = the sibling group mean of *EduYears-GPS*; ΔR^2 = change in r-squared.