Drug and Alcohol Dependence 219 (2021) 108480

Contents lists available at ScienceDirect

Drug and Alcohol Dependence

journal homepage: www.elsevier.com/locate/drugalcdep

Short communication

Predictive validity of genome-wide polygenic scores for alcohol use from adolescence to young adulthood

Radhika Kandaswamy [a,*], Andrea Allegrini [a], Robert Plomin [a], Sophie von Stumm [b]

[a] Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK
[b] Department of Education, University of York, Heslington, York, UK

ARTICLE INFO

Keywords:
Alcohol use
Adolescence
Polygenic score
TEDS
Alcohol intake frequency
Hazardous drinking

ABSTRACT

Background: Adolescence is a critical period for experimenting with alcohol, and these early experiences have long-term influences on alcohol-related behaviours throughout adulthood. This study examined the utility of genome-wide polygenic scores (GPS) for predicting alcohol use during adolescence and young adulthood.
Methods: We used GPS based on the Genome-wide association study and Sequencing Consortium of Alcohol and Nicotine use (GSCAN) study on drinks per week to predict alcohol use in a longitudinal, UK-representative sample of unrelated adolescents aged 16 through to 22 years ($N_{max} = 3390$).
Results: At age 16, the GSCAN GPS predicted variance in alcohol consumption on a typical day (0.58 %), intake frequency (0.89 %), and hazardous drinking (i.e. ≥6 units at one occasion) (1.07 %). At age 22, the predictive power of the GPS had increased, explaining variance in alcohol consumption (0.61 %), intake frequency (1.69 %), and hazardous drinking (1.19 %).
Conclusions: The predictive validity of GPS for phenotypic alcohol use was evident in adolescence and increased in young adulthood. The findings suggest that GPS, which are available from birth, may be potentially useful for identifying individuals at risk for harmful and hazardous alcohol use. However, because the overall effect sizes were small, the utility of the GPS that are currently available is limited for the prediction of individual-level alcohol use.

1. Introduction

Adolescence is a critical period for experimenting with alcohol, and these early experiences have long-term influences on alcohol-related behaviours throughout adulthood (Marshall, 2014). For example, young people who start drinking alcohol before the age of 15 years are reported to be four times more likely to meet the diagnostic criteria for alcohol dependence later in life (Grant and Dawson, 1997). However, few predictors of adolescent alcohol use have been identified (Dick et al., 2013; Stephenson et al., 2020). Family socioeconomic status (SES), which typically exerts long-term pervasive influence on lifespan development (Bradley and Corwyn, 2002), is inconsistently and weakly associated with adolescent alcohol use (Grittner et al., 2013; Heckley et al., 2017; Kendler et al., 2014). Behavioural risk factors that can be reliably observed in early life, such as smoking and substance use, tend to cluster and co-occur with adolescent alcohol drinking: because these risk factors are contemporaneous with alcohol use, they cannot predict problems before they occur.

A systematic source of influence on alcohol use is genetic factors. The heritability of alcohol use is estimated to be 43 % (Vrieze et al., 2013) and due to many thousand DNA variants that each have a very small effect size (Liu et al., 2019). It follows that individual differences in alcohol-related behaviours map onto a continuum with most people being at average genetic risk for adverse alcohol use. It has recently become possible to aggregate the DNA variants – known as single-nucleotide polymorphisms (SNPs) – that are associated with alcohol use into genome-wide polygenic scores (GPS) that capture an individual's genetic propensity for drinking. GPS are the sum of the effect alleles across an individual's genome, weighted by each allele's strength of association with the target trait in an independent genome-wide association study (GWAS). GPS based on GWAS of alcohol-related behaviours have been shown to successfully predict alcohol consumption, although their predictive validity varies as a function of GWAS sample size, SNP heritability of the trait and selection

* Corresponding author at: Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, Denmark Hill, London, SE5 8AF, UK.
E-mail address: radhika.kandaswamy@kcl.ac.uk (R. Kandaswamy).

https://doi.org/10.1016/j.drugalcdep.2020.108480
Received 15 September 2020; Received in revised form 12 November 2020; Accepted 3 December 2020
Available online 31 December 2020

thresholds for SNP inclusion (Dudbridge, 2013). For example, a GPS based on a GWAS of 941,280 individuals for the number of alcoholic drinks consumed per week, known as GSCAN (i.e. *G*WAS and *S*equencing *C*onsortium of *A*lcohol and *N*icotine use), accounted for 2.4 % and 2.2 % of the phenotypic variance in the drinks consumed per week in independent samples with 3712 and 5483 adults aged 24–32 and 55–64 years, respectively (Liu et al., 2019). Based on the same GSCAN, GPS in 1201 Finnish adults aged 20–26 years accounted for up to 1.1 %, 1.4 %, and 1.6 % of the variance in drinking frequency, intoxication frequency, and alcohol dependence, respectively (Barr et al., 2019). However, using summary statistics from a different GWAS of 67,000 individuals for alcohol intake in grams per day (Schumann et al., 2016), GPS created in 6705 Dutch adults aged 18–83 years predicted only 0.1 % of the variance in alcohol consumption (Mies et al., 2018).

Because alleles do not change during the life course, GPS can be used to signal an individual's behavioural predispositions from birth, long before problems occur. Yet, a GPS-based approach has not been implemented to study genetic influences on the onset and development of alcohol use in a longitudinal sample, although it was recently tested in the prediction of alcohol use disorders (Barr et al., 2020). Here, we created GPS based on the summary statistics from GSCAN (Liu et al., 2019) and tested their association with alcohol-related phenotypes – alcohol consumption on a typical day, intake frequency, and hazardous drinking (≥6 units of alcohol on one occasion) – in a longitudinal population sample that was assessed at ages16 and 22.

2. Methods

2.1. Study population

Participants were drawn from the Twins Early Development Study (TEDS) (Rimfeld et al., 2019), a longitudinal study that enrolled over 10,000 twin pairs born 1994–1996 in England and Wales. Despite considerable attrition, TEDS participants and their families are representative of families in the UK (Rimfeld et al., 2019). Project approval was granted by King's College London's ethics committee for the Institute of Psychiatry, Psychology and Neuroscience (05.Q0706/228). Genotypes were processed using stringent quality control procedures followed by SNP imputation using the Haplotype Reference Consortium (release 1.1) reference panels. Following imputation, we excluded variants with minor allele frequency <0.5 %, Hardy-Weinberg equilibrium *P*-values of $<1 \times 10^{-5}$. For a full description of quality control, imputation procedures, and attrition see Supporting information.

2.2. Alcohol measures

Alcohol use initiation (i.e. having ever consumed alcohol), alcohol consumption (i.e. units of alcohol consumed on a typical day), intake frequency, and hazardous drinking (i.e., ≥6 units of alcohol on one occasion) were assessed at ages 16 and 22 using questionnaires (see Supporting information). Data on alcohol measures and genotype was available for 3063 and 3390 unrelated individuals at ages 16 and 22, respectively. We excluded individuals who stated never to have consumed alcohol by the age of 16 ($N = 433$) or 22 ($N = 104$) from all analyses, except for the alcohol use initiation analysis.

2.3. Genome-wide polygenic scores

We computed GPS for TEDS participants from GSCAN based on the target phenotype "average number of alcoholic drinks consumed per week" (Liu et al., 2019). We obtained summary statistics excluding the 23andMe sample ($N = 537,349$) and used LDpred (Vilhjalmsson et al., 2015) for GPS construction, computing the score for a fraction of causal markers = 1 (i.e. allowing all SNPs to contribute to trait variation) (see Supporting information). The GPS were standardized using the *scale* function in R within each sample to ease interpretation of effect sizes.

2.4. Statistical analyses

Using logistic regression models, we tested whether GPS predicted alcohol use initiation at ages 16 and 22, adjusting for the covariates of age, gender, SES, first ten principal components (i.e. population stratification), and genotyping array (see Supporting information). Associations between GSCAN GPS and alcohol-related phenotypes were tested using linear regressions, adjusting for the same set of covariates. The predictive power of the GPS was calculated as the change in McFadden's pseudo-R^2 and R^2 (variance) for the logistic and the linear regression, respectively, between the full regression model that included GPS along with all the covariates and the baseline model without GPS. We calculated 95 % confidence intervals (CI) around all R^2 values with 10,000 bootstrapped repetitions. We employed a Bonferroni corrected *P*-value of 0.008 (0.05/6, 2 age groups and 3 phenotypes) to adjust for multiple testing, which might be an overly conservative approach when studying associations with small effect sizes, as in the current study. We also conducted supplementary analyses to investigate whether the GPS for educational attainment (EA), one of the most powerful GPS in the prediction of behavioural outcomes (Lee et al., 2018), influenced alcohol-related outcomes at ages 16 and 22. Here, we created EA GPS from the largest available GWAS on EA (Lee et al., 2018), before adding them to the regression model that also included the GSCAN GPS and the covariates (see Supporting information). All statistical analyses were preregistered as part of a wider set of analyses (https://osf.io/q4gkh/) and conducted using R statistical software (R Core Team, 2018).

3. Results

Descriptive statistics for our sample are presented in Table 1. As expected, alcohol consumption, intake frequency and hazardous drinking all increased substantially from age 16 to 22. Alcohol-related phenotypes were positively correlated with each other at age 16 ($r = 0.29–0.72$, $P < 0.001$) and age 22 ($r = 0.15–0.56$, $P < 0.001$) (Fig. S1). Males drank more frequently and were involved in hazardous drinking at both ages ($P < 0.001$).

We found that GSCAN GPS predicted alcohol initiation at age 16 (McFadden's pseudo-$R^2 = 0.27$ %), and this prediction improved at age 22 (McFadden's pseudo-$R^2 = 0.90$ %). At age 16, a standard deviation increase in GPS for alcohol use was associated with 15 % (95 % CI: 0.02, 0.26) greater odds to have ever consumed alcohol ($N_{Drinkers} = 2630$; 86 % of sample). At age 22, a standard deviation increase in the GPS was associated with 35 % (95 % CI: 0.08, 0.52) greater odds to have ever consumed alcohol ($N_{Drinkers} = 3286$; 97 % of the sample). In exploratory analyses, the interaction between gender and GPS in predicting whether alcohol was ever consumed was not significant at age 16 ($\beta = 0.128$, 95 % CI: -0.11, 0.36; 55 % females) or age 22 (β=-0.139, 95 % CI: -0.58, 0.31; 63 % females). This suggested that the GPS prediction did not

Table 1
Descriptive sample summary statistics.

Characteristic/Variable	*Age 16 (N = 3063)*		*Age 22 (N = 3390)*	
	Mean/ (N)	*SD/ (%)*	*Mean/ (N)*	*SD/ (%)*
Age	16.32	0.68	22.26	0.91
Male	1379	45.0 %	1257	37.0%
Alcohol use	2630	85.9%	3286	96.9%
Alcohol consumption (units in a typical day)	1.86	1.16	4.79	3.01
Alcohol intake frequency	1.21	0.69	1.91	1.02
Hazardous drinking (≥6 units on one occasion)	0.62	0.78	1.56	0.96

*One unit of alcohol is: ½ pint average strength beer/lager OR one glass of wine OR on single measure of spirits, SD = standard deviation.

differ for males and females for the initiation of alcohol use.

We also found that GSCAN GPS significantly predicted alcohol-related behaviour at age 16 in individuals who had ever consumed alcohol ($N_{Drinkers} = 2630$). As illustrated in Fig. 1, we observed positive linear associations between the GPS and age 16 alcohol-related phenotypes, including alcohol consumption ($R^2 = 0.58$ %, $\beta = 0.092$, $P = 5.78e{-}04$), alcohol intake frequency ($R^2 = 0.89$ %, $\beta = 0.067$, $P = 4.49e{-}06$), and hazardous drinking ($R^2 = 1.07$ %, $\beta = 0.082$, $P = 2.58e{-}06$) (Table S1). At age 22 ($N_{Drinkers} = 3286$), a consistent increase in the predictive validity of the GPS was observed for all three outcome measures, specifically for alcohol consumption ($R^2 = 0.61$ %, $\beta = 0.045$, $P = 3.14e{-}04$), alcohol intake frequency ($R^2 = 1.69$ %, $\beta = 0.133$, $P = 4.80e{-}12$), and hazardous drinking ($R^2 = 1.19$ %, $\beta = 0.107$, $P = 3.24e{-}08$) (Table S1). SES did not show any significant association with alcohol use outcomes at age 16, and was only associated with intake frequency at age 22 ($b = 0.186$, $P \leq 2.2e{-}16$, Table S1), indicating that higher-SES individuals drank more frequently.

Our exploratory analyses revealed that the EA GPS did not affect the association between the GSCAN GPS and the alcohol use phenotypes (see Supporting information, Table S2). We observed one significant association between the EA GPS and alcohol intake frequency at age 22 ($\beta = 0.063$, $P = 0.002$), with the EA GPS accounting for 0.3 % of the variance, independent from the GSCAN GPS (Table S2).

4. Discussion

This is the first study to use a GPS based approach to investigate developmental trends in the genomic prediction of alcohol use behaviours from adolescence to young adulthood. We report an increase in the predictive validity of a GPS for alcohol use (GSCAN) from age 16 to 22 years by 5% for alcohol consumption, 90 % for alcohol intake frequency, and 11 % for hazardous drinking, with overall small effect sizes (Fig. 1). This finding is in line with other studies that observed increases in the heritability of behavioural phenotypes from childhood to adulthood (Allegrini et al., 2019; Bergen et al., 2007; Briley and Tucker-Drob, 2014). The increase in the predictive validity of GPS for phenotypic alcohol use from adolescence through young adulthood suggests that GPS, which are just as predictive at birth as in later life, may be potentially useful for identifying individuals at risk for harmful and hazardous alcohol use.

The variance explained by the GSCAN GPS in alcohol consumption on a typical day at the ages of 16 and 22 years (0.58 % and 0.61 %) was lower than that reported in a sample of adults aged 24–32 years (2.4 %; (Liu et al., 2019). However, our estimates for alcohol intake frequency and hazardous drinking at age 22 (1.7 % and 1.2 %) approximated the predictive validity reported in a Finnish sample aged 20–26 years (1.1 % and 1.4 %). These findings suggest that even though the effect sizes associated with the GPS for alcohol consumption in the prediction of alcohol use outcomes are small, they are fairly consistent across samples.

Because drinking behaviours in adolescence have been shown to be influenced by family background (Marees et al., 2019), we controlled for family SES in all our analyses. We found no association between family SES and alcohol use outcomes at age 16, except a relation at age 22 with drinking frequency. These results suggest an unstable pattern of association between family SES and alcohol consumption, at least by age 22.

Because the overall effect sizes were small, the utility of the GPS that are currently available is limited in the prediction of individual-level alcohol use. That said, GWAS are widely expected to improve in the near future, for example by including even larger population samples and using whole-genome sequencing (Chatterjee et al., 2013; Khera et al., 2019). These studies will in turn beget GPS that are more powerful for identifying individuals at risk for harmful or hazardous alcohol use. Including more powerful GPS as a risk factor for alcohol use in adolescence may help to prevent individuals at risk from suffering long-term impairments from early alcohol use. However at present, the clinical utility of GPS for alcohol use appears limited.

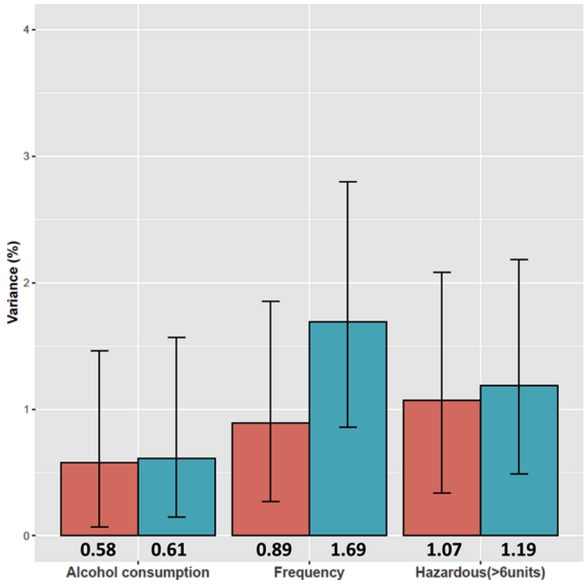

Fig. 1. Genome-wide polygenic score prediction of alcohol use at age 16 and 22 (pink and blue bars respectively). Error bars are 95 % confidence intervals estimated with 10,000 bootstrapped repetitions. Note. Alcohol consumption assessed the number of units consumed on a typical day when drinking; frequency referred to how often alcoholic beverages were consumed; and hazardous drinking assessed the frequency of having six or more units of alcohol on a single occasion (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Role of funding source

This study was supported by the European Foundation for Alcohol Research (ERAB) (grant EA 18 16, SvS & RP) and Jacobs Foundation (Early Career Fellowship 2017–2019, SvS). AA is funded by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 721567. TEDS is supported by a program grant from the UK Medical Research Council (grant G0901245 and previously G0500079, RP), with additional support from the US National Institutes of Health (grants HD044454 & HD059215, RP).

Contributors

RK and SvS were involved in the concept and design of the study. RK, AA and SvS contributed to data acquisition and analysis. RK and SvS wrote the manuscript and all authors critically reviewed the content and approved the final version for publication.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

We gratefully acknowledge the ongoing contribution of the participants in the Twins Early Development Study (TEDS) and their families.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.drugalcdep.2020.108480.

References

Allegrini, A.G., Selzam, S., Rimfeld, K., von Stumm, S., Pingault, J.B., Plomin, R., 2019. Genomic prediction of cognitive traits in childhood and adolescence. Mol. Psychiatry 24 (6), 819–827.

Barr, P.B., Kuo, S.I., Aliev, F., Latvala, A., Viken, R., Rose, R.J., Kaprio, J., Salvatore, J.E., Dick, D.M., 2019. Polygenic risk for alcohol misuse is moderated by romantic partnerships. Addiction 114 (10), 1753–1762.

Barr, P.B., Ksinan, A., Su, J., Johnson, E.C., Meyers, J.L., Wetherill, L., Latvala, A., Aliev, F., Chan, G., Kuperman, S., Nurnberger, J., Kamarajan, C., Anokhin, A., Agrawal, A., Rose, R.J., Edenberg, H.J., Schuckit, M., Kaprio, J., Dick, D.M., 2020. Using polygenic scores for identifying individuals at increased risk of substance use disorders in clinical and population samples. Transl. Psychiatry 10 (1), 196.

Bergen, S.E., Gardner, C.O., Kendler, K.S., 2007. Age-related changes in heritability of behavioral phenotypes over adolescence and young adulthood: a meta-analysis. Twin Res. Hum. Genet. 10 (3), 423–433.

Bradley, R.H., Corwyn, R.F., 2002. Socioeconomic status and child development. Annu. Rev. Psychol. 53, 371–399.

Briley, D.A., Tucker-Drob, E.M., 2014. Genetic and environmental continuity in personality development: a meta-analysis. Psychol. Bull. 140 (5), 1303–1331.

Chatterjee, N., Wheeler, B., Sampson, J., Hartge, P., Chanock, S.J., Park, J.H., 2013. Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. Nat. Genet. 45 (4), 400–405, 405e401-403.

Dick, D.M., Aliev, F., Latendresse, S.J., Hickman, M., Heron, J., Macleod, J., Joinson, C., Maughan, B., Lewis, G., Kendler, K.S., 2013. Adolescent alcohol use is predicted by childhood temperament factors before age 5, with mediation through personality and peers. Alcohol. Clin. Exp. Res. 37 (12), 2108–2117.

Dudbridge, F., 2013. Power and predictive accuracy of polygenic risk scores. PLoS Genet. 9 (3) e1003348.

Grant, B.F., Dawson, D.A., 1997. Age at onset of alcohol use and its association with DSM-IV alcohol abuse and dependence: results from the National Longitudinal Alcohol Epidemiologic Survey. J. Subst. Abuse 9, 103–110.

Grittner, U., Kuntsche, S., Gmel, G., Bloomfield, K., 2013. Alcohol consumption and social inequality at the individual and country levels–results from an international study. Eur. J. Public Health 23 (2), 332–339.

Heckley, G., Jarl, J., Gerdtham, U.G., 2017. Frequency and intensity of alcohol consumption: new evidence from Sweden. Eur. J. Health Econ. 18 (4), 495–517.

Kendler, K.S., Gardner, C.O., Hickman, M., Heron, J., Macleod, J., Lewis, G., Dick, D.M., 2014. Socioeconomic status and alcohol-related behaviors in mid- to late adolescence in the Avon Longitudinal Study of Parents and Children. J. Stud. Alcohol Drugs 75 (4), 541–545.

Khera, A.V., Chaffin, M., Zekavat, S.M., Collins, R.L., Roselli, C., Natarajan, P., Lichtman, J.H., D'Onofrio, G., Mattera, J., Dreyer, R., Spertus, J.A., Taylor, K.D., Psaty, B.M., Rich, S.S., Post, W., Gupta, N., Gabriel, S., Lander, E., Ida Chen, Y.D., Talkowski, M.E., Rotter, J.I., Krumholz, H.M., Kathiresan, S., 2019. Whole-genome sequencing to characterize monogenic and polygenic contributions in patients hospitalized with early-onset myocardial infarction. Circulation 139 (13), 1593–1602.

Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., et al., 2018. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat. Genet. 50 (8), 1112–1121.

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D.M., Chen, F., et al., 2019. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat. Genet. 51 (2), 237–244.

Marees, A.T., Smit, D.J.A., Ong, J.S., MacGregor, S., An, J., Denys, D., Vorspan, F., van den Brink, W., Derks, E.M., 2019. Potential influence of socioeconomic status on genetic correlations between alcohol consumption measures and mental health. Psychol. Med. 1–15.

Marshall, E.J., 2014. Adolescent alcohol use: risks and consequences. Alcohol Alcohol. 49 (2), 160–164.

Mies, G.W., Verweij, K.J.H., Treur, J.L., Ligthart, L., Fedko, I.O., Hottenga, J.J., Willemsen, G., Bartels, M., Boomsma, D.I., Vink, J.M., 2018. Polygenic risk for alcohol consumption and its association with alcohol-related phenotypes: Do stress and life satisfaction moderate these relationships? Drug Alcohol Depend. 183, 7–12.

Core Team, R., 2018. R: A language and environment for statistical computingR: A Language and Environment for Statistical Computing. R Found. Stat. Comput.

Rimfeld, K., Malanchini, M., Spargo, T., Spickernell, G., Selzam, S., McMillan, A., Dale, P. S., Eley, T.C., Plomin, R., 2019. Twins early development study: a genetically sensitive investigation into behavioral and cognitive development from infancy to emerging adulthood. Twin Res. Hum. Genet. 22 (6), 508–513.

Schumann, G., Liu, C., O'Reilly, P., Gao, H., Song, P., Xu, B., et al., 2016. KLB is associated with alcohol drinking, and its gene product beta-Klotho is necessary for FGF21 regulation of alcohol preference. Proc. Natl. Acad. Sci. U. S. A. 113 (50), 14372–14377.

Stephenson, M., Barr, P., Ksinan, A., Aliev, F., Latvala, A., Viken, R., Rose, R., Kaprio, J., Dick, D., Salvatore, J.E., 2020. Which adolescent factors predict alcohol misuse in young adulthood? A co-twin comparisons study. Addiction 115 (5), 877–887.

Vilhjalmsson, B.J., Yang, J., Finucane, H.K., Gusev, A., Lindstrom, S., Ripke, S., et al., 2015. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. Am. J. Hum. Genet. 97 (4), 576–592.

Vrieze, S.I., McGue, M., Miller, M.B., Hicks, B.M., Iacono, W.G., 2013. Three mutually informative ways to understand the genetic relationships among behavioral disinhibition, alcohol use, drug use, nicotine use/dependence, and their co-occurrence: twin biometry, GCTA, and genome-wide scoring. Behav. Genet. 43 (2), 97–107.