Identifying and Reducing Bias in Genome-Wide Association Studies of Alcohol-Related Traits

Henry R. Kranzler, M.D., Hang Zhou, Ph.D., Rachel L. Kember, Ph.D.

Because an extended period of heavy alcohol consumption is the sine qua non of a diagnosis of alcohol use disorder (AUD), heavy drinking and AUD are often viewed as interchangeable traits. However, despite their obvious phenotypic overlap, neither of the major diagnostic systems—DSM or ICD—includes a quantitative measure of alcohol consumption as a criterion for the diagnosis of AUD. These diagnostic systems focus on the biobehavioral elements of AUD that comprise the alcohol dependence syndrome (1), rather than alcohol consumption per se, and underscore the differences between the traits.

Recent large-scale genome-wide association studies (GWASs) in samples for which both a measure of alcohol consumption and an AUD diagnosis are available have identified multiple genetic variants that contribute to both traits—reflecting their shared genetic risk—as well as genetic variation unique to each (2, 3). In these studies, alcohol consumption and AUD, although positively genetically correlated, diverge in their genetic correlations with many other traits. These and other studies (e.g., 4, 5), show that AUD and the broader phenotype of problematic alcohol use are positively genetically correlated with a broad range of psychiatric (e.g., schizophrenia, major depression) and substance use (e.g., tobacco, cannabis) disorders and are negatively associated with intelligence and socioeconomic measures (e.g., educational level and material deprivation). In contrast, a measure of alcohol consumption (such as that derived from the Alcohol Use Disorders Identification Test [AUDIT]) (6), shows negative genetic correlations with some psychiatric disorders and positive correlations with socioeconomic traits (2, 3, 7). These divergent findings may reflect true biological effects or confounding by selection bias, genetic heterogeneity, or measurement error, including inaccurate self-reports and changes in alcohol consumption and symptoms over time (3, 8, 9).

The rapid pace of gene identification for alcohol-related traits was made possible by the availability of increasingly large study samples, including clinical samples for which electronic health record (EHR) and genetic data are available and community-based samples with phenotype data linked to genetic biobanks. However, EHR data can be biased by socioeconomic and other environmental characteristics unique to specific patient populations. They are also limited by a lack of structured phenotypes, as they were developed and are maintained to serve clinical and reimbursement-related, rather than research, aims. Similarly, individuals whose data come from direct-to-consumer commercial DNA testing companies (e.g., 23andMe) or public efforts aimed at promoting genetic research (e.g., the UK Biobank) differ from the general population on socioeconomic factors and other features, which can also contribute to confounding in genetic analyses.

Recent efforts to assess the impact of these confounders on the observed inconsistencies in genetic correlations between alcohol-related traits and other traits include two studies conducted in the UK Biobank (8, 9) and one in the Million Veteran Program (3). Xue et al. (8) showed that individuals in the UK Biobank with a higher disease burden—based on the number of common diseases they endorsed—are more likely to misreport or to reduce their level of alcohol consumption over time and proposed a method to mitigate the biases. Their application of the correction procedure led to the removal of metabolic and cardiovascular traits from the alcohol consumption GWAS and almost all previously reported negative genetic correlations between alcohol consumption and common diseases becoming either positive or nonsignificant. Dao et al. (9) assessed the effects of excluding former drinkers from the heterogeneous group of current nondrinkers in three overlapping GWASs in participants from the UK Biobank. These investigators compared groups stratified by alcohol consumption level and former drinker or lifetime abstinent status on the frequency of alleles in rs1229984, a polymorphism in the *ADH1B* gene consistently associated in multiple populations with both alcohol consumption and AUD (2–4, 7). Excluding former drinkers from a GWAS of alcohol consumption yielded a stronger association with rs1229984 and three novel single-nucleotide polymorphisms (SNPs) (9). Kember et al. (3) also found that excluding individuals from the Million Veteran Program who reported no drinking in the previous year increased the level of significance of the association of the *ADH1B* SNP with alcohol consumption (measured with

> The study by Mallard et al. complements previous efforts to mitigate biases in the genetic analysis of data from EHRs or other potentially biased sources.

the AUDIT-Consumption [AUDIT-C] score) and increased the proportion of phenotypic variance explained by all common SNPs (i.e., the SNP heritability) for both AUD and alcohol consumption. Further, this approach revealed positive genetic correlations of alcohol consumption with psychiatric traits and a negative correlation with intelligence.

The article by Mallard et al. (10) in this issue of the *Journal* sought to evaluate and account for heterogeneity in the AUDIT itself. This 10-item self-report instrument was developed and widely used to screen clinical and population groups for hazardous or harmful drinking (6). The AUDIT has also been widely used in research as a measure of alcohol involvement. Recently, AUDIT data have been used in GWASs of both alcohol consumption (based on responses to the first three items, i.e., the AUDIT-C subscale) (2, 3, 7) and alcohol-related problems (based on responses to items 4–10, i.e., the AUDIT-Problems or AUDIT-P subscale) (5, 7). Mallard et al. conducted an item-level GWAS of the 10 AUDIT items in 160,824 individuals from three population cohorts. Then, using genomic structural equation modeling to mitigate biases, they identified novel patterns among the items. Their findings showed a correlated genetic structure comprising two factors (alcohol consumption and alcohol-related problems factors) consistent with prior evidence of phenotypic and genetic differences between these traits (7). Using empirically derived weights for each item, they created latent factors (i.e., variables that are not directly observed but are inferred statistically) of consumption and problems. Importantly, these factors were positively genetically correlated with alcohol dependence and other measures of psychiatric and substance use disorders but not with measures of intelligence or socioeconomic status. Mallard et al. also observed key differences between the consumption and problems factors in gene- and transcriptome-based analyses, suggesting partially divergent biological mechanisms for the two latent factors. In a multivariate GWAS, they identified novel genetic variants associated with one, the other, or both factors, consistent with a model of overlapping but distinct genetic etiologies for alcohol consumption and alcohol-related problems.

The analyses by Mallard et al. also revealed that the items in the AUDIT are not equally informative. The first AUDIT item, which measures the frequency of alcohol consumption, was generally less genetically correlated with the other items and had a smaller factor loading and larger residual variance than they did. The resultant "frequency residual" was positively associated with intelligence and education level and negatively associated with psychopathology (e.g., major depressive disorder, alcohol dependence), suggesting that it may be particularly susceptible to biases introduced by environmental or socioeconomic factors. This finding is consistent with epidemiological studies of drinking behavior (11), which show that greater frequency of alcohol consumption is associated with higher socioeconomic status and lower risk of other psychiatric and substance use disorders. In contrast, greater intensity of drinking (i.e., the number of drinks per occasion) and frequency of intoxication—the other two items in the three-item AUDIT-C—show the opposite associations with socioeconomic and psychopathologic features.

The study by Mallard et al. complements the previous efforts (3, 8, 9) to mitigate biases in the genetic analysis of data from EHRs or other potentially biased sources. Each of these four approaches (3, 8–10) yielded novel genetic loci for alcohol-related traits. Thus, efforts are warranted to combine the approaches in a manner appropriate to the sample being studied to enhance phenotypic accuracy and increase the statistical power of GWASs, thereby substantially advancing gene discovery for alcohol-related traits. Confounding in the measurement of alcohol consumption and AUD phenotypes has exaggerated the differences between these traits. However, even if it were possible to account for all confounding in the measurement of these two traits, real genetic differences between them would remain.

AUTHOR AND ARTICLE INFORMATION

Center for Studies of Addiction, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia (Kranzler, Kember); Mental Illness Research, Education, and Clinical Center, Veterans Integrated Service Network 4, Crescenz Veterans Affairs Medical Center, Philadelphia (Kranzler, Kember); Department of Psychiatry, Yale School of Medicine, New Haven, Conn., and Veterans Affairs Connecticut Healthcare System, West Haven, Conn. (Zhou).

Send correspondence to Dr. Kranzler (kranzler@pennmedicine.upenn. edu).

REFERENCES

1. Edwards G, Gross MM: Alcohol dependence: provisional description of a clinical syndrome. BMJ 1976; 1:1058–1061
2. Kranzler HR, Zhou H, Kember RL, et al: Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. Nat Commun 2019; 10: 1499
3. Kember RL, Vickers-Smith R, Zhou H, et al: Genetic underpinnings of the transition from alcohol consumption to alcohol use disorder: shared and unique genetic architectures in a cross-ancestry sample. medRxiv, September 12, 2021 (https://doi.org/10.1101/2021.09.08.21263302)
4. Walters RK, Polimanti R, Johnson EC, et al: Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. Nat Neurosci 2018; 21:1656–1669
5. Zhou H, Sealock JM, Sanchez-Roige S, 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; 23:809–818

6. Saunders JB, Aasland OG, Babor TF, et al: Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons With Harmful Alcohol Consumption–II. Addiction 1993; 88:791–804

7. Sanchez-Roige S, Palmer AA, Fontanillas P, et al: Genome-wide association study meta-analysis of the Alcohol Use Disorders Identification Test (AUDIT) in two population-based cohorts. Am J Psychiatry 2019; 176:107–118

8. Xue A, Jiang L, Zhu Z, et al: Genome-wide analyses of behavioural traits are subject to bias by misreports and longitudinal changes. Nat Commun 2021; 12:20211

9. Dao C, Zhou H, Small A, et al: The impact of removing former drinkers from genome-wide association studies of AUDIT-C. Addiction 2021; 116:3044–3054

10. Mallard TT, Savage JE, Johnson EC, et al. Item-level genome-wide association study of the Alcohol Use Disorders Identification Test in three population-based cohorts. Am J Psychiatry 2022; 179:58–70

11. Marees AT, Smit DJA, Ong JS, et al: Potential influence of socio-economic status on genetic correlations between alcohol consumption measures and mental health. Psychol Med 2020; 50: 484–498