Polygenic Risk for Substance-Related Traits Predicts Substance Use Onset and Progression: Sex and Population Group Differences

Henry R. Kranzler,1,2 Richard Feinn,3 Heng Xu,1 Brendan L. Ho,1 Divya Saini,1 Olivia R. Nicastro,1 Anya Jacoby,1 Sylvanus Toikumo,1,2 Joel Gelernter,4 Emily E. Hartwell,1,3 and Rachel L. Kember,1,2

1Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104
2Mental Illness Research, Education and Clinical Center, Veterans Integrated Service Network 4, Crescenz Veterans Affairs Medical Center, Philadelphia, PA 19104
3Department of Medical Sciences, Frank H. Netter School of Medicine at Quinnipiac University, North Haven, CT 06473
4Department of Psychiatry, Yale University School of Medicine, New Haven, CT 06511 and VA Connecticut Healthcare System, West Haven, CT 06516

*Correspondence to Henry R. Kranzler, M.D., Center for Studies of Addiction, 3535 Market Street, Suite 500, Philadelphia, PA 19104; email: kranzler@pennmedicine.upenn.edu; telephone: 215-746-1943

Abstract

Background: Charting the clinical course of substance use disorders (SUDs) to identify etiologic contributors to milestone onset and progression could inform intervention efforts. Methods: We calculated polygenic risk scores (PRS) in 5,692 European-ancestry individuals (EUR) (56.2% male) and 4,918 African-ancestry (AFR) individuals (54.9% male) using genome-wide association studies (GWAS) of alcohol use disorder (AUD), opioid use disorder (OUD), and smoking trajectory (SMK). Using Cox regression, we examined the association of polygenic risk with age of first substance use, regular use, reported problems, and dependence diagnosis and with progression from regular use to onset of problems and dependence. Results: EUR and males reported earlier onset and shorter progression times than AFR and females, respectively. Among EUR, higher AUD PRS predicted earlier onset and more rapid progression to alcohol-related milestones (p’s<0.0001) and although a stronger moderator of problem onset among females (p=0.0165), it was more predictive of the progression to problems among males (p=0.0054). OUD and SMK PRS in EUR also predicted earlier onset of the respective milestones (p’s=0.0002). Among AFR, where power is lower, AUD PRS predicted age of regular alcohol use (p=0.039) and dependence (p=0.001) and progression from regular use to diagnosis (p=0.045), while SMK PRS predicted earlier age of initiation (p=0.036). Conclusions: Genetic risk for SUDs predicts milestones and symptom progression in EUR and, to a lesser extent, among AFR. Larger, diverse discovery GWAS and target samples are needed to enhance the power of PRS to personalize interventions for individuals at genetic risk of serious substance-related outcomes.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Disclosure: Dr. Kranzler is a member of advisory boards for Dicerna Pharmaceuticals, Sophrosyne Pharmaceuticals, and Enthion Pharmaceuticals; a consultant to Sobrera Pharmaceuticals; the recipient of research funding and medication supplies for an investigator-initiated study from Alkermes; and a member of the American Society of Clinical Psychopharmacology’s Alcohol Clinical Trials Initiative, which was supported in the last three years by Alkermes, Dicerna, Ethypharm, Lundbeck, Mitsubishi, and Otsuka. Drs. Gelernter and Kranzler hold U.S. Patent 10,900,082: Genotype-guided Dosing of Opioid Receptor Agonists, 26 Jan. 2021. The other authors have no disclosures to make.

Acknowledgments: Supported by the Veterans Integrated Service Network 4 Mental Illness Research, Education and Clinical Center and NIH grants P30 DA046345, R01 AA026364, and K01 AA028292 (to RLK).
Introduction

Substance use disorders (SUDs), characterized by the chronic use of alcohol or drugs, are common among adolescents and adults and result in clinically significant social impairments and medical and psychiatric disorders. SUDs develop in stages following initial substance use, often progressing across a series of sequential transitions, which can be conceptualized as a continuous trajectory marked by milestones of escalating use or severity. Charting the clinical course of SUDs using these developmental events can help to elucidate the factors that underlie symptom progression, a key example of which is the transition from regular substance use to substance dependence. The timing of milestones could also provide a personalized assessment of an individual’s risk of symptom progression and a more precise time window for a targeted intervention aimed at preventing the progression to a subsequent milestone.

The age at which substance-related milestones occur and the rate of progression through them are influenced by both genetic and environmental factors. These risk factors may be reflected in population-group differences in the initiation of substance use and the progression across substance-related milestones. For example, Black individuals report first consuming alcohol and tobacco products and initiating regular drinking and binge drinking, and use of illicit drugs, later than White individuals.

Although Black adolescents had a significantly lower risk of transitioning to regular use of alcohol and illicit drugs, Black adults 30 years of age and older have also been shown to have a more rapid progression from regular alcohol use to regular drinking and intoxication than White adults. These differences occur in the context of a general paucity of studies of population-group differences in substance-related symptom progression. There is a similar dearth of findings on population group differences in age-of-onset and progression measures for opioids, though in one study Black individuals had a more rapid progression to opioid dependence (OD) than White individuals.
Sex may also influence the developmental course of substance-related traits. In monozygotic and dizygotic twins, the initiation of alcohol and tobacco use occurred earlier among males than females with earlier initiation of use associated with a significantly increased likelihood of developing dependence on these substances. In contrast, in another twin study, between-twin comparisons of the rate of progression through alcohol-related milestones showed no overall pattern of sex differences. Thus, the relationship of sex to alcohol-related milestones and progression in SUDs is not fully understood. A controversial question in relation to sex differences is the validity of the phenomenon of telescoping, which posits that despite their later initiation of substance use, women manifest substance-related problems sooner than men.

Risk for SUDs is highly polygenic, involving potentially thousands of individual variants, each accounting for a very small proportion of trait variance. Polygenic risk scores (PRS) sum data from multiple genetic variants and account for greater proportions of trait variance than single polymorphisms. PRS are useful in evaluating the risk for disease progression in diverse medical disorders, including breast cancer and rheumatoid arthritis, and the prediction of sudden death in individuals with coronary disease. A recent large study of EUR in the UK Biobank showed the utility of a prostate cancer genetic risk score for triaging patients in primary care. Men in the highest quintile of risk had a prostate cancer incidence of 8.1% and could be fast-tracked for further investigation, while the incidence among those in the lowest risk quintile was <1% and they could more safely avoid invasive investigation. Thus, there currently are clinical applications of PRS to differentiate individuals based on their genetic risk for a disease.

For SUDs, a PRS for alcohol dependence (AD) was associated with the progression from onset of regular drinking to AD in a European-ancestry (EUR) sample. The ability to quantify the genetic risk of symptom progression could help to identify individuals at highest risk of developing more severe milestones (e.g., alcohol-related problems or AD) and who could benefit most from intensive interventions.
Here, we examined the previously reported association of an AD PRS with the progression from onset of regular drinking to onset of AD in EUR. We then extend the analyses to include additional milestones and measures of progression of alcohol-related symptoms and conduct parallel analyses of opioid-related and smoking-related traits. Finally, we refine our understanding of the effects of sex and population group differences on these features by studying males and females and EUR and African-ancestry (AFR) participants, all of whom are well represented in our sample, which was recruited and deeply phenotyped for genetic studies of SUDs.

**Methods**

**Discovery Samples**

We used GWAS summary statistics for alcohol use disorder (AUD), opioid use disorder (OUD), and smoking trajectories (SMK) as discovery samples for calculating PRS in the Yale-Penn sample. The three discovery samples are described in detail in Supplemental Methods.

To compare the power among the discovery samples, we calculated the genomic inflation factor ($\lambda_{gc}$). In the absence of population structure, $\lambda_{gc}$ is a function of sample size and the number of causal variants and thus reflects power to detect significant SNP-trait associations. We calculated $\lambda_{gc}$ using the expected median of SNP test statistics from each set of GWAS summary statistics, with greater $\lambda_{gc}$ denoting higher predictive power. These showed that for all three traits EUR samples had better predictive power than AFR samples. For AUD, the $\lambda_{gc}$ was 1.100 for EUR and 1.034 for AFR, while for OUD the respective values were 1.112 and 1.028, and for SMK they were 1.336 and 1.094.

**Target Sample**

We calculated PRS in the Yale-Penn sample, which comprises 16,715 individuals recruited at five U.S. sites for genetic studies of dependence on cocaine, opioids, and alcohol.
Of this number, genome-wide genotype data are available for 10,610 individuals, including 5,692 EUR (56.2% male) and 4,918 AFR (54.9% male) (Table 1). The study was approved by the institutional review board at each site and participants gave written informed consent for data collection. Cases were identified through addiction treatment facilities, inpatient and outpatient psychiatric services, and posters and advertisements in local media. Although some cocaine- or opioid-dependent individuals were recruited as probands of small nuclear families, data from family members are excluded in analyses presented here. Unaffected controls were recruited from non-psychiatric medical settings and through advertisements.

**Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA)**

The SSADDA is a comprehensive psychiatric interview that comprises 24 modules assessing the physical, psychological, social, and psychiatric manifestations of SUDs, psychiatric disorders, and environmental covariates considered likely to have an impact on SUDs. The SSADDA’s semi-structured format, accompanied by the rigorous training and quality control procedures used in the Yale-Penn sample,\textsuperscript{24} allow a carefully trained non-clinician interviewer to assess diagnostic criteria and disorders and their ages of onset, to yield DSM-IV diagnoses of AD, OD, and nicotine dependence (ND). Ascertainment of ages of onset of the different milestones is done using questions that elicit estimated ages (e.g., of initiation of substance and regular substance use) and clustering of criteria within a 12-month period (for DSM-IV diagnoses).

**Genotyping, Imputation, and Polygenic Risk Scores**

Yale-Penn samples were genotyped in three batches using the Illumina HumanOmni1-Quad microarray, the Illumina HumanCoreExome array, or the Illumina Multi-Ethnic Global array. Genotype data were filtered for individual call rates and excessive heterozygosity using PLINK v1.9 and were imputed using the Michigan Imputation Server\textsuperscript{25} and the Haplotype Reference Consortium Panel.\textsuperscript{26}
PRS were calculated for AUD,20 OUD,21 and SMK22 using Polygenic Risk Scores–
Continuous Shrinkage software (PRS-CS)27 and the 1000 Genomes Project phase 3 AFR and
EUR samples for estimates of linkage disequilibrium. Global shrinkage parameters were
obtained from each set of summary statistics by the PRS-CS package and effective sample
sizes were used to calculate the final PRS. We used matched, genetically determined, ancestral
summary statistics (e.g., an AFR GWAS for AUD was used to calculate AUD PRS in AFR Yale-
Penn individuals). Whereas PRS calculated with the effective or actual sample sizes did not
differ substantially, we report findings using effective sample sizes.

Statistical Analysis

We used PROC PHREG in SAS v9.4 to run Cox proportional hazards models for each of
three substances (alcohol, opioids, and smoking) and the following four measures: age of first
use, age of regular use, age of first bringing up problems with a healthcare professional, and
age of diagnosis of the disorder (AD, OD, and ND). Using these milestones, we also examined
two measures of progression: time from age of regular use to age of first bringing up problems
and time from age of regular use to age of diagnosis. Analyses were conducted separately for
EUR and AFR, for which there are comparable numbers in the Yale-Penn sample. Similarly,
analyses were conducted separately for females and males within each ancestry group, also
roughly equally distributed in the target sample.

All models included the respective PRS, age, and the first 10 principal ancestral
components as covariates. In analyses that did not examine sex as a factor, sex was included
as a covariate. The models for the progression outcomes also included the age of regular use
as a covariate. Because the proportional hazards assumption for this covariate was violated in
the progression models for all three substances, we added an interaction term for age of regular
use for the two progression measures, which was significant in all progression models. Thus,
we assessed the impact of age of onset of regular use on the two progression outcomes by
decomposing the effects for participants with early onset (≤18 years) and late onset (>18 years) of regular use.\textsuperscript{18}

For individuals who reported never having experienced a specific event, data were censored for the event in the Cox models and the age at interview was substituted for the missing age of onset. Thus, the age-of-onset outcomes for all substances have the same sample size. For the two progression outcomes, only individuals who reported regular use were included in the analysis, as censoring that age of onset would distort the analysis. Thus, the sample sizes for the two progression outcomes vary by substance.

For effect sizes, we report the hazard ratio (HR) with 95\% confidence intervals, reflecting the change in hazard for a one-standard-deviation increase in PRS. For individual milestones, a HR>1.0 reflects a greater likelihood that the event will occur as PRS increases. Similarly, with progression (or latency) measures, there is a greater likelihood of progression from onset of regular substance use to either first reported problems or first dependence diagnosis as PRS increases. We also report p-values and adjusted p-values using the Hommel correction for multiple testing, with adjustments made separately by population group and substance. We consider PRS as a significant predictor when the p-value, adjusted for multiple comparisons, is <0.05.

**Results**

**Rates of Endorsement of Substance Use Milestones**

Table 1 shows the prevalence of the milestones of first use, regular use, first reporting of problems, and onset of dependence, for alcohol, opioids, and nicotine in the Yale-Penn sample. More than 95\% of both population groups reported ever having used alcohol and over 80\% reported ever having used alcohol regularly. Lifetime opioid use was less common among AFR (33\%) than EUR (51\%), as was regular opioid use (21\% vs. 41\%). In both groups, nearly 90\% of individuals reported having smoked more than 100 cigarettes lifetime, with over 65\% endorsing
regular smoking. In both AFR and EUR, the event endorsed least commonly for all three substances was having brought up problems to a healthcare professional, which ranged from 15%-31% in AFR and 31%-38% in EUR.

Most individuals in both population groups met DSM-IV criteria for AD and ND, while the prevalence of OD was 39% among EUR and less than half that among AFR (18%). The high rate of SUD diagnoses in the Yale-Penn sample reflects its ascertainment for studies of SUD genetics. The sample prevalence of all three SUD diagnoses is lower among women and AFR and the onset of all milestones earlier than among men and EUR, respectively. Similarly, EUR and males reported earlier onset and shorter progression times for all substances than AFR and women, respectively (Table 2).

(Insert Tables 1 and 2 Here)

PRS Effects on Alcohol-related Milestones and Symptom Progression

Table 3 and Figure 1 (upper two panels) show the Cox regression model of the effects of the AUD PRS on alcohol-related milestones by population group. Significant associations reflect a younger age of onset and a shorter latency between milestones as a function of increasing PRS. Among EUR, AUD PRS was a significant predictor (adjusted p-value \(p_{\text{adj}}<0.001\)) of all four milestones, with HRs ranging from 1.06 (age of first use and age of regular use) to 1.19 (age at which alcohol-related problems were brought up to a healthcare professional). The AUD PRS was also a significant predictor of the progression from age of regular alcohol use both to bringing up alcohol-related problems (HR=1.14, \(p_{\text{adj}}<0.001\)) and age of AD diagnosis (HR=1.10, \(p_{\text{adj}}<0.001\)) (Figure 2).

(Insert Table 3 and Figures 1 and 2 Here)

Among AFR, the AUD PRS significantly predicted both age of regular alcohol use (HR=1.05, \(p_{\text{adj}}=0.039\)) and age of AD diagnosis (HR=1.09, \(p_{\text{adj}}=0.001\)) (Table 3). It was also a significant predictor of the progression from age of regular alcohol use to age of AD diagnosis (HR=1.06, \(p_{\text{adj}}=0.045\)) (Figure 2).
In analyses stratified by age-of-onset of regular drinking (Supplemental Table 1), in both population groups, early onset of regular drinking (<18 years) was associated with more rapid progression both to age of first reported alcohol-related problems and to age of initial AD diagnosis. Among EUR, the AUD PRS, irrespective of the age of onset of regular drinking, significantly predicted the progression to first reported alcohol-related problems and to an AD diagnosis (HRs = 1.09-1.18). Among AFR, the AUD PRS was positively associated with both progression measures (HRs=1.02-1.11), though the effect was significant only for progression to an AD diagnosis and only among individuals with a late onset of regular drinking (>18 years).

A sex-stratified analysis among EUR (Table 4 and Figure 3) showed that the association of AUD PRS with milestones was greater among women (HRs=1.06-1.30) than men (HRs=1.06-1.15). There was also a significant interaction effect of sex by AUD PRS on the age at which alcohol-related problems were raised with a health professional (\( \rho_{adj} = 0.0165 \)). The genetic risk for AUD was a stronger moderator in females than males despite men having brought up problems earlier than women. This is evidenced by greater separation between survival curves for low, medium, and high PRS tertiles among women than men (Figure 3). Similarly, although the effects of the AUD PRS on the two measures of progression were significant in both sexes, the effects were greater among women (HR=1.27 for progression to reporting problems and 1.14 for progression to AD) than among men (HR=1.09 for both). The progression from regular drinking to first reported alcohol-related problems in EUR differed significantly by sex (\( \rho_{adj} = 0.0054 \)), with men bringing up alcohol-related problems sooner after beginning regular drinking than women. As can be seen in Figure 4, there is overlap between the survival curves of females in the highest PRS tertile and males in the lowest PRS tertile.

Among AFR, the only significant effect of the alcohol PRS when stratified by sex was on age of AD diagnosis. The effect was significant in both men (HR=1.09) and women (HR=1.11) (Supplemental Table 2).
Opioid-related Milestones and Progression

Among EUR, the OUD PRS significantly predicted all four milestones (HRs=1.14-1.19) but neither of the progression outcomes. Among AFR, the OUD PRS was not associated with any opioid-related milestones or progression outcomes (Table 5). Figure 1 (middle two panels) shows the difference between population groups in the survival curves for age of first OD diagnosis as a function of OUD PRS strata.

(Insert Table 5 Here)

When stratified by sex, among EUR, the OUD PRS was significantly associated in both sexes with all four of the milestones but neither of the progression measures (Supplemental Table 3).

Among AFR, sex-stratified analyses of the effects of the OUD PRS show that the only milestone that was significant was age of onset of opioid use, an effect limited to women (Supplemental Table 4).

Smoking-related Milestones and Progression

Table 6 shows the results from the Cox models for smoking-related traits. Among EUR, as with OUD, the SMK PRS significantly predicted all four age-of-onset measures (HRs=1.15-1.25), while among AFR it predicted only age of first use (HR=1.05). Survival curves for the age of diagnosis of ND by PRS strata (Figure 1, lower two panels), show the population differences on this key milestone. In neither population group was SMK PRS a significant predictor of the progression from regular smoking to reporting smoking-related problems or to ND.

(Insert Table 6 here)

Among EUR, sex-stratified analyses showed that, in both sexes, the SMK PRS is associated with all four milestones and, among females, with the progression from regular smoking to onset of ND (Supplemental Table 5). Similar analyses among AFR (Supplemental Table 6) yielded only one significant effect on smoking-related traits: among males SMK PRS predicted the age of smoking onset.
Discussion

There is growing interest in the clinical utility of PRS for identifying individuals at high risk for a variety of disorders. In addition to case identification, estimates of genetic risk are increasingly being used to predict disease progression. We used such an approach to demonstrate polygenic effects on the age of onset of substance-related traits and the progression from regular substance use to substance-related problems and dependence. To do this, we used summary statistics from large GWAS of AUD, OUD, and SMK to calculate PRS in a sample of deeply phenotyped individuals either with alcohol and drug use disorders or who were screened controls. We also compared these effects by population group and by sex within population groups.

Our most consistent findings were for alcohol-related traits. Among EUR, an AUD PRS predicted the age of all four alcohol-related milestones and both measures of progression from the onset of regular drinking. The findings replicate a previous observation that an AD PRS predicted the progression from onset of regular drinking to AD diagnosis in a sample of EUR. We also extended the analysis to AFR, among whom an AUD PRS significantly predicted the age of regular alcohol use, the age of an AD diagnosis, and the progression from regular use to an AD diagnosis.

Stratifying the analyses on the age of onset of regular drinking did not substantially alter the findings. Sex-stratified analyses showed that for some outcomes, the effects of PRS were greater among women than men. These findings could be relevant to the phenomenon of telescoping, in which women who, despite initiating substance use at a later age than men have been reported to have a more rapid progression to developing problems and presenting for treatment than men. However, the findings supporting telescoping are inconsistent (e.g.,). Here we found that among EUR the AUD PRS predicted a significantly shorter time from onset of regular alcohol use to bringing up alcohol-related problems among men than
women and among AFR the effect of the AUD PRS on the progression from regular drinking to onset of AD was comparable for men and women.

Among EUR, there were also robust effects of OUD and SMK PRS on opioid- and tobacco-related milestones, respectively. However, neither PRS predicted the progression from age of onset of regular use either to bringing up problems related to these substances or dependence diagnoses. Among AFR, the only opioid-related milestone that was significantly associated with OUD PRS was the age of onset of opioid use among women. In this population group, the SMK PRS was associated with an earlier age of smoking initiation, a finding that was significant in men only.

Overall, we found consistently greater associations of PRS with substance-related milestones and symptom progression in EUR than AFR, attributable to the greater predictive value of the EUR summary statistics for all three substances, evidenced by greater genomic inflation factor values among EUR. Further, despite comparable numbers of EUR and AFR individuals with AD and ND in the Yale-Penn (i.e., target) sample, the number of AFR subjects in the target sample that endorsed opioid-related milestones and that met criteria for OD was about one-third the number among EUR. Thus, for opioids, differences in the target sample also likely contributed to the population-group difference in PRS effects. Differences in the sizes of the discovery and target samples by population group underscore the need, particularly in non-EUR populations, for larger GWAS samples and additional deeply phenotyped samples for more granular studies of genetic risk for substance use milestones and progression. This would require targeted recruitment efforts.

Larger discovery samples would increase the predictive power in both population groups, thereby accounting for more variance in the progression to problematic substance use. A more powerful PRS could thus aid in identifying individuals at greatest risk for developing more serious substance-related consequences and permit secondary preventive efforts. One potential test of the clinical utility of PRS in AUD would involve recruiting individuals at the upper
end of the polygenic risk continuum for the disorder to participate in clinical trials of interventions aimed at preventing the progression to problematic alcohol use. Existing biobanks could be used to accomplish such an effort.

Among EUR, the effects of the AUD PRS were more consistent and robust than were the OUD or SMK PRS. This could have resulted in part from there being approximately one-third fewer participants in the SMK GWAS than either the AUD or OUD GWAS and approximately one-quarter fewer Yale-Penn participants with an OD diagnosis than either an AD or ND diagnosis. Substance-specific differences have also been shown to exist in symptom progression. National survey data showed that the cumulative probability of progressing to dependence was 67.5% for nicotine users and 22.7% for alcohol users. A prospective study of adolescents showed that the shortest progression times (i.e., greatest addictive liabilities) were seen with opioids; tobacco and alcohol had the lowest liabilities. Despite these findings, there are countervailing biological effects. For example, the rate of absorption of nicotine from smoking is much higher than gastrointestinal absorption of alcohol given the extensive surface area of pulmonary alveoli. Thus, pharmacologic or other features specific to individual substances could add to or interact with genetic risk for dependence on them.

Although here we focus principally on genetic risk, environmental factors are also relevant to the age at which substance-related milestones occur. Among both AFR and EUR, we found the lowest HRs for the age of first alcohol use (1.02 and 1.06, respectively). This is consistent with the notion that the initiation of substance use is strongly influenced by social and environmental factors, whereas the progression from first use to heavy use and from heavy use to problematic use or dependence is influenced more by neurobiological, including genetic, factors. This is further supported by the finding that the age at first alcohol use is only modestly genetically correlated with AD ($r_g=18–29\%$), while for age of onset of regular drinking and AD the genetic correlation is moderate ($r_g=0.54$).
This study has limitations. We conducted analyses in only two population groups—EUR and AFR—the smaller numbers of Hispanic, East Asian, and South Asian individuals in both the discovery and target samples do not provide adequate statistical power to support these analyses. We also used DSM-IV SUD diagnoses in the target sample to ensure consistency across the substances, as we did not have all criteria required for a DSM-5 tobacco use disorder diagnosis. Third, we used a trajectory phenotype in the discovery GWAS for smoking, as it was the largest available GWAS for smoking in AFR. Nonetheless, the phenotype differs from the ICD-9/10 codes used in the AUD and OUD GWAS. Whereas the trajectories are a probabilistic categorization rather than a binary diagnosis, the trajectory-based groups are potentially more heterogeneous than AUD or OUD cases and controls. This variability could have diminished the association of the SMK PRS with smoking-related milestones or latency outcomes. Finally, the proportions of SUD diagnoses overall and by population reflect the recruitment strategy of the original studies, and therefore no conclusions may be drawn about these proportions per se.

Despite these limitations, the availability of large discovery GWAS and the deeply phenotyped target sample made it possible to generate sex- and population-specific profiles of the association of polygenic risk for SUDs with both milestones and symptom progression over time. These findings provide an initial characterization of the genetic risk of these more granular features and a foundation for efforts to identify higher-risk individuals who can be targeted with interventions aimed at preventing symptom progression.
Figure Legends

Figure 1: Age of Alcohol, Opioid, and Nicotine Dependence Diagnoses by European or African Ancestry and Low, Medium, and High Polygenic Risk Scores for Alcohol Use Disorder, Opioid Use Disorder, and Smoking Trajectory, Respectively

Figure 2: Years from Regular Alcohol Use to Alcohol Dependence Diagnosis by European or African Ancestry and Low, Medium, and High Alcohol Use Disorder Polygenic Risk Scores

Figure 3: Age First Brought Up Alcohol-Related Problems to a Healthcare Professional Among European-Ancestry Individuals by Sex Alcohol Use Disorder Polygenic Risk Score

Figure 4: Years from First Regular Alcohol Use to First Brought Up Alcohol-Related Problems to a Healthcare Professional Among European-Ancestry Individuals by Sex Alcohol Use Disorder Polygenic Risk Score
References


11. Sartor, C. E., Kranzler, H. R. & Gelernter, J. Rate of progression from first use to 
dependence on cocaine or opioids: A cross-substance examination of associated 

12. Alvanzo, A. A. H. *et al.* Race/ethnicity and sex differences in progression from drinking 
(2011).

13. Lewis, C. M. & Vassos, E. Polygenic risk scores: from research tools to clinical instruments. 

14. Lakeman, I. M. M. *et al.* The predictive ability of the 313 variant–based polygenic risk score 
for contralateral breast cancer risk prediction in women of European ancestry with a 

15. Honda, S. *et al.* Association of Polygenic Risk Scores With Radiographic Progression in 

16. Sandhu, R. K. *et al.* Polygenic Risk Score Predicts Sudden Death in Patients With Coronary 

17. Green, H. D. *et al.* Applying a genetic risk score for prostate cancer to men with lower 
urinary tract symptoms in primary care to predict prostate cancer diagnosis: a cohort study in 

drinking on the progression to alcohol dependence: A polygenic risk score approach. *Drug 

deeply phenotyped sample. *medRxiv* 2022.02.09.22270737 (2022) 

20. Kember, R. L. *et al.* Genetic underpinnings of the transition from alcohol consumption to 
alcohol use disorder: shared and unique genetic architectures in a cross-ancestry sample.

doi:10.1038/s41593-022-01160-z.

22. Xu, K. et al. Genome-wide association study of smoking trajectory and meta-analysis of
smoking status in 842,000 individuals. Nat. Commun. 11, 5302 (2020).


1284–1287 (2016).


27. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via

dependent women show more rapid progression to substance abuse treatment. Drug Alcohol

alcohol dependence among adolescent and young-adult newly incident drinkers in the United


Table 1: Prevalence of Substance Use Disorder Milestones by Ancestral Group and Sex

<table>
<thead>
<tr>
<th>Substances</th>
<th>All (n=4,918)</th>
<th>Male (n=2,704)</th>
<th>Female (n=2,214)</th>
<th>All (n=5,692)</th>
<th>Male (n=3,200)</th>
<th>Female (n=2,492)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First use</td>
<td>4,679 (95.1%)</td>
<td>2,582 (95.5%)</td>
<td>2,097 (94.7%)</td>
<td>5,476 (96.2%)</td>
<td>3,079 (96.2%)</td>
<td>2,397 (96.2%)</td>
</tr>
<tr>
<td>Regular use</td>
<td>4,034 (82.1%)</td>
<td>2,390 (88.4%)</td>
<td>1,644 (74.3%)</td>
<td>4,792 (84.2%)</td>
<td>2,848 (89.0%)</td>
<td>1,944 (78.0%)</td>
</tr>
<tr>
<td>Brought up problems</td>
<td>1,878 (38.2%)</td>
<td>1219 (45.1%)</td>
<td>659 (29.8%)</td>
<td>2,174 (38.2%)</td>
<td>1,479 (46.2%)</td>
<td>695 (27.9%)</td>
</tr>
<tr>
<td>Dependence diagnosis</td>
<td>2,631 (53.5%)</td>
<td>1,693 (62.6%)</td>
<td>938 (42.4%)</td>
<td>2,945 (51.7%)</td>
<td>1,942 (60.7%)</td>
<td>1,003 (40.2%)</td>
</tr>
<tr>
<td>Opioids</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First use</td>
<td>1,642 (33.4%)</td>
<td>1109 (41.0%)</td>
<td>533 (24.1%)</td>
<td>2,913 (51.2%)</td>
<td>1,927 (60.2%)</td>
<td>986 (39.6%)</td>
</tr>
<tr>
<td>Regular use</td>
<td>1,042 (21.2%)</td>
<td>682 (25.2%)</td>
<td>360 (16.3%)</td>
<td>2,307 (40.5%)</td>
<td>1,518 (47.4%)</td>
<td>789 (31.7%)</td>
</tr>
<tr>
<td>Brought up problems</td>
<td>759 (15.4%)</td>
<td>476 (17.6%)</td>
<td>283 (12.8%)</td>
<td>2,039 (35.8%)</td>
<td>1,338 (41.8%)</td>
<td>701 (28.1%)</td>
</tr>
<tr>
<td>Dependence diagnosis</td>
<td>889 (18.1%)</td>
<td>573 (21.2%)</td>
<td>316 (14.3%)</td>
<td>2,202 (38.7%)</td>
<td>1,437 (44.9%)</td>
<td>765 (30.7%)</td>
</tr>
<tr>
<td>Nicotine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First use</td>
<td>4,269 (86.8%)</td>
<td>2,448 (90.5%)</td>
<td>1,821 (82.2%)</td>
<td>4,998 (87.8%)</td>
<td>2,895 (90.5%)</td>
<td>2,103 (84.4%)</td>
</tr>
<tr>
<td>Regular use</td>
<td>3,324 (67.6%)</td>
<td>2,015 (74.5%)</td>
<td>1,309 (59.1%)</td>
<td>3,709 (65.2%)</td>
<td>2,269 (70.9%)</td>
<td>1,440 (57.8%)</td>
</tr>
<tr>
<td>Brought up problems</td>
<td>1,017 (20.7%)</td>
<td>608 (22.5%)</td>
<td>409 (18.5%)</td>
<td>1,751 (30.8%)</td>
<td>1,073 (33.5%)</td>
<td>678 (27.2%)</td>
</tr>
<tr>
<td>Dependence diagnosis</td>
<td>2,679 (54.5%)</td>
<td>1,601 (59.2%)</td>
<td>1,078 (48.7%)</td>
<td>3,156 (55.4%)</td>
<td>1,966 (61.4%)</td>
<td>1,190 (47.8%)</td>
</tr>
</tbody>
</table>
Table 2: Age (Mean ± SD) of Substance Use Disorder Milestones by Ancestral Group and Sex

<table>
<thead>
<tr>
<th></th>
<th>African Ancestry</th>
<th>European Ancestry</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Male</td>
</tr>
<tr>
<td>Alcohol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First use</td>
<td>15.8 ± 3.8</td>
<td>14.9 ± 3.4</td>
</tr>
<tr>
<td>Regular use</td>
<td>18.6 ± 4.5</td>
<td>17.8 ± 4.0</td>
</tr>
<tr>
<td>Brought up problems</td>
<td>30.9 ± 9.2</td>
<td>31.2 ± 9.5</td>
</tr>
<tr>
<td>Dependence diagnosis</td>
<td>24.7 ± 7.1</td>
<td>24.3 ± 6.9</td>
</tr>
<tr>
<td>Opioids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First use</td>
<td>22.0 ± 5.9</td>
<td>21.5 ± 5.8</td>
</tr>
<tr>
<td>Regular use</td>
<td>23.2 ± 6.7</td>
<td>22.7 ± 6.7</td>
</tr>
<tr>
<td>Brought up problems</td>
<td>31.0 ± 8.8</td>
<td>31.1 ± 9.3</td>
</tr>
<tr>
<td>Dependence diagnosis</td>
<td>26.0 ± 7.8</td>
<td>25.8 ± 8.0</td>
</tr>
<tr>
<td>Nicotine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First use</td>
<td>14.9 ± 4.3</td>
<td>14.8 ± 4.4</td>
</tr>
<tr>
<td>Regular use</td>
<td>19.0 ± 5.6</td>
<td>18.9 ± 5.5</td>
</tr>
<tr>
<td>Brought up problems</td>
<td>36.8 ± 10.0</td>
<td>37.2 ± 10.3</td>
</tr>
<tr>
<td>Dependence diagnosis</td>
<td>26.6 ± 8.4</td>
<td>26.2 ± 8.3</td>
</tr>
</tbody>
</table>

Note: The number of subjects in each cell is shown in Table 1.
Table 3: Effect of Alcohol Use Disorder Polygenic Risk Score on Age of Onset and Progression of Alcohol-Related Measures by Population Group

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>European Ancestry</th>
<th></th>
<th>African Ancestry</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
<td>HR</td>
<td>95% CI</td>
</tr>
<tr>
<td>Age of First Use</td>
<td>0.057</td>
<td>0.014</td>
<td>1.06</td>
<td>1.03-1.09</td>
</tr>
<tr>
<td>Age of Regular Use</td>
<td>0.060</td>
<td>0.015</td>
<td>1.06</td>
<td>1.03-1.10</td>
</tr>
<tr>
<td>Age First Brought Up Problems</td>
<td>0.173</td>
<td>0.023</td>
<td>1.19</td>
<td>1.14-1.24</td>
</tr>
<tr>
<td>Age of Diagnosis of AD</td>
<td>0.136</td>
<td>0.019</td>
<td>1.15</td>
<td>1.10-1.19</td>
</tr>
<tr>
<td>Regular Use to Brought Up Problems</td>
<td>0.133</td>
<td>0.024</td>
<td>1.14</td>
<td>1.09-1.20</td>
</tr>
<tr>
<td>Regular Use to AD Diagnosis</td>
<td>0.096</td>
<td>0.021</td>
<td>1.10</td>
<td>1.06-1.15</td>
</tr>
</tbody>
</table>

Note: All models include sex, age, and the first 10 ancestry principal components. Progression models (Regular Use to ...) also include Age of Regular Use as a covariate and the interaction between time and Age of Regular Use because the proportional hazards assumption was violated. Events=number of individuals meeting the milestone or progression event, SE=standard error, 95%CI=95% confidence interval, HR=hazard ratio, AD=alcohol dependence, PRS=polygenic risk score.
Table 4: Effect of Alcohol Use Disorder Polygenic Risk Score in Cox Proportional Hazard Models of Alcohol-Related Time-to-Event and Progression Measures by Sex among Individuals of European Ancestry

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Male</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Female</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
<td>HR</td>
<td>95% CI</td>
<td>P-Value</td>
<td>Adjusted P-Value</td>
<td>B</td>
<td>SE</td>
<td>HR</td>
<td>95% CI</td>
<td>P-Value</td>
</tr>
<tr>
<td>Age of First Use</td>
<td>0.057</td>
<td>0.019</td>
<td>1.06</td>
<td>1.02-1.10</td>
<td>0.003</td>
<td><strong>0.005</strong></td>
<td>0.060</td>
<td>0.022</td>
<td>1.06</td>
<td>1.02-1.10</td>
<td>0.006</td>
</tr>
<tr>
<td>Age of Regular Use</td>
<td>0.056</td>
<td>0.020</td>
<td>1.06</td>
<td>1.02-1.10</td>
<td>0.005</td>
<td><strong>0.005</strong></td>
<td>0.070</td>
<td>0.024</td>
<td>1.07</td>
<td>1.02-1.13</td>
<td>0.004</td>
</tr>
<tr>
<td>Age First Brought Up Problems</td>
<td>0.136</td>
<td>0.027</td>
<td>1.15</td>
<td>1.09-1.21</td>
<td>&lt;0.0001</td>
<td><strong>0.0005</strong></td>
<td>0.264</td>
<td>0.041</td>
<td>1.30</td>
<td>1.20-1.41</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age of AD Diagnosis</td>
<td>0.123</td>
<td>0.024</td>
<td>1.13</td>
<td>1.08-1.19</td>
<td>&lt;0.0001</td>
<td><strong>0.0005</strong></td>
<td>0.171</td>
<td>0.034</td>
<td>1.19</td>
<td>1.11-1.27</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Regular Use to Brought Up Problems</td>
<td>0.089</td>
<td>0.029</td>
<td>1.09</td>
<td>1.03-1.16</td>
<td>0.002</td>
<td><strong>0.0045</strong></td>
<td>0.234</td>
<td>0.044</td>
<td>1.27</td>
<td>1.16-1.38</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Regular Use to AD Diagnosis</td>
<td>0.085</td>
<td>0.026</td>
<td>1.09</td>
<td>1.04-1.15</td>
<td>0.001</td>
<td><strong>0.004</strong></td>
<td>0.127</td>
<td>0.039</td>
<td>1.14</td>
<td>1.05-1.23</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Note: All models include sex, age, and the first 10 ancestry principal components. Progression models (Regular Use to …) also include Age of Regular Use as a covariate and the interaction between time and Age of Regular Use because the proportional hazards assumption was violated. Events=number of individuals meeting the milestone or progression event, SE=standard error, 95% CI=95% confidence interval, HR=hazard ratio, AD=alcohol dependence.
Table 5: Effect of Opioid Use Disorder Polygenic Risk Score in Cox Proportional Hazard Models of Opioid-Related Time-to-Event and Progression Measures by Population Group

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>European Ancestry</th>
<th>African Ancestry</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
</tr>
<tr>
<td>Age of First Use</td>
<td>0.135</td>
<td>0.019</td>
</tr>
<tr>
<td>Age of Regular Use</td>
<td>0.150</td>
<td>0.021</td>
</tr>
<tr>
<td>Age First Brought Up Problems</td>
<td>0.177</td>
<td>0.023</td>
</tr>
<tr>
<td>Age of OD Diagnosis</td>
<td>0.173</td>
<td>0.022</td>
</tr>
<tr>
<td>Regular Use to Problems</td>
<td>0.054</td>
<td>0.027</td>
</tr>
<tr>
<td>Regular Use to OD Diagnosis</td>
<td>0.035</td>
<td>0.034</td>
</tr>
</tbody>
</table>

Note: All models include sex, age, and the first 10 ancestry principal components. Progression models (Reg Use to …) also include Age of Regular Use as a covariate and the interaction between time and Age of Regular Use because the proportional hazards assumption was violated. Events=number of individuals meeting the milestone or progression event, SE=standard error, 95%CI=95% confidence interval, HR=hazard ratio, OD=opioid dependence.
<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>European Ancestry</th>
<th>African Ancestry</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
</tr>
<tr>
<td>Age of First Use</td>
<td>0.139</td>
<td>0.015</td>
</tr>
<tr>
<td>Age of Regular Use</td>
<td>0.217</td>
<td>0.017</td>
</tr>
<tr>
<td>Age First Brought Up Problems</td>
<td>0.220</td>
<td>0.025</td>
</tr>
<tr>
<td>Age of ND Diagnosis</td>
<td>0.223</td>
<td>0.019</td>
</tr>
<tr>
<td>Regular Use to Problems</td>
<td>0.039</td>
<td>0.026</td>
</tr>
<tr>
<td>Regular Use to ND Diagnosis</td>
<td>0.041</td>
<td>0.020</td>
</tr>
</tbody>
</table>

Note: All models include sex, age, and the first 10 ancestry principal components. Progression models (Reg Use to ...) also include Age of Regular Use as a covariate and the interaction between time and Age of Regular Use because the proportional hazards assumption was violated. Events=number of individuals meeting the milestone or progression event, SE=standard error, 95%CI=95% confidence interval, HR=hazard ratio, ND=tobacco dependence.
Figure 1

African Ancestry

European Ancestry

Proportion Diagnosis Free

Age of Alcohol Dependence Diagnosis

Proportion Diagnosis Free

Age of Alcohol Dependence Diagnosis

Proportion Diagnosis Free

Age of Opioid Dependence Diagnosis

Proportion Diagnosis Free

Age of Opioid Dependence Diagnosis

Proportion Diagnosis Free

Age of Nicotine Dependence Diagnosis

Proportion Diagnosis Free

Age of Nicotine Dependence Diagnosis
Figure 2

The graph shows the proportion of diagnosis-free years from regular alcohol use to Alzheimer's disease (AD) diagnosis for individuals of African and European ancestry, stratified by alcohol use disorder polygenic risk score (PRS) levels (low, medium, high).
Figure 3

Proportion Not Reporting Problems

Age First Brought Up Alcohol-Related Problems

- Female Low
- Female Medium
- Female High
- Male Low
- Male Medium
- Male High