Clinical, environmental, and genetic risk factors for substance use disorders : characterizing combined effects across multiple cohorts

Barr, Peter B.

2022

Barr , P B , Driver , M N , Kuo , S I-C , Stephenson , M , Aliev , F , Linner , R K , Marks , J , Anokhin , A P , Bucholz , K , Chan , G , Edenberg , H J , Edwards , A C , Francis , M W , Hancock , D B , Harden , K P , Kamarajan , C , Kaprio , J , Kinreich , S , Kramer , J , Kuperman , S , Latvala , A , Meyers , J L , Palmer , A A , Plawecki , M H , Porjesz , B , Rose , R J , Schuckit , M A , Salvatore , J E & Dick , D M 2022 , ' Clinical, environmental, and genetic risk factors for substance use disorders : characterizing combined effects across þÿ m u l t i p l e c o h o r t s ' , M o l e c u l a r P s y c h i a t r y , v o l . 2 7 , p p . 4 6 3 3 4 6 4 1 . h t

http://hdl.handle.net/10138/356809
https://doi.org/10.1038/s41380-022-01801-6

unspecified
acceptedVersion

Clinical, Environmental, and Genetic Risk Factors for Substance Use Disorders: Characterizing Combined Effects across Multiple Cohorts

Peter B. Barr, Ph.D. [1,2], Morgan N. Driver, B.S. [3], Sally I-Chun Kuo, Ph.D. [4], Mallory Stephenson, M.S. [5], Fazil Aliev, Ph.D. [4,6], Richard Karlsson Linnér, Ph.D. [7], Jesse Marks [8], Andrey P. Anokhin, Ph.D. [9], Kathleen Bucholz, Ph.D. [9], Grace Chan, Ph.D. [10,11], Howard J. Edenberg, Ph.D. [12, 13], Alexis C. Edwards, Ph.D. [5], Meredith W. Francis, Ph.D. [9], Dana B. Hancock [8], K. Paige Harden, Ph.D. [14, 15], Chella Kamarajan, Ph.D. [1], Jaakko Kaprio M.D., Ph.D. [16], Sivan Kinreich, Ph.D. [1], John Kramer, Ph.D. [11], Samuel Kuperman, M.D. [11], Antti Latvala, Ph.D. [17], Jacquelyn L. Meyers, Ph.D. [1, 2], Abraham A. Palmer, Ph.D. [18, 19], Martin H. Plawecki M.D., Ph.D. [20], Bernice Porjesz, Ph.D. [1], Richard J. Rose, Ph.D. [21], Marc A. Schuckit, M.D. [18], Jessica E. Salvatore, Ph.D. [4], and Danielle M. Dick, Ph.D. [4,6]

[1] Department of Psychiatry and Behavioral Sciences, SUNY Downstate Health Sciences University, Brooklyn, NY, USA.
[2] VA New York Harbor Healthcare System, Brooklyn, NY, USA.
[3] Department of Human and Molecular Genetics, Virginia Commonwealth University, Richmond, VA, USA.
[4] Department of Psychiatry, Robert Wood Johnson Medical School, Rutgers University, Piscataway, NJ, USA.
[5] Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA.
[6] Rutgers Addiction Research Center, Rutgers University, Piscataway, NJ, USA.
[7] Department of Economics, Leiden University, Leiden, The Netherlands.
[8] Biostatistics and Epidemiology Division, RTI International, Research Triangle Park, NC, USA
[9] Department of Psychiatry, School of Medicine, Washington University in St. Louis, St Louis, MO, USA.
[10] Department of Psychiatry, School of Medicine, University of Connecticut, Farmington, CT, USA.
[11] Department of Psychiatry, Carver College of Medicine, University of Iowa, Iowa City, IA, USA
[12] Department of Medical and Molecular Genetics, School of Medicine, Indiana University, Indianapolis, IN, USA
[13] Department of Biochemistry and Molecular Biology, School of Medicine, Indiana University, Indianapolis, IN, USA
[14] Department of Psychology, University of Texas at Austin, Austin, TX, USA.
[15] Population Research Center, University of Texas at Austin, Austin, TX, USA.
[16] Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland
[17] Institute of Criminology and Legal Policy, University of Helsinki, Helsinki, Finland
[18] Department of Psychiatry, University of California San Diego, La Jolla, CA, USA.
[19] Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA.
[20] Department of Psychiatry, School of Medicine, Indiana University, Indianapolis, IN, USA
[21] Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA

Corresponding author: Peter B. Barr
Department of Psychiatry and Behavioral Sciences
SUNY Downstate Health Sciences University
450 Clarkson Ave, MSC 1203, Brooklyn, NY 11203
e-mail: peter.barr@downstate.edu.

Running Header: Characterizing Risk for SUDs Across Cohorts
Word count abstract: 256; Word count text body: 3,548
Figures and Tables: 3 Tables, 2 Figures

ABSTRACT

Substance use disorders (SUDs) incur serious social and personal costs. Risk for SUDs is complex, with risk factors ranging from social conditions to individual genetic variation. We examined whether models that include a clinical/environmental risk index (CERI) and polygenic scores (PGS) are able to identify individuals at increased risk of SUD in young adulthood across four longitudinal cohorts for a combined sample of N = 15,134. Our analyses included participants of European (N_{EUR} = 12,659) and African (N_{AFR} = 2,475) ancestries. SUD outcomes included: 1) alcohol dependence, 2) nicotine dependence; 3) drug dependence, and 4) any substance dependence. In the models containing the PGS and CERI, the CERI was associated with all three outcomes (ORs = 1.37 – 1.67). PGS for problematic alcohol use, externalizing, and smoking quantity were associated with alcohol dependence, drug dependence, and nicotine dependence, respectively (OR = 1.11 – 1.33). PGS for problematic alcohol use and externalizing were also associated with any substance dependence (ORs = 1.09 – 1.18). The full model explained 6% - 13% of the variance in SUDs. Those in the top 10% of CERI and PGS had relative risk ratios of 3.86 - 8.04 for each SUD relative to the bottom 90%. Overall, the combined measures of clinical, environmental, and genetic risk demonstrated modest ability to distinguish between affected and unaffected individuals in young adulthood. PGS were significant but added little in addition to the clinical/environmental risk index. Results from our analysis demonstrate there is still considerable work to be done before tools such as these are ready for clinical applications.

INTRODUCTION

Substance use disorders (SUDs) are associated with substantial costs to affected individuals, their families, and society. An estimated 107,000 Americans died as the result of an overdose in 2021 [1]. In 2016, alcohol use contributed 4.2% to the global disease burden and other drug use contributed 1.3% [2]. Excessive alcohol use and illicit drug use cost the United States an annual $250 billion [3] and $190 billion [4] respectively. Given the substantial human and economic costs of substance misuse and disorders, understanding the combined impact of important risk factors across multiple levels of analysis has important public health implications.

Substance use disorders are complex phenomena, and the development of substance related problems can be attributed to factors ranging from broader social and economic conditions to individual genetic variation [5–10]. Prior research using a multifactorial index of clinical and environmental risk factors (e.g., childhood disadvantage, family history of SUD, childhood conduct problems, childhood depression, early exposure to substances, frequent use during adolescence) found it useful in identifying those with persistent SUDs [11].

More recently, polygenic scores (PGS), which aggregate risk for a trait across the genome using information from genome-wide association studies (GWAS), were robustly associated with substance use [12] and substance related problems [13] across adolescence and into young adulthood. However, though robustly associated, current PGS do poorly in identifying individuals affected by SUDs [14]. To date, there is limited work on the combined impact of genetic, environmental, and clinical risk factors for SUDs. Prior work combining individual genetic variants and clinical features outperformed clinical features alone [15], but individual variants have limited predictive power. In other medical conditions, such as melanoma [16] or ischemic stroke [17], combining clinical and genetic risk factors showed improvement predicting risk for a specific outcome over models using individual risk factors.

In the current study, we examine the joint association of early life clinical/environmental risk factors and PGSs with SUDs in early adulthood across four longitudinal cohorts: the National Longitudinal Study of Adolescent to Adult Health (Add Health); the Avon Longitudinal Study of Parents and Children (ALSPAC); the Collaborative Study on the Genetics of Alcoholism (COGA); and the youngest cohort of the Finnish Twin Cohort Study (FinnTwin12). These samples include population-based cohorts from three countries (United States, England, and Finland) and a predominantly high-risk sample. Two of the samples (COGA and Add Health) are ancestrally diverse. We focus on early adulthood as this is a critical period for the development and onset of SUDs [18]. Our research questions are guided by the understanding that risk factors for SUDs range across multiple levels of analysis.

METHODS

Samples

Add Health is a nationally representative longitudinal study of adolescents followed into adulthood in the United States [19]. Data have been collected from Wave I when respondents were between 11-18 (1994-1995) to Wave V (2016-2018) when respondents were 35-42. The current analysis uses data from Waves I, II, and Wave IV.

ALSPAC is an ongoing, longitudinal population-based study of a birth cohort in the (former) Avon district of Southwest England [20–23]. Pregnant female residents with an expected date of delivery between April 1, 1991 and December 31, 1992 were invited to participate (N = 14,541 pregnant women, 80% of those eligible). This analysis uses data up to the age 24 assessment (details of all the data that is available through a searchable, web-based tool: http://www.bristol.ac.uk/alspac/researchers/our-data/).

COGA is a family-based sample consisting of alcohol dependent individuals (identified through treatment centers across the United States), their extended families, and community

controls (N ~16,000) [24, 25]. We use a prospective sample of offspring of the original COGA participants (baseline ages 12-22, N = 3,573) that have been assessed biennially since recruitment (2004-2019) [26].

FinnTwin12 is a population-based study of Finnish twins born 1983–1987 identified through Finland's Central Population Registry. A total of 2,705 families (87% of all identified) returned the initial family questionnaire late in the year in which twins reached age 11 [27]. Twins were invited to participate in follow-up surveys when they were ages 14, 17, and approximately 22.

Each cohort includes a wide range of social, behavioral, and phenotypic data measured across the life course. The SUD measures were derived from the corresponding young adult phases of data collection in each cohort (mean ages ~ 22 - 28). A full description of each sample is presented in the supplementary information (section 2).

Measures

Lifetime Diagnosis of Substance Use Disorder

We constructed measures of lifetime SUD diagnosis based on the data that were available in each of the samples, defined as meeting criteria for four, non-mutually exclusive categories of substance dependence: 1) alcohol dependence; 2) nicotine dependence; 3) drug dependence (inclusive of drugs such as cannabis, cocaine, opioids, sedatives, etc.); and 4) any substance dependence (alcohol, nicotine, or drug). Our analyses focused primarily on DSM-IV as this diagnostic system was most consistently used across all samples. There was one exception: in each of the samples, nicotine dependence was measured using a cutoff of 7 or higher on the Fagerstrom Test for Nicotine Dependence (FTND) [28]. Where possible, we drew measures of substance dependence from data collected during young adulthood to try and maintain temporal ordering between SUD diagnoses and measured risk factors.

Clinical/Environmental Risk Index

We created a clinical/environmental risk index (CERI) considering a variety of established risk factors for SUD (Table 1). The CERI included ten validated early life risk factors associated with later development of SUDs, including: low childhood socioeconomic status (SES), family history of SUD, early initiation of substance use, childhood internalizing problems, childhood externalizing problems, frequent drinking in adolescence, frequent smoking in adolescence, frequent cannabis use in adolescence, peer substance use, and exposure to trauma/traumatic experiences [11, 29, 30]. We dichotomized each risk factor (present vs not present) and summed them into an index for each person ranging from 0 to 10, providing a single measure of aggregate risk. Dichotomizing these items allowed us to harmonize measures across each sample in an interpretable manner. A full list of how each measure is defined within each of the samples is available in the supplementary information (section 3).

Polygenic Scores

We constructed polygenic scores (PGS), which are aggregate measures of the number of risk alleles individuals carry weighted by effect sizes from GWAS summary statistics, from six recent GWAS of SUDs and comorbid conditions including: 1) externalizing problems (EXT) [31]; 2) depression (DEP) [32]; 3) problematic alcohol use [33] (ALCP); 4) alcohol consumption (drinks per week, ALCC) [34, 35]; 5) cigarettes per day/FTND (CPD) [34, 36]; and 6) schizophrenia (SCZ) [37, 38] . We focused on these PGS, specifically, because: 1) SUDs show strong genetic overlap with other externalizing [39–41], internalizing [32, 42], and psychotic disorders [33, 43, 44]; 2) both shared and substance-specific genetic risk are associated with later SUDs [45–47]; and 3) substance use and SUDs have only partial genetic overlap [48, 49]. Therefore, our PGS cover a spectrum of genetic risk for SUDs, using the most current and well-powered results for each of the listed domains (see supplementary information section 4 for a detailed description).

GWAS have been overwhelmingly limited to individuals of European ancestries [50, 51]. Importantly, PGS derived from GWAS of one ancestry do not always transport into other ancestral populations [52, 53]. We therefore used PRS-CSx [54], a new method that combines information from well-powered GWAS (typically of European ancestries) and ancestrally matched GWAS to improve the predictive power of PGS in the African ancestry samples from Add Health and COGA. PRS-CSx integrates GWAS summary statistics across multiple input populations and employs a Bayesian approach to correct GWAS summary statistics for the non-independence of SNPs in linkage disequilibrium (LD) with one another [54]. For participants of European ancestries, we used the EUR derived PRS-CSx results, while we used the EUR+AFR meta-analyzed results for the African ancestry participants. See the supplementary information (section 5) for details.

Analytic Strategy

We pooled all the data for analysis using a fixed effects integrative data analytic (IDA) approach [55]. The IDA approach is more powerful than traditional meta-analyses when one has access to raw data for each of the contributing samples. Our approach to harmonization and pooling was as follows. First, we defined the measures and cutoffs to be used in each of the samples, creating the CERI, PGS, and SUD outcomes at the cohort level. Second, within each cohort, we regressed each PGS on age, age^2, sex, sex*age, sex*age^2, and the first 10 ancestral PCs (specific to each sample) to account for population stratification in the PGS. Next, we pooled all the data for analysis. We included cohort as a fixed effect for each of the six cohorts (4 samples, of which two were split by ancestry) in subsequent analyses. Additionally, we included age of last observation and sex as covariates.

We estimated a series of nested logistic regression models with the pooled data: 1) a baseline model (sex, age, and cohort), 2) a genetic risk model (baseline + PGS), 3) a clinical/environmental risk model (baseline + CERI), and 4) a combined risk model (baseline +

PGS + CERI). Because COGA and FT12 included a large number of related individuals, we adjusted for familial clustering using cluster-robust standard errors [56]. To assess the predictive accuracy of each model, we took the difference in pseudo-R^2 (Δ*Pseudo*-R^2) [57], between the baseline and corresponding models. Finally, we calculated the discriminatory power of the combined model using the area under the curve (AUC) from a receiver operating characteristic (ROC) curve. We included a variety of robustness checks to ensure that no single cohort in the IDA was unduly influencing the results. Our analytic strategy was preregistered on the Open Science Framework (https://osf.io/etbw8). Deviations from the preregistration are described in the supplementary information (section 6).

RESULTS

Table 2 contains the descriptive statistics for each of the cohorts and ancestries. Each cohort had similar proportions of females (~51% - 56%). The mean ages ranged from ~22 to ~29 years of age. The COGA cohorts (both European and African ancestries) reported the highest rates of SUD, an expected finding given the nature of the sample (highly selected for SUDs). Add Health participants generally had higher rates of SUD than ALSPAC or FinnTwin12, but lower than COGA. Finally, ALSPAC and FinnTwin12 reported similar levels of alcohol, nicotine, drug, and any substance dependence. COGA participants reported higher mean values on the CERI. The remaining cohorts report relatively similar rates of risk factor exposure.

Table 3 presents the results from the *PGS only*, *CERI only*, and *combined* models for each outcome. Three of the six PGS were associated with the SUD outcomes in the *PGS only* model. EXT was associated with each of the SUD outcomes (EXT OR = 1.18 – 1.50); ALCP was associated with alcohol dependence and any substance dependence (ALCP OR = 1.10 – 1.13); and CPD was associated with nicotine dependence (CPD OR = 1.33). In the *CERI only* models, the CERI was consistently associated across each of the SUD categories (ORs = 1.37 – 1.67).

When we combined the PGS and CERI into the same model, the CERI remained significant across SUDs and was largely unchanged (ORs = 1.35 – 1.65). EXT remained associated with drug dependence (OR = 1.11) and nicotine dependence (OR = 1.33), ALCP remained associated alcohol dependence (OR = 1.12), and CPD remained associated with nicotine dependence (OR = 1.31). Both EXT and ALCP remained associated with any substance dependence diagnosis (ORs = 1.09 – 1.18). Overall, the combined model explained 5.9%, 12.6%, 13.1%, and 12.8% of the variance in alcohol dependence, nicotine dependence, drug dependence, and any substance dependence, respectively.

Figure 1 (Panel A) presents the raw prevalence for each outcome across counts of the CERI. The proportion of those meeting criteria for SUDs among those reporting 3 or more, 5 or more, and 7 or more risk factors surpassed lifetime prevalence estimates from nationally representative samples for drug dependence, alcohol dependence, and nicotine dependence, respectively [58]. Panel B depicts the prevalence of each category of SUD across several mutually exclusive categories: 1) those in the bottom 90% of both the CERI and all PGS (averaged across the six scores); 2) those in the top 10% of the CERI but the bottom 90% of the PGS distribution; 3) those in the top 10% of the PGS distribution and the bottom 90% of the CERI; and 4) those in the top 10% of both PGS and the CERI. There is an increase in risk across those with elevated genetic risk, clinical/environmental risk, and both. Those in the top 10% of both PGS and CERI had the highest prevalence of each of the SUDs, though the error bars overlap with the estimates from those in the top 10% of the risk index, alone. Compared to those in the bottom 90% on both, those in the to the top 10% of both have a relative risk of 3.86 (95% CI = 3.20, 4.65) for alcohol dependence, 6.11 (95% CI = 4.84, 7.72) for nicotine dependence, 8.04 (95% CI = 6.92, 9.36) for drug dependence, and 4.05 (95% CI = 3.64, 4.51) for any substance dependence.

Finally, we considered the AUC for the combined model for each of the SUD categories. Figure 2 presents the ROC curves for the full (CERI and PGS) and baseline (covariates only)

models for each SUD category. The AUC for each combined model was 0.74 for alcohol dependence, 0.82 for nicotine dependence, 0.86 for drug dependence, and 0.78 for any substance dependence. The overall change in AUC (from the baseline to the full model) that we achieve when adding the CERI and PGS was modest (ΔAUC = 0.05 – 0.10), and this improvement was due in large part to the explanatory power of the CERI. ROC curves for the CERI only and PGS only models are presented in Supplemental Figure 6.

Sensitivity Analyses

We performed a variety of sensitivity analyses. Results from leave-one-out (LOO) and sex-stratified analyses were largely similar to those from the main results. In ancestry stratified analyses, results in the cohorts of European ancestries largely mirrored the main results. None of the PGS were associated with SUDs in the cohorts of African ancestries. Effect sizes for the CERI were largely similar across European and African ancestries (see Supplemental Tables S1-S3) and were mostly stable when removing individual risk factors (supplemental information section 7).

We also tested for interactions between the PGS and CERI and cohort (Add Health EUR as the reference group). There were few significant interactions and no consistent patterns in variation for PGS, though the CERI did show considerable variation across cohort (Supplemental Table S4). Finally, we fit complimentary models using a random effects approach, allowing the slopes for the PGS and CERI to vary randomly across cohort. Random slopes for PGS did not consistently improve model fit, though a random slope for the CERI consistently improved model fit (Supplemental Table S5). We compared the parameter estimates from the random effect models to the main analyses and results were largely consistent (Supplemental Table S6).

DISCUSSION

Substance use disorders remain a serious threat to public health. In the current analysis, we examined the combination of clinical, environmental, and genetic risk factors for determining who is more likely to develop a SUD in early adulthood. We used previously validated measures of environmental and clinical risk [11, 29, 30] and polygenic scores for externalizing problems [31], depression [32], problematic alcohol use [33, 35], alcohol consumption [34, 35], cigarettes per day/nicotine dependence [34, 36], and schizophrenia [37, 38]. The combination of genetic and social-environmental measures was significantly associated with the development of SUDs. The overall association was strongest for drug dependence, followed by any substance dependence, nicotine dependence, and alcohol dependence.

The CERI was the strongest association with each outcome. The proportion of those meeting criteria for each SUD surpassed lifetime estimates in persons with 3 or more, 5 or more, and 7 or more risk factors for drug dependence, alcohol dependence, and nicotine dependence, respectively. The discriminatory power of the combined model (AUC = .74 - .86) was similar to AUC estimates published in the original paper from which many of the risk index items were derived (AUC ~ 0.80) [11]. Interestingly, this risk index was originally developed for identifying persons with persistent SUD through early mid-life (~age 40). In the current analysis we demonstrated that the CERI in conjunction with demographic covariates and PGS does equally well for those who meet criteria for any SUD by young adulthood.

The overall predictive power of the PGS alone was in the range of 1.1 – 3.7%. Only the PGS for externalizing problems, problematic alcohol use, and cigarettes per day were consistently associated with SUD outcomes. The PGS for externalizing problems was associated with drug dependence and nicotine dependence, the PGS for problematic alcohol use PGS was associated with alcohol dependence, and both were associated with any substance dependence. The PGS for cigarettes per day was only associated with nicotine dependence. Overall, these results

support prior evidence that genetic risk for SUDs consists of a both shared and substance-specific variance [31, 41, 47].

Interestingly, even though the effect sizes were attenuated in the model, the PGS for externalizing problems, problematic alcohol use, and cigarettes per day remained significantly associated when we included the CERI, though the additional information the PGS provided was minimal. Since the CERI also included many of the phenotypes each of the PGS measured (e.g., childhood conduct disorder for externalizing, childhood depression for depression; and frequent alcohol use for alcohol consumption), part of this attenuation is likely due to the inclusion of the actual phenotypes through which risk for some of these disorders is expressed. PGS are also confounded by environmental variance [59] and the reduction in effect sizes could be accounting for some of that confounding. PGS may add information beyond well-known risk factors, which could prove useful when information on certain exposures or behaviors is unavailable.

Further refinement of risk measures may improve our ability to develop screening protocols for those at greater risk of developing substance-related problems. Early detection has the potential to improve prevention efforts, as prior work suggests that those at highest risk of substance misuse stand to benefit the most from prevention efforts [60]. Ideally, screening tools for SUD risk would include measures of social, clinical, and genetic risk factors, as each impacts the development of SUDs [5–10]. In the push for precision medicine, the focus is often on biological information, but social determinants of health are also critically important.

Currently, these tools are not ready for clinical use. If we reach the point where social, clinical, and genetic information become sufficiently powerful, we must recognize that identifying persons for early intervention carries a significant risk. Screening for social determinants has the potential for unintended consequences, including further stigmatization [61]. Genetic information has even more potential for abuse. Policy makers must ensure that there is comprehensive legal protection against discrimination using any form of information. Additionally, any attempt to use social, clinical, or genetic information for targeted intervention or identification in a clinical setting

must be done so in a patient-centered approach, rather than any “one-size fits all” that exclude patients from their own healthcare decisions [62].

Our analysis has several important limitations. First, although we included individuals of diverse ancestries, the PGS for our samples of African ancestries were severely underpowered due to the small size of the discovery sample. Large-scale GWAS in diverse cohorts are vital to ensuring that any benefit of precision medicine is shared equitably across the population [63]. Second, while distinct, ancestry is related to race-ethnicity, and with it, racism and racial discrimination, some of the most profound social determinants of health [64]. Our measure of environmental risk was crude and may not fully capture risk factors that contribute to SUDs in populations beyond non-Hispanic Whites. Future studies should include racially relevant measures of risk (e.g., experiences of interpersonal racism/discrimination, racial residential segregation) as well as other social and environmental measures that are known risk factors for SUDs (e.g., neighborhood social conditions, alcohol outlet density). Further refinement of known risk factors may allow for better prediction of those at risk of developing an SUD. We did observe variation in the predictive ability of the CERI across cohorts, suggesting the observed effect may differ in magnitude across populations. We therefore urge caution in overinterpreting study results. Finally, while we tried to ensure time order between risk factors and onset of disorder, some risk factors (particularly adolescent substance use) could have occurred concurrently with diagnosis. Future work in samples with risk factors measured before the initiation of substance use (such as the Adolescent Brain Cognitive Development Study) will be important for replication efforts.

Recognizing that multiple social, clinical, and genetic factors contribute to risk for SUDs is important as we move towards the goal precision medicine that benefits all segments of the population. There is still much work to be done before tools such as these are useful in a clinical setting. However, the results of this integrative data analysis provide initial evidence *each* of these risk factors contribute unique information to SUDs in early adulthood. Expanding our sources of

information (such as electronic health records, census data from home of record) and making use of increasingly well-powered PGS will continue to improve our ability to understand how SUDs develop.

ACKNOWLEDGEMENTS

Research reported in this publication was supported by the National Institute on Alcohol Abuse and Alcoholism and the National Institute of Drug Abuse of the National Institutes of Health under award numbers R01AA015416, R01DA050721, R01DA042090, and K02AA018755; the Academy of Finland (grants 100499, 205585, 118555, 141054, 265240, 308248, 308698 and 312073); and the Scientific and Technological Research Council of Turkey (TÜBİTAK) under award number 114C117 (FA); and the Sigrid Juselius Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding bodies. This research also used summary data from the Psychiatric Genomics Consortium (PGC), the Million Veterans Program (MVP), the GWAS and Sequencing Consortium for Alcohol and Nicotine (GSCAN), UK Biobank, the Genomic Psychiatry Cohort (GPC) and 23andMe, Inc. We would like to thank the many studies that made these consortia possible, the researchers involved, and the participants in those studies, without whom this effort would not be possible. We would also like to thank the research participants and employees of 23andMe.

The Externalizing Consortium: Principal Investigators: Danielle M. Dick, Philipp Koellinger, K. Paige Harden, Abraham A. Palmer. Lead Analysts: Richard Karlsson Linnér, Travis T. Mallard, Peter B. Barr, Sandra Sanchez-Roige. Significant Contributors: Irwin D. Waldman. The Externalizing Consortium has been supported by the National Institute on Alcohol Abuse and Alcoholism (R01AA015416 - administrative supplement), and the National Institute on Drug Abuse (R01DA050721). Additional funding for investigator effort has been provided by K02AA018755, U10AA008401, P50AA022537, as well as a European Research Council

Consolidator Grant (647648 EdGe to Koellinger). The content is solely the responsibility of the authors and does not necessarily represent the official views of the above funding bodies. **Add Health:** Add Health is directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Waves I-V data are from the Add Health Program Project, grant P01 HD31921 (Harris) from *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill. **ALSPAC:** We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses. The UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors, and Peter Barr and Danielle Dick will serve as guarantors for the contents of this paper. A comprehensive list of grants funding is available on the ALSPAC website (http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf); This research was specifically funded by the Medical Research Council (MRC) under grants MR/L022206/1, MR/M006727/1, and G0800612/86812; the Wellcome Trust under grant 086684; and the National Institute on Alcohol Abuse and Alcoholism under 5R01AA018333-05. GWAS data was generated by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. **COGA:** We thank The Collaborative Study on the Genetics of Alcoholism (COGA), Principal Investigators B. Porjesz, V. Hesselbrock, T. Foroud; Scientific Director, A. Agrawal; Translational Director, D. Dick, includes eleven different centers: University of Connecticut (V. Hesselbrock); Indiana University (H.J. Edenberg, T. Foroud, Y. Liu, M.H. Plawecki); University of Iowa Carver College of Medicine

(S. Kuperman, J. Kramer); SUNY Downstate Health Sciences University (B. Porjesz, J. Meyers, C. Kamarajan, A. Pandey); 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, R. Hart, J. Salvatore); The Children's Hospital of Philadelphia, University of Pennsylvania (L. Almasy); Virginia Commonwealth University (D. Dick); Icahn School of Medicine at Mount Sinai (A. Goate, P. Slesinger); and Howard University (D. Scott). Other COGA collaborators include: L. Bauer (University of Connecticut); J. Nurnberger Jr., L. Wetherill, X., Xuei, D. Lai, S. O'Connor, (Indiana University); G. Chan (University of Iowa; University of Connecticut); D.B. Chorlian, J. Zhang, P. Barr, S. Kinreich, G. Pandey (SUNY Downstate); N. Mullins (Icahn School of Medicine at Mount Sinai); A. Anokhin, S. Hartz, E. Johnson, V. McCutcheon, S. Saccone (Washington University); J. Moore, Z. Pang, S. Kuo (Rutgers University); A. Merikangas (The Children's Hospital of Philadelphia and University of Pennsylvania); F. Aliev (Virginia Commonwealth University); H. Chin and A. Parsian 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). All code necessary to replicate this study is available upon request. We are grateful Drs. Amy Taylor and Joëlle A. Pasman for providing GWAS summary statistics with ALSPAC samples removed.

ETHICS DECLARATIONS

The authors have no conflicts of interest to declare.

CONTRIBUTIONS

P.B.B., S.I.K., M.N.D., and D.M.D. conceived the study. D.M.D. oversaw the study. P.B.B. led the writing of the manuscript, with substantive contributions to the writing from D.M.D., S.I.K., and M.N.D. P.B.B. was the lead analyst and prepared data in Add Health and FinnTwin12. S.I.K. prepared data in COGA. M.N.D. prepared data in ALSPAC. R.K.L., F.A., and J.M. provided GWAS summary statistics. M.S., K.P.H, B.P., K.B., J.K., A.L., H.J.E., M.H.P., and A.A.P provided helpful advice and feedback on various aspects of the study design. All authors contributed to and critically reviewed the manuscript.

CORRESPONDING AUTHORS

Correspondence to Peter B. Barr or Danielle M. Dick.

REFERENCES

1. U.S. Overdose Deaths In 2021 Increased Half as Much as in 2020 - But Are Still Up 15%. https://www.cdc.gov/nchs/pressroom/nchs_press_releases/2022/202205.htm. Accessed 15 May 2022.
2. Degenhardt L, Charlson F, Ferrari A, Santomauro D, Erskine H, Mantilla-Herrara A, et al. The global burden of disease attributable to alcohol and drug use in 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Psychiatry. 2018;5:987–1012.
3. Sacks JJ, Gonzales KR, Bouchery EE, Tomedi LE, Brewer RD. 2010 National and State Costs of Excessive Alcohol Consumption. Am J Prev Med. 2015;49:e73–e79.
4. National Drug Intelligence Center. National drug threat assessment. vol. 2019. Washington, DC: United States Department of Justice; 2011.
5. Verhulst B, Neale MC, Kendler KS. The heritability of alcohol use disorders: a meta-analysis of twin and adoption studies. Psychol Med. 2015;45:1061–1072.
6. Verweij KJH, Zietsch BP, Lynskey MT, Medland SE, Neale MC, Martin NG, et al. Genetic and environmental influences on cannabis use initiation and problematic use: A meta-analysis of twin studies. Addiction. 2010;105:417–430.
7. Kendler KS, Jacobson KC, Prescott CA, Neale MC. Specificity of Genetic and Environmental Risk Factors for Use and Abuse/Dependence of Cannabis, Cocaine, Hallucinogens, Sedatives, Stimulants, and Opiates in Male Twins. American Journal of Psychiatry. 2003;160:687–695.
8. Galea S, Nandi A, Vlahov D. The Social Epidemiology of Substance Use. Epidemiol Rev. 2004;26:36–52.
9. Barr PB. Neighborhood conditions and trajectories of alcohol use and misuse across the early life course. Health Place. 2018;51:36–44.

444 10. Barr PB, Silberg J, Dick DM, Maes HH. Childhood socioeconomic status and longitudinal
445 patterns of alcohol problems: Variation across etiological pathways in genetic risk. Soc Sci
446 Med. 2018;209:51–58.
447 11. Meier MH, Hall W, Caspi A, Belsky DW, Cerda M, Harrington HL, et al. Which adolescents
448 develop persistent substance dependence in adulthood? Using population-representative
449 longitudinal data to inform universal risk assessment. Psychol Med. 2016;46:877–889.
450 12. Schaefer JD, Jang SK, Clark DA, Deak JD, Hicks BM, Iacono WG, et al. Associations
451 between polygenic risk of substance use and use disorder and alcohol, cannabis, and
452 nicotine use in adolescence and young adulthood in a longitudinal twin study. Psychol Med.
453 2021:1–11.
454 13. Deak JD, Clark DA, Liu M, Schaefer JD, Jang SK, Durbin CE, et al. Alcohol and nicotine
455 polygenic scores are associated with the development of alcohol and nicotine use problems
456 from adolescence to young adulthood. Addiction. 2022;117:1117–1127.
457 14. Barr PB, Ksinan A, Su J, Johnson EC, Meyers JL, Wetherill L, et al. Using polygenic scores
458 for identifying individuals at increased risk of substance use disorders in clinical and
459 population samples. Transl Psychiatry. 2020;10:196.
460 15. Kinreich S, Meyers JL, Maron-Katz A, Kamarajan C, Pandey AK, Chorlian DB, et al.
461 Predicting risk for Alcohol Use Disorder using longitudinal data with multimodal biomarkers
462 and family history: a machine learning study. Mol Psychiatry. 2021;26:1133–1141.
463 16. Gu F, Chen TH, Pfeiffer RM, Fargnoli MC, Calista D, Ghiorzo P, et al. Combining common
464 genetic variants and non-genetic risk factors to predict risk of cutaneous melanoma. Hum
465 Mol Genet. 2018. 2018. https://doi.org/10.1093/hmg/ddy282.
466 17. O'Sullivan JW, Shcherbina A, Justesen JM, Turakhia M, Perez M, Wand H, et al.
467 Combining Clinical and Polygenic Risk Improves Stroke Prediction among Individuals with
468 Atrial Fibrillation. Circ Genom Precis Med. 2021;14:339–347.

18. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE, et al. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry. 2005;62:593.

19. Harris KM, Halpern CT, Haberstick BC, Smolen A. The National Longitudinal Study of Adolescent Health (Add Health) sibling pairs data. Twin Research and Human Genetics. 2013;16:391–398.

20. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al. Cohort profile: The 'Children of the 90s'-The index offspring of the avon longitudinal study of parents and children. Int J Epidemiol. 2013. 2013. https://doi.org/10.1093/ije/dys064.

21. Fraser A, Macdonald-wallis C, Tilling K, Boyd A, Golding J, Davey smith G, et al. Cohort profile: The avon longitudinal study of parents and children: ALSPAC mothers cohort. Int J Epidemiol. 2013;42:97–110.

22. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42:377–381.

23. Northstone K, Lewcock M, Groom A, Boyd A, Macleod J, Timpson N, et al. The Avon Longitudinal Study of Parents and Children (ALSPAC): an update on the enrolled sample of index children in 2019 [version 1; peer review: 2 approved]. Wellcome Open Res. 2019;4.

24. Edenberg HJ. The collaborative study on the genetics of alcoholism: An update. Alcohol Research and Health. 2002;26:214–218.

25. Begleiter H. The Collaborative Study on the Genetics of Alcoholism. Alcohol Health Res World. 1995;19:228.

26. Bucholz KK, McCutcheon V V., Agrawal A, Dick DM, Hesselbrock VM, Kramer JR, et al. Comparison of Parent, Peer, Psychiatric, and Cannabis Use Influences Across Stages of

494 Offspring Alcohol Involvement: Evidence from the COGA Prospective Study. Alcohol Clin
495 Exp Res. 2017;41:359–368.
496 27. Rose RJRJ, Salvatore JEJE, Aaltonen S, Barr PBPB, Bogl LHLH, Byers HAHA, et al.
497 FinnTwin12 Cohort: An Updated Review. vol. 22. 2019.
498 28. Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO. The Fagerstrom Test for
499 Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. Br J Addict.
500 1991;86:1119–1127.
501 29. Hughes K, Bellis MA, Hardcastle KA, Sethi D, Butchart A, Mikton C, et al. The effect of
502 multiple adverse childhood experiences on health: a systematic review and meta-analysis.
503 Lancet Public Health. 2017. 2017. https://doi.org/10.1016/S2468-2667(17)30118-4.
504 30. Sher KJ, Grekin ER, Williams NA. The development of alcohol use disorders. Annu Rev
505 Clin Psychol. 2005;1:493–523.
506 31. Karlsson Linner R, Mallard TT, Barr PB, Sanchez-Roige S, Madole JW, Driver MN, et al.
507 Multivariate genomic analysis of 1.5 million people identifies genes related to addiction,
508 antisocial behavior, and health. Nat Neurosci. January .
509 32. Levey DF, Stein MB, Wendt FR, Pathak GA, Zhou H, Aslan M, et al. Bi-ancestral
510 depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million
511 individuals highlight new therapeutic directions. Nat Neurosci. 2021. 27 May 2021.
512 https://doi.org/10.1038/s41593-021-00860-2.
513 33. Zhou H, Sealock JM, Sanchez-Roige S, Clarke TK, Levey DF, Cheng Z, et al. Genome-
514 wide meta-analysis of problematic alcohol use in 435,563 individuals yields insights into
515 biology and relationships with other traits. Nat Neurosci. 2020. 2020.
516 https://doi.org/10.1038/s41593-020-0643-5.
517 34. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to 1.2
518 million individuals yield new insights into the genetic etiology of tobacco and alcohol use.
519 Nat Genet. 2019;51:237–244.

35. Kranzler HR, Zhou H, Kember RL, Vickers Smith R, Justice AC, Damrauer S, et al. Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. Nat Commun. 2019;10:1499.

36. Quach BC, Bray MJ, Gaddis NC, Liu M, Palviainen T, Minica CC, et al. Expanding the genetic architecture of nicotine dependence and its shared genetics with multiple traits. Nat Commun. 2020;11.

37. Trubetskoy V, Pardiñas AF, Qi T, Panagiotaropoulou G, Awasthi S, Bigdeli TB, et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature 2022. 2022:1–13.

38. Bigdeli TB, Fanous AH, Li Y, Rajeevan N, Sayward F, Genovese G, et al. Genome-Wide Association Studies of Schizophrenia and Bipolar Disorder in a Diverse Cohort of US Veterans. Schizophr Bull. 2020. 2020. https://doi.org/10.1093/schbul/sbaa133.

39. Barr PB, Dick DM. The Genetics of Externalizing Problems. Curr Top Behav Neurosci. 2020;47:93–112.

40. Krueger RF, Hicks BM, Patrick CJ, Carlson SR, Iacono WGWG, McGue M. Etiological connections among substance dependence, antisocial behavior and personality: Modeling the externalizing spectrum. J Abnorm Psychol. 2002;111:411–424.

41. Kendler KS, Myers J. The boundaries of the internalizing and externalizing genetic spectra in men and women. Psychol Med. 2014;44:647–655.

42. Polimanti R, Peterson RE, Ong JS, MacGregor S, Edwards AC, Clarke TK, et al. Evidence of causal effect of major depression on alcohol dependence: Findings from the psychiatric genomics consortium. Psychol Med. 2019. 2019. https://doi.org/10.1017/S0033291719000667.

43. Johnson EC, Demontis D, Thorgeirsson TE, Walters RK, Polimanti R, Hatoum AS, et al. A large-scale genome-wide association study meta-analysis of cannabis use disorder. Lancet Psychiatry. 2020. 2020. https://doi.org/10.1016/S2215-0366(20)30339-4.

546 44. Zhou H, Rentsch CT, Cheng Z, Kember RL, Nunez YZ, Sherva RM, et al. Association of
547 OPRM1 Functional Coding Variant With Opioid Use Disorder: A Genome-Wide Association
548 Study. JAMA Psychiatry. 2020. June 2020.
549 https://doi.org/10.1001/jamapsychiatry.2020.1206.
550 45. Kendler KS, Gardner C, Dick DM. Predicting alcohol consumption in adolescence from
551 alcohol- specific and general externalizing genetic risk factors, key environmental
552 exposures and their interaction. Psychol Med. 2011;41:1507–1516.
553 46. Meyers JL, Salvatore JE, Vuoksimaa E, Korhonen T, Pulkkinen L, Rose RJ, et al. Genetic
554 Influences on Alcohol Use Behaviors Have Diverging Developmental Trajectories : A
555 Prospective Study Among Male and Female Twins. Alcohol Clin Exp Res. 2014;38:2869–
556 2877.
557 47. Barr PB, Mallard TT, Sanchez-Roige S, Poore HE, Linnér RK, Collaborators C, et al.
558 Parsing Genetically Influenced Risk Pathways: Genetic Loci Impact Problematic Alcohol
559 Use Via Externalizing and Specific Risk. MedRxiv. 2021:2021.07.20.21260861.
560 48. Sanchez-Roige S, Palmer AA, Clarke TK. Recent Efforts to Dissect the Genetic Basis of
561 Alcohol Use and Abuse. Biol Psychiatry. 2020.
562 49. Walters RK, Polimanti R, Johnson EOECEO, McClintick JN, Adams MJ, Adkins AE, et al.
563 Trans-ancestral GWAS of alcohol dependence reveals common genetic underpinnings
564 with psychiatric disorders. Nat Neurosci. 2018;21:1656–1669.
565 50. Dick DM, Barr P, Guy M, Nasim A, Scott D. Review: Genetic research on alcohol use
566 outcomes in African American populations: A review of the literature, associated
567 challenges, and implications. American Journal on Addictions. 2017;26:486–493.
568 51. Mills MC, Rahal C. A scientometric review of genome-wide association studies. Commun
569 Biol. 2019;2:9.

570 52. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human
571 Demographic History Impacts Genetic Risk Prediction across Diverse Populations. Am J
572 Hum Genet. 2017;100:635–649.
573 53. Duncan L, Shen H, Gelaye B, Meijsen J, Ressler K, Feldman M, et al. Analysis of polygenic
574 risk score usage and performance in diverse human populations. Nat Commun. 2019.
575 2019. https://doi.org/10.1038/s41467-019-11112-0.
576 54. Ruan Y, Lin Y-F, Feng Y-CA, Chen C-Y, Lam M, Guo Z, et al. Improving Polygenic
577 Prediction in Ancestrally Diverse Populations. Nat Genet. 2022. 2022.
578 https://doi.org/10.1038/s41588-022-01054-7.
579 55. Curran PJ, Hussong AM. Integrative Data Analysis: The Simultaneous Analysis of Multiple
580 Data Sets. Psychol Methods. 2009;14:81–100.
581 56. Cameron CA, Gelbach JB, Miller DL. Robust inference with multiway clustering. Journal of
582 Business and Economic Statistics. 2011;29:238–249.
583 57. Nagelkerke NJD. A note on a general definition of the coefficient of determination.
584 Biometrika. 1991;78:691–692.
585 58. Hasin DS, Grant BF. The National Epidemiologic Survey on Alcohol and Related
586 Conditions (NESARC) Waves 1 and 2: review and summary of findings. Soc Psychiatry
587 Psychiatr Epidemiol. 2015;50:1609–1640.
588 59. Kong A, Thorleifsson G, Frigge ML, Vilhjalmsson BJ, Young AI, Thorgeirsson TE, et al.
589 The nature of nurture: Effects of parental genotypes. Science (1979). 2018;359:424–428.
590 60. Conrod PJ, O'Leary-Barrett M, Newton N, Topper L, Castellanos-Ryan N, MacKie C, et al.
591 Effectiveness of a Selective, Personality-Targeted Prevention Program for Adolescent
592 Alcohol Use and Misuse: A Cluster Randomized Controlled Trial. JAMA Psychiatry.
593 2013;70:334–342.
594 61. Garg A, Boynton-Jarrett R, Dworkin PH. Avoiding the Unintended Consequences of
595 Screening for Social Determinants of Health. JAMA. 2016;316:813–814.

62. Davidson KW, McGinn T. Screening for Social Determinants of Health: The Known and Unknown. JAMA. 2019;322:1037–1038.

63. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 2019;51:584–591.

64. Williams DR, Mohammed SA, Leavell J, Collins C. Race, socioeconomic status, and health: complexities, ongoing challenges, and research opportunities. Ann N Y Acad Sci. 2010;1186:69–101.

DATA AVAILABILITY

All data sources are described in the manuscript and supplemental information. No new data were collected. Only data from existing studies or study cohorts were analyzed, some of which have restricted access to protect the privacy of the study participants. Add Health genetic data obtained through dbGaP (Study Accession: phs001367.v1.p1). Instructions on gaining access to Add Health restricted use data can be found at: https://data.cpc.unc.edu/projects/2/view. COGA genetic data available through dbGaP (Study Accession: phs000763.v1.p1). Instructions for access to ALSPAC data available at: http://www.bristol.ac.uk/alspac/researchers/access/. The process for obtaining the GWAS summary statistics used in these analyses are described in the corresponding original GWAS publications.

CODE AVAILABILITY

No custom algorithms or software was developed in this study. All code is available by request from the corresponding author. Polygenic scores generated using PRS-CSx (https://github.com/getian107/PRScsx). All primary analyses completed in R 4.1.0 using the

619 *data.table* (1.14.0), *pROC* (1.18.0), *lme4* (1.1-27.1), *DescTools* (0.99.45), *sandwich* (3.0-2), and

620 base packages.

621

Table 1: Items included in the Clinical/Environmental Risk Index (CERI)

Measure	Definition
1) Low childhood SES	Parent(s) report having less than basic level of education [culturally dependent]; having a low-skill or menial occupation; income at or below the poverty line; or receipt of government assistance.
2) Family history of SUD	Biological parent self-reports history of SUD for themselves or other biological parent or meets criteria for SUD from clinical interview/AUDIT threshold of 8 or higher.
3) Childhood externalizing problems	Respondent meets criteria for conduct disorder or oppositional defiant disorder from a clinical interview or computer-based prediction; or has a behavior problems score at or above the 90th percentile at 15 or younger.
4) Childhood internalizing problems	Respondent reports diagnosis of depression/anxiety or panic disorder; meets criteria for internalizing disorder in clinical interview/computer-based prediction; or has a CES-D score above a threshold of 16 at 15 or younger.
5) Early initiation of substance use	Respondent reports age of first whole alcoholic drink, smoked whole cigarette, or tried cannabis before the age of 15.
6) Adolescent alcohol use	Frequency of self-reported use 5 or more days per week at age 18 and below.
7) Adolescent tobacco use	Frequency of self-reported use at daily use at age 18 and below.
8) Adolescent cannabis use	Frequency of self-reported use 5 or more days per week at age 18 and below.
9) Peer substance use	Respondent reports the majority of their best friends use alcohol/tobacco/cannabis; their three best friends smoke daily/drink once a month/use cannabis once a month; or more than one friend smokes/drinks alcohol/has tried other drugs.
10) Traumatic events	Respondent reports exposure to any traumatic event.

622 Full description of sample specific definitions available in the supplementary information.

Table 2: Prevalence of SUDs and CERI by Cohort

	Add Health AFR (N = 1,605)*		*Add Health* EUR (N = 4,855)*		*ALSPAC* EUR (N = 4,733)*		*COGA* AFR (N = 870)*		*COGA* EUR (N = 1,878)*		*FinnTwin12* EUR (N = 1,193)*	
	Mean (SD)/%		**Mean (SD)/%**		**Mean (SD)/%**		**Mean (SD)/%**		**Mean (SD)/%**		**Mean (SD)/%**	
Female	55.26%	-	53.59%	-	56.71%	-	51.38%	-	51.33%	-	53.73%	-
Age (at last observation)	28.89	(1.69)	28.84	(1.70)	22.47	(2.20)	24.13	(5.12)	24.24	(5.26)	22.44	(0.72)
Alcohol dependence	3.93%	-	12.75%	-	5.92%	-	11.49%	-	21.14%	-	8.55%	-
Nicotine dependence	2.74%	-	10.28%	-	1.54%	-	3.91%	-	7.83%	-	2.26%	-
Drug dependence	6.73%	-	10.79%	-	0.78%	-	26.44%	-	23.59%	-	1.34%	-
Any substance dependence†	11.21%	-	25.81%	-	8.87%	-	30.69%	-	34.66%	-	10.98%	-
CERI	1.95	(1.48)	2.07	(1.65)	2.08	(1.19)	3.98	(2.24)	3.65	(2.38)	2.62	(1.27)

* Available samples with genotypic, phenotypic, and environmental risk data

† Any substance dependence includes those who meet criteria for alcohol, nicotine, or drug dependence.

AFR = African ancestries; EUR = European ancestries; CERI = clinical/environmental risk index

Table 3: Estimates for PGS Only, CERI Only, and Combined Models

		Alcohol Dependence			Nicotine Dependence			Drug Dependence			Any substance dependence		
		OR	**95% CI**		**OR**	**95% CI**		**OR**	**95% CI**		**OR**	**95% CI**	
PGS Only Model*	ALCC PGS	1.05	(0.99,	1.11)	0.96	(0.89,	1.04)	1.05	(0.98,	1.12)	1.00	(0.96,	1.05)
	ALCP PGS	**1.13**	**(1.06,**	**1.20)**	1.01	(0.93,	1.10)	1.07	(1.00,	1.15)	**1.10**	**(1.05,**	**1.16)**
	EXT PGS	**1.18**	**(1.11,**	**1.26)**	**1.50**	**(1.38,**	**1.63)**	**1.27**	**(1.19,**	**1.36)**	**1.31**	**(1.25,**	**1.38)**
	DEP PGS	1.00	(0.94,	1.06)	1.06	(0.98,	1.15)	1.08	(1.02,	1.15)	1.02	(0.98,	1.07)
	SCZ PGS	1.04	(0.97,	1.10)	0.98	(0.90,	1.06)	1.03	(0.96,	1.11)	1.00	(0.96,	1.05)
	CPD PGS	1.00	(0.94,	1.06)	**1.33**	**(1.24,**	**1.43)**	1.01	(0.95,	1.08)	1.08	(1.03,	1.13)
ΔPseudo-R^2			*0.011*			*0.037*			*0.014*			*0.022*	
CERI Only Model*	CERI	**1.37**	**(1.33,**	**1.41)**	**1.63**	**(1.57,**	**1.70)**	**1.67**	**(1.61,**	**1.72)**	**1.58**	**(1.54,**	**1.63)**
ΔPseudo-R^2			*0.054*			*0.107*			*0.129*			*0.120*	
Combined Model*	CERI	**1.35**	**(1.31,**	**1.40)**	**1.58**	**(1.52,**	**1.65)**	**1.65**	**(1.59,**	**1.70)**	**1.55**	**(1.51,**	**1.60)**
	ALCC PGS	1.04	(0.97,	1.10)	0.94	(0.87,	1.03)	1.03	(0.96,	1.11)	0.99	(0.94,	1.04)
	ALCP PGS	**1.12**	**(1.05,**	**1.19)**	0.99	(0.91,	1.08)	1.06	(0.98,	1.14)	**1.09**	**(1.04,**	**1.15)**
	EXT PGS	1.08	(1.01,	1.15)	**1.33**	**(1.22,**	**1.45)**	**1.11**	**(1.03,**	**1.20)**	**1.18**	**(1.12,**	**1.24)**
	DEP PGS	0.97	(0.91,	1.03)	1.02	(0.94,	1.10)	1.03	(0.96,	1.10)	0.98	(0.93,	1.03)
	SCZ PGS	1.03	(0.97,	1.10)	0.96	(0.88,	1.05)	1.01	(0.94,	1.08)	1.00	(0.95,	1.05)
	CPD PGS	0.98	(0.92,	1.04)	**1.31**	**(1.22,**	**1.42)**	0.98	(0.92,	1.04)	1.06	(1.01,	1.11)
ΔPseudo-R^2			*0.059*			*0.126*			*0.131*			*0.128*	

* All models included age, sex, and cohort as covariates. See Supplementary Table 7 for all parameter estimates. PGS residualized on age, sex, and first 10 ancestral principal components.
Bolded estimates = $p < .05$ after correction for multiple testing ($p < .05/4 = 0.0125$)
ΔPseudeo-R^2 denotes pseudo-R^2 above model including age, sex, and cohort. CI = confidence interval; PGS = polygenic score; CERI = clinical/environmental risk index

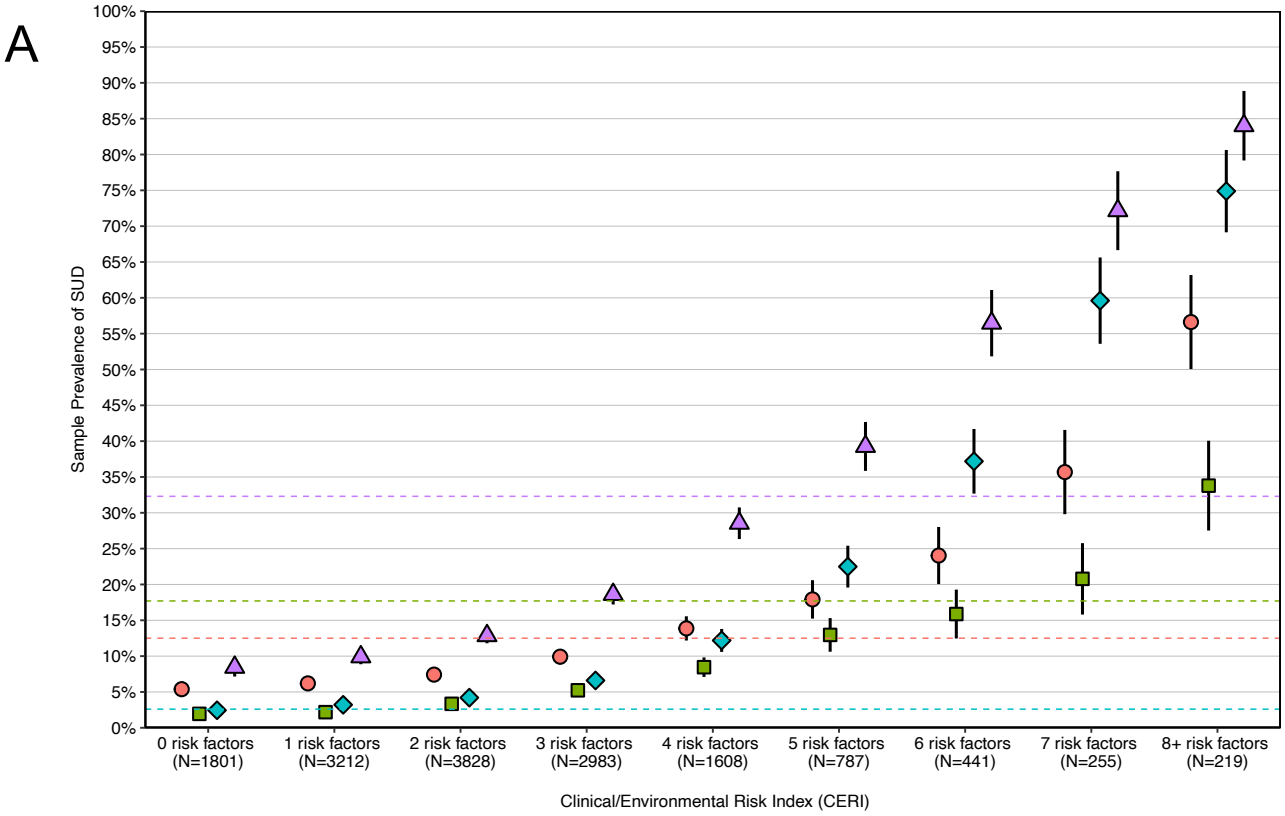
FIGURE CAPTIONS

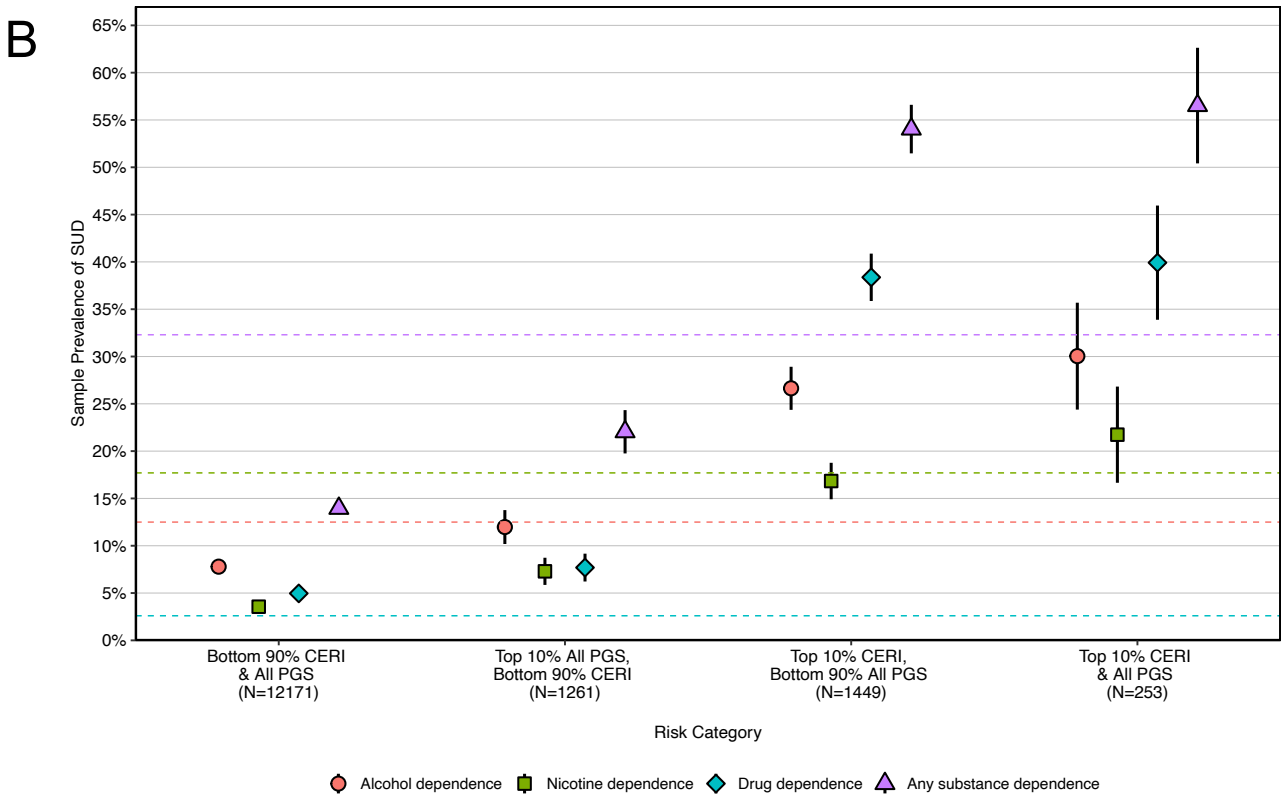
Figure 1: SUD Prevalence Across Genetic and Environmental Risk Factors

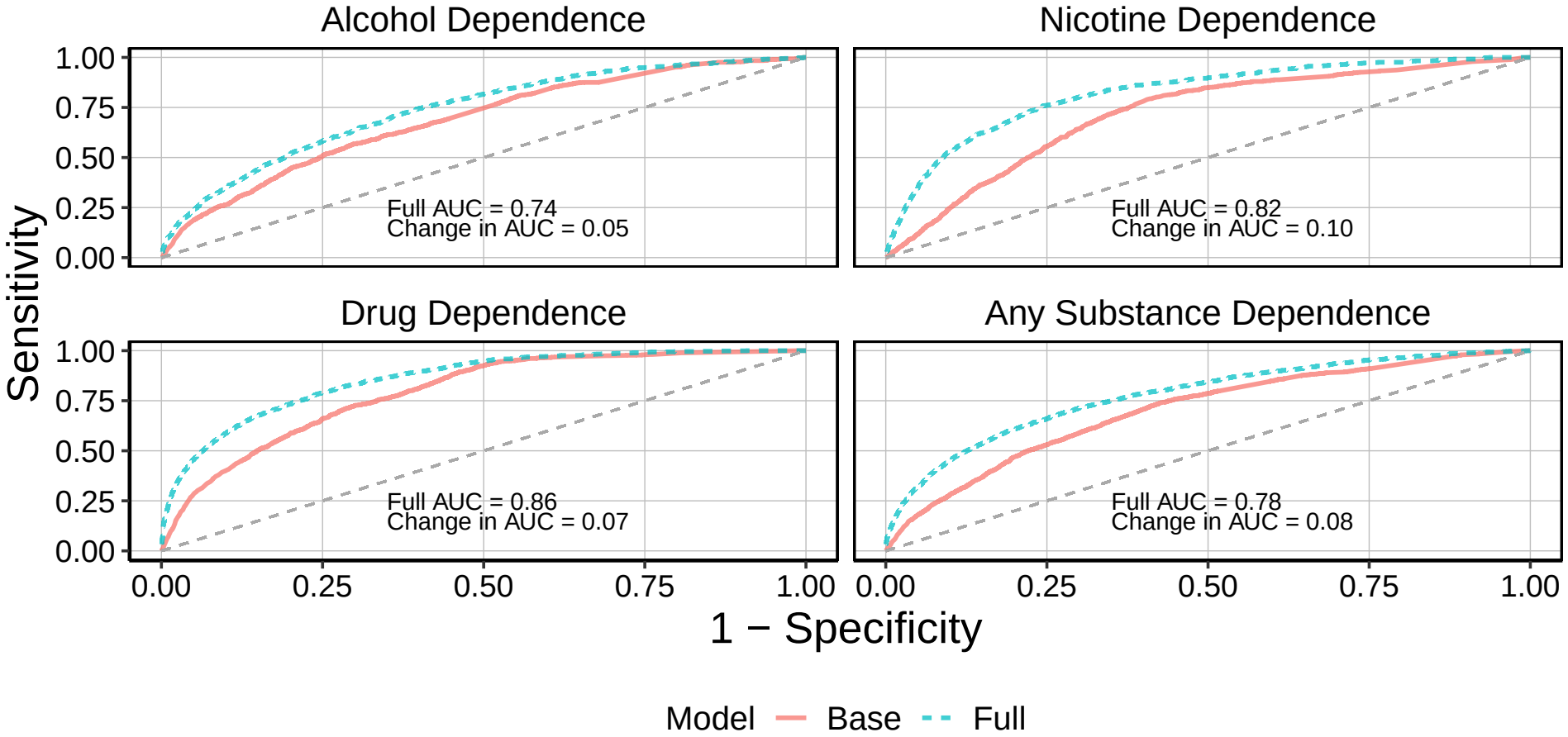
Panel A: Prevalence (and 95% confidence intervals) of those who meet criteria for alcohol, nicotine, drug, or any substance dependence across counts for items in the risk index. Panel B: Prevalence (and 95% confidence intervals) of those who meet criteria for alcohol, nicotine, drug, or any substance dependence across four categories: 1) those below the 90^{th} percentile for all PGS and the CERI; 2) those at or above the 90^{th} percentile for the CERI; 3) those at or above the 90^{th} percentile for all PGS; and 4) those at or above the 90^{th} percentile for both the CERI and PGS. PGS and risk index were first residualized on sex, age, age^2, cohort, sex*age, sex*age^2, sex*cohort, cohort*age, cohort*age^2, sex*cohort*age, and sex*cohort*age^2. Dotted colored lines represent corresponding lifetime prevalence estimates for alcohol dependence (red), nicotine dependence (green), drug dependence (blue), and any substance use disorder (purple) from nationally representative data [58].

Figure 2: ROC Curves for Combined and Baseline Models

Receiver operating characteristic (ROC) curves for baseline models (red line, covariates only) and the full models (blue line, PGS + CERI + covariates) for each substance use disorder. Area under the curve (AUC) is presented for the PGS model in each cell. Change in AUC represents value of the difference between AUC from the full model and AUC from the base model.


A
100%
95%
90%
85%
80%
75%
70%
65%
60%
55%
50%
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
Sample Prevalence of SUD
0 risk factors (N=1801)
1 risk factors (N=3212)
2 risk factors (N=3828)
3 risk factors (N=2983)
4 risk factors (N=1608)
5 risk factors (N=787)
6 risk factors (N=441)
7 risk factors (N=255)
8+ risk factors (N=219)
Clinical/Environmental Risk Index (CERI)


B
65%
60%
55%
50%
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
Sample Prevalence of SUD
Bottom 90% CERI & All PGS (N=12171)
Top 10% All PGS, Bottom 90% CERI (N=1261)
Top 10% CERI, Bottom 90% All PGS (N=1449)
Top 10% CERI & All PGS (N=253)
Risk Category
Alcohol dependence
Nicotine dependence
Drug dependence
Any substance dependence


Alcohol Dependence
Nicotine Dependence
Drug Dependence
Any Substance Dependence
Full AUC = 0.74
Change in AUC = 0.05
Full AUC = 0.82
Change in AUC = 0.10
Full AUC = 0.86
Change in AUC = 0.07
Full AUC = 0.78
Change in AUC = 0.08
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
Model
Base
Full

Supplementary Information

Clinical, Environmental, and Genetic Risk Factors for Substance Use Disorders: Characterizing Combined Effects across Multiple Cohorts

Correspondence to: peter.barr@downstate.edu, danielle.m.dick@rutgers.edu

Table of contents

1 Study introduction 3

2 Samples 4

2.1 The National Longitudinal Study of Adolescent to Adult Health (Add Health) 4

2.2 Avon Longitudinal Study of Parents and Children (ALSPAC) 4

2.3 The Collaborative Study of the Genetics of Alcoholism (COGA) 5

2.4 The Finnish Twin Cohort (FinnTwin12) 5

3 Clinical/environmental risk index measures 7

3.1 Low childhood socioeconomic status (SES) 7

3.2 Family history of substance use disorders (SUD) 8

3.3 Childhood behavior/externalizing problems 9

3.4 Childhood internalizing problems 9

3.5 Early substance use initiation 10

3.6 Frequent adolescent alcohol use 10

3.7 Frequent adolescent tobacco use 11

3.8 Frequent adolescent cannabis use 12

3.9 Adolescent peer substance use 12

3.10 Exposure to stressful/traumatic events 13

4 GWAS selection and inclusion 17

4.1 GWAS of externalizing (EXT) 17

4.2 GWAS of depression (DEP) 17

4.3 GWAS of problematic alcohol use (ALCP) 17

4.4 GWAS of alcohol consumption (ALCC) 18

4.5 GWAS of schizophrenia (SCZ) 18

4.6 GWAS of cigarettes per day/nicotine dependence (CPD) 18

5 Polygenic Score Creation 19

6 Deviations from preregistration 21

7 Variation in effect of clinical/environmental risk index (CERI) 22

8 ROC Curves for CERI only and PGS only Models 24

9 Random-effects Integrative Data Analysis (RE IDA) 26

10 References 27

1 Study introduction

Substance use disorders (SUD) are associated with substantial cost to society, affected individuals, and their families. In 2016, alcohol use contributed 4.2% to the global disease burden and other drug use contributed 1.3% [1]. Given the substantial human and economic costs of misuse and disorders, developing methods of identifying persons at heightened risk for SUD is a vital public health concern.

Ideally, screening tools for SUD risk would include measures of environmental, clinical, and genetic risk factors, as each are known to impact the development of substance use disorders [2–6]. Previous research using an index of established clinical and environmental risk factors related to adult SUD (e.g., childhood disadvantage, family history of SUD, childhood conduct problems, childhood depression, early exposure to substances, frequent use during adolescence) found this risk index to be useful (AUC ~ .80) in differentiating between individuals that were affected and unaffected with SUDs [7]. For measures of genetic risk, recent analyses evaluating the potential for polygenic risk scores, or PGS, which aggregate risk for a trait across the genome using information from genome-wide association studies (GWAS), have found current PGS alone provide little additional information to differentiate between individuals affected and unaffected by SUDs [8]. However, no research has examined these genetic, environmental, and clinical risk factors for SUD together. For other medical conditions, such as melanoma[9] or ischemic stroke [10], models using combined clinical and genetic risk factors showed improvement over models using individual risk factors in isolation.

The current proposal builds upon prior work developing risk indices for SUDs. We examined the joint effect of early life (defined as the periods of childhood and adolescence) risk factors and genetic liability (in the form of polygenic risk scores) to build prediction models for lifetime diagnosis of SUDs (alcohol dependence, drug dependence, and/or any substance dependence) using four longitudinal cohorts: the Collaborative Study on the Genetics of Alcoholism (COGA); the National Longitudinal Study of Adolescent to Adult Health (Add Health); the Avon Longitudinal Study of Parents and Children (ALSPAC); and the younger cohort of the Finnish Twin Study (FinnTwin12; FT12). We performed all analyses according to a preregistered analysis plan, which was time-stamped on December 3, 2020 (https://osf.io/etbw8).

2 Samples

2.1 The National Longitudinal Study of Adolescent to Adult Health (Add Health)

Add Health is an ongoing, nationally representative longitudinal study of adolescents followed into adulthood in the United States[11]. Data has been collected ranging from Wave I when respondents were between 11-18 (1994-1995) to Wave V (2016-2018) when respondents were 35-42. Add Health participants were selected from a stratified sample of 132 schools resulting in an initial, nationally representative sample of 90,118 students in grades 7-12. Of the original sample, 20,745 were selected for additional in-home interviews. Of those who completed the Wave I interview (1994-1995), 14,738 (71%) completed Wave II (1996); 15,197 (73%) completed Wave III (2001-2002); and 15,701 (75%) completed Wave IV (2007-2008). Most respondents completed the majority of the waves, with 16,278 (78%) completing three or more waves. Wave V (ages 32-42) data collection is underway, with a target sample of 19,828 (data for N = 3,872 is already released). In total, 15,159 individuals interviewed during Wave IV (ages 24-32) provided samples for genotyping, conducted using the Illumina Omni1 and Omni2.5 arrays. After quality control, genotypic data are available for 9,974 individuals (5,896 non-Hispanic White; 2,081 African American; 1,448 Hispanic; 550 Other). Genotypes for European ancestry participants were imputed to the Haplotype Reference Consortium (HRC) reference panel [12], and data for the African ancestry were imputed to the 1000 Genomes, Phase III reference panel [13]. The current analysis uses data from Waves I and II, when respondents were adolescents, and Wave IV, when respondents received a clinical interview assessing lifetime SUD diagnosis. We removed those who were >18 years old at Wave I to ensure timing of childhood/adolescent risk factors. Our final analytic sample consisted of 4,855 individuals of European ancestries and 1,605 individuals of African ancestries.

2.2 Avon Longitudinal Study of Parents and Children (ALSPAC)

ALSPAC is an ongoing, longitudinal population-based study of a birth cohort in the (former) Avon district of Southwest England[14–16]. Study data were collected and managed using REDCap electronic data capture tools hosted at the University of Bristol. REDCap (Research Electronic Data Capture) is a secure, web-based software platform designed to support data capture for research studies [17]. Pregnant women resident in Avon, UK with expected dates of delivery 1st April 1991 to 31st December 1992 were invited to take part in the study. The initial number of pregnancies enrolled is 14,541 (for these at least one questionnaire has been returned or a "Children in Focus" clinic had been attended by 19/07/99). Of these initial pregnancies, there was a total of 14,676 fetuses, resulting in 14,062 live births and 13,988 children who were alive at 1 year of age.

When the oldest children were approximately 7 years of age, an attempt was made to bolster the initial sample with eligible cases who had failed to join the study originally. As a result, when considering variables collected from the age of seven onwards (and potentially abstracted from obstetric notes) there are data available for more than the 14,541 pregnancies mentioned above. The number of new pregnancies not in the initial sample (known as Phase I enrollment) that are currently represented on the built files and reflecting enrollment status at the age of 24 is 913 (456, 262 and 195 recruited during Phases II, III and IV respectively), resulting in an additional 913 children being enrolled. The phases of enrollment are described in more detail in the cohort profile paper and its update (see footnote 4 below). The total sample size for analyses using any

data collected after the age of seven is therefore 15,454 pregnancies, resulting in 15,589 fetuses. Of these 14,901 were alive at 1 year of age.

A 10% sample of the ALSPAC cohort, known as the Children in Focus (CiF) group, attended clinics at the University of Bristol at various time intervals between 4 to 61 months of age. The CiF group were chosen at random from the last 6 months of ALSPAC births (1432 families attended at least one clinic). Excluded were those mothers who had moved out of the area or were lost to follow-up, and those partaking in another study of infant development in Avon.

Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committee. Consent for biological samples has been collected in accordance with the Human Tissue Act (2004). Informed consent for the use of data collected via questionnaires and clinics was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time. Children from the ALSPAC cohort were genotyped using the Illumina HumanHap550 quad chip genotyping platform[18]. Genotype data were imputed to the Haplotype Reference Consortium (HRC) reference panel [12]. Our final analytic sample consisted of 4,733 individuals of European ancestries.

2.3 The Collaborative Study of the Genetics of Alcoholism (COGA)

COGA, initiated in 1989 to identify genes associated with vulnerability for AUD, ascertained *high-risk families* through adult probands in treatment for alcohol dependence [19]. Probands along with all willing first-degree relatives were assessed; recruitment was extended to include additional relatives in families that contained 2 or more first degree relatives with alcohol dependence and community- ascertained comparison families (n = 16,848). Data collection included a psychiatric interview (the Semi-Structured Assessment for the Genetics of Alcoholism, or SSAGA [20]), neurophysiological and neuropsychological protocols, and collection of blood for DNA. We currently have genome wide data on 12,145 individuals (8,038 individuals of European ancestry; 3,655 individuals of African ancestry). In 2004, COGA began the prospective study of adolescents and young adults, targeting assessment of youth aged 12-22 from COGA families where at least one parent had been interviewed [21]. These subjects were re-assessed every two years; currently, 89% of individuals have 2+ interviews. COGA is *racially/ethnically diverse* (60.6% non-Hispanic White, 24.9% African American, 11.1% Hispanic, and 3.4% Other). Genotyping of the COGA samples was conducted across different phases of data collection. European ancestry (EA) samples were genotyped at multiple sites, including: (1) Center for Inherited Disease Research using the Illumina HumanHap1M array; (2) Genome Technology Access Center at Washington University School of Medicine using the Illumina OmniExpress; and (3) Rutgers University using the Affymetrix Smokescreen array. In addition, the two datasets genotyped on the Smokescreen genotyping array were also imputed separately, due to different processing pipelines used by the genotyping laboratory. Principal components were computed from GWAS data using Eigenstrat and 1000 Genomes, Phase III reference panel [13]. Individual ancestry was assigned using the YRI, CEU, JPT and CHB populations to set reference points. We limited our focus to the prospective sample of adolescent and young adult offspring (bassline ages 12-22; N = 3,573) of the original phases of COGA adult participants in the current analyses. Our final analytic sample consisted of 1,878 individuals of European ancestries and 870 individuals of African ancestries.

2.4 The Finnish Twin Cohort (FinnTwin12)

FinnTwin12 is the youngest cohort of the Finnish Twin Cohort Study, a population-based study of Finnish twins born 1983–1987 identified through Finland's Central Population Registry. A total of

2,705 families (87% of all identified) returned the initial family questionnaire late in the year in which twins reached age 11 [22]. Twins were invited to participate in follow-up surveys when they were ages 14, 17, and approximately 22 (during young adulthood). An intensively studies sample was selected as 1035 families, among whom 1854 twins were interviewed at age 14. The interviewed twins were invited as young adults to complete the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) [20] interview (n = 1,347) and provide DNA samples [23]. Genotyping was conducted using the Human670-QuadCustom Illumina BeadChip at the Wellcome Trust Sanger Institute. Quality control steps included removing SNPs with minor allele frequency (MAF) <1%, genotyping success rate <95%, or Hardy–Weinberg equilibrium $p < 1 \times 10^{-6}$, and removing individuals with genotyping success rate <95%, a mismatch between phenotypic and genotypic gender, excess relatedness (outside of known families), and heterozygosity outliers. Genotypes were imputed to the Haplotype Reference Consortium (HRC) reference panel [12]. The current analysis uses data from the intensive sub sample with available DNA and diagnostic data across each wave of data collection. Our final analytic sample consisted of 1,193 individuals of European ancestries.

3 Clinical/environmental risk index measures

The environmental/clinical risk index was based on a previously validated index of risk factors for persistent SUD[7], including low childhood socioeconomic status (SES), family history of SUD, early initiation of substance use, childhood internalizing problems, childhood externalizing problems, frequent drinking in adolescence, frequent smoking in adolescence, frequent cannabis use in adolescence, along with other known risk factors, such as peer substance use [24], and exposure to trauma/traumatic experiences [25]. We dichotomized each risk factor (present vs not present) and summed them into an index for each person ranging from 0 to 10, providing a single measure of aggregate risk. In order to ensure that constructs were comparable across each of the four samples, we compared and harmonized the available measures. Below, we present the exact measurement for each of the ten items in each sample. Supplemental Figure 2 depicts the breakdown of each risk factor across each of the cohorts. Supplemental Figure 3 presents the tetrachoric correlations between each of the risk factors, by cohort and pooled into one sample. While there is variation in the strength of the correlations, overwhelmingly we see that many of these risk factors are weakly-to-modestly, positively correlated with one another. The strongest correlations (~.7) are between frequent tobacco and cannabis use in adolescence. Even this relatively strong correlation suggests that, at most, ~50% of the variance is shared between any given item in the risk index. The lack of consistent, strong correlations indicate that these items are not mere proxies for one another.

3.1 Low childhood socioeconomic status (SES)

3.1.1 Add Health

Participants were classified as experiencing low SES in childhood if they met criteria for any of the below items:

(i) Parental education: both residential parents reported having less than a high school.

(ii) Parental occupation: both residential parents reported occupations that were manual/low wage/low skill.

(iii) Household poverty: respondents report household income at or below the 1994 Federal Poverty threshold (Poverty Status: 1 person/Per extra person/4 person HH example = 7360/2480/14800).

(iv) Receipt of public assistance: respondent or parents report receipt of public assistance.

3.1.2 ALSPAC

Participants were classified as experiencing low SES in childhood if they met criteria for any of the below items:

(i) Parental education: mother and partner (if present) report no educational qualifications.

(ii) Household poverty: mother reported weekly income less than 100 pounds a week at ages 2.5, 4, or 7.

3.1.3 COGA

Participants were classified as experiencing low SES in childhood if their parent(s) reported having less than a high school.

3.1.4 FinnTwin12

Participants were classified as experiencing low SES in childhood if they met criteria for any of the below items:

(i) Parental education: parent(s) reported having less than a basic level education (minimum in Finland).

(ii) Parental occupation: both parents reported occupations that were manual/low wage/low skill.

3.2 Family history of substance use disorders (SUD)

3.2.1 Add Health

Respondents were classified as having a family history of SUD if parents reported yes to either of the following questions:

(i) “Does {NAME]'s biological mother currently have the following health problem (check all that apply): Alcoholism”

(ii) “Does {NAME]'s biological father currently have the following health problem (check all that apply): Alcoholism”

3.2.2 ALSPAC

Respondents were classified as having a family history of SUD if parents met criteria for any of the below items:

(i) Mother/Father - AUDIT total score greater than a threshold of 8.

(ii) Mother/Father - Self-reported having alcoholism or a drug addiction.

3.2.3 COGA

Respondents were classified as having a family history of SUD if parents met criteria for an alcohol use disorder based on parent SSAGA interviews. In instances where direct parent SSAGA interview is not available, collateral parental alcohol use disorder information collected as part of family history reports was used[21, 26].

3.2.4 FinnTwin12

Respondents were classified as having a family history of SUD if parents met criteria for any substance use disorder based on parent SSAGA interviews.

3.3 Childhood behavior/externalizing problems

3.3.1 Add Health

Respondents were classified as having childhood behavior problems if their score on a list of antisocial behaviors was at or above the 90th percentile.

3.3.2 ALSPAC

Respondents were classified as having childhood behavior problems if participants met DSM-IV clinical diagnostic criteria for any oppositional-conduct disorder.

3.3.3 COGA

Respondents were classified as having childhood behavior problems if they met criteria for conduct disorder (CD) or oppositional defiant disorder (ODD) from the SSAGA/C-SSAGA interview.

3.3.4 FinnTwin12

Respondents were classified as having childhood behavior problems if they met criteria for conduct disorder (CD) or oppositional defiant disorder (ODD) from the age 14 SSAGA interview.

3.4 Childhood internalizing problems

3.4.1 Add Health

Respondents were classified as having childhood internalizing problems if their score on the Center for Epidemiological Study Depression Scale (CES-D) was above 16 before age 15 or they retrospectively reported a diagnosis of depression from before age 15 at Wave IV.

3.4.2 ALSPAC

Respondents were classified as having childhood internalizing problems based on the Short Mood and Feelings Questionnaire (SMFQ) scores and Strengths and Difficulties Questionnaire (SDQ) emotional symptoms scores.

3.4.3 COGA

Respondents were classified as having childhood internalizing problems if they reported an onset age below age 15 on the following item across the SSAGA/C-SSAGA interview:

> (i) "Think about the time in your life that stands out as the "worst" time in your life of feeling (MOOD ENDORSED ABOVE). I'm interested in periods that lasted at least two weeks."

3.4.4 FinnTwin12

Respondents were classified as having childhood internalizing problems if they met criteria for major depressive disorder (MDD) from the age 14 SSAGA interview.

3.5 Early substance use initiation

3.5.1 Add Health

Respondents were classified as having initiated substance use early if they reported an age below 15 for any of the following Wave I items, or reported use in the Wave II follow up and their age was below 15:

(i) "How old were you when you smoked a whole cigarette for the first time?"

(ii) "Think about the first time you had a drink of beer, wine, or liquor... How old were you then?"

(iii) "How old were you when you tried marijuana for the first time?"

3.5.2 ALSPAC

Respondents were classified as having initiated substance use early if they reported an age below 15 for any of the following items across the ages 12.5, 13.5, 15.5, 17.5, or 24 follow-ups:

(i) Age of respondent when first smoked a cigarette

(ii) Age when respondent had first whole alcoholic drink

(iii) Age of respondent when first tried cannabis

3.5.3 COGA

Respondents were classified as having initiated substance use early if they reported an age below 15 for any of the following items across from the SSAGA/C-SSAGA interviews:

(i) "How old were you the first time you had your very first whole drink?"

(ii) "How old were you the first time you smoked a full cigarette?"

(iii) "How old were you the first time you used marijuana?"

3.5.4 FinnTwin12

Respondents were classified as having initiated substance use early if they reported an age below 15 for any of the following items across the ages 12, 14, and 17.5 interviews:

(i) Age of respondent when first smoked a cigarette

(ii) Age when respondent had first whole alcoholic drink

(iii) Age of respondent when first tried cannabis.

3.6 Frequent adolescent alcohol use

3.6.1 Add Health

Respondents were classified as regular users if they reported drinking on most days (≥ 5 days a week) before age 18 (Waves I and II), using the following question:

(i) "During the past 12 months, on how many days did you drink alcohol?"

3.6.2 ALSPAC

Respondents were classified as regular users if they reported drinking on most days (≥ 5 days a week) before age 18, using the following question:

(i) "How often do you have a drink containing alcohol?"

3.6.3 COGA

Respondents were classified as regular users if they reported drinking on most days (≥ 5 days a week) before age 18, using any of the following questions:

(i) “On how many days did you drink any beverages containing alcohol during the last 12 months?” (from C-SSAGA interview)

(ii) If respondents reported an onset age before age 18 on the following SSAGA question: “Was there ever a time when you drank almost every day for a week or more?”

3.6.4 FinnTwin12

Respondents were classified as regular users if they reported drinking on most days (≥ 5 days a week) before age 18 (age 14 and 17 survey), using the following question:

(i) “How often do you drink any amount of alcohol?”

3.7 Frequent adolescent tobacco use

3.7.1 Add Health

Respondents were classified as regular users if they reported smoking daily before age 18 (Waves I and II), using the following question:

(i) “During the past 30 days, on how many days did you smoke cigarettes?”

3.7.2 ALSPAC

Respondents were classified as regular users if they reported smoking daily before age 18, using the following questions:

(i) "Please mark the box next to the statement which describes you the best: - I usually smoke one or more cigarettes every day” **"**

(i) "Do you smoke every day?"

3.7.3 COGA

Respondents were classified as regular users if they reported smoking daily before age 18, using the following question:

(i) “When were you smoking regularly, how many days per week did you usually smoke cigarettes?”

3.7.4 FinnTwin12

Respondents were classified as regular users if they reported smoking daily before age 18 (age 14 and 17 survey), using the following question:

(i) “Which of the following best describes your present smoking habits: I smoke at least once each day”

3.8 Frequent adolescent cannabis use

3.8.1 Add Health

Respondents were classified as regular users if they reported cannabis use on most days (≥ 5 days a week) before age 18 (Waves I and II), using the following question:

(i) “During the past 30 days, how many times did you use marijuana?”

3.8.2 ALSPAC

Respondents were classified as regular users if they reported cannabis use on most days (≥ 5 days a week) before age 18, using the following questions:

(i) Frequency respondent uses or takes cannabis (example response option "I sometimes use or take cannabis but less than once a week"),

(ii) "How many times per week? (over the last 6 months)"

3.8.3 COGA

Respondents were classified as regular users if they reported regular use before age 18, using the following question from SSAGA/C-SSAGA:

(i) “How old were you the (first/last) time you used marijuana almost every day for at least two weeks?

3.8.4 FinnTwin12

Respondents were classified as regular users if they reported cannabis use on most days (≥ 5 days a week) before age 18 from the cannabis section of the age 22 SSAGA (retrospective).

3.9 Adolescent peer substance use

3.9.1 Add Health

Respondents were classified as having substance using peers if they reported 3 or more of their best friends used substances from the following questions at Waves I and II”

(i) “Of your three best friends, how many smoke at least 1 cigarette a day?”

(ii) “Of your three best friends, how many drink alcohol at least once a month?”

(iii) “Of your three best friends, how many use marijuana at least once a month?”

3.9.2 ALSPAC

Respondents were classified as having substance using peers if they reported most or all of their friends' used substances from the following items:

(i) Number of friends that drank alcohol during the last year

(ii) Number of friends that smoked cigarettes during the last year

(iii) Number of friends that took illegal drugs during the last year

3.9.3 COGA

Respondents were classified as having substance using peers if they reported most of their friends' used substances from the following SSAGA/C-SSAGA questions (ages 12 – 17):

(i) C-SSAGA: "How many of your best friends smoke?"; "How many of your best friends use alcohol?"; "How many of your best friends use marijuana?"; and "How many of your best friends use other drugs (like cocaine, uppers, or any of the other drugs we've talked about)?"

(ii) SSAGA (retrospective reports): "When you were 12-17, how many of your best friends smoked?"; "how many of your best friends used alcohol?"; "how many of your best friends used marijuana?"; and "how many of your best friends used other drugs (like cocaine, uppers, or any of the other drugs we've talked about)?"

3.9.4 FinnTwin12

Respondents were classified as having substance using peers if they reported most of their friends' used substances from the following questions at ages 14 and 17:

(i) "Do any of your friends smoke?"

(ii) "Do any of your friends drink?"

(iii) "Have any of your acquaintances tried drugs?"

3.10 Exposure to stressful/traumatic events

3.10.1 Add Health

Respondents were classified as having been exposed to a stressful/traumatic event if they reported any of the following:

(i) Friend or family member committed suicide

(ii) Victim of a violent assault, sexual assault (females only), or other violent crime

(iii) Witness violence

(iv) Serious injury

(v) Experience intimate partner violence

(vi) Loss of a child

(vii) Loss of a parent

3.10.2 ALSPAC

Respondents were classified as having been exposed to a stressful/traumatic event if they reported any of the following:

(ii) ever been physically or sexually abused as a child

(iii) ever been bullied

(iv) ever had a serious illness, injury, or hospitalization

(i) ever experienced the death of a parent, sibling, close friend

3.10.3 COGA

Respondents were classified as having been exposed to a stressful/traumatic event if they reported any of the following:

(i) Ever been shot

(ii) Ever been stabbed

(iii) Ever been mugged or threatened with a weapon or experienced a break-in or robbery

(iv) Ever been raped or sexually assaulted by a relative

(v) Ever been raped or sexually assaulted by someone not related to you

(vi) Ever been in military combat

(vii) Ever wounded in combat

(viii) Ever been held captive, tortured, or kidnapped

(ix) Ever been in a natural disaster like a fire, flood, earthquake, tornado, mudslide, or hurricane

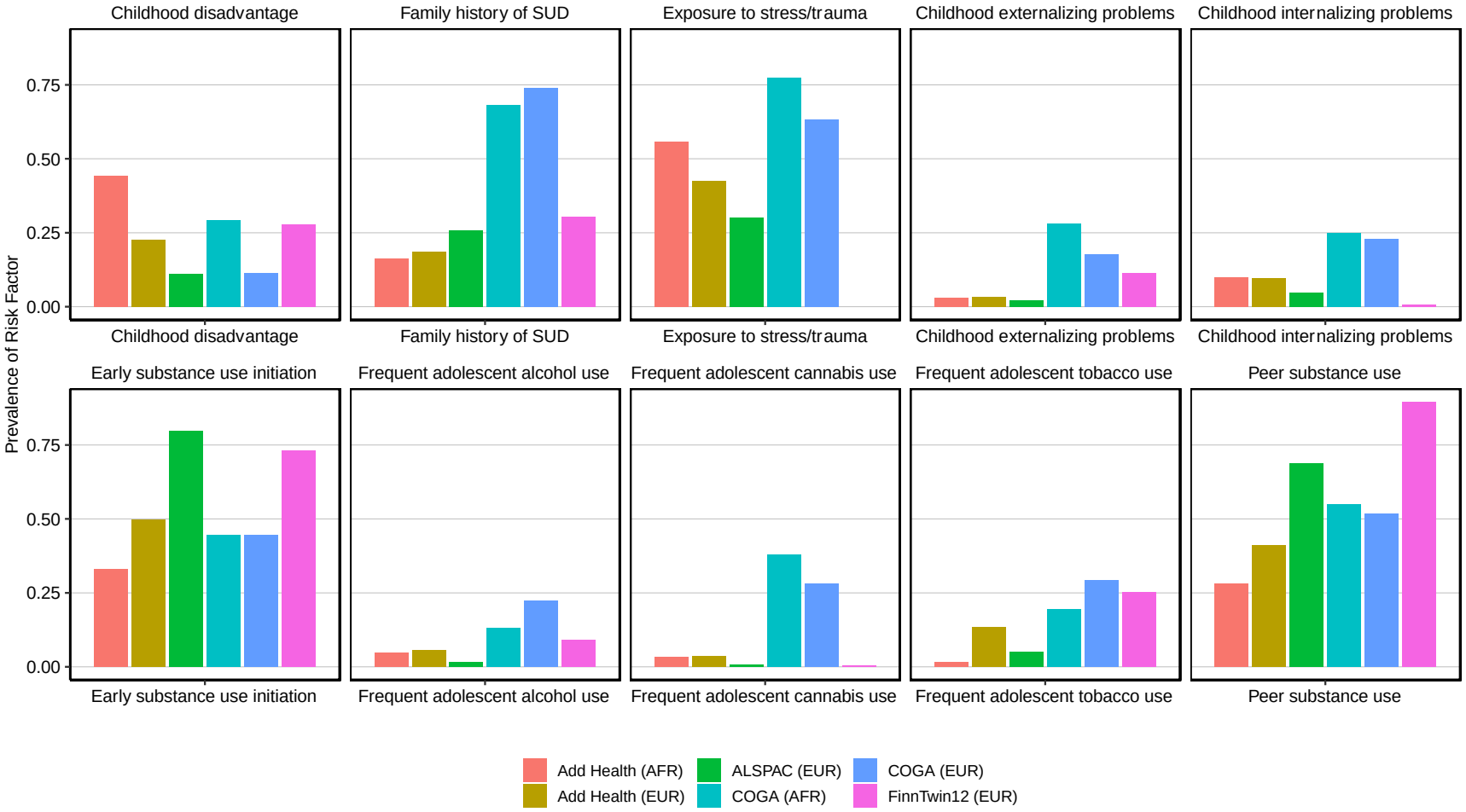
(x) Ever been in a serious accident

(xi) Ever seen someone being seriously injured or killed

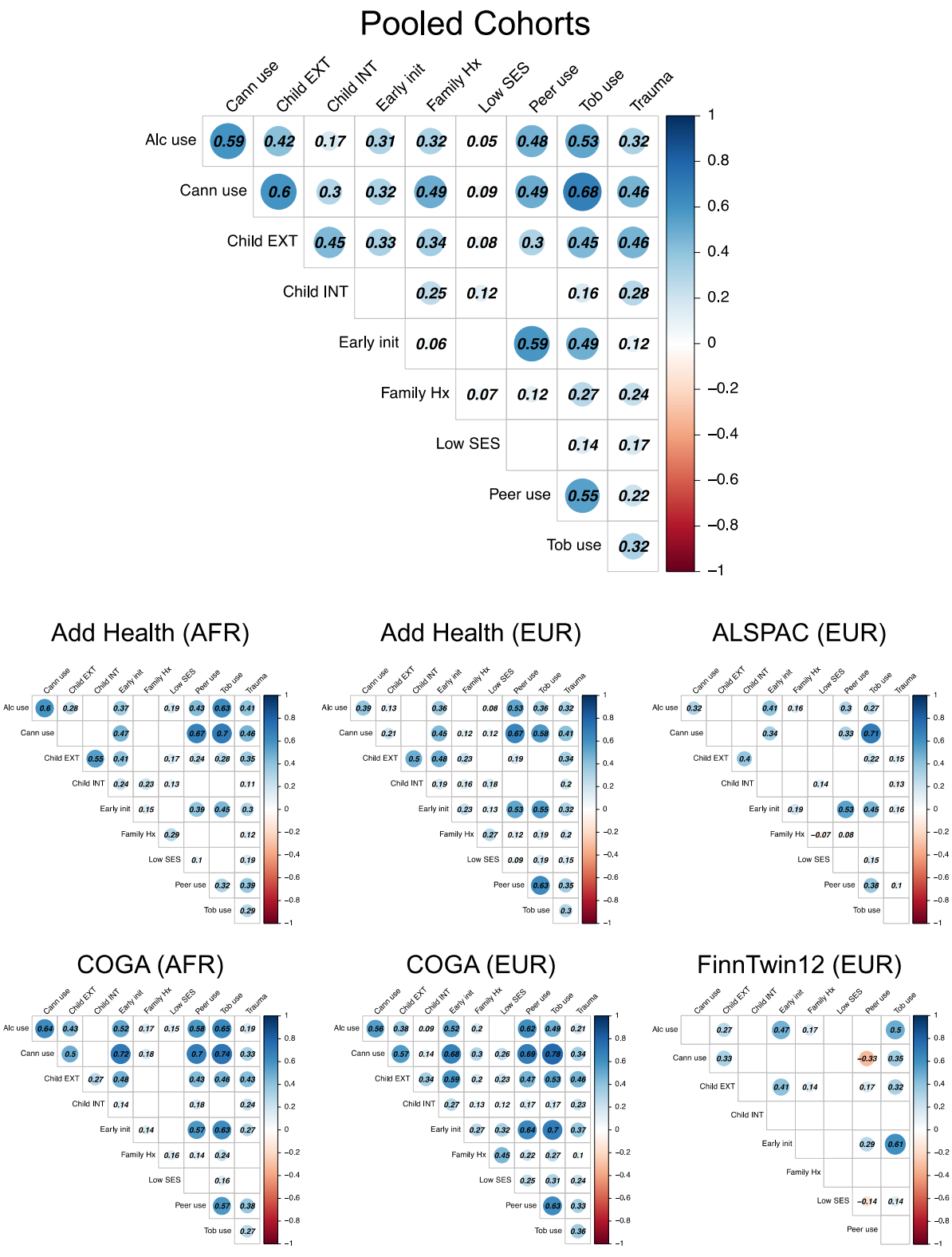
(xii) Ever unexpectedly discovered a dead body

3.10.4 FinnTwin12

FinnTwin12 did not contain measures related to stressful or traumatic events.



Supplemental Figure 1: Prevalence of Risk Factors by Cohort



Supplemental Figure 2: Tetrachoric Correlations Among Risk Index Items in Combined and Individual Cohorts

4 GWAS selection and inclusion

We used summary statistics from recent genome wide association studies (GWAS) to create polygenic scores (PGS) in the four holdout samples. We chose GWAS for inclusion based on the fact that: 1) SUD show strong genetic overlap with other externalizing [27–29], internalizing [30, 31], and psychotic disorders [32–34]; 2) both shared and substance specific genetic risk are associated with later SUDs [35–37]; 3) substance use and SUDs have only partial genetic overlap [38]; and 4) these samples had available results in both European and African ancestry cohorts.

4.1 GWAS of externalizing (EXT)

Summary statistics used for EXT in the European ancestry cohorts come from the recent multivariate GWAS of externalizing problems by the Externalizing Consortium[39]. The Externalizing Consortium analyses focused on a GWAS of a latent factor for externalizing derived from seven input GWAS theorized to be part of the externalizing spectrum, including ADHD [40], problematic alcohol use [41, 42], lifetime cannabis use [43], age of first sexual intercourse [44], number of sexual partners [44], general risk tolerance [44] and lifetime smoking initiation [45].These analyses converged onto a single factor. Polygenic scores for the latent externalizing factor were associated with externalizing factor scores in two holdout cohorts and with a variety of exploratory traits, including multiple substance use outcomes (both substance use and SUD).

For EXT in African ancestry cohorts, there is not an available multivariate GWAS that corresponds to the GWAS in European ancestries. Therefore, we performed a GWAS of an observed factor score in the COGA African ancestry cohort, derived from the same seven phenotypes used in the original Externalizing Consortium paper (and used for replication in the within family results in the European ancestry cohort). In order to ensure that there was no overlap between the discovery sample and COGA sample used in PGS analyses, we performed a ten-fold cross validation with leaving 10% of the sample out in every fold. GWAS from this analysis were used for PGS creation in the 10% not included in that run.

4.2 GWAS of depression (DEP)

Results for both the European and African ancestry GWAS come from a recent meta-analysis of large-scale depression GWAS using data from the Psychiatric Genomics Consortium (PGC), UK Biobank (UKB), Million Veterans Program (MVP), FinnGen, and 23andMe [31]. While the original meta-analysis includes all of these samples (N ~1.2 million), we restricted the current analysis to the PGC, UKB, and MVP cohorts only in European ancestries (N ~720K) as we did not have access to the 23andMe data, and we wanted to eliminate the possibility of sample overlap between FinnGen and the FinnTwin12 sample. GWAS for the African ancestry cohorts come exclusively from the African ancestry results for DEP in MVP (N = 59,600).

4.3 GWAS of problematic alcohol use (ALCP)

GWAS for problematic alcohol use (ALCP) in European ancestries is from a recent meta-analysis of GWAS for the PGC GWAS of alcohol dependence, the UKB GWAS of the problem subscale of the Alcohol Use Disorder Identification Test (AUDIT-P), and the MVP GWAS of alcohol use disorders (N ~ 430K)[32]. As Add Health, COGA, and FinnTwin12 were included in the original meta-analysis, we obtained GWAS results with each of those cohorts excluded for creating polygenic scores. Results for African ancestry come from the GWAS of AUD in MVP[46] (N ~ 56K).

4.4 GWAS of alcohol consumption (ALCC)

We used results from the GWAS and Sequencing Consortium for Alcohol and Nicotine's (GSCAN) meta-analysis of drinks per week for alcohol consumption (ALCC) in European ancestries[45]. These results included the publicly available GSCAN results as well was the 23andMe data (N ~900K). Both ALSPAC and FinnTwin12 were included in the original meta-analysis, and we obtained GWAS results with each of those cohorts excluded. Results for African ancestry come from the GWAS of the consumption subscale of the Alcohol Use Disorder Identification Test (AUDIT-C) in MVP[46] (N ~ 56K).

4.5 GWAS of schizophrenia (SCZ)

PGS for schizophrenia in the European ancestry cohorts were derived from the most recent iteration of the PGC's GWAS of SCZ (N ~130K) [47]. African ancestry results come from a meta-analysis of GWAS in the Genomic Psychiatry Cohort (GPC)[48] and Cooperative Studies Program (CSP) #572 [49].

4.6 GWAS of cigarettes per day/nicotine dependence (CPD)

For our smoking PGS in European ancestries, we used the publicly available GSCAN meta-analysis of cigarettes per day (CPD, N ~250K) [45]. These results again included ALSPAC and FinnTwin12, and we obtained GWAS results with each of those cohorts excluded. Results for PGS in African ancestries come from the most current GWAS of nicotine dependence [50] (N ~ 12K). While CPD and nicotine dependence are different phenotypes, the genetic correlation between the two is indistinguishable from one [50]. The GWAS of nicotine dependence included some COGA participants, and we obtained results with COGA excluded.

5 Polygenic Score Creation

5.1.1 Adjustment of GWAS effect sizes for linkage disequilibrium (LD)

We adjusted GWAS effect sizes for the non-independence of nearby SNPs in the genome (referred to as linkage disequilibrium, or LD) using PRS-CSx [51], which employs a Bayesian continuous shrinkage parameter to correct for LD. We used ancestry matched samples from 1KG as a reference panel for both European (EUR) and African (AFR) ancestries.

Rather than using each of the target samples for the training sample, we utilized the 1KG ancestry matched samples and restricted to the ~1.3 million SNPs in the high-quality consensus genotype set defined by the HapMap 3 Consortium [52, 53]. We generated polygenic scores using HapMap 3 SNPs that overlapped with the corresponding 1KG sample and UKB reference panel.

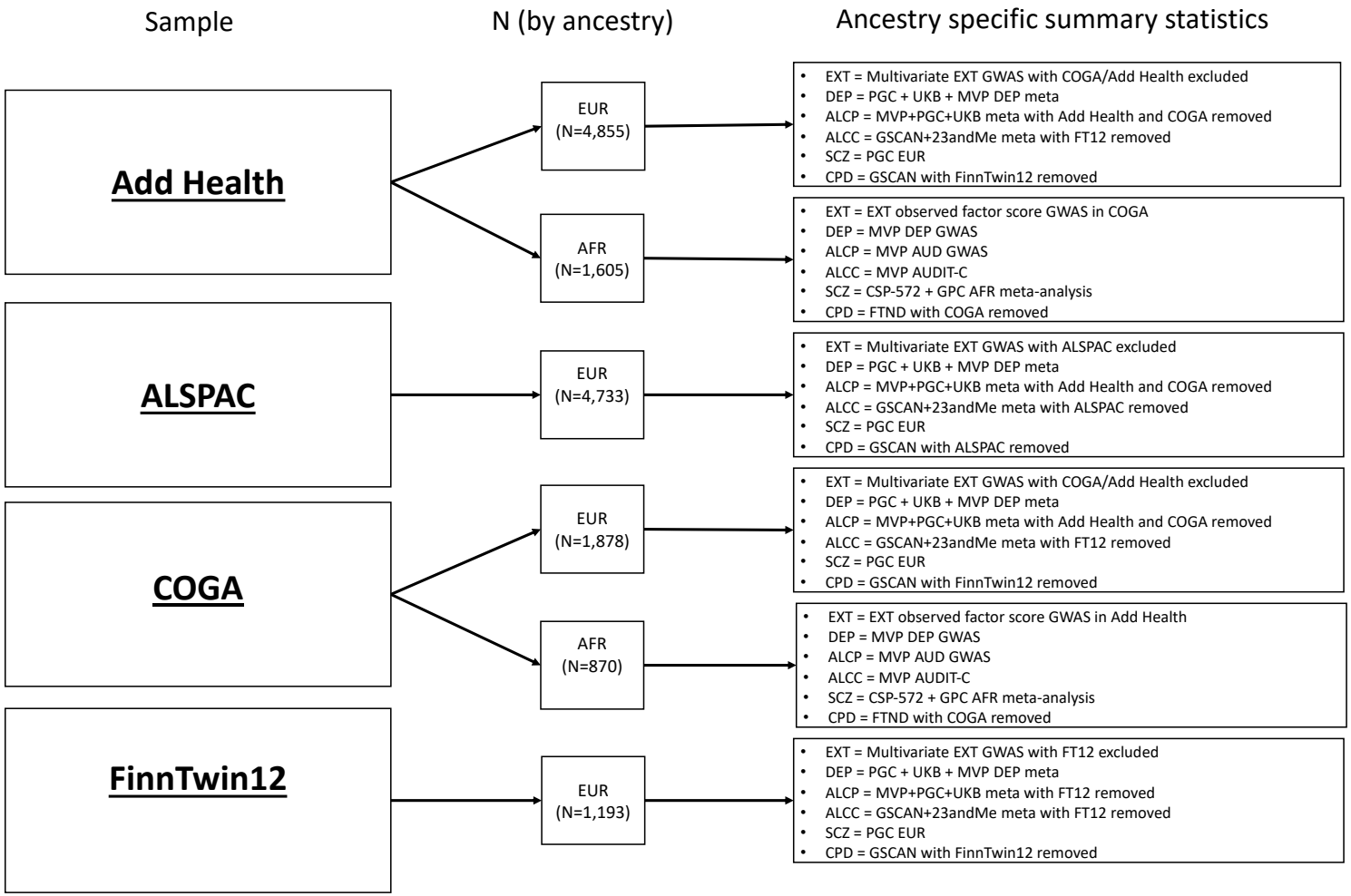
5.1.2 Polygenic scores

We computed polygenic scores from the weighted sum of the effect-coded alleles for a given individual *i*:

$$S_i = \sum_{j=1}^{M} \hat{\beta}_j g_{ij}$$

where S_i is the polygenic score, $\hat{\beta}_j$ is the estimated additive effect of the effect-coded allele at SNP j, and g_{ij} is the genotype at SNP j. The polygenic scores were standardized within each study cohort. Because PRS-CSx improves predictive power for non-European ancestry samples with smaller GWAS, we utilized the "meta" option for the AFR ancestries, creating scores that were derived from the meta-analyzed EUR and AFR specific weights. In the European ancestries, we derived scores from the EUR weights alone (not meta-analyzed). In each cohort, this provided us with one PGS per phenotype in each cohort to carry forward include in the models for the pooled analyses.

To account for population stratification, we regressed each PGS on age, age^2, sex, sex*age, sex* age^2, and the first 10 ancestral PC's. We then calculated the standardized residuals from these regression models for each of the six PGS (per cohort) and carried those forwards into the joint models that pooled the data from each cohort. Supplemental Figure 1, below, shows the GWAS matched for each PGS within each of the cohorts



Supplemental Figure 3: GWAS used for PGS creation in each cohort

6 Deviations from preregistration

We made several important deviations from the preregistration that are worth noting, in the interest of transparency. These changes were added to the analysis plan, posted on the open science framework, along with date and time stamps.

For each of the changes from the original plan, our motivations were driven by ways to either improve the analysis or address a problem we did not foresee in the original preregistration.

(a) **Amendment (04/22/21):** The polygenic scores for depression were expanded into a broader risk for internalizing after meta-analyzing with a GWAS of generalized anxiety disorder[54]. These GWAS showed relatively strong genetic overlap using bivariate LDSC[55] (rG ~.66).

(b) **Amendment (04/22/21):** We changed the PGS to those derived from PRS-CSx[51] (an extension of the original PRS-CS) as these allowed us to incorporate summary statistics from African ancestry GWAS and therefore create scores for the AFR subsamples in COGA and Add Health.

(c) **Amendment (04/22/21):** Due to issues with model convergence, we will use logistic regression in models with standard errors corrected for clustering at the family level[56].

(d) **Amendment (09/20/21):** Based on expert advice we will use an integrative data analysis approach[57] where we pool data and include cohort as a fixed effect. This approach is superior to meta-analysis because we have access to raw data.

(e) **Amendment (09/20/21):** We reverted to our original plan to use polygenic scores for depression as a new GWAS with AFR ancestry results became available[31]. We will also include a polygenic score for schizophrenia based on the overlap between psychotic disorders and SUD[47], and the availability of ancestry matched results[48].

(f) **Amendment (09/28/21):** We changed our focus on SUDs from including both abuse and dependence to dependence only. This change was driven by the fact that some of the samples (specifically Add Health) had a large number of people meeting criteria for alcohol abuse, and the sample prevalence for AUD was particularly high (over 40%). We therefore used the more restrictive measure of dependence for each of the substances to ensure we were not incorrectly categorizing people as having an SUD when they do not. We also omitted count of substances for which people meet criteria for the sake of space (these models were never run).

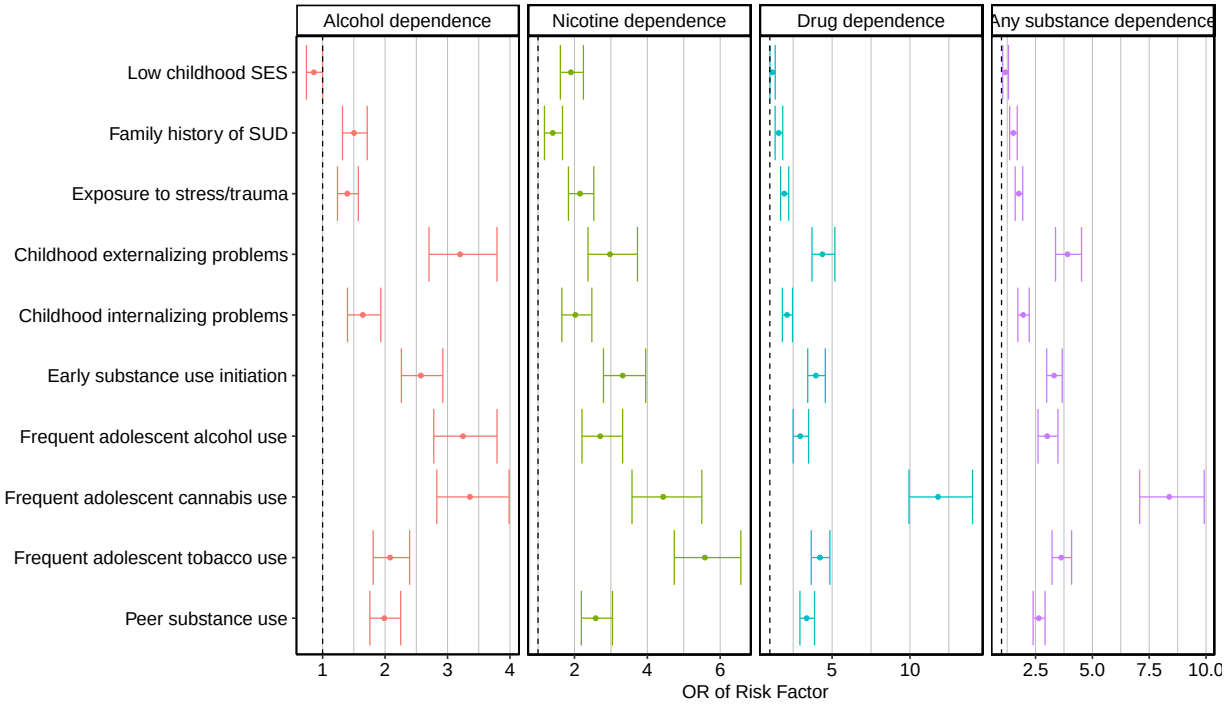
(g) **Amendment (05/12/22):** We made the following changes based on requests from reviewers:

(i) Added nicotine dependence as its own independent outcome to fully cover the range of SUD phenotypes.

(ii) Included PGS for nicotine dependence/cigarettes per day [45, 50] in addition to original PGS.

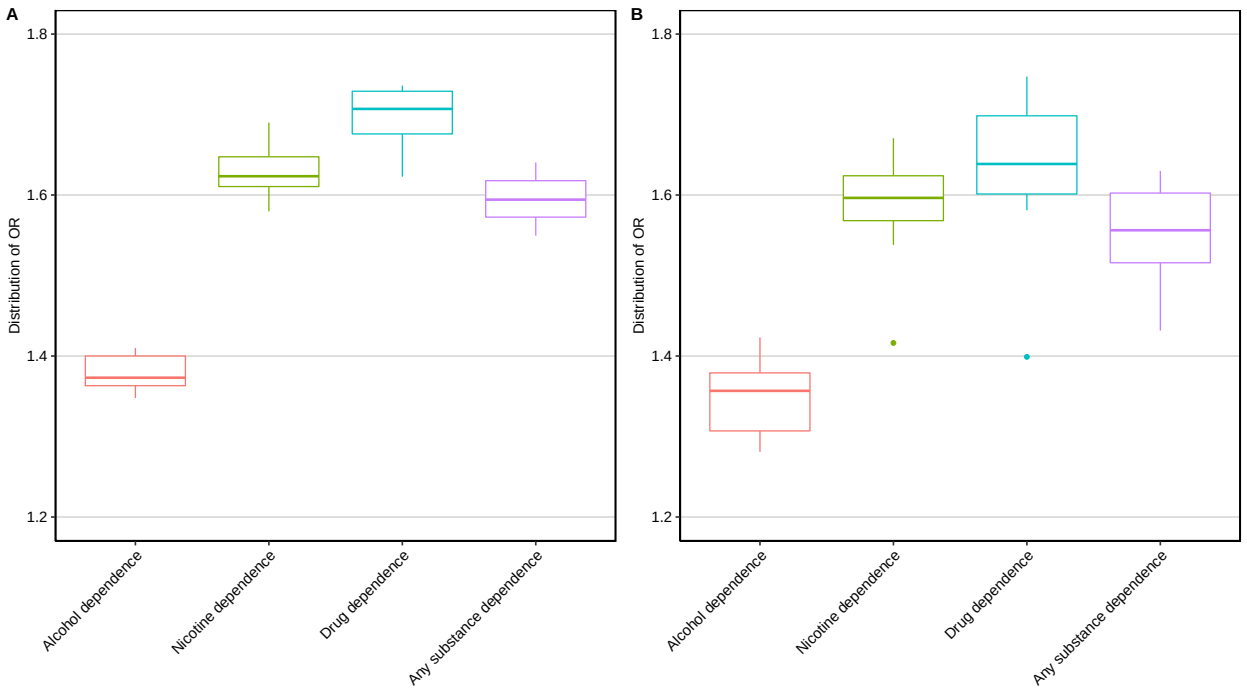
7 Variation in effect of clinical/environmental risk index (CERI)

To assess the relative impact of individual items, we ran a series of sensitivity analyses. The goal of these analyses was to ensure that the association between the CERI and each of the SUD phenotypes was not driven by any single item included in the CERI. We first estimated the association between individual risk factors and each of the SUD outcomes (Supplemental Figure 4). With the exception of the association between low childhood SES and alcohol dependence, each individual item is associated with increased odds of each of the SUD outcomes to varying degree. The one outlier for effect sizes of individual items was in regard to frequent adolescent cannabis use and both the drug dependence and any substance dependence outcomes.



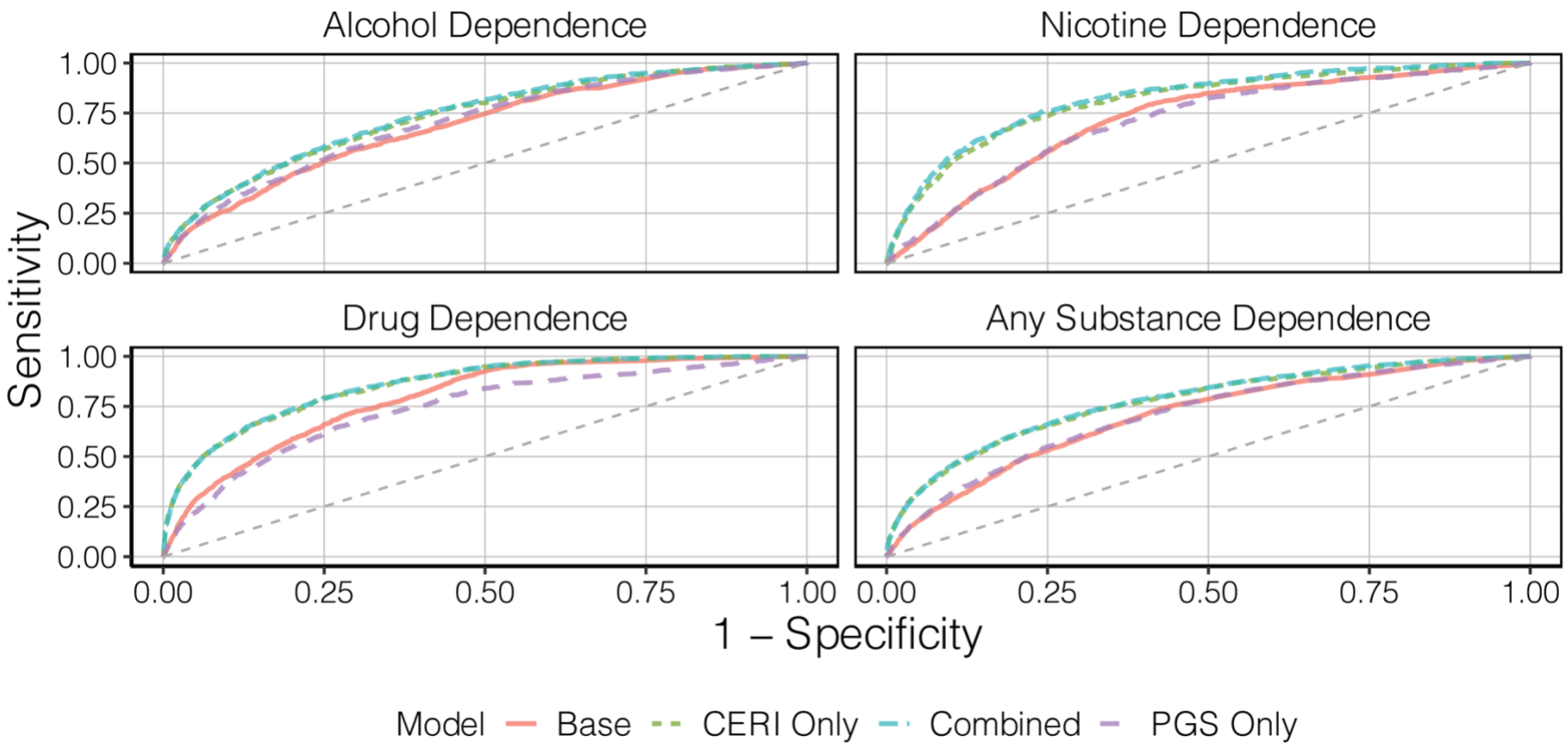
Supplemental Figure 4: Associations between individual risk factors and SUDs

In addition to testing the relative impact of each individual risk factor, we also evaluated the impact of removing one of the risk factors from the overall index to see relative change in the effect size. Supplemental Figure 5 presents the distribution of effect sizes for the CERI removing one of the risk factors for each of the SUD phenotypes. In each model, we also included sex, age, cohort, and all of the six PGSs (the Combined Risk Model). Panel A (Supplemental Figure 5) presents the distribution of the CERI effect sizes for each outcome. Overall, the effect sizes are relatively stable even when leaving one of the risk factors out. Panel B presents the same model, but with the removed risk factor included as a separate covariate. Again, the effect sizes are relatively stable, with two notable exceptions. The outlier for nicotine dependence is the effect size for the CERI when frequent adolescent tobacco use is included as a covariate. Similarly, the outlier for drug dependence is the effect size for the CERI when frequent adolescent cannabis use is included as a covariate. Even with these two outliers, to CERI is still significant and strongly associated with each SUD outcome.



Supplemental Figure 5: Effect Sizes for CERI with Individual Risk Factors Omitted

8 ROC Curves for CERI only and PGS only Models



Supplemental Figure 6: ROC Curves for Baseline (covariates only), CERI Only, PGS Only, and Combined Models

AUC Estimates for Baseline, PGS, CERI, and Combined Models

Phenotype	Model	AUC
Alcohol Dependence	Baseline (covariates only)	0.688
	CERI + covariates	0.732
	PGS + covariates	0.701
	Combined (PGS + CERI + covariates)	0.738
Nicotine Dependence	Baseline (covariates only)	0.721
	CERI + covariates	0.811
	PGS + covariates	0.763
	Combined (PGS + CERI + covariates)	0.824
Drug Dependence	Baseline (covariates only)	0.793
	CERI + covariates	0.857
	PGS + covariates	0.806
	Combined (PGS + CERI + covariates)	0.860
Any Substance Dependence	Baseline (covariates only)	0.702
	CERI + covariates	0.772
	PGS + covariates	0.720
	Combined (PGS + CERI + covariates)	0.777

9 Random-effects Integrative Data Analysis (RE IDA)

In order to ensure the robustness of our results to and between-sample heterogeneity, we ran a complementary set of analyses alongside our fixed-effects (FE) IDA approach. The random effects (RE) approach assumes that the samples in the analysis represent random draws from a larger population distribution ~$N(0, \sigma^2)$, as opposed to treating the effect of each cohort as known (and fixed, as in the fixed-effects approach). While there are more assumptions to the RE approach, the added advantages are that one can explicitly model between-study variation.

In our supplemental analyses, we tested for both random intercepts (for both study and family-unit) as well as testing for random slopes for each of the main predictors included in our analyses: the six polygenic scores (PGS) and the clinical/environmental risk index (CERI). In deciding the random-effects structure, we tested a series of nested models, adding random slopes and comparing the change in model fit using a χ^2 difference test ($\chi^2_{Full} - \chi^2_{Reduced}$). Once we identified the best fitting structure of the random effects, we estimated the models from the main analysis and compared the point estimates from the fixed effects and random effects models.

Supplemental Table 5 presents the tests for random slopes for the corresponding risk factors (6 PGS + CERI) with each of the SUD outcomes, compared to a baseline model which already includes a random intercept for cohort and family unit. We tested each random slope with each outcome, individually, as fitting all the random slopes at once was not possible. The model that included the random slope for the ALCC PGS showed improvement in overall fit above the baseline model in both alcohol dependence and any substance dependence. Likewise, for drug dependence, the model with a random slope for the EXT PGS showed significant improvement in fit. However, for each of the SUD outcomes, the biggest improvement in fit was gained by including a random slope for the CERI. We therefore included a random slope for the CERI, a random intercept for cohort, a random intercept for family unit, and a correlation between the random slope and the random intercept for cohort moving forward.

Supplemental Table 6 presents the parameter effects estimates from the models. Overwhelmingly, the parameter estimates from the random effects IDA approach, which explicitly models the between sample heterogeneity in the effect of the CERI, were consistent with the results from the main analysis (e.g., the fixed-effects IDA). Overall, these results support the findings from the main analyses and demonstrate that between-sample heterogeneity is not the reason for the associations between either the CERI or PGSs and each of SUD outcomes.

10 References

1. Degenhardt L, Charlson F, Ferrari A, Santomauro D, Erskine H, Mantilla-Herrara A, et al. The global burden of disease attributable to alcohol and drug use in 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Psychiatry. 2018;5:987–1012.
2. Verhulst B, Neale MC, Kendler KS. The heritability of alcohol use disorders: a meta-analysis of twin and adoption studies. Psychol Med. 2015;45:1061–1072.
3. Verweij KJH, Zietsch BP, Lynskey MT, Medland SE, Neale MC, Martin NG, et al. Genetic and environmental influences on cannabis use initiation and problematic use: A meta-analysis of twin studies. Addiction. 2010;105:417–430.
4. Burt SA. Are there meaningful etiological differences within antisocial behavior? Results of a meta-analysis. Clin Psychol Rev. 2009;29:163–178.
5. Rhee SH, Waldman ID. Genetic and environmental influences on antisocial behavior: A meta-analysis of twin and adoption studies. Psychol Bull. 2002;128:490–529.
6. Kendler KS, Jacobson KC, Prescott CA, Neale MC. Specificity of Genetic and Environmental Risk Factors for Use and Abuse/Dependence of Cannabis, Cocaine, Hallucinogens, Sedatives, Stimulants, and Opiates in Male Twins. American Journal of Psychiatry. 2003;160:687–695.
7. Meier MH, Hall W, Caspi A, Belsky DW, Cerda M, Harrington HL, et al. Which adolescents develop persistent substance dependence in adulthood? Using population-representative longitudinal data to inform universal risk assessment. Psychol Med. 2016;46:877–889.
8. Barr PB, Ksinan A, Su J, Johnson EC, Meyers JL, Wetherill L, et al. Using polygenic scores for identifying individuals at increased risk of substance use disorders in clinical and population samples. Transl Psychiatry. 2020;10:196.
9. Gu F, Chen TH, Pfeiffer RM, Fargnoli MC, Calista D, Ghiorzo P, et al. Combining common genetic variants and non-genetic risk factors to predict risk of cutaneous melanoma. Hum Mol Genet. 2018. 2018. https://doi.org/10.1093/hmg/ddy282.
10. O'Sullivan JW, Shcherbina A, Justesen JM, Turakhia M, Perez M, Wand H, et al. Combining Clinical and Polygenic Risk Improves Stroke Prediction among Individuals with Atrial Fibrillation. Circ Genom Precis Med. 2021;14:339–347.
11. Harris KM, Halpern CT, Haberstick BC, Smolen A. The National Longitudinal Study of Adolescent Health (Add Health) sibling pairs data. Twin Research and Human Genetics. 2013;16:391–398.
12. The Haplotype Reference Consortium, McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;48:1279–1283.
13. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, et al. A global reference for human genetic variation. Nature. 2015;526:68–74.
14. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al. Cohort profile: The 'Children of the 90s'-The index offspring of the avon longitudinal study of parents and children. Int J Epidemiol. 2013. 2013. https://doi.org/10.1093/ije/dys064.
15. Fraser A, Macdonald-wallis C, Tilling K, Boyd A, Golding J, Davey smith G, et al. Cohort profile: The avon longitudinal study of parents and children: ALSPAC mothers cohort. Int J Epidemiol. 2013;42:97–110.
16. Northstone K, Lewcock M, Groom A, Boyd A, Macleod J, Timpson N, et al. The Avon Longitudinal Study of Parents and Children (ALSPAC): an update on the enrolled sample of index children in 2019 [version 1; peer review: 2 approved]. Wellcome Open Res. 2019;4.

17. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42:377–381.
18. Taylor AE, Jones HJ, Sallis H, Smith GD, Lawlor DA, Davies NM, et al. Exploring the association of genetic factors with participation in the Avon Longitudinal Study of Parents and Children. Int J Epidemiol. 2018;47:dyy060.
19. Edenberg HJ. The collaborative study on the genetics of alcoholism: An update. Alcohol Research and Health. 2002;26:214–218.
20. Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI, 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;55:149–158.
21. Bucholz KK, McCutcheon V V., 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;41:359–368.
22. Rose RJRJ, Salvatore JEJE, Aaltonen S, Barr PBPB, Bogl LHLH, Byers HAHA, et al. FinnTwin12 Cohort: An Updated Review. vol. 22. 2019.
23. Kaprio J. The Finnish Twin Cohort Study: an update. Twin Res Hum Genet. 2013;16:157–162.
24. Sher KJ, Grekin ER, Williams NA. The development of alcohol use disorders. Annu Rev Clin Psychol. 2005;1:493–523.
25. Hughes K, Bellis MA, Hardcastle KA, Sethi D, Butchart A, Mikton C, et al. The effect of multiple adverse childhood experiences on health: a systematic review and meta-analysis. Lancet Public Health. 2017. 2017. https://doi.org/10.1016/S2468-2667(17)30118-4.
26. McCutcheon V V., Schuckit MA, Kramer JR, Chan G, Edenberg HJ, Smith TL, et al. Familial association of abstinent remission from alcohol use disorder in first-degree relatives of alcohol-dependent treatment-seeking probands. Addiction. 2017;112:1909–1917.
27. Barr PB, Dick DM. The Genetics of Externalizing Problems. Curr Top Behav Neurosci. 2020;47:93–112.
28. Krueger RF, Hicks BM, Patrick CJ, Carlson SR, Iacono WGWG, McGue M. Etiological connections among substance dependence, antisocial behavior and personality: Modeling the externalizing spectrum. J Abnorm Psychol. 2002;111:411–424.
29. Kendler KS, Myers J. The boundaries of the internalizing and externalizing genetic spectra in men and women. Psychol Med. 2014;44:647–655.
30. Polimanti R, Peterson RE, Ong JS, MacGregor S, Edwards AC, Clarke TK, et al. Evidence of causal effect of major depression on alcohol dependence: Findings from the psychiatric genomics consortium. Psychol Med. 2019. 2019. https://doi.org/10.1017/S0033291719000667.
31. Levey DF, Stein MB, Wendt FR, Pathak GA, Zhou H, Aslan M, et al. Bi-ancestral depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million individuals highlight new therapeutic directions. Nat Neurosci. 2021. 27 May 2021. https://doi.org/10.1038/s41593-021-00860-2.
32. Zhou H, Sealock JM, Sanchez-Roige S, Clarke TK, Levey DF, Cheng Z, et al. Genome-wide meta-analysis of problematic alcohol use in 435,563 individuals yields insights into biology and relationships with other traits. Nat Neurosci. 2020. 2020. https://doi.org/10.1038/s41593-020-0643-5.
33. Johnson EC, Demontis D, Thorgeirsson TE, Walters RK, Polimanti R, Hatoum AS, et al. A large-scale genome-wide association study meta-analysis of cannabis use disorder. Lancet Psychiatry. 2020. 2020. https://doi.org/10.1016/S2215-0366(20)30339-4.
34. Zhou H, Rentsch CT, Cheng Z, Kember RL, Nunez YZ, Sherva RM, et al. Association of OPRM1 Functional Coding Variant With Opioid Use Disorder: A Genome-Wide Association

Study. JAMA Psychiatry. 2020. June 2020. https://doi.org/10.1001/jamapsychiatry.2020.1206.
35. Kendler KS, Gardner C, Dick DM. Predicting alcohol consumption in adolescence from alcohol- specific and general externalizing genetic risk factors, key environmental exposures and their interaction. Psychol Med. 2011;41:1507–1516.
36. Meyers JL, Salvatore JE, Vuoksimaa E, Korhonen T, Pulkkinen L, Rose RJ, et al. Genetic Influences on Alcohol Use Behaviors Have Diverging Developmental Trajectories : A Prospective Study Among Male and Female Twins. Alcohol Clin Exp Res. 2014;38:2869–2877.
37. Barr PB, Mallard TT, Sanchez-Roige S, Poore HE, Linnér RK, Collaborators C, et al. Parsing Genetically Influenced Risk Pathways: Genetic Loci Impact Problematic Alcohol Use Via Externalizing and Specific Risk. MedRxiv. 2021:2021.07.20.21260861.
38. Sanchez-Roige S, Palmer AA, Clarke TK. Recent Efforts to Dissect the Genetic Basis of Alcohol Use and Abuse. Biol Psychiatry. 2020.
39. Karlsson Linnér R, Mallard TT, Barr PB, Sanchez-Roige S, Madole JW, Driver MN, et al. Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction. Nat Neurosci. 2021:1–10.
40. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet. 2019;51:63–75.
41. Walters RK, Polimanti R, Johnson EOECEO, McClintick JN, Adams MJ, Adkins AE, et al. Trans-ancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. Nat Neurosci. 2018;21:1656–1669.
42. Sanchez-Roige S, Palmer AA, Fontanillas P, Elson SL, Adams MJ, Howard DM, et al. Genome-wide association study meta-analysis of the alcohol use disorders identification test (AUDIT) in two population-based cohorts. American Journal of Psychiatry. 2019;176:107–118.
43. Pasman JA, Verweij KJH, Gerring Z, Stringer S, Sanchez-Roige S, Treur JL, et al. GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. Nat Neurosci. 2018;21:1161–1170.
44. Karlsson Linnér R, Biroli P, Kong E, Meddens SFW, Wedow R, Fontana MA, et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. Nat Genet. 2019;51:245–257.
45. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat Genet. 2019;51:237–244.
46. Kranzler HR, Zhou H, Kember RL, Vickers Smith R, Justice AC, Damrauer S, et al. Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. Nat Commun. 2019;10:1499.
47. Trubetskoy V, Pardiñas AF, Qi T, Panagiotaropoulou G, Awasthi S, Bigdeli TB, et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature 2022. 2022:1–13.
48. Bigdeli TB, Genovese G, Georgakopoulos P, Meyers JL, Peterson RE, Iyegbe CO, et al. Contributions of common genetic variants to risk of schizophrenia among individuals of African and Latino ancestry. Mol Psychiatry. 2019. 2019. https://doi.org/10.1038/s41380-019-0517-y.
49. Bigdeli TB, Fanous AH, Li Y, Rajeevan N, Sayward F, Genovese G, et al. Genome-Wide Association Studies of Schizophrenia and Bipolar Disorder in a Diverse Cohort of US Veterans. Schizophr Bull. 2020. 2020. https://doi.org/10.1093/schbul/sbaa133.

50. Quach BC, Bray MJ, Gaddis NC, Liu M, Palviainen T, Minica CC, et al. Expanding the genetic architecture of nicotine dependence and its shared genetics with multiple traits. Nat Commun. 2020;11.
51. Ruan Y, Lin Y-F, Feng Y-CA, Chen C-Y, Lam M, Guo Z, et al. Improving Polygenic Prediction in Ancestrally Diverse Populations. Nat Genet. 2022. 2022. https://doi.org/10.1038/s41588-022-01054-7.
52. Altshuler DM, Gibbs RA, Peltonen L, Schaffner SF, Yu F, Dermitzakis E, et al. Integrating common and rare genetic variation in diverse human populations. Nature. 2010;467:52–58.
53. Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat Commun. 2019;10:1776.
54. Levey DF, Gelernter J, Polimanti R, Zhou H, Cheng Z, Aslan M, et al. Reproducible Genetic Risk Loci for Anxiety: Results From ∼200,000 Participants in the Million Veteran Program. Am J Psychiatry. 2020. 2020. https://doi.org/10.1176/appi.ajp.2019.19030256.
55. Bulik-Sullivan BK, Loh P-RR, Finucane HK, Ripke S, Yang J, Psychiatric Genomics Consortium SWG, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47:291–295.
56. Zeileis A, Köll S, Graham N. Various versatile variances: An object-oriented implementation of clustered covariances in r. J Stat Softw. 2020;95:1–36.
57. Curran PJ, Hussong AM. Integrative Data Analysis: The Simultaneous Analysis of Multiple Data Sets. Psychol Methods. 2009;14:81–100.

Supplementary Table 1: Leave-one-out analyses for Combined Risk Models

		ALCOHOL DEPENDENCE					NICOTINE DEPENDENCE					DRUG DEPENDENCE					ANY DEPENDENCE				
		OR	Beta	SE	P		OR	Beta	SE	P		OR	Beta	SE	P		OR	Beta	SE	P	
	CERI	1.35	0.298	0.017	6.05E-72	*	1.60	0.467	0.021	2.26E-105	*	1.67	0.514	0.019	1.85E-167	*	1.56	0.448	0.015	5.90E-206	*
	ALCC PGS	1.04	0.038	0.031	2.28E-01		0.94	-0.066	0.044	1.32E-01		1.03	0.032	0.039	4.23E-01		0.99	-0.014	0.026	6.02E-01	
	ALCP PGS	1.13	0.120	0.033	2.28E-04	*	1.00	0.001	0.046	9.88E-01		1.05	0.046	0.040	2.55E-01		1.10	0.091	0.028	9.33E-04	*
No Add Health AFR	EXT PGS	1.08	0.078	0.033	1.89E-02		1.37	0.315	0.047	1.80E-11	*	1.13	0.125	0.040	1.56E-03	*	1.20	0.184	0.028	2.63E-11	*
	DEP PGS	0.98	-0.020	0.032	5.16E-01		1.02	0.019	0.044	6.69E-01		1.02	0.021	0.038	5.91E-01		0.98	-0.021	0.027	4.31E-01	
	CPD PGS	0.98	-0.022	0.031	4.70E-01		1.33	0.287	0.041	2.80E-12	*		-0.011	0.035	7.50E-01			0.070	0.025	5.43E-03	*
	SCZ PGS	1.03	0.032	0.033	3.40E-01		0.95	-0.051	0.046	2.69E-01		1.01	0.015	0.040	7.17E-01		1.01	0.006	0.027	8.33E-01	
	CERI	1.49	0.396	0.022	6.56E-71	*	1.71	0.534	0.031	7.58E-68	*	1.84	0.609	0.025	1.32E-134	*	1.68	0.519	0.019	1.04E-170	*
	ALCC PGS	0.95	-0.054	0.040	1.75E-01		1.01	0.010	0.063	8.73E-01		1.00	0.000	0.050	9.93E-01		0.95	-0.051	0.032	1.16E-01	
	ALCP PGS	1.16	0.146	0.040	2.53E-04	*	1.03	0.034	0.064	5.99E-01		1.04	0.042	0.052	4.21E-01		1.13	0.125	0.034	2.44E-04	*
No Add Health EUR	EXT PGS	1.07	0.066	0.044	1.30E-01		1.32	0.279	0.066	2.65E-05	*	1.06	0.061	0.051	2.27E-01		1.14	0.129	0.035	1.88E-04	*
	DEP PGS	0.96	-0.044	0.041	2.88E-01		1.05	0.046	0.063	4.69E-01		0.99	-0.008	0.050	8.80E-01		0.97	-0.028	0.034	4.13E-01	
	CPD PGS	0.96	-0.037	0.041	3.70E-01		1.32	0.279	0.061	4.77E-06	*		-0.022	0.048	6.51E-01			0.053	0.033	1.12E-01	
	SCZ PGS	1.01	0.010	0.044	8.17E-01		0.96	-0.041	0.068	5.49E-01		1.00	-0.004	0.052	9.31E-01		0.99	-0.014	0.035	6.82E-01	
	CERI	1.35	0.297	0.017	1.46E-71	*	1.60	0.471	0.021	9.55E-114	*	1.64	0.498	0.018	9.25E-175	*	1.56	0.443	0.014	1.06E-212	*
	ALCC PGS	1.04	0.036	0.032	2.57E-01		0.95	-0.054	0.043	2.09E-01		1.04	0.035	0.037	3.45E-01		0.99	-0.010	0.026	6.86E-01	
	ALCP PGS	1.13	0.124	0.033	1.64E-04	*	0.98	-0.017	0.045	6.99E-01		1.06	0.059	0.038	1.20E-01		1.10	0.092	0.027	6.18E-04	*
No FinnTwin12	EXT PGS	1.07	0.068	0.034	4.21E-02		1.31	0.272	0.045	1.53E-09	*	1.10	0.095	0.037	1.03E-02	*	1.17	0.157	0.027	4.29E-09	*
	DEP PGS	0.97	-0.026	0.031	3.99E-01		1.02	0.024	0.043	5.69E-01		1.03	0.034	0.036	3.46E-01		0.99	-0.009	0.026	7.22E-01	
	CPD PGS	0.97	-0.031	0.031	3.30E-01		1.33	0.284	0.040	1.53E-12	*		-0.028	0.034	4.01E-01			0.055	0.025	2.58E-02	
	SCZ PGS	1.03	0.029	0.033	3.94E-01		0.96	-0.044	0.045	3.27E-01		1.00	0.002	0.038	9.55E-01		0.99	-0.007	0.026	7.82E-01	
	CERI	1.36	0.309	0.017	3.15E-72	*	1.58	0.458	0.021	2.36E-108	*	1.64	0.497	0.018	4.53E-173	*	1.57	0.449	0.015	6.36E-204	*
	ALCC PGS	1.15	0.140	0.038	1.91E-04	*	0.95	-0.051	0.046	2.65E-01		1.05	0.051	0.038	1.82E-01		1.06	0.063	0.030	3.70E-02	
	ALCP PGS	1.01	0.010	0.037	7.85E-01		0.98	-0.020	0.048	6.76E-01		1.04	0.039	0.039	3.16E-01		1.01	0.015	0.031	6.29E-01	
No ALSPAC	EXT PGS	1.09	0.085	0.037	2.14E-02		1.30	0.260	0.047	3.01E-08	*	1.10	0.093	0.038	1.35E-02		1.16	0.146	0.029	4.46E-07	*
	DEP PGS	0.96	-0.040	0.034	2.50E-01		1.01	0.009	0.044	8.38E-01		1.02	0.022	0.036	5.49E-01		0.97	-0.032	0.028	2.54E-01	
	CPD PGS	0.98	-0.016	0.034	6.35E-01		1.28	0.245	0.041	2.89E-09	*		-0.035	0.034	3.11E-01			0.041	0.027	1.25E-01	
	SCZ PGS	1.02	0.021	0.037	5.75E-01		0.98	-0.020	0.047	6.63E-01		1.01	0.011	0.038	7.74E-01		1.00	0.003	0.029	9.12E-01	
	CERI	1.33	0.284	0.017	1.25E-61	*	1.57	0.448	0.021	5.39E-100	*	1.61	0.477	0.019	5.06E-136	*	1.52	0.418	0.015	3.64E-176	*
	ALCC PGS	1.02	0.024	0.031	4.48E-01		0.94	-0.059	0.043	1.66E-01		1.02	0.022	0.039	5.66E-01		0.98	-0.022	0.026	3.92E-01	
	ALCP PGS	1.14	0.133	0.033	4.57E-05	*	0.99	-0.007	0.045	8.73E-01		1.08	0.075	0.039	5.55E-02		1.10	0.099	0.027	2.22E-04	*
No COGA AFR	EXT PGS	1.07	0.064	0.034	5.94E-02		1.33	0.287	0.046	2.88E-10	*	1.16	0.147	0.040	2.22E-04	*	1.20	0.183	0.027	1.42E-11	*
	DEP PGS	0.97	-0.035	0.032	2.69E-01		1.02	0.015	0.043	7.22E-01		1.06	0.056	0.038	1.37E-01		0.99	-0.011	0.026	6.80E-01	
	CPD PGS		-0.004	0.030	8.84E-01		1.31	0.272	0.040	1.30E-11	*		-0.025	0.035	4.82E-01			0.066	0.025	7.84E-03	*
	SCZ PGS	1.05	0.045	0.033	1.77E-01		0.97	-0.032	0.045	4.73E-01		1.00	0.002	0.040	9.54E-01		1.01	0.006	0.027	8.18E-01	
	CERI	1.28	0.247	0.020	8.13E-36	*	1.51	0.414	0.024	4.11E-67	*	1.51	0.413	0.021	2.57E-88	*	1.47	0.386	0.016	8.22E-127	*
	ALCC PGS	1.02	0.025	0.034	4.65E-01		0.92	-0.083	0.045	6.45E-02		1.03	0.025	0.042	5.45E-01		0.97	-0.032	0.027	2.35E-01	
	ALCP PGS	1.13	0.119	0.035	7.18E-04	*	0.99	-0.006	0.048	9.00E-01		1.07	0.067	0.043	1.19E-01		1.10	0.095	0.028	7.43E-04	*
No COGA EUR	EXT PGS	1.10	0.093	0.036	9.29E-03	*	1.33	0.286	0.047	1.61E-09	*	1.11	0.105	0.042	1.22E-02	*	1.18	0.167	0.028	2.28E-09	*
	DEP PGS	0.96	-0.041	0.034	2.27E-01		0.99	-0.006	0.045	8.84E-01			0.011	0.040	7.76E-01			-0.019	0.027	4.81E-01	
	CPD PGS	1.00	-0.005	0.034	8.85E-01		1.31	0.271	0.042	1.47E-10	*	0.99	-0.014	0.038	7.03E-01		1.08	0.073	0.026	4.80E-03	*
	SCZ PGS	1.05	0.050	0.035	1.59E-01		0.96	-0.046	0.046	3.22E-01		1.01	0.006	0.041	8.88E-01		1.00	0.003	0.027	9.23E-01	

All models included age, sex, and cohort as covariates. PGS residualized on age, sex, and first 10 ancestral principal components

* $p < .05/4$; Beta = log(OR)

Supplementary Table 2: Sex-Stratified analyses for Combined Risk Models

		Males					Females				
		Beta	**OR**	**SE**	**P**		**Beta**	**OR**	**SE**	**P**	
Alcohol dependence	CERI	0.283	1.328	0.022	1.47E-37	*	0.32815156	1.388	0.02288387	1.23E-46	*
	ALCC PGS	0.036	1.037	0.042	3.91E-01		0.03292647	1.033	0.0444117	0.4584559	
	ALCP PGS	0.145	1.156	0.045	1.17E-03	*	0.07491772	1.078	0.04499584	0.09591433	
	EXT PGS	0.066	1.069	0.044	1.33E-01		0.08947406	1.094	0.0481803	0.06330159	
	DEP PGS	-0.003	0.997	0.041	9.36E-01		-0.0706639	0.932	0.04489902	0.11552414	
	CPD PGS	0.020	1.021	0.042	6.30E-01		-0.068182	0.934	0.04235254	0.10742665	
	SCZ PGS	0.016	1.016	0.043	7.15E-01		0.05242869	1.054	0.04832036	0.27791163	
Nicotine dependence	CERI	0.431	1.539	0.028	1.81E-53	*	0.49568449	1.642	0.03053261	2.87E-59	*
	ALCC PGS	-0.017	0.983	0.056	7.59E-01		-0.0910376	0.913	0.06298771	0.14836702	
	ALCP PGS	-0.074	0.929	0.059	2.10E-01		0.06768966	1.070	0.06776457	0.31784578	
	EXT PGS	0.197	1.218	0.061	1.28E-03	*	0.37140196	1.450	0.06328484	4.39E-09	*
	DEP PGS	0.038	1.039	0.055	4.85E-01		-0.0108971	0.989	0.06304038	0.86276192	
	CPD PGS	0.290	1.336	0.055	1.26E-07	*	0.25692835	1.293	0.0564349	5.30E-06	*
	SCZ PGS	-0.023	0.978	0.058	6.96E-01		-0.0527067	0.949	0.06660301	0.42873659	
Drug dependence	CERI	0.470	1.600	0.023	3.49E-91	*	0.52920269	1.698	0.02526292	1.96E-97	*
	ALCC PGS	0.036	1.037	0.048	4.52E-01		0.02331416	1.024	0.05553366	0.67461663	
	ALCP PGS	0.070	1.073	0.049	1.54E-01		0.03763326	1.038	0.05598582	0.501461	
	EXT PGS	0.105	1.110	0.049	3.30E-02		0.10887446	1.115	0.05494001	0.04751315	
	DEP PGS	0.031	1.031	0.046	5.04E-01		0.01845898	1.019	0.05362384	0.73067271	
	CPD PGS	-0.008	0.992	0.045	8.65E-01		-0.0441254	0.957	0.05047871	0.3820432	
	SCZ PGS	-0.024	0.976	0.049	6.18E-01		0.04476757	1.046	0.05752524	0.43643641	
Any substance dependence	CERI	0.407	1.502	0.019	2.94E-101	*	0.47521178	1.608	0.02003297	2.16E-124	*
	ALCC PGS	0.018	1.018	0.035	6.07E-01		-0.0420929	0.959	0.03545877	0.23519065	
	ALCP PGS	0.084	1.087	0.036	2.12E-02		0.09116808	1.095	0.03695126	0.01361545	
	EXT PGS	0.135	1.145	0.036	1.55E-04	*	0.19630639	1.217	0.03733965	1.46E-07	*
	DEP PGS	-0.011	0.989	0.034	7.39E-01		-0.028661	0.972	0.03634202	0.43031916	
	CPD PGS	0.075	1.078	0.033	2.50E-02		0.04300678	1.044	0.03444317	0.21180036	
	SCZ PGS	-0.014	0.986	0.035	6.89E-01		0.01710526	1.017	0.03806379	0.65315467	

All models included age, sex, and cohort as covariates. PGS residualized on age, sex, and first 10 ancestral principal components

* $p < .05/4$; Beta = log(OR)

Supplementary Table 3: Ancestry Stratified analyses for Combined Risk Models

		EUR Ancestry Results					AFR Ancestry Results				
		Beta	**OR**	**SE**	**P**		**Beta**	**OR**	**SE**	**P**	
Alcohol dependence	CERI	0.280	1.323	0.018	2.34E-57	*	0.453	1.573	0.048	2.80E-21	*
	ALCC PGS	0.027	1.027	0.032	3.97E-01		0.179	1.196	0.111	1.07E-01	
	ALCP PGS	0.142	1.153	0.034	2.37E-05	*	-0.124	0.883	0.095	1.89E-01	
	EXT PGS	0.063	1.065	0.035	6.96E-02		0.145	1.156	0.098	1.39E-01	
	DEP PGS	-0.020	0.980	0.033	5.31E-01		-0.145	0.865	0.085	8.66E-02	
	CPD PGS	-0.007	0.993	0.031	8.14E-01		-0.112	0.894	0.109	3.06E-01	
	SCZ PGS	0.046	1.047	0.034	1.82E-01		-0.083	0.920	0.091	3.61E-01	
Nicotine dependence	CERI	0.455	1.576	0.022	1.27E-93	*	0.475	1.609	0.055	4.45E-18	*
	ALCC PGS	-0.068	0.934	0.044	1.24E-01		0.028	1.029	0.133	8.32E-01	
	ALCP PGS	0.000	1.000	0.047	9.99E-01		-0.153	0.858	0.137	2.63E-01	
	EXT PGS	0.321	1.379	0.048	3.29E-11	*	-0.013	0.987	0.118	9.12E-01	
	DEP PGS	0.018	1.018	0.045	6.85E-01		-0.062	0.940	0.115	5.86E-01	
	CPD PGS	0.286	1.331	0.042	1.06E-11	*	0.164	1.179	0.106	1.22E-01	
	SCZ PGS	-0.045	0.956	0.047	3.46E-01		-0.050	0.951	0.118	6.72E-01	
Drug dependence	CERI	0.492	1.636	0.021	4.37E-127	*	0.496	1.642	0.035	1.50E-45	*
	ALCC PGS	0.021	1.021	0.042	6.24E-01		0.057	1.059	0.076	4.52E-01	
	ALCP PGS	0.064	1.066	0.042	1.29E-01		-0.024	0.977	0.084	7.76E-01	
	EXT PGS	0.176	1.192	0.043	4.78E-05	*	-0.114	0.892	0.072	1.13E-01	
	DEP PGS	0.053	1.054	0.041	1.99E-01		-0.095	0.909	0.070	1.75E-01	
	CPD PGS	-0.012	0.988	0.037	7.51E-01		-0.075	0.928	0.075	3.19E-01	
	SCZ PGS	0.010	1.010	0.044	8.24E-01		-0.045	0.956	0.071	5.27E-01	
Any substance dependence	CERI	0.424	1.529	0.016	7.85E-164	*	0.497	1.644	0.033	1.31E-52	*
	ALCC PGS	-0.024	0.976	0.027	3.70E-01		0.087	1.090	0.070	2.14E-01	
	ALCP PGS	0.102	1.108	0.028	2.90E-04	*	-0.064	0.938	0.071	3.73E-01	
	EXT PGS	0.207	1.229	0.029	1.04E-12	*	-0.071	0.931	0.062	2.49E-01	
	DEP PGS	-0.012	0.988	0.028	6.61E-01		-0.106	0.899	0.062	8.56E-02	
	CPD PGS	0.077	1.080	0.026	3.03E-03	*	-0.045	0.956	0.065	4.88E-01	
	SCZ PGS	0.012	1.012	0.029	6.64E-01		-0.106	0.899	0.061	8.32E-02	

All models included age, sex, and cohort as covariates. PGS residualized on age, sex, and first 10 ancestral principal components

* $p < .05/4$; Beta = log(OR)

Supplementary Table 4: Testing for Variation in Predictors Across Cohorts

	Beta (log OR)	SE	P	P < .05/4	Beta (log OR)	SE	P	P < .05/4	Beta (log OR)	SE	P	P < .05/4	Beta (log OR)	SE	P	P < .05/4
Intercept	-2.470	0.087	7.27E-177	*	-3.332	0.121	7.90E-167	*	-2.841	0.100	3.08E-179	*	-1.901	0.072	3.92E-153	*
Female	-0.419	0.057	2.63E-13	*	-0.321	0.078	3.55E-05	*	-0.565	0.067	2.68E-17	*	-0.436	0.046	5.45E-21	*
Age	0.426	0.038	1.75E-28	*	0.250	0.061	3.60E-05	*	0.154	0.041	1.50E-04	*	0.327	0.033	1.52E-23	*
Add Health AFR	-1.797	0.281	1.64E-10	*	-1.084	0.285	1.44E-04	*	-0.355	0.211	9.31E-02		-0.972	0.164	3.40E-09	*
FinnTwin12 EUR	-0.637	0.308	3.85E-02		-0.239	0.473	6.13E-01		-3.270	0.755	1.48E-05	*	-0.841	0.266	1.59E-03	*
ALSPAC EUR	-0.419	0.172	1.51E-02		-2.029	0.351	7.59E-09	*	-3.316	0.387	1.08E-17	*	-0.948	0.150	2.54E-10	*
COGA AFR	-1.948	0.416	2.88E-06	*	-3.523	0.716	8.75E-07	*	-0.688	0.276	1.26E-02		-1.511	0.271	2.38E-08	*
COGA EUR	-0.417	0.188	2.65E-02		-2.316	0.309	6.47E-14	*	-1.195	0.240	6.39E-07	*	-0.925	0.166	2.59E-08	*
ALCC PGS	0.184	0.050	2.63E-04	*	-0.102	0.057	7.53E-02		0.052	0.055	3.50E-01		0.046	0.040	2.49E-01	
ALCP PGS	0.025	0.053	6.38E-01		-0.024	0.060	6.90E-01		0.069	0.056	2.17E-01		0.023	0.042	5.82E-01	
EXT PGS	0.091	0.050	7.08E-02		0.300	0.060	4.85E-07	*	0.185	0.055	8.07E-04	*	0.217	0.040	5.24E-08	*
DEP PGS	-0.011	0.046	8.05E-01		-0.008	0.055	8.91E-01		0.057	0.051	2.66E-01		-0.005	0.038	9.04E-01	
SCZ PGS	0.067	0.048	1.65E-01		-0.039	0.057	4.97E-01		0.007	0.055	8.93E-01		0.016	0.039	6.90E-01	
CPD PGS	0.014	0.045	7.56E-01		0.266	0.051	2.13E-07	*	-0.019	0.047	6.87E-01		0.074	0.035	3.40E-02	
CERI	0.160	0.026	3.34E-10	*	0.404	0.028	1.47E-46	*	0.337	0.027	1.51E-35	*	0.336	0.021	6.57E-56	*
ALCC PGS: Add Health AFR	-0.189	0.173	2.75E-01		0.112	0.172	5.16E-01		-0.052	0.120	6.65E-01		-0.052	0.098	5.92E-01	
ALCC PGS: FinnTwin12 EUR	-0.167	0.140	2.31E-01		-0.058	0.228	8.01E-01		-0.300	0.269	2.64E-01		-0.095	0.122	4.35E-01	
ALCC PGS: ALSPAC EUR	-0.376	0.081	3.30E-06	*	0.163	0.139	2.41E-01		-0.172	0.167	3.04E-01		-0.185	0.067	5.99E-03	*
ALCC PGS: COGA AFR	0.176	0.164	2.82E-01		0.104	0.250	6.76E-01		0.073	0.124	5.54E-01		0.178	0.119	1.36E-01	
ALCC PGS: COGA EUR	-0.111	0.092	2.27E-01		0.179	0.128	1.61E-01		-0.035	0.098	7.23E-01		0.048	0.081	5.55E-01	
ALCP PGS: Add Health AFR	-0.074	0.153	6.30E-01		-0.234	0.197	2.35E-01		0.008	0.123	9.48E-01		-0.043	0.099	6.67E-01	
ALCP PGS: FinnTwin12 EUR	-0.072	0.131	5.80E-01		0.298	0.248	2.30E-01		-0.262	0.184	1.56E-01		-0.003	0.117	9.81E-01	
ALCP PGS: ALSPAC EUR	0.288	0.086	7.67E-04	*	0.133	0.148	3.71E-01		0.158	0.173	3.60E-01		0.223	0.071	1.67E-03	*
ALCP PGS: COGA AFR	-0.259	0.148	8.05E-02		0.001	0.174	9.97E-01		-0.195	0.140	1.64E-01		-0.174	0.127	1.70E-01	
ALCP PGS: COGA EUR	0.060	0.093	5.18E-01		0.009	0.121	9.43E-01		-0.055	0.097	5.69E-01		0.005	0.083	9.50E-01	
EXT PGS: Add Health AFR	-0.121	0.158	4.44E-01		-0.484	0.150	1.28E-03	*	-0.264	0.127	3.68E-02		-0.304	0.094	1.17E-03	*
EXT PGS: FinnTwin12 EUR	0.118	0.146	4.18E-01		0.298	0.240	2.14E-01		0.677	0.294	2.12E-02		0.108	0.126	3.95E-01	
EXT PGS: ALSPAC EUR	-0.176	0.086	4.14E-02		0.161	0.151	2.86E-01		0.110	0.182	5.46E-01		-0.072	0.073	3.27E-01	
EXT PGS: COGA AFR	0.192	0.140	1.70E-01		-0.056	0.226	8.03E-01		-0.330	0.113	3.43E-03	*	-0.275	0.104	8.06E-03	*
EXT PGS: COGA EUR	-0.084	0.096	3.80E-01		-0.031	0.135	8.21E-01		-0.083	0.098	3.96E-01		-0.080	0.081	3.28E-01	
DEP PGS: Add Health AFR	-0.291	0.120	1.51E-02		-0.112	0.161	4.86E-01		-0.032	0.111	7.73E-01		-0.060	0.088	4.93E-01	
DEP PGS: FinnTwin12 EUR	-0.125	0.148	3.96E-01		-0.193	0.210	3.59E-01		-0.478	0.225	3.37E-02		-0.181	0.125	1.49E-01	
DEP PGS: ALSPAC EUR	0.041	0.081	6.10E-01		0.083	0.135	5.37E-01		0.106	0.188	5.72E-01		0.062	0.068	3.62E-01	
DEP PGS: COGA AFR	-0.020	0.133	8.80E-01		0.009	0.196	9.65E-01		-0.249	0.117	3.37E-02		-0.154	0.110	1.64E-01	
DEP PGS: COGA EUR	0.005	0.088	9.58E-01		0.144	0.128	2.62E-01		0.009	0.096	9.29E-01		-0.016	0.078	8.34E-01	
SCZ PGS: Add Health AFR	-0.052	0.145	7.20E-01		0.078	0.151	6.06E-01		-0.089	0.114	4.32E-01		-0.098	0.088	2.65E-01	
SCZ PGS: FinnTwin12 EUR	0.018	0.138	8.95E-01		0.128	0.220	5.60E-01		0.223	0.269	4.07E-01		0.109	0.120	3.63E-01	
SCZ PGS: ALSPAC EUR	0.032	0.085	7.04E-01		-0.152	0.135	2.61E-01		-0.083	0.174	6.35E-01		-0.012	0.069	8.61E-01	
SCZ PGS: COGA AFR	-0.271	0.133	4.18E-02		-0.186	0.202	3.57E-01		-0.031	0.118	7.94E-01		-0.183	0.109	9.30E-02	
SCZ PGS:cohortCOGA_EUR	-0.150	0.100	1.34E-01		0.052	0.145	7.20E-01		0.006	0.107	9.58E-01		-0.050	0.088	5.71E-01	
CPD PGS: Add Health AFR	0.025	0.161	8.78E-01		-0.186	0.151	2.18E-01		-0.097	0.118	4.12E-01		-0.126	0.093	1.75E-01	
CPD PGS: FinnTwin12 EUR	0.102	0.118	3.86E-01		-0.216	0.199	2.78E-01		0.235	0.251	3.49E-01		0.054	0.099	5.89E-01	
CPD PGS: ALSPAC EUR	-0.025	0.079	7.48E-01		0.254	0.132	5.41E-02		0.297	0.148	4.46E-02		0.079	0.065	2.28E-01	
CPD PGS: COGA AFR	-0.252	0.168	1.33E-01		0.071	0.159	6.56E-01		0.010	0.115	9.29E-01		-0.082	0.106	4.38E-01	
CPD PGS: COGA EUR	-0.092	0.082	2.64E-01		0.008	0.119	9.46E-01		-0.042	0.087	6.33E-01		-0.084	0.076	2.64E-01	
CERI: Add Health AFR	0.209	0.089	1.86E-02		-0.083	0.087	3.43E-01		-0.029	0.067	6.66E-01		-0.007	0.055	8.99E-01	
CERI: FinnTwin12 EUR	0.258	0.088	3.56E-03	*	-0.423	0.154	6.02E-03	*	0.220	0.199	2.68E-01		0.028	0.079	7.27E-01	
CERI: ALSPAC EUR	0.094	0.056	9.48E-02		0.100	0.107	3.52E-01		0.248	0.124	4.63E-02		0.052	0.049	2.90E-01	
CERI: COGA AFR	0.346	0.071	9.43E-07	*	0.268	0.106	1.12E-02	*	0.279	0.056	5.90E-07	*	0.312	0.055	1.36E-08	*
CERI: COGA EUR	0.258	0.041	4.41E-10	*	0.215	0.053	4.37E-05	*	0.361	0.048	4.14E-14	*	0.277	0.038	4.44E-13	*

Supplementary Table 5: Random Effects Model Fitting

	Model	# of parameters	AIC	BIC	LL	Deviance (-2*LL)	Chi-squared	ΔDf	P	P < .05/4
Alcohol dependence	Base (RI only)	12	8975.86	9067.36	-4475.93	8951.86	-	-	-	
	RS for ALCC	14	8966.95	9073.69	-4469.47	8938.95	12.92	2	1.57E-03	*
	RS for ALCP	14	8973.41	9080.16	-4472.71	8945.41	6.45	2	3.97E-02	
	RS for CPD	14	8979.84	9086.59	-4475.92	8951.84	0.02	2	9.90E-01	
	RS for EXT	14	8979.86	9086.60	-4475.93	8951.86	0.01	2	9.97E-01	
	RS for DEP	14	8977.53	9084.27	-4474.76	8949.53	2.33	2	3.11E-01	
	RS for SCZ	14	8979.86	9086.61	-4475.93	8951.86	0.00	2	1.00E+00	
	RS for CERI	14	8935.70	9042.44	-4453.85	8907.70	44.16	2	2.57E-10	*
Nicotine dependence	Base (RI only)	12	5241.87	5333.37	-2608.94	5217.87	-	-	-	
	RS for ALCC	14	5243.52	5350.27	-2607.76	5215.52	2.35	2	3.09E-01	
	RS for ALCP	14	5243.86	5350.60	-2607.93	5215.86	2.02	2	3.65E-01	
	RS for CPD	14	5245.72	5352.47	-2608.86	5217.72	0.15	2	9.30E-01	
	RS for EXT	14	5243.81	5350.56	-2607.90	5215.81	2.06	2	3.57E-01	
	RS for DEP	14	5245.33	5352.08	-2608.67	5217.33	0.54	2	7.64E-01	
	RS for SCZ	14	5245.87	5352.61	-2608.93	5217.87	0.00	2	9.98E-01	
	RS for CERI	14	5230.05	5336.80	-2601.03	5202.05	15.82	2	3.67E-04	*
Drug dependence	Base (RI only)	12	6628.00	6719.49	-3302.00	6604.00	-	-	-	
	RS for ALCC	14	6629.52	6736.26	-3300.76	6601.52	2.48	2	2.90E-01	
	RS for ALCP	14	6629.87	6736.62	-3300.94	6601.87	2.13	2	3.45E-01	
	RS for CPD	14	6628.21	6734.95	-3300.10	6600.21	3.79	2	1.50E-01	
	RS for EXT	14	6620.98	6727.73	-3296.49	6592.98	11.01	2	4.06E-03	*
	RS for DEP	14	6630.59	6737.34	-3301.30	6602.59	1.40	2	4.96E-01	
	RS for SCZ	14	6631.82	6738.56	-3301.91	6603.82	0.18	2	9.14E-01	
	RS for CERI	14	6562.70	6669.45	-3267.35	6534.70	69.29	2	8.98E-16	*
Any substance dependence	Base (RI only)	12	12229.74	12321.23	-6102.87	12205.74	-	-	-	
	RS for ALCC	14	12223.39	12330.13	-6097.69	12195.39	10.35	2	5.65E-03	*
	RS for ALCP	14	12226.01	12332.75	-6099.00	12198.01	7.73	2	2.10E-02	
	RS for CPD	14	12233.05	12339.79	-6102.52	12205.05	0.69	2	7.08E-01	
	RS for EXT	14	12226.72	12333.46	-6099.36	12198.72	7.02	2	2.99E-02	
	RS for DEP	14	12232.40	12339.15	-6102.20	12204.40	1.33	2	5.13E-01	
	RS for SCZ	14	12233.72	12340.47	-6102.86	12205.72	0.01	2	9.94E-01	
	RS for CERI	14	12170.27	12277.01	-6071.13	12142.27	63.47	2	1.65E-14	*

RI = random intercept; RS = random slope

Supplementary Table 6: Comparison of Fixed and Random Effects Integrative Data Analysis Results

FIXED EFFECTS RESULTS †

	Alcohol dependence					Nicotine dependence				Drug dependence				Any substance dependence			
	OR	**Beta**	**SE**	**P**		**Beta**	**SE**	**P**		**OR**	**Beta**	**SE**	**P**	**OR**	**Beta**	**SE**	**P**
Female	0.622	-0.475	0.056	1.78E-17	*	0.690	-0.371	0.074	5.41E-07	*	0.536	-0.624	0.062	1.15E-23	*	0.619	-0.480	0.044	7.44E-28	*
Age	1.586	0.461	0.035	4.16E-39	*	1.390	0.329	0.050	3.20E-11	*	1.303	0.265	0.033	2.60E-15	*	1.462	0.380	0.029	2.14E-38	*
ALCC PGS	1.049	0.047	0.030	1.17E-01		0.959	-0.041	0.041	3.07E-01		1.049	0.048	0.035	1.67E-01		1.002	0.002	0.024	9.33E-01	
ALCP PGS	1.127	0.119	0.031	1.35E-04	*	1.011	0.011	0.043	7.94E-01		1.074	0.072	0.036	4.66E-02		1.100	0.095	0.025	1.28E-04	*
EXT PGS	1.182	0.167	0.031	1.02E-07	*	1.501	0.406	0.042	6.79E-22	*	1.272	0.241	0.035	3.67E-12	*	1.310	0.270	0.025	1.16E-27	*
DEP PGS	0.996	-0.004	0.030	8.96E-01		1.061	0.059	0.040	1.40E-01		1.083	0.079	0.033	1.54E-02		1.023	0.023	0.024	3.30E-01	
SCZ PGS	1.038	0.037	0.032	2.47E-01		0.975	-0.025	0.043	5.63E-01		1.032	0.032	0.036	3.83E-01		1.004	0.004	0.025	8.85E-01	
CPD PGS	1.000	0.000	0.030	9.95E-01		1.330	0.285	0.038	3.83E-14	*	1.011	0.011	0.032	7.41E-01		1.081	0.078	0.023	7.77E-04	*
Female	0.653	-0.426	0.057	6.25E-14	*	0.733	-0.311	0.076	4.63E-05	*	0.570	-0.562	0.065	8.12E-18	*	0.646	-0.437	0.046	1.59E-21	*
Age	1.515	0.415	0.036	4.82E-30	*	1.275	0.243	0.055	1.19E-05	*	1.176	0.162	0.036	6.07E-06	*	1.372	0.316	0.030	7.64E-26	*
CERI	1.369	0.314	0.016	2.93E-88	*	1.632	0.490	0.020	1.72E-134	*	1.666	0.510	0.017	1.76E-193	*	1.582	0.459	0.014	1.64E-245	*
Female	0.651	-0.430	0.057	4.29E-14	*	0.722	-0.325	0.077	2.38E-05	*	0.567	-0.568	0.066	4.71E-18	*	0.641	-0.445	0.046	5.09E-22	*
Age	1.522	0.420	0.037	2.65E-30	*	1.290	0.255	0.056	4.90E-06	*	1.182	0.167	0.036	3.65E-06	*	1.381	0.323	0.030	2.93E-26	*
CERI	1.352	0.302	0.016	2.42E-76	*	1.584	0.460	0.020	6.64E-112	*	1.646	0.498	0.018	1.68E-176	*	1.553	0.440	0.014	3.11E-218	*
ALCC PGS	1.035	0.035	0.031	2.61E-01		0.945	-0.057	0.042	1.74E-01		1.031	0.031	0.037	4.01E-01		0.988	-0.012	0.025	6.21E-01	
ALCP PGS	1.119	0.113	0.032	3.69E-04	*	0.993	-0.007	0.044	8.79E-01		1.057	0.056	0.038	1.39E-01		1.092	0.088	0.026	7.67E-04	*
EXT PGS	1.081	0.078	0.033	1.69E-02		1.326	0.283	0.044	1.61E-10	*	1.112	0.106	0.037	3.94E-03	*	1.179	0.165	0.026	2.37E-10	*
DEP PGS	0.967	-0.034	0.031	2.65E-01		1.016	0.016	0.042	7.09E-01		1.026	0.026	0.035	4.68E-01		0.981	-0.019	0.025	4.39E-01	
SCZ PGS	1.032	0.032	0.032	3.24E-01		0.962	-0.039	0.044	3.76E-01		1.007	0.007	0.037	8.60E-01		1.000	0.000	0.026	9.92E-01	
CPD PGS	0.981	-0.019	0.030	5.33E-01		1.314	0.273	0.039	3.56E-12	*	0.977	-0.023	0.033	4.86E-01		1.062	0.060	0.024	1.27E-02	

RANDOM EFFECTS RESULTS ††

	OR	**Beta**	**SE**	**P**		**OR**	**Beta**	**SE**	**P**		**OR**	**Beta**	**SE**	**P**		**OR**	**Beta**	**SE**	**P**	
Intercept	0.131	-2.036	0.760	7.39E-03	*	0.045	-3.103	0.682	5.31E-06	*	0.098	-2.324	1.724	1.78E-01		0.143	-1.946	0.500	1.01E-04	*
Female	0.648	-0.433	0.057	2.18E-14	*	0.724	-0.323	0.077	2.45E-05	*	0.555	-0.589	0.067	1.26E-18	*	0.639	-0.448	0.047	1.82E-21	*
Age	1.547	0.436	0.041	4.83E-26	*	1.297	0.260	0.060	1.38E-05	*	1.179	0.165	0.043	1.12E-04	*	1.409	0.343	0.035	7.32E-23	*
ALCC PGS	1.032	0.031	0.030	3.02E-01		0.941	-0.061	0.042	1.46E-01		1.023	0.022	0.037	5.47E-01		0.984	-0.016	0.025	5.12E-01	
ALCP PGS	1.127	0.119	0.031	1.48E-04	*	0.994	-0.006	0.043	8.98E-01		1.065	0.063	0.038	9.64E-02		1.096	0.092	0.026	4.12E-04	*
EXT PGS	1.097	0.093	0.032	3.25E-03	*	1.338	0.291	0.043	1.42E-11	*	1.129	0.121	0.037	9.05E-04	*	1.194	0.178	0.026	9.36E-12	*
DEP PGS	0.964	-0.037	0.031	2.32E-01		1.015	0.015	0.042	7.26E-01		1.025	0.025	0.036	4.84E-01		0.980	-0.020	0.025	4.33E-01	
SCZ PGS	1.032	0.031	0.031	3.15E-01		0.959	-0.042	0.042	3.16E-01		1.006	0.006	0.037	8.74E-01		0.997	-0.003	0.026	8.99E-01	
CPD PGS	0.984	-0.016	0.029	5.75E-01		1.320	0.277	0.040	2.53E-12	*	0.983	-0.017	0.034	6.18E-01		1.068	0.066	0.024	5.85E-03	*
Intercept	0.034	-3.385	0.282	4.28E-33	*	0.008	-4.892	0.443	2.14E-28	*	0.012	-4.457	0.530	4.34E-17	*	0.059	-2.830	0.181	7.27E-55	*
Female	0.650	-0.430	0.057	2.98E-14	*	0.734	-0.309	0.076	4.60E-05	*	0.559	-0.582	0.067	2.49E-18	*	0.644	-0.440	0.047	5.56E-21	*
Age	1.538	0.430	0.041	1.45E-25	*	1.281	0.247	0.059	2.78E-05	*	1.173	0.159	0.042	1.79E-04	*	1.399	0.336	0.035	3.27E-22	*
CERI	1.437	0.363	0.049	8.28E-14	*	1.622	0.483	0.069	3.28E-12	*	1.759	0.565	0.076	1.23E-13	*	1.604	0.473	0.053	6.56E-19	*
Intercept	0.034	-3.371	0.290	3.36E-31	*	0.008	-4.868	0.453	6.06E-27	*	0.012	-4.432	0.524	2.62E-17	*	0.061	-2.805	0.186	1.22E-51	*
Female	0.648	-0.434	0.057	2.05E-14	*	0.725	-0.321	0.077	2.76E-05	*	0.555	-0.588	0.067	1.43E-18	*	0.639	-0.448	0.047	1.87E-21	*
Age	1.546	0.435	0.041	5.22E-26	*	1.297	0.260	0.060	1.26E-05	*	1.179	0.165	0.043	1.07E-04	*	1.408	0.342	0.035	6.66E-23	*
CERI	1.424	0.353	0.050	1.37E-12	*	1.572	0.453	0.073	6.52E-10	*	1.735	0.551	0.075	2.63E-13	*	1.579	0.457	0.055	1.07E-16	*
ALCC PGS	1.032	0.032	0.030	2.94E-01		0.941	-0.061	0.042	1.48E-01		1.023	0.023	0.037	5.43E-01		0.984	-0.016	0.025	5.17E-01	
ALCP PGS	1.126	0.119	0.031	1.55E-04	*	0.994	-0.006	0.043	8.91E-01		1.065	0.063	0.038	9.63E-02		1.096	0.092	0.026	4.22E-04	*
EXT PGS	1.095	0.091	0.032	4.11E-03	*	1.336	0.290	0.043	1.93E-11	*	1.128	0.121	0.037	9.68E-04	*	1.193	0.177	0.026	1.22E-11	*
DEP PGS	0.964	-0.037	0.031	2.30E-01		1.014	0.014	0.042	7.37E-01		1.025	0.025	0.036	4.91E-01		0.980	-0.020	0.025	4.27E-01	
SCZ PGS	1.032	0.032	0.031	3.08E-01		0.960	-0.041	0.042	3.26E-01		1.006	0.006	0.037	8.70E-01		0.997	-0.003	0.026	9.12E-01	
CPD PGS	0.984	-0.017	0.029	5.67E-01		1.317	0.276	0.040	3.38E-12	*	0.983	-0.017	0.034	6.09E-01		1.068	0.066	0.024	6.21E-03	*

† All models included cohort as a covariate. SE's adjusted for clustering at the family level.

†† All models included cohort, family unit, and CERI as random effects.

PGS residualized on age, sex, and first 10 ancestral principal components

Supplementary Table 7: Full Results from Fixed Effects Integrative Data Analysis

Baseline Model	OR	Beta	SE	P		OR	Beta	SE	P		OR	Beta	SE	P		OR	Beta	SE	P	
Female	0.624	-0.472	0.056	2.34E-17	*	0.694	-0.366	0.073	5.55E-07	*	0.540	-0.617	0.062	1.48E-23	*	0.624	-0.472	0.043	1.61E-27	*
Age	1.581	0.458	0.035	7.22E-39	*	1.382	0.323	0.049	4.27E-11	*	1.298	0.261	0.033	2.54E-15	*	1.453	0.373	0.029	2.10E-38	*
Add Health (AFR)	0.035	-3.363	0.135	9.53E-137	*	0.026	-3.662	0.165	1.68E-109	*	0.078	-2.553	0.108	4.40E-124	*	0.116	-2.152	0.086	3.34E-137	*
Add Health (EUR)	0.125	-2.083	0.060	4.66E-268	*	0.105	-2.258	0.074	3.65E-205	*	0.130	-2.039	0.059	4.16E-258	*	0.323	-1.130	0.046	5.80E-132	*
ALSPAC (EUR)	0.111	-2.198	0.074	2.91E-193	*	0.024	-3.732	0.124	7.93E-199	*	0.013	-4.348	0.168	6.36E-147	*	0.163	-1.814	0.059	1.59E-204	*
COGA (AFR)	0.168	-1.782	0.137	8.33E-39	*	0.050	-2.992	0.174	1.30E-66	*	0.516	-0.661	0.101	5.25E-11	*	0.611	-0.493	0.099	6.88E-07	*
COGA (EUR)	0.355	-1.036	0.069	2.14E-50	*	0.105	-2.258	0.104	1.83E-105	*	0.438	-0.826	0.074	5.71E-29	*	0.730	-0.314	0.063	7.29E-07	*
FinnTwin12 (EUR)	0.166	-1.796	0.115	5.86E-55	*	0.035	-3.342	0.204	3.98E-60	*	0.022	-3.817	0.268	6.11E-46	*	0.206	-1.580	0.100	4.40E-56	*
PGS Only Model	OR	Beta	SE	P		OR	Beta	SE	P		OR	Beta	SE	P		OR	Beta	SE	P	
Female	0.622	-0.475	0.056	1.78E-17	*	0.690	-0.371	0.074	5.41E-07	*	0.536	-0.624	0.062	1.15E-23	*	0.619	-0.480	0.044	7.44E-28	*
Age	1.586	0.461	0.035	4.16E-39	*	1.390	0.329	0.050	3.20E-11	*	1.303	0.265	0.033	2.60E-15	*	1.462	0.380	0.029	2.14E-38	*
Add Health (AFR)	0.034	-3.387	0.136	4.87E-137	*	0.023	-3.789	0.171	3.40E-109	*	0.075	-2.586	0.109	1.56E-123	*	0.112	-2.188	0.088	9.40E-136	*
Add Health (EUR)	0.121	-2.110	0.060	8.35E-269	*	0.094	-2.369	0.077	1.94E-208	*	0.125	-2.081	0.060	7.30E-260	*	0.314	-1.158	0.047	5.70E-134	*
ALSPAC (EUR)	0.109	-2.219	0.074	1.01E-197	*	0.021	-3.870	0.125	2.08E-211	*	0.012	-4.390	0.168	1.11E-150	*	0.157	-1.853	0.060	6.76E-212	*
COGA (AFR)	0.165	-1.800	0.139	3.18E-38	*	0.045	-3.103	0.177	1.03E-68	*	0.509	-0.676	0.105	1.31E-10	*	0.603	-0.505	0.104	1.24E-06	*
COGA (EUR)	0.349	-1.052	0.068	1.64E-53	*	0.094	-2.367	0.102	1.49E-118	*	0.429	-0.847	0.073	1.47E-31	*	0.723	-0.325	0.062	1.72E-07	*
FinnTwin12 (EUR)	0.162	-1.819	0.115	3.08E-56	*	0.031	-3.463	0.206	1.87E-63	*	0.021	-3.862	0.268	3.97E-47	*	0.199	-1.613	0.100	2.61E-58	*
ALCC PGS	1.049	0.047	0.030	1.17E-01		0.959	-0.041	0.041	3.07E-01		1.049	0.048	0.035	1.67E-01		1.002	0.002	0.024	9.33E-01	
ALCP PGS	1.127	0.119	0.031	1.35E-04	*	1.011	0.011	0.043	7.94E-01		1.074	0.072	0.036	4.66E-02		1.100	0.095	0.025	1.28E-04	*
EXT PGS	1.182	0.167	0.031	1.02E-07	*	1.501	0.406	0.042	6.79E-22	*	1.272	0.241	0.035	3.67E-12	*	1.310	0.270	0.025	1.16E-27	*
DEP PGS	0.996	-0.004	0.030	8.96E-01		1.061	0.059	0.040	1.40E-01		1.083	0.079	0.033	1.54E-02		1.023	0.023	0.024	3.30E-01	
SCZ PGS	1.038	0.037	0.032	2.47E-01		0.975	-0.025	0.043	5.63E-01		1.032	0.032	0.036	3.83E-01		1.004	0.004	0.025	8.85E-01	
CPD PGS	1.000	0.000	0.030	9.95E-01		1.330	0.285	0.038	3.83E-14	*	1.011	0.011	0.032	7.41E-01		1.081	0.078	0.023	7.77E-04	*
CERI Only Model	OR	Beta	SE	P		OR	Beta	SE	P		OR	Beta	SE	P		OR	Beta	SE	P	
Female	0.653	-0.426	0.057	6.25E-14	*	0.733	-0.311	0.076	4.63E-05	*	0.570	-0.562	0.065	8.12E-18	*	0.646	-0.437	0.046	1.59E-21	*
Age	1.515	0.415	0.036	4.82E-30	*	1.275	0.243	0.055	1.19E-05	*	1.176	0.162	0.036	6.07E-06	*	1.372	0.316	0.030	7.64E-26	*
Add Health (AFR)	0.017	-4.070	0.141	4.91E-182	*	0.008	-4.861	0.185	1.78E-151	*	0.023	-3.760	0.128	6.05E-190	*	0.041	-3.198	0.099	2.26E-227	*
Add Health (EUR)	0.060	-2.821	0.078	3.84E-290	*	0.030	-3.497	0.105	4.09E-243	*	0.036	-3.317	0.086	0.00E+00	*	0.114	-2.171	0.061	1.86E-274	*
ALSPAC (EUR)	0.052	-2.964	0.087	1.73E-253	*	0.007	-5.005	0.137	1.30E-290	*	0.003	-5.692	0.176	3.02E-230	*	0.053	-2.946	0.073	0.00E+00	*
COGA (AFR)	0.037	-3.285	0.150	9.53E-106	*	0.004	-5.604	0.206	2.86E-162	*	0.049	-3.008	0.135	9.90E-110	*	0.080	-2.526	0.117	8.67E-103	*
COGA (EUR)	0.093	-2.380	0.099	5.85E-129	*	0.009	-4.679	0.145	2.35E-228	*	0.046	-3.083	0.108	3.02E-180	*	0.115	-2.166	0.083	9.95E-151	*
FinnTwin12 (EUR)	0.064	-2.744	0.125	7.78E-107	*	0.007	-4.917	0.226	9.45E-105	*	0.004	-5.483	0.274	6.49E-89	*	0.051	-2.981	0.113	1.41E-153	*
CERI	1.369	0.314	0.016	2.93E-88	*	1.632	0.490	0.020	1.72E-134	*	1.666	0.510	0.017	1.76E-193	*	1.582	0.459	0.014	1.64E-245	*
Combined Risk Model	OR	Beta	SE	P		OR	Beta	SE	P		OR	Beta	SE	P		OR	Beta	SE	P	
Female	0.651	-0.430	0.057	4.29E-14	*	0.722	-0.325	0.077	2.38E-05	*	0.567	-0.568	0.066	4.71E-18	*	0.641	-0.445	0.046	5.09E-22	*
Age	1.522	0.420	0.037	2.65E-30	*	1.290	0.255	0.056	4.90E-06	*	1.182	0.167	0.036	3.65E-06	*	1.381	0.323	0.030	2.93E-26	*
Add Health (AFR)	0.017	-4.050	0.141	8.38E-181	*	0.008	-4.863	0.188	9.83E-148	*	0.024	-3.733	0.127	8.82E-189	*	0.042	-3.166	0.099	1.51E-223	*
Add Health (EUR)	0.060	-2.811	0.078	1.41E-285	*	0.030	-3.508	0.107	2.71E-237	*	0.037	-3.308	0.087	0.00E+00	*	0.117	-2.148	0.062	1.92E-266	*
ALSPAC (EUR)	0.053	-2.944	0.088	4.35E-248	*	0.007	-5.021	0.138	6.85E-292	*	0.003	-5.672	0.176	3.10E-228	*	0.054	-2.921	0.073	0.00E+00	*
COGA (AFR)	0.040	-3.230	0.153	1.62E-99	*	0.004	-5.472	0.202	3.37E-162	*	0.052	-2.952	0.136	7.67E-104	*	0.087	-2.445	0.119	4.16E-94	*
COGA (EUR)	0.096	-2.343	0.100	9.97E-122	*	0.010	-4.624	0.146	2.82E-219	*	0.048	-3.046	0.109	3.63E-171	*	0.122	-2.101	0.084	8.41E-138	*
FinnTwin12 (EUR)	0.066	-2.722	0.126	4.41E-103	*	0.007	-4.919	0.229	3.88E-102	*	0.004	-5.467	0.275	6.25E-88	*	0.052	-2.948	0.114	2.12E-148	*
CERI	1.352	0.302	0.016	2.42E-76	*	1.584	0.460	0.020	6.64E-112	*	1.646	0.498	0.018	1.68E-176	*	1.553	0.440	0.014	3.11E-218	*
ALCC PGS	1.035	0.035	0.031	2.61E-01		0.945	-0.057	0.042	1.74E-01		1.031	0.031	0.037	4.01E-01		0.988	-0.012	0.025	6.21E-01	
ALCP PGS	1.119	0.113	0.032	3.69E-04	*	0.993	-0.007	0.044	8.79E-01		1.057	0.056	0.038	1.39E-01		1.092	0.088	0.026	7.67E-04	*
EXT PGS	1.081	0.078	0.033	1.69E-02		1.326	0.283	0.044	1.61E-10	*	1.112	0.106	0.037	3.94E-03	*	1.179	0.165	0.026	2.37E-10	*
DEP PGS	0.967	-0.034	0.031	2.65E-01		1.016	0.016	0.042	7.09E-01		1.026	0.026	0.035	4.68E-01		0.981	-0.019	0.025	4.39E-01	
SCZ PGS	1.032	0.032	0.032	3.24E-01		0.962	-0.039	0.044	3.76E-01		1.007	0.007	0.037	8.60E-01		1.000	0.000	0.026	9.92E-01	
CPD PGS	0.981	-0.019	0.030	5.33E-01		1.314	0.273	0.039	3.56E-12	*	0.977	-0.023	0.033	4.86E-01		1.062	0.060	0.024	1.27E-02	

* $p < .05/4$; Beta = log(OR)